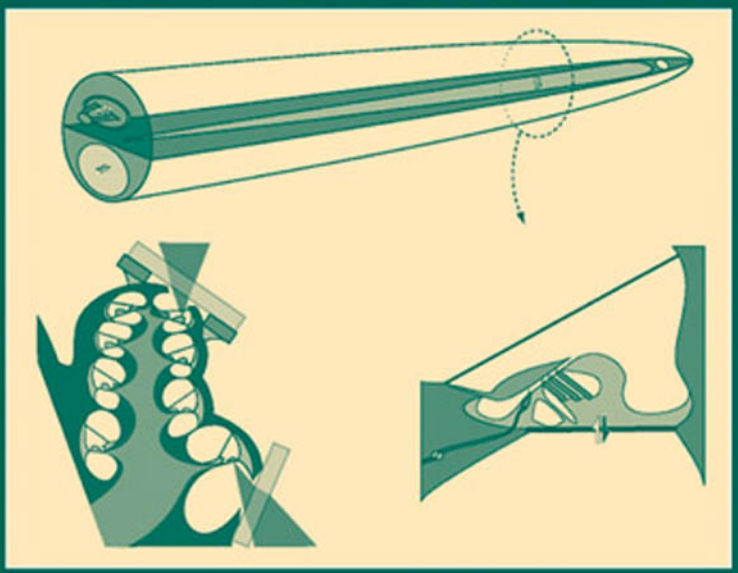
SPRINGER HANDBOOK OF AUDITORY RESEARCH
Series Editors: Richard R. Fay and Arthur N. Popper

Sid P. Bacon
Richard R. Fay
Arthur N. Popper Editors

Compression

From Cochlea to Cochlear Implants



Springer

Springer Handbook of
Auditory Research

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

Springer
New York
Berlin
Heidelberg
Hong Kong
London
Milan
Paris
Tokyo

This page intentionally left blank

Sid P. Bacon
Richard R. Fay
Arthur N. Popper
Editors

Compression: From Cochlea to Cochlear Implants

With 62 Illustrations

Springer

Sid P. Bacon
Department of Speech and Hearing Science and
Psychoacoustics Laboratory
Arizona State University
Tempe, AZ 85287-1908, USA
spb@asu.edu

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

Arthur N. Popper
Department of Biology and
Neuroscience and Cognitive Science Program and
Center for Comparative and Evolutionary Biology of Hearing
University of Maryland
College Park, MD 20742–4415, USA
ap17@umail.umd.edu

Cover illustration: Functional anatomy of the mammalian cochlea. This figure appears on p. 20 of the text.

Library of Congress Cataloging-in-Publication Data
Compression : from cochlea to cochlear implants / editor, Sid P. Bacon, Richard R. Fay, Arthur N. Popper.
p. cm.—(Springer handbook of auditory research ; v. 17)
Includes bibliographical references and index.
ISBN 0-387-00496-3 (alk. paper)
1. Cochlea. 2. Compression (Audiology). 3. Auditory perception. 4. Cochlear implants. I. Bacon, Sid P. II. Fay, Richard R. III. Popper, Arthur N. IV. Series.
QP471.2.C66 2003
612.8′58—dc21 2003042430

ISBN 0-387-00496-3 Printed on acid-free paper.

9 8 7 6 5 4 3 2 1 SPIN 10913048

www.springer-ny.com

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

This book is dedicated to Professor Neal F. Viemeister, whose numerous scientific contributions have led to a deeper understanding of auditory perception and its intricate relation to physiological processing at the auditory periphery.

This page intentionally left blank

Series Preface

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

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

Each volume in the series consists of five to eight substantial chapters on a particular topic. In some cases, the topics will be ones of traditional interest for which there is a substantial body of data and theory, such as auditory neuroanatomy (Vol. 1) and neurophysiology (Vol. 2). Other volumes in the series will 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 will be joined by a co-editor having special expertise in the topic of the volume.

Richard R. Fay, Chicago, Illinois
Arthur N. Popper, College Park, Maryland

This page intentionally left blank

Preface

One truly remarkable aspect of the auditory system is its ability to process sounds over an incredibly wide range of levels, on the order of 120 dB. For example, the loudness of a sound will increase with increasing level over this entire range. That the auditory system has an extraordinarily large dynamic range and is acutely sensitive to small changes in level throughout that range is an especially important aspect of hearing because it enables humans to process speech over a wide range of levels and to extract important information from the small changes in level that occur over that rather wide range.

An interesting and important problem for hearing scientists has been to understand how the auditory system is able to process sound over such a wide dynamic range. As discussed in this volume, it is now believed that the enormous psychophysical dynamic range is accomplished via a form of compression that exists at the level of the cochlea. This peripheral auditory compression is the focus of this book. In the first chapter, Bacon provides a brief overview of peripheral compression, the perceptual consequences of this compression, the perceptual consequences of reduced or absent compression after cochlear hearing loss, and the signal-processing strategies used to compensate for this reduced compression. This chapter sets the stage for the remaining chapters, where those topics are covered in detail. In Chapter 2, Cooper describes peripheral auditory compression as observed in the mechanical responses of the basilar membrane as well as in the electrical responses of the auditory nerve. In Chapter 3, Oxenham and Bacon describe a wide range of perceptual consequences that are thought to result from peripheral auditory compression. Then in Chapter 4, Bacon and Oxenham discuss how perceptual processing can be affected by a reduction in or loss of compression after cochlear damage. In Chapter 5, Levitt discusses a form of signal processing (known as compression amplification) that is often used in hearing aids as a means to at least partially compensate for reduced peripheral compression in individuals with cochlear hearing loss. Finally, in Chapter 6, Zeng describes psychophysical research on individuals fitted with cochlear implants and discusses the the-

oretical implications of that research for peripheral compression and the practical implications for signal processing in cochlear implants.

This volume can serve as a companion to several others in the *Springer Handbook of Auditory Research*. Many aspects of psychophysics were reviewed in *Human Psychophysics* (Vol. 3). A number of other chapters in the series are related to similar topics in other animal species. For example, psychophysics of fish and amphibians is considered in a chapter by Fay and Megela Simmons in *Comparative Hearing: Fish and Amphibians* (Vol. 11), on birds and reptiles in a chapter by Dooling et al. in *Comparative Hearing: Birds and Reptiles* (Vol. 13), and on dolphins and whales in a chapter by Nachtigall et al. in *Hearing by Whales and Dolphins* (Vol. 12).

Sid P. Bacon, Tempe, Arizona
Richard R. Fay, Chicago, Illinois
Arthur N. Popper, College Park, Maryland

Contents

This page intentionally left blank

Contributors

Sid P. Bacon
Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, Tempe, AZ 85287-1908, USA *spb@asu.edu*

Nigel P. Cooper
MacKay Institute of Communication and Neuroscience, Keele University, Keele, Staffordshire, ST5 5BG, UK *n.p.cooper@cns.keele.ac.uk*

Harry Levitt
998 SeaEagle Loop, Bodega Bay, CA 94923-0610, USA *hlevitt@starband.net*

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

Fan-Gang Zeng
University of California, 364 Med Surge II, Irvine, CA 92697, USA *fzeng@uci.edu*

This page intentionally left blank

1
Overview of Auditory Compression

Sid P. Bacon

1. Introduction

One truly remarkable aspect of the auditory system is its ability to process sounds over an incredibly wide range of levels. This range for humans is illustrated in Figure 1.1. The bottom curve shows the sound pressure levels at which pure tones of a wide range of frequencies can just be detected. In other words, it shows the absolute thresholds for human hearing. This curve is sometimes referred to as an audibility curve. Everything below the curve is inaudible, whereas everything above it is audible.[1] The top curve shows the sound pressure levels at which the pure tones are uncomfortably loud. The area between these two curves represents the so-called dynamic range of hearing. Although the dynamic range varies with frequency, the dynamic range for a tone with a frequency of around 1,000 Hz is at least 120 dB. This corresponds to a truly impressive dynamic range of 10^{12} intensity units (W/m^2). The loudness of a sound will increase with increasing stimulus level throughout the range, and it may continue to increase at even higher levels, although few are willing to find out! Not only does the sensation of loudness grow with increasing level over this range but humans can also detect small (approximately 1 dB) changes in level over this entire range (Viemeister and Bacon 1988). Thus the auditory system has an extraordinarily large dynamic range and is acutely sensitive to small changes in level throughout that range. This is an especially important aspect of human hearing. It enables humans, for example, to accommodate the wide range of levels in speech that, when one considers both the changes in average level and the changes in instantaneous level, is on the order of 60–80 dB for everyday listening situations (e.g., Levitt 1982). Furthermore, it enables humans to extract important information from the small changes in level that occur over that wide range.

[1] This is, however, not strictly true. The absolute threshold is defined statistically as the sound pressure level necessary to detect the sound a certain percentage of times (e.g., 70%).

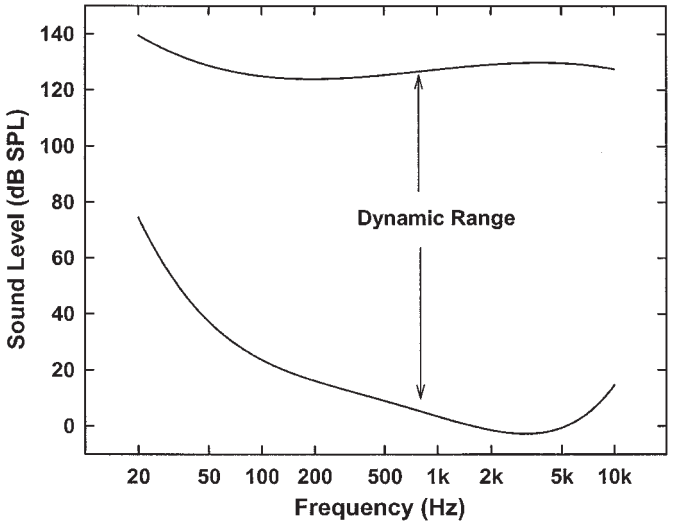


FIGURE 1.1. Bottom solid line: levels at which pure tones of various frequencies can be just detected; top solid line: levels at which the tones become uncomfortably loud. The area in between the 2 curves represents the dynamic range of hearing. SPL, sound pressure level.

An interesting problem for hearing scientists has been to understand how the auditory system is able to process or encode sound over such a large range of levels. The problem is particularly apparent when one considers that the typical dynamic range of auditory nerve fibers is on the order of 20–40 dB (for reviews, see Ruggero 1992; Cooper, Chapter 2). This substantial discrepancy between the behavioral or psychophysical dynamic range and the physiological dynamic range has been referred to as the "dynamic range problem" (Evans 1981). It has been the focus of both psychophysical and physiological research (for reviews, see Viemeister 1988a,b; Delgutte 1996). As discussed below (Section 2), it is now believed that the enormous psychophysical dynamic range is accomplished via a form of compression that exists at the level of the cochlea. This peripheral auditory compression is the focus of this book. The various chapters describe not only the cochlear processing underlying compression (Cooper, Chapter 2) but also the wide range of perceptual consequences that are thought to result from this compression (Oxenham and Bacon, Chapter 3) as well as the perceptual consequences of reduced or absent compression (Bacon and Oxenham, Chapter 4). This book also provides a discussion of the rehabilitative or signal-processing strategies used to compensate for the lack of peripheral auditory compression in individuals fitted with either hearing aids (Levitt, Chapter 5) or cochlear implants (Zeng, Chapter 6). The

purpose of this chapter is to provide a brief overview of these topics and hence set the stage for the more detailed discussions that follow.

2. Cochlear Compression

The mammalian cochlea consists of two distinct types of sensory cells that run almost its entire length (for a review of cochlear anatomy, see Slepecky 1996). There is one row of inner hair cells and generally three rows of outer hair cells. The hair cells and their supporting cells lie directly above the basilar membrane. The vast majority (90–95%) of the auditory nerve fibers synapse directly with the inner hair cells, whereas the remainder synapse with the outer hair cells. It is generally believed that all recordings of neural activity from the auditory nerve fibers reflect the responses of fibers innervating inner hair cells (e.g., Liberman 1982; Liberman and Oliver 1984). Furthermore, those fibers that synapse with outer hair cells are apparently unresponsive to sound (Robertson 1984). Thus the neural code at the auditory nerve directly reflects the response properties solely of the inner hair cells. One might ask, then: What is the role of the considerably more numerous outer hair cells in hearing? Over the past 20 or so years, tremendous insight into this question has been gained (for reviews of cochlear mechanics and hair cell physiology, see Dallos 1988, 1992; Kros 1996; Patuzzi 1996; Robles and Ruggero 2001; Cooper, Chapter 2). The results from a wide range of scientific approaches clearly indicate that the outer hair cells are crucial for normal auditory function. Indeed, it is now understood that these cells are responsible for our exquisite sensitivity and superb frequency-resolving capabilities. Furthermore, the importance of these cells is underscored by the often severe auditory-processing difficulties experienced by individuals with outer hair cell damage (for a review of the perceptual consequences of hair cell damage, see Moore 1998).

The outer hair cells are also responsible for various nonlinear phenomena that can be observed, for instance, in the mechanical response of the basilar membrane and the neural response of the auditory nerve. Moreover, psychophysical correlates of these nonlinear phenomena can be measured in human listeners. These nonlinear phenomena include the generation of distortion products (summation and difference tones) and two-tone suppression (see Cooper, Chapter 2) as well as a compressive growth of response, which is the focus of this book. The remainder of this section considers this compressive growth as it is observed at the basilar membrane.

The basilar membrane vibrates in a characteristic manner in response to sound (for a review of cochlear mechanics, see Patuzzi 1996). In particular, there is a wave of displacement that travels from the base of the cochlea to the apex. It increases in magnitude until it reaches a peak at some place on the basilar membrane, and then it dies off rather abruptly. This is known as

the traveling wave. Much of what is known about the traveling wave comes from the pioneering work on cochlear mechanics by Georg von Békésy, for which he won the Nobel Prize in Physiology or Medicine in 1961 (for a summary of that work, see von Békésy 1960). The place where the traveling wave reaches its peak depends on the frequency of the sound. High frequencies peak near the base of the cochlea, whereas low frequencies peak near the apex. Indeed, there is a frequency-to-place mapping along the basilar membrane that is known as tonotopic organization. This organization is maintained throughout the auditory system. Thus each place along the basilar membrane responds best to only one frequency, although it will respond to other frequencies as well. This frequency is usually referred to as the characteristic frequency (CF).[2]

To gain an appreciation of the compression that exists at the basilar membrane, it is necessary to consider how the magnitude of basilar membrane vibration at a given point along the membrane grows as a function of stimulus level. An example of this is shown in Figure 1.2, where the velocity of basilar membrane movement is plotted as a function of stimulus level. This type of function is generally referred to as a basilar membrane input-output function. These results are from the basal region of a chinchilla cochlea, a region that responds best to high frequencies. The CF of this particular recording site was 10,000 Hz. The results in Figure 1.2 are in response to a 10,000-Hz tone (i.e., a tone at CF). The magnitude of the response generally increases with increasing stimulus level, but the growth is quite compressive. This is clear by comparing the input-output function to the linear function (dashed line). Throughout its most compressive region (at moderate to high stimulus levels), the input-output function has a slope of about 0.2 dB/dB, which corresponds to a compression ratio of about 5:1. Interestingly, the input-output function is compressive only for stimulus frequencies near the CF of the recording site. The input-output function is linear for stimulus frequencies well below or well above the CF (e.g., Ruggero et al. 1997; Cooper, Chapter 2). As a result of this compression, any given point along the basilar membrane is able to respond to an extremely large range of stimulus levels. Because basilar membrane motion serves as the proximal stimulus for the inner hair cells and, subsequently, the auditory nerve fibers, the compression that is observed at the basilar membrane greatly extends the dynamic range of the peripheral auditory system.

---

[2] The CF for a given place is typically defined as the frequency that requires the least sound pressure level to generate a criterion response; thus it is defined using relatively low stimulus levels. At high sound pressure levels, a given place may respond best to a frequency lower than its nominal CF (see Cooper, Chapter 2). This corresponds to a basal shift in the peak of the traveling wave with increasing stimulus level (e.g., Ruggero et al. 1997; Rhode and Recio 2000). Throughout this chapter, CF refers to the frequency that is most effective at eliciting a response at a given place in the cochlea at relatively low levels.

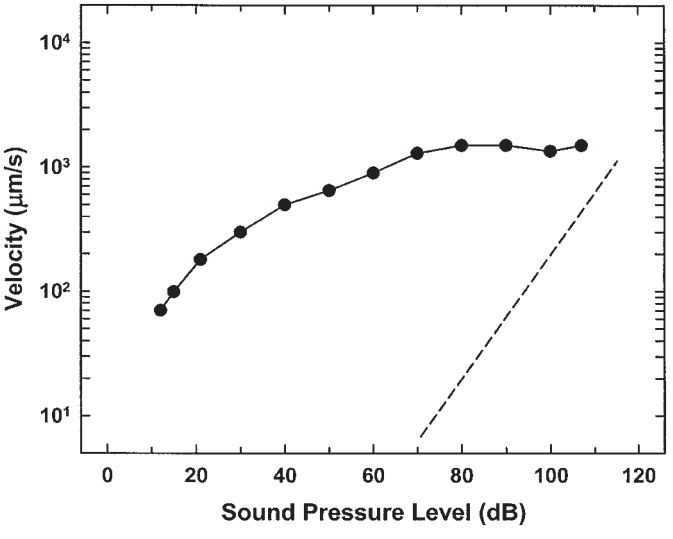


FIGURE 1.2. An example of a basilar membrane input-output function (circles connected by a solid line). The stimulus frequency (10,000 Hz) was equal to the characteristic frequency (CF) of the recording site. Dashed line: linear growth of response. Note that the input-output function is highly compressive. (Data from Ruggero et al. 1997.)

2.1 The Role of Outer Hair Cells

Although compression can be measured at the basilar membrane, it is not due to the mechanics of the membrane per se. In other words, the basilar membrane by itself is not compressive. Where does this compression come from? Some insight into that question can be gained from the results in Figure 1.3 that show basilar membrane input-output functions measured at the 11,000-Hz place both before and after an intravenous injection of the drug quinine. Quinine is ototoxic and is thought to directly affect the outer hair cells in the cochlea (Karlsson and Flock 1990; Jarboe and Hallworth 1999; Zheng et al. 2001). In terms of the response to a tone at CF, quinine reduces the magnitude of the basilar membrane response at the lower stimulus levels but not at the higher levels. In other words, a drug that adversely affects the outer hair cells can result in a smaller mechanical response at the basilar membrane. Because the effect of quinine is relatively large at low levels and essentially nonexistent at high levels, the consequence of this temporary outer hair cell damage is a more linear input-output function. Quinine has no effect on the response to a relatively low-frequency tone (data not shown) where the growth of response at the basilar membrane is

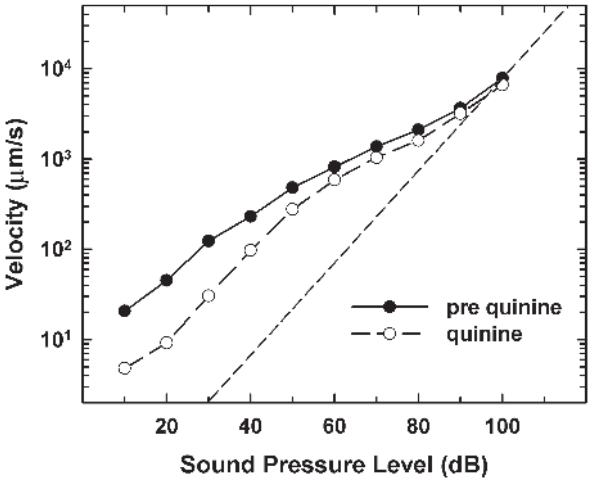


FIGURE 1.3. The effects of quinine on the basilar membrane input-output function for a tone at which the frequency (11,000 Hz) was equal to the CF of the recording site. The input-output function was measured before and after administration of quinine. (Data are from Recio and Ruggero 1995.)

already linear (see Fig. 2.9 in Ruggero et al. 1996). This suggests that the outer hair cells only affect the response to a tone at (or near) the CF where the growth is normally compressive.

A variety of manipulations that have resulted in various degrees of temporary or permanent hearing loss have yielded results similar to those described above. The main finding is summarized schematically in Figure 1.4. For a tone at CF, the growth of response is compressive under normal conditions but is linear (or at least more linear) when the outer hair cells are damaged or functioning abnormally. Although the precise way in which the outer hair cells affect the motion of the basilar membrane is unclear, it is likely the result of outer hair cell electromotility. Indeed, a fascinating finding from research on cochlear hair cells is that the outer hair cells have motor capability resulting in their being motile and, in isolation, being capable of changing shape at rates in the audio frequency range (Brownell et al. 1985; Kachar et al. 1986; Zenner 1986; Ashmore 1987; for a review, see Holley 1996). Recently, Zheng et al. (2000) have identified the motor protein (prestin) responsible for this electromotility. These shape changes are thought to alter the micromechanical properties of the cochlea so as to increase the vibration of the basilar membrane in a frequency-selective way. In other words, they provide local mechanical amplification in the form of feedback. For this reason, the outer hair cells are often referred to as the "cochlear amplifier" (Davis 1983; for a review, see Robles and Ruggero

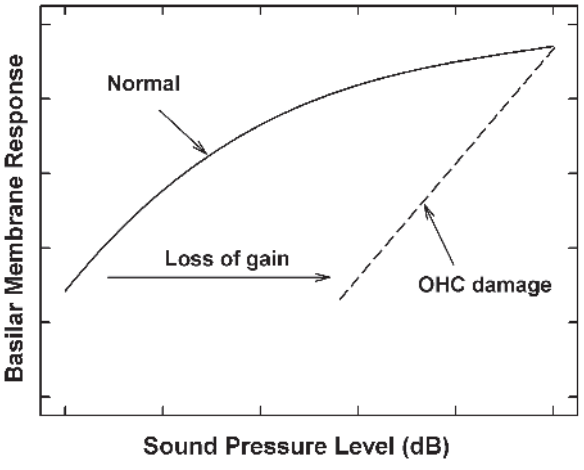


FIGURE 1.4. Summary of the effects of outer hair cell (OHC) damage on the basilar membrane input-output function for a tone at the CF of the recording site. Under normal conditions, the function is compressive. After outer hair cell damage, the function becomes linear (or at least more linear) and there is a concomitant loss of gain, particularly at lower SPLs.

2001). Damage to the outer hair cells results in a loss of that amplification or gain, as shown in Figure 1.4. The amount of gain that normally exists has been estimated to be as large as 50–80 dB for lower sound pressure levels (Ruggero et al. 1997). The gain decreases with increasing level and is negligible at high levels (as evidenced by the horizontal difference between the solid and dashed lines in Fig. 1.4). This amplification is responsible for our incredible sensitivity to sound. That is, this cochlear amplification is what enables us to detect low-level sounds, thereby lending strong support to the claim that the outer hair cells are largely responsible for our enormous dynamic range of hearing.

3. Some Perceptual Consequences of Basilar Membrane Compression

An important goal in hearing science is to explain auditory perceptual phenomena in terms of underlying physiological mechanisms. Although the following point may be obvious, it is nevertheless worth noting that psychophysical measurements do not reflect solely the response of any one part of the auditory system but instead reflect the response of the entire system. This clearly makes it difficult to conclude with absolute certainty that a particular psychophysical phenomenon reflects largely or primarily the response of a specific site in the system. Nevertheless, it is possible to

observe psychophysical correlates of processing that exist at a specific site in the auditory system. Furthermore, the greater the similarity between the psychophysical and physiological observations or measurements, the stronger the putative connection between the two. Sufficient evidence exists linking our exquisite sensitivity to functioning outer hair cells that few would argue against the claim that the extremely large dynamic range of hearing is one perceptual consequence of basilar membrane compression. There are many other perceptual phenomena that likewise appear to be closely linked to basilar membrane compression. Oxenham and Bacon (Chapter 3) discuss these in detail, whereas the remainder of this section provides an overview of just a few such phenomena.

Masking experiments have been used for well over a century to gain insight into the way in which the auditory system processes sound (Mayer 1876). The focus often has been on the spectral characteristics of masking so that these masking experiments have concentrated primarily on revealing the limits of our frequency-resolving (or filtering) capabilities. It is likely that the filtering that exists along the basilar membrane is reflected in the psychophysical estimates of auditory filtering (Fletcher 1940; Moore 1993; Shera et al. 2002). Indeed, it has often been argued that the auditory filter (or critical band) corresponds to a certain distance along the basilar membrane (Fletcher 1940; Zwicker et al. 1957; Greenwood 1961, 1990; Moore 1986; Glasberg and Moore 1990). This distance is usually estimated at about 0.9 mm. In other words, the auditory filter bandwidth is thought to correspond to a distance along the basilar membrane of about 0.9 mm.

Given the apparently close tie between psychophysical masking and basilar membrane mechanics, it is perhaps not surprising that the growth of masking might be intimately linked to the growth of response at the basilar membrane. It seems likely, for example, that the well-known "upward spread of masking" (e.g., Wegel and Lane 1924; Egan and Hake 1950; see Fig. 3.2 in Oxenham and Bacon, Chapter 3) can be explained by basilar membrane compression. This is where the masking produced by a narrowband masker spreads considerably more to higher frequencies than it does to lower frequencies, growing expansively at the higher frequencies as the masker level increases to high sound pressure levels. Moreover, it is now believed that the rate at which this upward spread of masking grows with masker level can provide an estimate of the degree of compression measured at the basilar membrane (Oxenham and Plack 1997; Oxenham and Bacon, Chapter 3; Bacon and Oxenham, Chapter 4), at least when the masker and signal do not overlap in time, such as in forward masking. To the extent to which this is true, these behavioral experiments can be used to obtain noninvasive estimates of the growth of basilar membrane response and, by implication, insight into the functioning of the outer hair cells in the cochlea. Whether these types of measures will ever make their way into the audiology clinic as a diagnostic test of outer hair cell function remains to be seen, but their utility in the laboratory is clear.

The apparent influence of basilar membrane compression can also be seen in a variety of measures of temporal processing such as forward masking and the masking by temporally fluctuating maskers. This is particularly interesting given that temporal measures are often thought to reflect processing primarily at a site in the auditory system that is central to the auditory nerve. Of course, peripheral processing is obligatory and could obviously have an effect on any measure, even those that reflect significant amounts of central processing. As discussed by Oxenham and Bacon (Chapter 3), the role of basilar membrane compression in these types of measures has been evaluated within the context of a temporal window model (see Fig. 3.5 in Oxenham and Bacon, Chapter 3). This model incorporates a static nonlinearity that is thought to represent basilar membrane compression. The degree of compression that is needed to account for various psychophysical data is generally consistent with the amount of compression observed directly via mechanical measurements at the basilar membrane. Furthermore, the results from individuals with cochlear hearing loss (Bacon and Oxenham, Chapter 4) often can be predicted rather well with the complete elimination of compression in the model, providing converging evidence in support of an important role for basilar membrane compression in these measures.

4. Some Perceptual Consequences of Reduced or Absent Compression

Most individuals with a permanent sensorineural hearing loss suffer from damage to the outer hair cells in the cochlea. As discussed in Section 2.1, this damage results in a reduction in or loss of basilar membrane compression. Thus these individuals provide an opportunity to evaluate the potential role of basilar membrane compression in hearing. In addition, it is possible to adversely affect outer hair cells temporarily via well-controlled exposures to intense sound or the ingestion of moderately high doses of aspirin. Individuals with a temporary loss induced by either of these agents provide an especially interesting and informative subject population for studying the role of basilar membrane compression. This is because the temporary loss is restricted to the outer hair cells, whereas with a permanent loss, the inner hair cells may be damaged as well. Bacon and Oxenham (Chapter 4) discuss research on individuals with either a temporary or permanent cochlear hearing loss and evaluate the extent to which the various perceptual consequences of cochlear damage can be understood in terms of a loss of basilar membrane compression. An example of some of that research is highlighted below.

As noted in Section 3, the growth of masking may provide an estimate of the growth of response at the basilar membrane. Supporting evidence

for that comes from masking experiments in individuals with a temporary, aspirin-induced hearing loss (Hicks and Bacon 1999, 2000). Some results are shown in Figure 1.5. In these types of experiments, the masker is usually much lower in frequency than the signal so that the response to the masker at the signal frequency place (where the subject is likely to detect the signal) will be linear, whereas the response to the signal at that same place will be compressive (for more details, see Oxenham and Bacon, Chapter 3; Bacon and Oxenham, Chapter 4). In this particular experiment, the signal was fixed in level and the masker was varied in level to mask the signal. The experimental paradigm was forward masking, and thus the brief signal occurred immediately after the offset of the masker. The slope of the masking function is more gradual (i.e., more compressive) before aspirin ingestion than it is after aspirin ingestion. This indicates that the aspirin has reduced the amount of compression in the auditory system. Note that aspirin has its greatest effect at lower signal levels and has little or no effect at higher levels. Importantly, this mimics the effect of quinine on the growth of response at the basilar membrane (see Fig. 1.3), thus strengthening the argu-

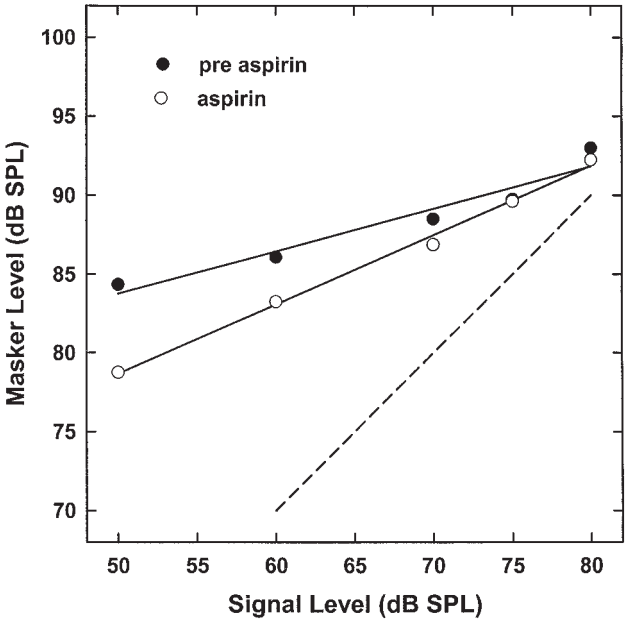


FIGURE 1.5. An example of the effect of aspirin on growth-of-masking functions in forward masking. The frequency of the masker (2,222 Hz) was well below the frequency of the signal (4,000 Hz). The masking function was measured before and after administration of aspirin, which resulted in a temporary hearing loss of about 13 dB. Dashed line: linear growth. (Data are from Hicks and Bacon 2000.)

ment that the behavioral experiments estimate the growth of response measured directly at the basilar membrane. Similar masking experiments have been used to examine how permanent sensorineural hearing loss affects compression (Oxenham and Plack 1997; Nelson et al. 2001). These results suggest that relatively severe, permanent loss can completely eliminate the compressive growth of response.

As discussed in Section 2.1, the reduction in or loss of basilar membrane compression as a consequence of outer hair cell damage is accompanied by a decrease in or loss of gain that is normally observed at low stimulus levels. The most obvious perceptual consequence of this damage is an elevation of the absolute threshold for sound, which represents the hearing loss. Because the level at which sounds become uncomfortably loud usually does not change much with hearing loss, the elevated absolute thresholds result in a reduced dynamic range of hearing. For individuals with a cochlear hearing loss, the loudness of a sound goes from relatively soft to uncomfortably loud over a smaller range of sound pressure levels than it does in individuals with normal hearing. This is referred to as "loudness recruitment." The reduced dynamic range in individuals with a cochlear hearing loss poses a problem in terms of rehabilitation: a large range of stimulus levels must be "squeezed" or compressed into the relatively small dynamic range.

5. Hearing Aids

The most common form of rehabilitation for individuals with a cochlear hearing loss is amplification. The intent is to amplify sounds so that the individual with the hearing loss can hear them. The reduced dynamic range, however, provides a considerable challenge. This is illustrated in Figure 1.6. In Figure 1.6*A*, the growth of loudness is shown schematically for a normal (solid line) and an impaired (dashed line) ear. The specific goal of a hearing aid might be to amplify low-level sounds a great deal but high-level sounds only a little, if at all, so as to shift the response of the impaired ear to be more in line with the response of the normal ear (as indicated by the arrows). As noted in Section 2.1, outer hair cells normally provide this type of level-dependent amplification. That is, the amount by which they amplify the vibration of the basilar membrane decreases with increasing input level. This goal of level-dependent amplification will not be accomplished by simple linear amplification where all sounds are amplified by the same amount. Instead, compression amplification has become an increasingly more popular type of amplification for individuals with cochlear hearing loss in order to deal successfully with their reduced dynamic range. An illustration of one type of compression amplification is shown in Figure 1.6*B*, which shows an example of the input-output function of a compression hearing aid. In this case, the function is linear up to an input level of 40 dB

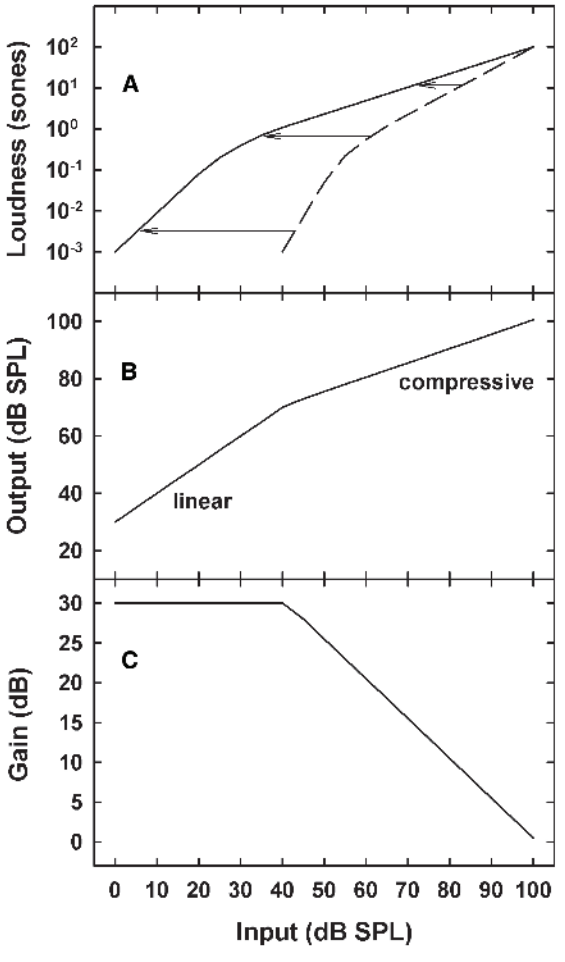


FIGURE 1.6. *A*: growth of loudness (in sones) for a normal (solid line) and an impaired (dashed line) ear. The length of the arrows indicates the amount by which the sound would need to be amplified to shift the impaired ear to be equal to the normal ear. *B*: input-output function for a hearing aid that is linear at input levels of 40 dB SPL and below and compressive at levels above that. *C*: gain provided by that hearing aid.

and compressive above that. The gain (the difference between the output and the input) for this function is shown in Figure 1.6*C*. Note that this type of amplification has accomplished the specific goal of providing considerable gain at low levels and increasingly smaller amounts of gain at higher levels. In a broad sense, compression hearing aids are attempting to "restore" the compression that is normally observed and, consequently, extend the dynamic range of people with hearing loss. As discussed by

Levitt (Chapter 5), hearing aids can provide various degrees of compression and can provide different amounts of compression in different frequency regions in order to accommodate different severities and configurations of hearing loss.

6. Cochlear Implants

Some individuals have a hearing loss that is so severe that they will not benefit from acoustic amplification. Those individuals are often fitted with a cochlear implant, which bypasses the damaged cochlea and stimulates the auditory nerve directly with electric current. The dynamic range of individuals with a cochlear implant is typically on the order of only about 5–20 dB in electric current, which is much smaller than the typical acoustic dynamic range of a hearing-impaired individual fitted with a hearing aid. Thus the amount of compression required to map the acoustic amplitude to electric current may be even greater than the compression required in typical hearing aids. Although cochlear implants are relatively new, a considerable amount of psychophysical research has been conducted in individuals fitted with cochlear implants over the past 20 years. Some of that research has focused on determining the appropriate compression required in the signal-processing stage to go from acoustic amplitude to electric current. Zeng (Chapter 6) summarizes this research and discusses, among other things, the important issues faced by scientists and clinicians when trying to map the large range of acoustic levels in the auditory environment to an auditory system that does not benefit from the compression that normally exists at the basilar membrane in the cochlea.

7. Summary

One function of the cochlea is to transduce mechanical vibrations along the basilar membrane into neural impulses that can be processed by the auditory nervous system. The auditory nerve fibers that convey neural information from the cochlea synapse directly with the inner hair cells. Fibers that synapse with the outer hair cells are apparently unresponsive to sound and thus do not contribute directly to the neural code. Nevertheless, research over the last 20 or so years has clearly elucidated the importance of outer hair cells for hearing. Amazingly, these sensory cells have a motor function. As a result of their motile properties, they affect basilar membrane mechanics and hence, ultimately, the neural information conveyed by the auditory nerve fibers. The outer hair cells enhance or amplify the vibration of the basilar membrane in a frequency-selective way at low stimulus levels, but the amount of that amplification decreases with increasing stimulus level, leading to a compressive growth of response at the basilar membrane.

The amplification accounts for the exquisite sensitivity of the auditory system and, consequently, its impressively large dynamic range of hearing. And because the amplification is frequency selective, it is responsible for the excellent frequency-resolving (or filtering) capabilities of the system. Psychophysical research in subjects with normal hearing and subjects with either temporary or permanent cochlear hearing loss, together with various modeling efforts, suggests that the perceptual consequences of basilar membrane compression may be far-reaching.

Temporary or permanent damage to the outer hair cells in the cochlea results in a decrease in or loss of the amplification provided by those cells and hence a decrease in or loss of basilar membrane compression (among other things). The perceptual consequences of this damage are significant and lie at the heart of the auditory processing difficulties experienced by hearing-impaired individuals. Hearing aids and cochlear implants are important rehabilitative devices for individuals with a cochlear hearing loss. These devices must cope with the loss of basilar membrane compression and the subsequent reduction in the dynamic range of hearing. By incorporating compression in their signal-processing schemes, they can map the large level variations in the acoustic environment into the significantly reduced dynamic range of hearing experienced by these patients. In a sense, they restore some of the important compression that normally exists at the basilar membrane of the cochlea.

The importance of peripheral auditory compression is becoming increasingly clear as auditory scientists from various disciplines focus on the role of the outer hair cells in hearing. Considerable advances in understanding the nature and perceptual consequences of basilar membrane compression have occurred primarily over the past two decades. It is fair to say that future research will provide important new insights into this compression and the extent to which it influences auditory behavior. Some of this future research undoubtedly will focus on whether the degree and nature of compression vary along the length of the cochlea from base to apex. Most of the direct measurements of basilar membrane motion to date have been restricted to the basal region. There is, however, some evidence that there is less compression at the apical region and that the compression there is less frequency selective (see Cooper, Chapter 2). Similarly, there is psychophysical evidence that the degree of compression is greatly reduced at low frequencies, although the possibility that compression may be less frequency selective at low frequencies complicates most psychophysical estimates of compression at those frequencies. This, in turn, complicates the comparisons of compression across frequency (see Oxenham and Bacon, Chapter 3).

Research on compression also might provide important insight into why hearing-impaired individuals with similar amounts of hearing loss often perform quite differently on a range of auditory tasks. One possibility is that the differences in performance might be explained by differences in

the degree of residual compression and, by implication, differences in the amount of residual outer hair cell function.

Advances in the rehabilitation of cochlear hearing loss will also benefit from additional research on compression. In this regard, the important issues for both hearing aids and cochlear implants revolve around determining the optimum compression settings that will maximize the understanding of speech while at the same time providing the listener with a high level of overall satisfaction with the device.

These and many other issues will be at the forefront of future research as scientists continue to strive not only to understand the complex and pervasive role that basilar membrane compression plays in hearing but also to determine the best ways to compensate for the reduction in or loss of compression that accompanies cochlear hearing loss.

Acknowledgments. The writing of this chapter was partially supported by National Institute on Deafness and Other Communication Disorders Grant DC-01376. I thank Li Liu for her assistance with the figures.

References

Ashmore JF (1987) A fast motile response in guinea-pig outer hair cells: the cellular basis of the cochlear amplifier. J Physiol (Lond) 388:323–347.

Brownell WE, Bader CR, Bertrand D, de Ribaupierre Y (1985) Evoked mechanical responses of isolated cochlear outer hair cells. Science 227:194–196.

Dallos P (1988) Cochlear neurobiology: some key experiments and concepts of the past two decades. In: Edelman GM, Gall WE, Cowan WM (eds) Auditory Function: Neurobiological Bases of Hearing. New York: John Wiley and Sons, pp. 153–189.

Dallos P (1992) The active cochlea. J Neurosci 12:4575–4585.

Davis H (1983) An active process in cochlear mechanics. Hear Res 9:79–90.

Delgutte B (1996) Physiological models for basic auditory percepts. In: Hawkins HL, McMullen TA, Popper AN, Fay RR (eds) Auditory Computation. New York: Springer-Verlag, pp. 157–220.

Egan JP, Hake HW (1950) On the masking pattern of a simple auditory stimulus. J Acoust Soc Am 22:622–630.

Evans EF (1981) The dynamic range problem: place and time coding at the level of cochlear nerve and nucleus. In: Syka J, Aitkin L (eds) Neuronal Mechanisms in Hearing. New York: Plenum Press, pp. 69–85.

Fletcher H (1940) Auditory patterns. Rev Mod Phys 12:47–65.

Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. Hear Res 47:103–138.

Greenwood DD (1961) Critical bandwidth and the frequency coordinates of the basilar membrane. J Acoust Soc Am 33:1344–1356.

Greenwood DD (1990) A cochlear frequency-position function for several species—29 years later. J Acoust Soc Am 87:2592–2605.

Hicks ML, Bacon SP (1999) Effects of aspirin on psychophysical measures of frequency selectivity, two-tone suppression, and growth of masking. J Acoust Soc Am 106:1436–1451.
Hicks ML, Bacon SP (2000) The effects of aspirin on a psychophysical estimate of basilar membrane compression. J Acoust Soc Am 107:2914.
Holley MC (1996) Outer hair cell motility. In: Dallos P, Popper AN, Fay RR (eds) The Cochlea. New York: Springer-Verlag, pp. 386–434.
Jarboe JK, Hallworth R (1999) The effect of quinine on outer hair cell shape, compliance and force. Hear Res 132:43–50.
Kachar B, Brownell WE, Altschuler R, Fex J (1986) Electrokinetic shape changes of cochlear outer hair cells. Nature 322:365–368.
Karlsson KK, Flock Å (1990) Quinine causes isolated outer hair cells to change length. Neurosci Lett 116:101–105.
Kros CJ (1996) Physiology of mammalian cochlear hair cells. In: Dallos P, Popper AN, Fay RR (eds) The Cochlea. New York: Springer-Verlag, pp. 318–385.
Levitt H (1982) Speech discrimination ability in the hearing impaired: spectrum considerations. In: Studebaker GA, Bess FH (eds) The Vanderbilt Hearing Aid Report (Monographs in Contemporary Audiology). Upper Darby, Pa, pp. 32–43.
Liberman MC (1982) Single-neuron labeling in the cat auditory nerve. Science 216:1239–1241.
Liberman MC, Oliver ME (1984) Morphometry of intracellularly labeled neurons of the auditory nerve: correlations with functional properties. J Comp Neurol 223: 163–176.
Mayer AM (1876) Researches in acoustics. Philos Mag 2:500–507.
Moore BCJ (1986) Parallels between frequency selectivity measured psychophysically and in cochlear mechanics. Scand Audiol Suppl 25:139–152.
Moore BCJ (1993) Frequency analysis and pitch perception. In: Yost WA, Popper AN, Fay RR (eds) Human Psychophysics. New York: Springer-Verlag, pp. 56–115.
Moore BCJ (1998) Cochlear Hearing Loss. London: Whurr Publishers.
Nelson DA, Schroder AC, Wojtczak M (2001) A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners. J Acoust Soc Am 110:2045–2064.
Oxenham AJ, Plack CJ (1997) A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. J Acoust Soc Am 101:3666–3675.
Patuzzi R (1996) Cochlear micromechanics and macromechanics. In: Dallos P, Popper AN, Fay RR (eds) The Cochlea. New York: Springer-Verlag, pp. 186–257.
Recio A, Ruggero MA (1995) Effects of quinine on basilar-membrane responses to sound. Assoc Res Otolaryngol Midwinter Mtg Abstr 18:200.
Rhode WS, Recio A (2000) Study of mechanical motions in the basal region of the chinchilla cochlea. J Acoust Soc Am 107:3317–3332.
Robertson D (1984) Horseradish peroxidase injection of physiologically characterised afferent and efferent neurons in the guinea pig spiral ganglion. Hear Res 15:113–121.
Robles L, Ruggero MA (2001) Mechanics of the mammalian cochlea. Physiol Rev 81:1305–1352.
Ruggero MA (1992) Physiology and coding of sound in the auditory nerve. In: Popper AN, Fay RR (eds) The Mammalian Auditory Pathway: Neurophysiology. New York: Springer-Verlag, pp. 34–93.

Ruggero MA, Rich NC, Robles L, Recio A (1996) The effects of acoustic trauma, other cochlear injury, and death on basilar-membrane responses to sound. In: Axelsson A, Borchgrevink H, Hamernik RP, Hellström P, Henderson D, Salvi RJ (eds) Scientific Basis of Noise-Induced Hearing Loss. New York: Thieme, pp. 23–35.

Ruggero MA, Rich NC, Recio A, Narayan SS, Robles L (1997) Basilar-membrane responses to tones at the base of the chinchilla cochlea. J Acoust Soc Am 101:2151–2163.

Shera CA, Guinan JJ Jr, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc Natl Acad Sci USA 99:3318–3323.

Slepecky NB (1996) Structure of the mammalian cochlea. In: Dallos P, Popper AN, Fay RR (eds) The Cochlea. New York: Springer-Verlag, pp. 44–129.

Viemeister NF (1988a) Intensity coding and the dynamic range problem. Hear Res 34:267–274.

Viemeister NF (1988b) Psychophysical aspects of auditory intensity coding. In: Edelman GM, Gall WE, Cowan WM (eds) Auditory Function: Neurobiological Bases of Hearing. New York: John Wiley and Sons, pp. 213–242.

Viemeister NF, Bacon SP (1988) Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones. J Acoust Soc Am 84:172–178.

von Békésy G (1960) Experiments In Hearing. New York: McGraw-Hill.

Wegel RL, Lane CE (1924) The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear. Phys Rev 23:266–285.

Zenner HP (1986) Motile responses in outer hair cells. Hear Res 22:83–90.

Zheng J, Shen W, He DZ, Long KB, Madison LD, Dallos P (2000) Prestin is the motor protein of cochlear outer hair cells. Nature 405:149–155.

Zheng J, Ren T, Parthasarathi A, Nuttall AL (2001) Quinine-induced alterations of electrically evoked otoacoustic emissions and cochlear potentials in guinea pigs. Hear Res 154:124–134.

Zwicker E, Flottorp G, Stevens SS (1957) Critical band width in loudness summation. J Acoust Soc Am 29:548–557.

2
Compression in the Peripheral Auditory System

Nigel P. Cooper

1. Introduction

1.1 Overview

The peripheral auditory system converts sound into an informative ensemble of neural signals. The range of sounds that it deals with is immense. Most mammals can accurately distinguish spectral features spanning at least two decades of frequency and six decades of intensity. To a large extent, this performance is made possible through the biologically and physically active processing of sound that occurs in the cochlea (the acoustic part of the inner ear). The purpose of this chapter is to describe this processing in some detail, with particular emphasis on the coding of the intensity of a sound. The coding of the spectrum of a sound, through cochlear frequency analysis, is the subject of detailed reviews elsewhere (e.g., Patuzzi and Robertson 1988; Patuzzi 1996).

As far as intensity coding is concerned, the primary function of the peripheral auditory system is to combine high sensitivity with a large dynamic range. It achieves this by combining amplification with compression. For the purposes of this review, it is important to make the meanings of these two terms very clear. Amplification simply means "making something larger," whereas compression means "fitting more into an available space than would otherwise be possible." The two terms are not reciprocal; hence something that is more compressed is not necessarily less amplified, and something that is more amplified is not, of necessity, less compressed. The idea that amplification and compression are reciprocal is a common misunderstanding, and one that can lead to much confusion when interpreting data from the peripheral auditory system. As we shall see, a healthy peripheral auditory system fits a lot more information into a limited amount of output space by the selective amplification of its responses to low-level stimuli.

1.2 A Brief Review of the Anatomy and Physiology of the Peripheral Auditory System

The peripheral auditory system comprises the external ear, the middle ear, the cochlea, and the auditory nerve (the acoustic part of the eighth cranial nerve). The external and middle ears are primarily concerned with the efficient transmission of sound from an animal's environment to its inner ear. This transmission is generally regarded to be a linear process, such that the spectral and temporal features of the sound are preserved with high fidelity at the input to the cochlea (Guinan and Peake 1967; see Rosowski 1994 for a review). However, the middle ear does afford some means of amplitude control on the transmitted sounds. Two of the small, articulated bones within the middle ear are attached to a pair of muscles known as the tensor tympani and the stapedius muscles. The bones in question are the malleus and the stapes, which connect to the eardrum and the oval window of the cochlea, respectively. The muscles can alter the position of both of these bones as well as the tension under which they operate. The muscles can therefore alter the acoustic impedance[1] of the middle ear, and this, in turn, affects the efficiency of a sounds transmission to the cochlea.

The middle ear muscles are normally activated only in response to very loud sounds, and their primary function is therefore commonly considered to be protective (Borg et al. 1984; Pang and Peake 1986). However, the muscles can also be activated voluntarily, and they are thought to contract routinely during vocalization (Borg and Zakrisson 1975). The muscles reduce the transmission of low-frequency sounds more than high-frequency sounds, so they can often be used to improve the detection of signals in background noise (Pang and Guinan 1997). Contraction of the middle ear muscles is also known to improve the intelligibility of speech at high sound pressure levels (SPLs) (Borg and Counter 1989). Because the middle ear muscles effectively operate by turning down the volume of high-level sounds, they can be considered as one means of achieving compression in the peripheral auditory system. When compared to the amount of compression to be achieved within the cochlea, however, their significance is lessened considerably.

The anatomical structure of the mammalian cochlea is fairly complex, and has been described in great detail elsewhere (e.g., see Slepecky 1996). For the purposes of this chapter, the gross structure of the cochlea can be simplified into the form illustrated in Figure 2.1*A*. All of the space within the cochlea is filled with fluids. These fluids are split into three chambers or scalae by the structures of the cochlear partition (only two of the chambers are evident in Fig. 2.1*A*, but all three are shown in Fig. 2.1*B*,*C*). The fluids in two of the chambers, the scala vestibuli and the scala tympani, can

[1] The acoustic impedance of a material determines how much sound pressure is needed to cause a particular amount of movement within that material.

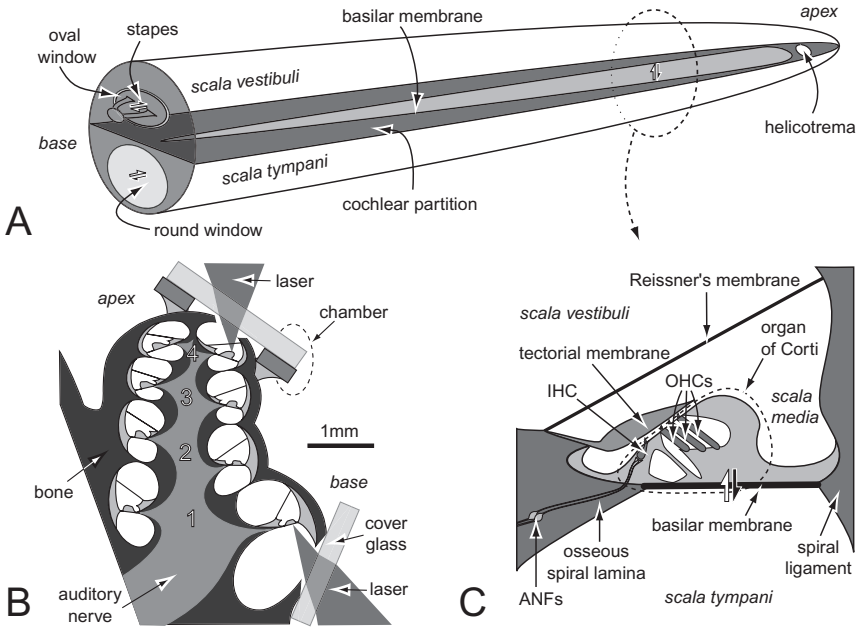


FIGURE 2.1. Functional anatomy of the mammalian cochlea. *A*: layout of a hypothetically straightened or uncoiled cochlea. The cochlear partition is represented as a flat plate that separates the fluid-filled spaces of the cochlea into 2 main chambers. *B*: schematic cross section through a real mammalian cochlea, showing 2 experimental approaches to the cochlear partition. The diagram is based on a cross section of a guinea pig cochlea in which the cochlear partition spirals four times around the trunk of the auditory nerve. *C*: schematic cross section through the apical turn of a real cochlea showing some of the cellular detail of the partition. IHC, inner hair cells; OHCs, outer hair cells; ANFs, auditory nerve fibers. (*B* and *C* adapted with permission from Cooper 1999b. Copyright © 1999 Elsevier Science.)

communicate with one another through the helicotrema, a small hole close to one end of the cochlear partition. The fluids can also communicate with the outside world through the two "windows" in the cochlea. The oval window is closely apposed to the stapes of the middle ear and allows sounds to be transferred into the cochlea very efficiently. The round window, on the other hand, vents the pressure in the cochlear fluid back to that in the air-filled cavities of the middle ear. This arrangement means that the middle ear does not have to work too hard to pressurize the fluid-filled inner ear in order to transmit a sound into it; it merely has to move the fluid back and forth in synchrony with the sound (see Patuzzi 1996). When the stapes pushes the oval window into (out of) the cochlea, the fluid moves almost instantaneously along the cochlear ducts and the round window bulges outward (inward) to preserve the volume of the fluid. The movements of

the fluid are driven by the pressure gradient that exists between the oval and round windows. The details of this gradient are determined by the flexible parts of the cochlear partition, which act to set up a secondary series of pressure waves that travel along the length of the cochlea (see below).

The difference between the pressures on the two sides of the cochlear partition (between the scala vestibuli and scala media on the one side and the scala tympani on the other) forces the flexible parts of the partition (including the basilar membrane, the organ of Corti, the tectorial membrane, and Reissner's membrane; see Fig. 2.1*C*) to move up and down in synchrony with the sounds (see black and white half-arrows in Fig. 2.1*A*,*C*). Not all of the cross sections of the partition move at once, however. The time that the "traveling" pressure waves take to reach a particular cross section depends on the position of the cross section along the length of the cochlea as well as both the frequency and intensity of the sound. In general, both the pressure waves and the vibration waves that they set up tend to travel from the base to the apex of the cochlea, with the low-frequency components of the waves traveling further and slightly faster than the high-frequency components, and the components of the low-intensity sounds traveling slightly further and slightly more slowly than those of the higher intensity sounds (for reviews see Davis 1983; Patuzzi 1996). The mapping of the spectral characteristics of a sound onto the spatial coordinates of the cochlear partition is known as tonotopy, and is a feature of the fundamental importance to the performance of the entire auditory system. It is brought about by variations in the physical characteristics of the cochlear partition, such as its stiffness, its mass, and its architecture. All of these characteristics change systematically from one end of the cochlea to the other (Fernàndez 1952; von Békésy 1960; Lim 1980; see Patuzzi 1996 for a review).

The structures that actually detect the presence of a sound within the cochlea are a group of highly specialized receptor cells known as hair cells. These cells can be subdivided into two groups depending on their location across the width of the cochlear partition (see Fig. 2.1*C*): The inner hair cells line up in a single row at the edge of the organ of Corti, directly above the border between the flexible basilar membrane and the inflexible osseous spiral lamina (see Fig. 2.1*C*), whereas the outer hair cells are distributed across three (or occasionally four) rows in the center of the organ of Corti, directly above the most flexible region of the basilar membrane. The inner hair cells are heavily innervated by the afferent fibers that form the auditory nerve, and their main function is to signal the presence of a sound to the central nervous system (Spoendlin 1967). In stark contrast, the outer hair cells receive relatively little afferent innervation, but are the targets of numerous efferent nerve fibers that run from the brainstem to the organ of Corti (see Guinan 1996 for a review). The main function of the outer hair cells appears to be to select, amplify, and compress the acoustic signals that actually reach the inner hair cells. They do this by affecting the mechanics of the cochlear partition.

Despite the clear division of labor that exists between the two types of hair cells in the mammalian cochlea, the mechanism by which each cell transduces a sound is essentially the same. Each hair cell possesses many tens, or possibly hundreds, of mechanically sensitive transducer channels at the tips of a highly specialized bundle of "hairs" or stereocilia. The stereocilia project from the uppermost surfaces of the hair cells and respond to stimuli that deflect their tips sideways in the plane in Figure 2.1*C* (cf. Hudspeth and Corey 1977). The conversion of a sound into such a stimulus is a result of the peculiar architecture of the organ of Corti: upward (downward) motion of the basilar membrane causes the innermost parts of the organ of Corti to rotate about their attachment to the osseous spiral lamina and leads to inward (outward) motion of the bases of the stereocilia with respect to the overlying tectorial membrane (cf. ter Kuile 1900).

The amount of movement that is necessary to change the electrical properties of the transducer channels in the stereocilia of each hair cell is miniscule; displacements of less than ±100 nm are sufficient to change almost all of the channels in a given hair cell from an open, highly conductive state into a closed, nonconductive state (see Kros 1996). The state changes of the channels alter the electrical conductivity of the hair cells and give rise to sound-evoked receptor potentials between the inside and outside of the hair cells (Russell and Sellick 1978; Dallos 1985). In the case of the inner hair cells, the receptor potentials can trigger or modulate the release of chemical neurotransmitters onto the innervating fibers of the auditory nerve (Palmer and Russell 1986; Siegel 1992). In the case of the outer hair cells, the receptor potentials affect the mechanical properties of the basolateral wall of the cells, leading to changes in the length of the cells (Brownell et al. 1985; Ashmore 1987; Zheng et al. 2000) and the mechanical properties of the entire cochlear partition (see de Boer 1996; Holley 1996; Patuzzi 1996 for reviews).

The mechanical sensitivity of the hair cells in the mammalian cochlea appears to have evolved at a considerable cost: as well as being remarkably sensitive, the transduction mechanisms are incredibly fragile and highly nonlinear. The fragility of the mechanism is thought to be the source of many forms of hearing loss (see Patuzzi et al. 1989) and can lead to all sorts of problems in experimental studies of cochlear function. The nonlinearity is a consequence of the molecular nature of the transducer (see Kros 1996 for a review) and is thought to be the source of almost all of the compression that occurs in the cochlea (see Section 2.2).

2. Compression in the Mechanics of the Cochlea

2.1 Observations

Direct observations of the cochlea's mechanical responses to sound have provided the most detailed, wide-ranging, and readily interpretable infor-

mation that is available to date on the intensity-coding performance of the peripheral auditory system. These observations have been made by numerous investigators using numerous techniques over numerous years, and it is not the intention of this chapter to review them all. The rationale behind the exclusion of many observations is quite straightforward: direct observations made in early mechanical studies revealed very little about the way in which we now believe a normal cochlea processes the intensity of a sound because they were made using techniques that caused too much trauma to the extremely vulnerable cochlea. The first observations to have overcome the problems of trauma were only provided in the 1970s, and these were neither confirmed nor commonly accepted until the early 1980s (for historical reviews, see Patuzzi 1996; Robles and Ruggero 2001). Most of the observations to be described in this section have hence been made in the last two decades.

There are two major limitations to all of the direct observations of cochlear mechanics that have been made to date. The first is that the observations have been limited to only a few locations within the cochlea. This is a straightforward consequence of the peculiar anatomy of the cochlea. Instead of having their fluid-filled ducts arranged in a straight line, as shown in Figure 2.1*A*, real mammalian cochleae consist of tightly coiled helices that are at least partially embedded in one of the hardest bones of the body (cf. Fig. 2.1*B*). Making observations from even the most accessible parts of the cochlea therefore involves fairly major surgery, which often leads to trauma and pathophysiological conditions. The way that most investigators get around this problem is to minimize the amount of surgery and to perform it with extreme caution. The parts of the cochlea that can be accessed under these conditions include (1) those parts that can be seen through the round window at the very base of the cochlea (this is commonly referred to as the hook region of the cochlea) and (2) those parts that can have holes drilled into them without upsetting the rest of the cochlea too much. There are basically two of these, as illustrated in Figure 2.1*B*: one is partway around the first or most basal turn of the cochlea, and the other is partway around the last or most apical turn. Because the surgery that is necessary to expose even these parts of the cochlea is so invasive, all of the measurements that have been made to date have had to be made in deeply anesthetized animals. The anesthesia is thought to have little effect on the performance of the peripheral auditory system, however, and there are good reasons to believe that the best measurements made to date reflect almost "normal" mammalian hearing. Perhaps the most important of these reasons is that the best mechanical measurements have been supported by careful physiological controls. These show that it is possible to make mechanical measurements in cochleae that have undergone little or no change in their sensitivity to sound. And as we shall see later, the sensitivity of the cochlea is tightly linked to its amplification and compression of a sound.

The second limitation on direct measurements of cochlear function is that they invariably involve making holes into the cochlea and usually require the placement of foreign bodies (e.g., radioactive sources or tiny light-reflecting mirrors) on the cochlear partition. The effects of making these holes and introducing these objects have been considered in both theoretical and practical terms. At present, the consensus view appears to be that holes made into the basal-turn scala tympani have little or no effect on cochlear function, whereas those made into the scala vestibuli (to access the apical turn of the cochlea, for example) can alter the tuning of the cochlear partition quite dramatically (e.g., see de Boer 1991; Ulfendahl et al. 1991; Cooper and Rhode 1996a). The foreign bodies are thought to have little influence on the results in most modern studies (e.g., see Sellick et al. 1983; Cooper 1999a), although they remain a cause for concern in some instances (e.g., see Khanna et al. 1998).

2.1.1 Input-Output Functions

Perhaps the most straightforward way to illustrate the compression that occurs in the mechanics of the cochlea is to consider the variations in the amplitude of a mechanical response that occur when the intensity of a sound is varied. These variations are best illustrated in the form of input-output functions such as those shown in Figure 2.2.

Each of the curves in Figure 2.2 shows the amplitude of a mechanical response to a sinusoidal stimulus (a tone) as a function of the intensity of that stimulus (measured in dB SPL; i.e., dB re: 20 μPa). The different curves in each panel of Figure 2.2 illustrate the responses to different frequencies of stimulation (relatively low-frequency data are plotted with open symbols and dashed lines and relatively high-frequency ones with solid symbols and solid lines), and the different panels illustrate responses observed in different species or in different regions of these species, cochleae. It should be evident immediately that there is a fair amount of variation in the observations made to date. Nonetheless, closer inspection of the data reveals many similarities across both species and region (or the "place" along the length of the partition).

If we concentrate first on the data that were collected at the most sensitive or characteristic frequency (CF; as indicated by the solid circles and thick lines in Fig. 2.2) of each preparation, for example, we can see a fairly consistent pattern, varying only in quantitative detail from one location to another. All but one of the thick CF curves in Figure 2.2 begins by growing almost linearly (i.e., at a rate of 1 dB/dB) with increasing intensity and then gradually bends over to grow at a rate of less than 1 dB/dB. That is, the CF responses become compressively nonlinear once either the sound level or the response amplitude grows above some criterion level or breakpoint. The breakpoints for the CF data are marked with arrows in Figure 2.2*A*–*E*. In some cases, the CF curves show a tendency to bend over once again at even

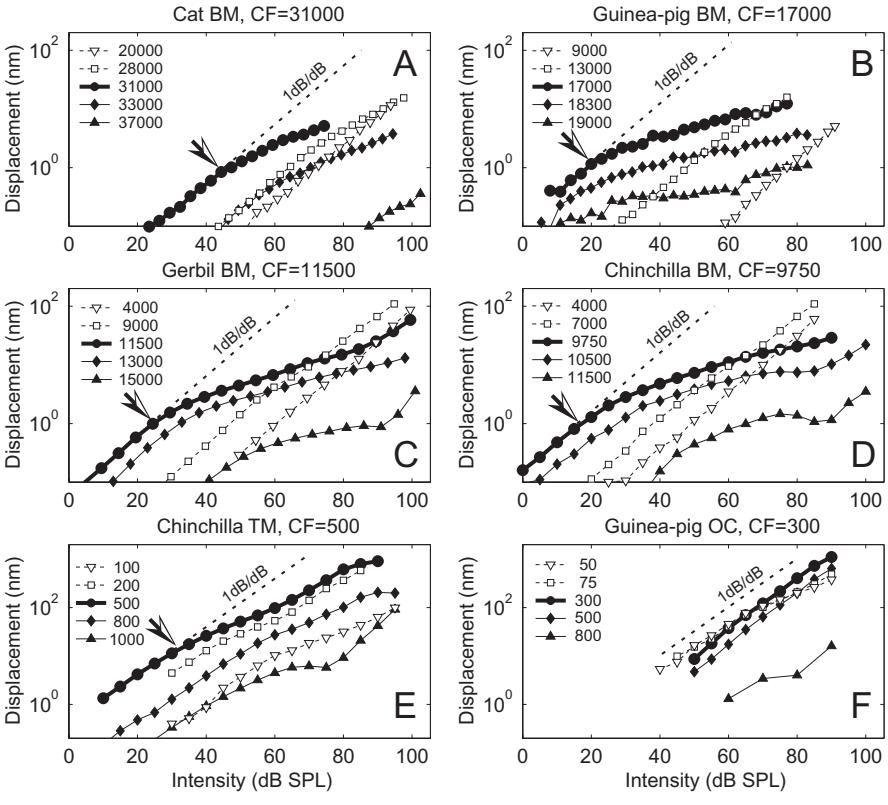


FIGURE 2.2. Mechanical input-output functions in various species and regions of the cochlea. *A*: from the basilar membrane (BM) at the extreme base of the cat cochlea. A–F, keys specify stimulus frequencies in Hertz. Characteristic frequency (CF) data are shown by bold lines. Dashed lines indicate growth rates of 1 dB/dB, and arrows indicate compression thresholds or breakpoints. (Reprinted with permission from Cooper and Rhode 1992. Copyright © 1992 World Scientific Publishing Co. Pte. Ltd.) *B*: from the BM in the basal turn of the guinea pig cochlea. (Reprinted with permission from Nuttall and Dolan 1996. Copyright © 1996 Acoustical Society of America.) *C*: from the BM in the basal turn of the gerbil cochlea. (Data are from Cooper 2000.) *D*: from the BM in the basal turn of the chinchilla cochlea. (Reprinted with permission from Rhode and Recio 2000. Copyright © 2000 Acoustical Society of America.) *E*: from the tectorial membrane (TM) in the apical turn of the chinchilla cochlea. (Reprinted with permission from Cooper and Rhode 1997. Copyright © 1997 American Physiological Society.) *F*: from the organ of Corti (OC) in the apical turn of the guinea pig cochlea. (Data are from Cooper and Dong 2001.)

higher intensities, but this time in the opposite direction to that observed at lower intensities (e.g., see Fig. 2.2*C*,*E*). That is, the slopes of some of the CF curves tend to increase back toward 1 dB/dB at the highest intensities shown (cf. Johnstone et al. 1986; Ruggero et al. 1997).

At this point, it may be worth considering the relationship between compressive nonlinearity and compression as defined in Section 2.1. Not surprisingly, compression is a straightforward consequence of compressive nonlinearity: response growth rates of less than 1 dB/dB necessarily imply that a greater range of input signals are being processed to fit into a smaller amount of "output space" than would otherwise be possible. We can illustrate this by taking the CF input-output function in Figure 2.2*C* as an example. In this case, the input signals are quantified in terms of SPL, and the "output space" comes in the form of basilar membrane displacements. The CF responses in Figure 2.2*C* grow from around 3 nm at 30 dB SPL to just less than 30 nm at 90 dB SPL. Hence, a 60-dB range of input signals is being compressed into just less than a 20-dB (i.e., 10-fold) range of outputs. Both the degree (i.e., the shallow slope) and the wide dynamic range of this compression appear to be essential for normal sound processing in the cochlea.

The absolute intensities at which the breakpoints between the linear and compressively nonlinear sections of the mechanical input-output functions occur vary considerably from species to species and from place to place in a given species (see arrows in Fig. 2.2). However, the basic pattern of the curves is the same in all cases. The only matter of real debate at present is how the compressive regions of the input-output functions relate to the overall, either physiologically or psychophysically relevant, intensity ranges. Arguably, the best mechanical data that are available to date imply that the lowest intensity breakpoints occur within 5 dB of the threshold of hearing (Nuttall and Dolan 1996; Ruggero et al. 1997; see Robles and Ruggero 2001). However, It should be noted that these implications are based on data from only two research laboratories, and that these laboratories make observations in just one region of the cochlea. The majority of observations made elsewhere suggest that the lowest intensity breakpoints occur around 20–25 dB above the threshold of hearing (this finding also proves to be more consistent with the implications that can be drawn from studies in the auditory nerve, as will be discussed later). There is a slightly better consensus regarding the highest intensity limit of the CF compression. Compression that extends to at least 100 dB SPL has been observed in at least five laboratories, and definitely appears to be a normal feature of hearing (Ruggero et al. 1997; Cooper 1998; de Boer and Nuttall 2000; Rhode and Recio 2000; Ren and Nuttall 2001; see Robles and Ruggero 2001 for a review).

If we now look at the input-output functions for a given cochlear location as a function of frequency relative to the location's CF, several other generalizations can be drawn from the data in Figure 2.2. In all but one case, for example, the curves relating to the below CF responses (the dashed lines

with open symbols) exhibit less nonlinearity than is seen in the CF responses. That is, the below CF curves do not bend over either as much or at as low an intensity as the CF curves do. In fact, when the stimulus frequency falls sufficiently far below the CF of each site, most of the input-output functions become linear (i.e., grow at rates of 1 dB/dB) across the entire range of intensities observed. The only exceptions to this finding are seen in the data in Figure 2.2*E*,*F*, which come from the relatively little-studied apical turn of the cochlea (these data are noteworthy in many respects, and are discussed further in Section 2.1.2).

In contrast to the case for below CF responses, many of the input-output functions for tones presented just above the CF of each preparation show just as much, if not more, compression than is evident in the CF curves. In some cases, the above CF responses actually decrease in magnitude as the stimulus intensity increases, at least over a limited range of intensities (e.g., Fig. 2.2*D*,*E*). Some of the above CF input-output functions show a distinct transition from compressive to near linear or perhaps even slightly expansive behavior above around 80–90 dB SPL (cf. Fig. 2.2*C*–*F*, solid triangles). However, these transitions seem to depend on the exact frequencies tested, with responses at other frequencies exhibiting compression to the highest intensities tested (cf. Fig. 2.2B,*C*,*E*, solid diamonds). These characteristics resemble those expected of a system in which the responses originate from multiple pathways, as will be described in Section 2.2.

A final response region, which is not illustrated in the data in Figure 2.2, is observed at frequencies well above the CF of each recording site. This region is known as the plateau region. The phases of the responses in this region show very little dependence on frequency. The amplitudes of the responses in the plateau region are very small, but they grow almost linearly (i.e., at a rate of 1 dB/dB) with increasing intensity. As discussed later, there is some debate over exactly what happens in the transition region between the frequencies illustrated in Figure 2.2 and those in the plateau region proper, but this debate is probably of little importance (see Section 2.1.2).

2.1.2 Quantitative Measurement of Compression

The input-output functions in Section 2.1.1 illustrate the compression that occurs in cochlear mechanics in a very straightforward way, but they do not quantify this compression directly. To quantify the compression, the data have to be transformed into rate-of-growth functions. This could be achieved by simply differentiating the input-output functions. However, more information can be revealed to the human eye if the rates of growth are plotted as a function of stimulus frequency at each observation site. Examples of such plots are shown in Figures 2.3*C,D* and 2.4*C,D*, along with another series of readily interpretable plots (Figs. 2.3*A*,B and 2.4*A,B*) that show the tuning characteristics of the sites under study.

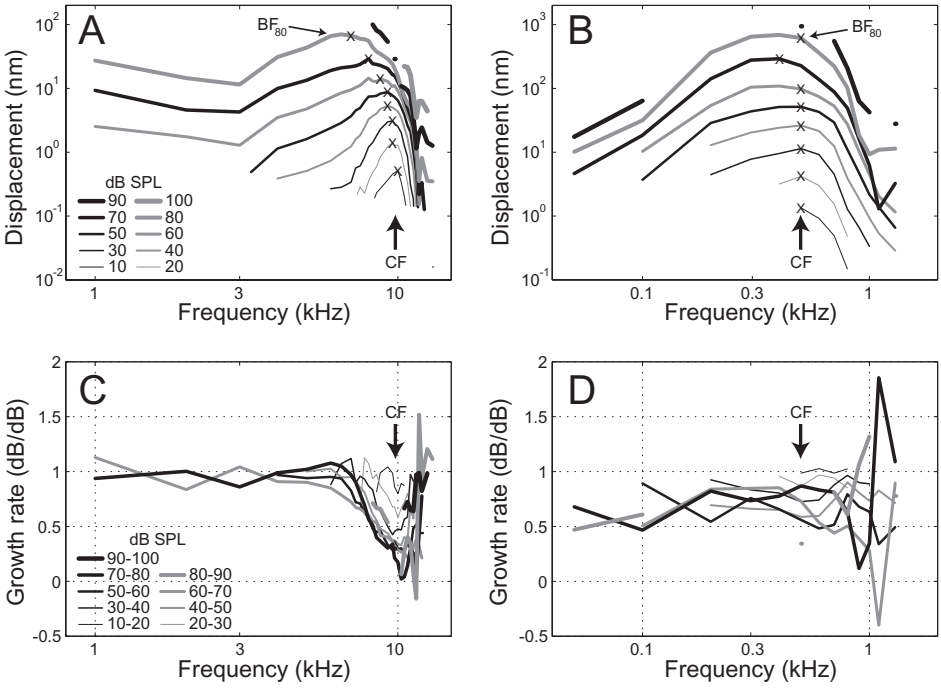


FIGURE 2.3. Tuning and compression characteristics at 2 sites in the chinchilla cochlea. *A* and *C*: from the BM in the basal turn of the cochlea. (Adapted with permission from Rhode and Recio 2000. Copyright © 2000 Acoustical Society of America.) *B* and *D*: from the TM in the apical turn of the cochlea. (Adapted with permission from Cooper and Rhode 1997. Copyright © 1997 American Physiological Society.) *A* and *B*: tuning characteristics of each site shown by plotting the amplitudes of the vibrations evoked by various frequencies at intensities of between 10 and 100 dB SPL. ×: Best frequencies (BFs) for each site as defined by the stimuli evoking the largest vibration velocity (not displacement) at each intensity. (e.g., BF_{80}) *C* and *D*: quantification of the degree of compression that is evident across different intensity ranges (e.g., 50–60 dB SPL) at each site. Strong compression is indicated by low (i.e., closer to zero) growth rates, whereas linearity (i.e., the complete absence of compression) is indicated by growth rates of 1 dB/dB.

The data in Figures 2.3 and 2.4 were selected to illustrate the differences in both tuning and compression characteristics between sites in the two halves of the cochlea (sites near the base of the cochlea and sites near the apex). The data come from studies of the chinchilla and guinea pig cochlea. These are the most commonly used species for cochlear mechanics studies and are the only species where modern in vivo measurements have been made at both ends of the cochlea. In each case, there are notable differences between characteristics observed at the base and the apex of the

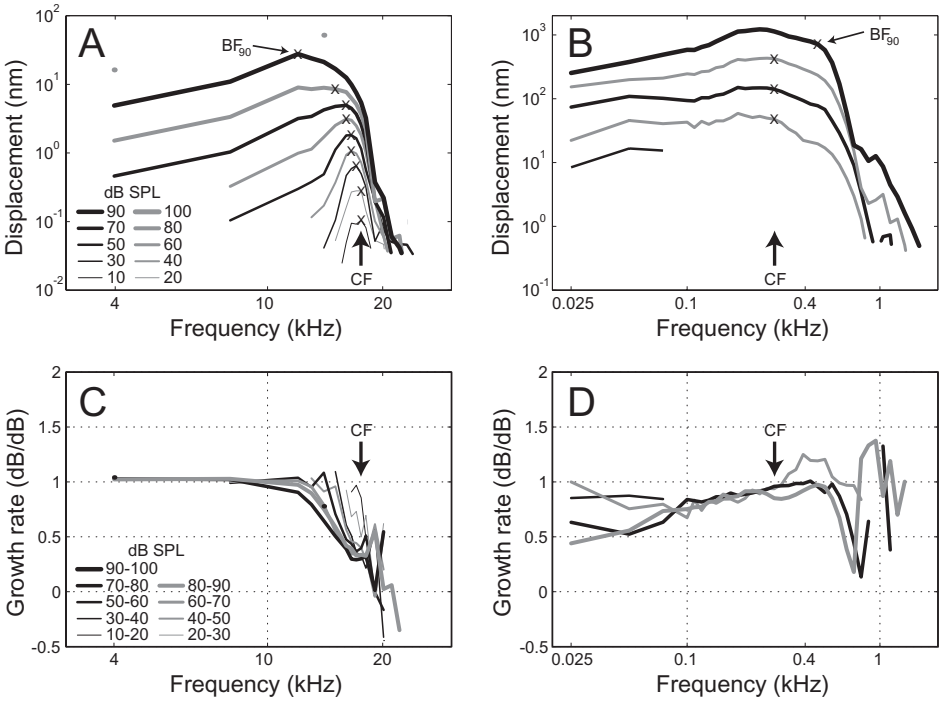


FIGURE 2.4. Tuning and compression characteristics at 2 sites in the guinea pig cochlea. *A* and *C*: from the BM in the basal turn of the cochlea. (Adapted with permission from Cooper 1998. Copyright © 1998 The Physiological Society. [London]) *B* and *D*: from the organ of Corti in the apical turn of the cochlea. (Data are from Cooper and Dong 2001.)

cochlea; in general terms, the basal sites, which are tuned to high frequencies, reveal much stronger, much more extensive (i.e., covering a wider range of intensities), and much more sharply tuned compression than the apical sites. There is some concern about the validity of all the apical-turn data observed to date, however, so comparisons between the apical- and basal-turn characteristics must be made with caution. This is particularly so in the case of the guinea pig data in Figure 2.4*B*,*D*, where the signs of compression are very weak indeed. In the case of the apical-turn chinchilla data, the signs of compression are sufficiently strong and sufficiently widespread across frequency, that even if the amount of compression has been underestimated severely (by compromising the physiological condition of the cochlea in order to make the observations), the pattern of the compression can still be seen to differ from that in the more basal parts of the cochlea.

The data in Figures 2.3*C* and 2.4*C* illustrate both the frequency and intensity dependence of the compression that occurs in the basal regions of the

mammalian cochlea. The compression is indicated by a response growth rate of less than 1 dB/dB, and only becomes evident above a frequency that is around half an octave below the CF for the location under study (e.g., above approximately 6 kHz in Fig. 2.3*C* and approximately 12 kHz in Fig. 2.4*C*). The amount of compression generally increases with increasing frequency, albeit in an intensity-dependent manner, up to a point that is around one-quarter of an octave above the CF (e.g., approximately 12 kHz in Fig. 2. 3*C* and approximately 20 kHz in Fig. 2.4*C*). What happens beyond this point is a matter of some debate, but perhaps little importance. The reason for the debate is that different investigators find different results: some find that growth rates return quite rapidly to 1 dB/dB (e.g., Ruggero et al. 1997) as the response plateau region (cf. Section 2.1.1) is approached, whereas others find areas where the growth rates exceed 1 dB/dB (i.e., there is evidence of expansive nonlinearity) before they return to 1 dB/dB (e.g., Cooper and Rhode 1992; Rhode and Recio 2000). Whether this discrepancy is real, or whether it is simply a consequence of the different frequency resolutions that have been used in the different laboratories, remains to be seen. The reasons that any debate over "who is right" here can be viewed as unimportant in any case are that (1) the magnitudes of the mechanical displacements that are being considered are miniscule, and (2) the responses in question do not seem to be passed on to subsequent stages in the auditory periphery (cf. Cooper and Rhode 1996a; Narayan et al. 1998; see Robles and Ruggero 2001 for a review). The latter point raises an interesting question for those who study cochlear mechanics, but it is clearly unrelated to the issue of compression.

The data in Figures 2.3*D* and 2.4*D* illustrate the frequency and intensity dependence of the compression that occurs in the apical regions of the mammalian cochlea. In the case of the chinchilla data in Figure 2.3*D*, some compression is evident across the entire frequency range of the measurements, extending from at least two octaves below the CF to at least one octave above the CF. At any given intensity of stimulation, the amount of compression varies only slightly with frequency below approximately 900 Hz. The amount of compression does vary systematically with intensity, however; in general, the compression is nonexistent (the growth rate approximates 1 dB/dB) at low intensities, maximal (around 0.5 dB/dB) at moderate intensities (e.g., between 40 and 60 dB SPL near the CF), and then moderate (between 0.5 and 1 dB/dB) at high intensities. It has been suggested that the amount of compression in the apical turn of the cochlea depends more on the level of the displacements of the partition rather than on the SPL per se (Rhode and Cooper 1996). However, this suggestion is not consistent with observations made in the more basal regions of the cochlea.

In the case of the apical-turn guinea-pig data shown in Figure 2.4*D*, much less compression is evident at most frequencies within about one octave of the CF of the preparation, but there are clear hints of the patterns seen in

both the apex of the chinchilla cochlea and the basal turns of many other cochleae. There is, for example, some compression even at the lowest frequencies tested (again over two octaves below the CF), just as there is in the apex of the chinchilla cochlea. However, little or no compression is evident at the CF of the guinea pig preparations (Cooper and Rhode 1995; Ulfendahl et al. 1996; Zinn et al. 2000). On the high-frequency side of the CF, there is evidence of strong compression (e.g., growth rates of just 0.2 dB/dB) over a very restricted frequency region, but this rapidly turns into a region where both compression and expansion (i.e., growth rates of greater than 1 dB/dB) can be seen at different intensity levels (Cooper and Rhode 1995). This is somewhat akin to the situation seen in the more basal turns of the cochlea.

It is noteworthy that the growth rate of the responses to a few tones around one octave above the CF in Figures 2.3*D* and 2.4*D* is significantly greater than one. That is, there is clear evidence of expansive nonlinearity at sites near the apex of the cochlea (in both chinchillas and guinea pigs). The expansion in the apex cannot be "ignored" like that in the basal turns was because the actual amplitudes of the responses involved are not trivial. The expansion in the apical turns is almost always associated with input-output functions that contain sharp discontinuities, or notches, at a particular intensity (cf. Fig. 2.2*E*). This observation has been used by some investigators (e.g., Cooper and Rhode 1995, 1996a) to suggest that the expansion may be caused by the intensity-dependent interference of a compressively nonlinear process with a linear process. However, other investigators have taken their own observations of expansive nonlinearity as direct evidence for negative feedback in the mechanics of the cochlea (Zinn et al. 2000). These ideas are discussed at greater length in Section 2.2.

2.1.3 Dynamic Aspects of Compression

All of the data considered so far have related to stationary properties of the peripheral auditory system. That is, they have been based on measurements of the performance of the system that were made without regard to the time course of the responses. As will become apparent later (cf. Fig. 2.11), the amount of compression that can be seen at subsequent stages in the periphery (e.g., in the auditory nerve) depends on the temporal as well as the spectral characteristics of a sound or a sound-evoked response. There is very little evidence to suggest that such time-dependent characteristics originate in the mechanics of the cochlea, however. In fact, there is strong evidence to the contrary. This evidence comes from several lines of experimentation.

The first two lines of evidence come from looking at the time courses of the mechanical responses to single tones or click stimuli. The waveforms of the responses to individual tone bursts almost invariably show that the instantaneous responses of the system (measured while the tonal stimuli

are rising or falling in amplitude, for example) undergo compression in just the same way that the steady-state responses do (e.g., see Rhode and Cooper 1996; Ruggero et al. 1997; Rhode and Recio 2000). The individual cycles of the response waveforms are not particularly distorted (Ruggero et al. 1997; Cooper 1998), but their envelopes seem to trace out exactly the same input-output functions as the time-averaged responses discussed in Section 2.1.1.

The waveforms of the cochlear partition's responses to click stimuli are somewhat more complicated than those of its responses to tones, but in essence they show the same characteristics. This is illustrated explicitly in Figure 2.5, which shows two series of click responses from the two halves of the chinchilla cochlea. The peak-to-peak (or, more accurately, peak-to-trough) amplitudes of the individual cycles of the click responses have been plotted against the peak-equivalent intensity level of the click in Figure 2.5*C,D*, and compressive nonlinearity (i.e., growth rates of less than 1 dB/dB) is seen for every cycle. In the apical-turn data in Figure 2.5*D*, the degree of compression is small, but the compression is basically similar regardless of time. This is most probably a consequence of two factors: first, the frequency composition of each cycle in the apical-turn click responses is fairly similar, and second, as we saw in Figs. 2.2*E* and 2.3*D*, some compression occurs at almost all frequencies in the apex of the chinchilla cochlea. In the basal-turn data of Figure 2.5*C*, there are clear and systematic differences between the input-output functions for the individual cycles of the click-evoked responses, but these do not mean that the compression depends on time *per se*. The initial cycles of the basal-turn click responses show much less compression than the later cycles, but the differences turn out to be explicable entirely in terms of the frequency composition of the responses. The earliest cycles of the basal-turn click responses are driven by the lowest frequency components of the stimulus because these components travel along the cochlear partition much faster than the high-frequency components (there is a progressive shift in the frequency of the components that drive the click-evoked responses, from well below the CF of the preparation at the onset of the responses to somewhere very close to the CF after between 3 and 4 cycles; cf. Robles et al. 1976; de Boer and Nuttall 1997; Recio et al. 1998; Recio and Rhode 2000). The earliest cycles of the basal-turn click responses therefore grow at a relatively high rate against stimulus intensity (just as the responses to below CF tones do; cf. Sections 2.1.1 and 2.1.2), whereas the later cycles grow at a much lower rate, exhibiting a greater degree of compression over a much wider dynamic range.

Further lines of evidence that mechanical compression does not depend on time come from the phenomenon of two-tone suppression. This is a phenomenon whereby the sensitivity of the responses to one tone (a probe tone, which is normally placed at the CF of the site under study) can be affected by the presence of a second tone (a suppressor tone, which may be either higher or lower in frequency than the probe tone). Findings from

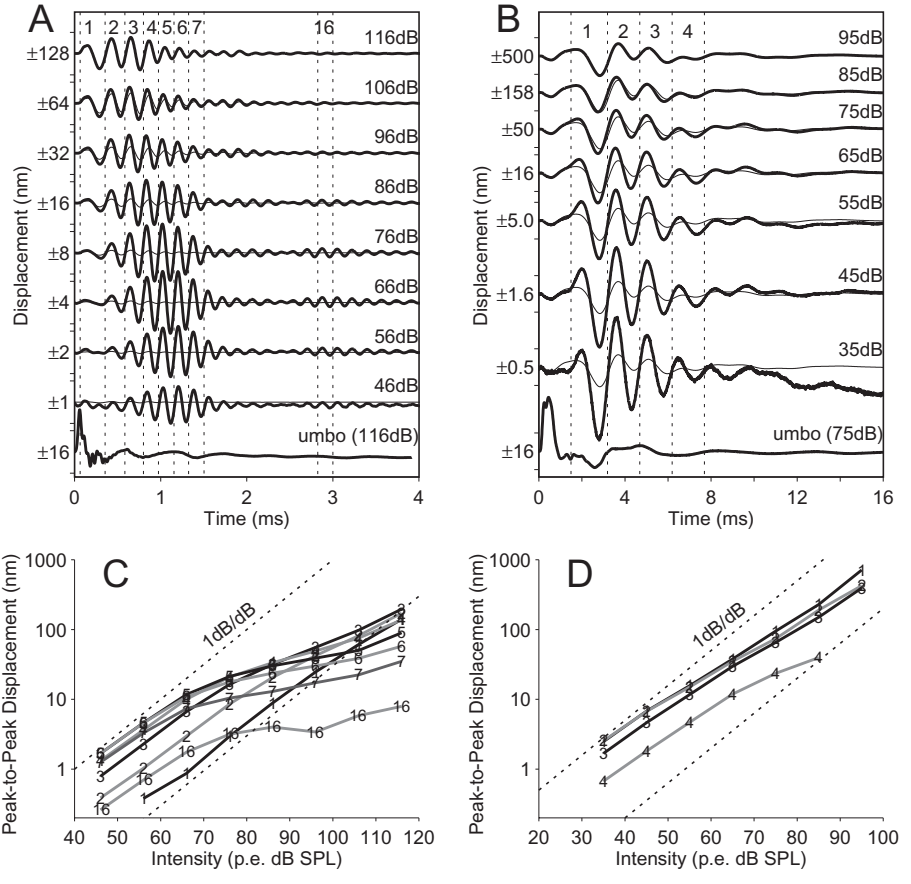


FIGURE 2.5. Compression characteristics revealed in the responses of the BM to click stimuli at 2 sites in the chinchilla cochlea. *A* and *C*: basal-turn data from Recio and Rhode (2000). (Reprinted with permission. Copyright © 2000 Acoustical Society of America.) *B* and *D*: apical-turn data from Cooper and Rhode (1996a). (Reprinted with permission. Copyright © 1996 Taylor & Francis Ltd [http://www.tandf.co.uk/journals].) *A* and *B*: BM waveforms in response to 10- and 100-μs condensation clicks, respectively, at the peak-equivalent (p.e.) intensities indicated. Bold lines: responses of the BM; thin lines: response patterns predicted by linear extrapolation from the highest level responses. The waveforms of the mechanical stimuli to the cochleae, as recorded from the malleus in the middle ear, are shown in the lowermost traces (umbo). *C* and *D*: input-output functions for the individual cycles of the click responses, with numbers indicating the different time periods of the responses illustrated in *A* and *B*, respectively. Dashed lines indicate growth rates of 1 dB/dB.

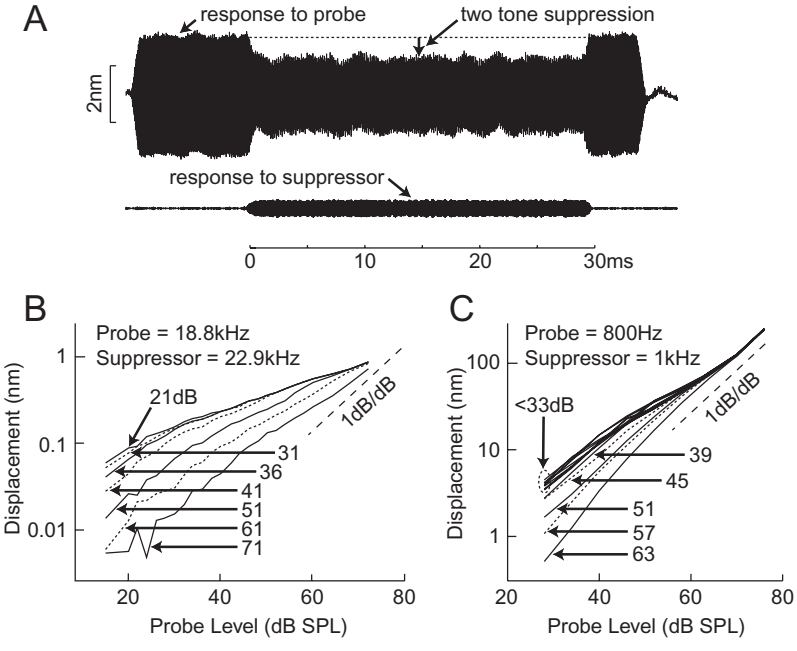


FIGURE 2.6. Two-tone suppression in the mechanics of the cochlea. *A*: responses of the BM to a 2-tone stimulus in the hook region of the guinea pig cochlea illustrating the rapid decrease in response to 1 tone (the probe; a 45-ms 26-kHz tone at 56 dB SPL) during the presentation of a second tone (the suppressor; a 30-ms 32-kHz tone at 76 dB SPL). (Data reprinted with permission from Cooper 1996. Copyright © 1996 Acoustical Society of America.) *B* and *C*: input-output functions for CF probe tones in the presence of suppressor tones at various intensities (in dB SPL). Increasing amounts of suppression, caused by increases in the intensity of the suppressor, are seen to decrease both the degree (i.e., the slope) and the extent (i.e., the dynamic range) of the compression in the probe tone responses. Dashed lines: growth rates of 1 dB/dB. *B*: basal-turn guinea pig data from Nuttall and Dolan (1993). (Reprinted with permission. Copyright © 1993 Acoustical Society of America.) *C*: apical-turn chinchilla data from Cooper and Rhode (1996b). (Reprinted with permission. Copyright © 1996 Taylor & Francis Ltd. [http://www.tandf.co.uk/journals].).

three studies of this phenomenon are shown in Figure 2.6. The amount of two-tone suppression that is observed depends on many factors, including both the frequency and intensity of the probe and the suppressor tones (Robles et al. 1991; Nuttall and Dolan 1993; Rhode and Cooper 1993; Cooper 1996; Cooper and Rhode 1996b; Geisler and Nuttall 1997; Rhode and Recio 2001a). In general, only probes that fall within the frequency region where compression can be observed (cf. Figs. 2.3 and 2.4) are susceptible to two-tone suppression, and the amount of suppression that can

be achieved increases with decreasing probe intensity (at least up to a point). This latter finding has a very simple consequence as far as compression is concerned: two-tone suppression leads to decreases in both the degree (i.e., the slope) and the dynamic range of the compression that can be observed in the mechanics of the cochlea. This finding is well illustrated in the data in Figure 2.6*B*,*C*. In the basal-turn data in Figure 2.6*B*, for example, the unsuppressed responses to probe tones ranging between 20 and 70 dB SPL are compressed into a range of displacements that spans from about 0.1 nm to just less than 1 nm (so an input range of 50 dB is compressed into an "output space" of less than 20 dB). In the presence of a 71-dB SPL suppressor tone, however, the same range of response amplitudes is evoked by probe tones ranging from 46- to 72-dB SPL (so an input range of only 26 dB occupies the same, slightly less than 20-dB, range of "output space"). In other words, the presence of the 71-dB suppressor roughly halves the amount of compression seen in the probe tone responses. The apical-turn data in Figure 2.6*C* illustrate a similar pattern, albeit with a lower amount of compression (both in degree and in range) to start with.

As far as we have been able to resolve to date, the time course of the two-tone suppression that is seen in the mechanics of the cochlea is instantaneous (cf. Fig. 2.6*A*; Rhode and Cooper 1993; Cooper 1996). The amount of two-tone suppression even varies periodically within the individual cycles of sufficiently low-frequency suppressor tones (cf. Patuzzi et al. 1984; Ruggero et al. 1992; Rhode and Cooper 1993; Cooper 1996; Geisler and Nuttall 1997). Once again, it should be emphasized that despite the great speed of its effects, two-tone suppression is not achieved by distorting the individual cycles of the responses to the probe tones. Two-tone distortion does exist in cochlear mechanics (cf. Robles et al. 1991, 1997; Cooper and Rhode 1997), but it is not the cause of the instantaneous compression. The cochlea appears to act more in the manner of an automatic gain control system than an instantaneous compressor, but it retains an ability to change its state almost instantaneously.

2.1.4 Other Factors That Influence Compression

The characteristics of the mechanical compression discussed in Sections 2.1.1–2.1.3 are thought to reflect those that occur in normal cochleae in normal-hearing animals. There are numerous factors that can affect this compression, including the effects of ototoxic drugs (Ruggero and Rich 1991; Murugasu and Russell 1995; Ruggero et al. 1996b), noise (Ruggero et al. 1996a), and feedback from the central nervous system (via the fibers of the olivocochlear efferent system; cf. Murugasu and Russell 1996; Dolan et al. 1997; Russell and Murugasu 1997; Cooper and Guinan 2002). At the systems level, almost all these factors are thought to act in similar ways (as explained in Section 2.2) and almost all lead to decreases in the amount of

compression seen. In general terms, anything that affects the function of the outer hair cells in the organ of Corti also appears to affect how much compression can be seen in the mechanics of the cochlea.

The only treatment that has been seen to increase the amount of compression in the mechanics of the cochlea is the application of direct electrical current into the cochlea (Nuttall et al. 1995). Injecting negative current was found to decrease both the sensitivity and the amount of compression observed on the basilar membrane, in much the same way as other manipulations do. However, injecting positive current increased the sensitivity of the responses of the basilar membrane to low-level tones as well as increasing both the degree and extent of the compression that could be seen at low-to-moderate sound levels. The possibility that the positive currents were merely reenabling some "normal" compression that had been lost during the preparation of the electrical injection experiments cannot be ruled out, of course.

2.2 Interpretation

Almost all the observations of compression in the mechanics of the cochlea can be interpreted very simply in the terms of a nonlinear, positive feedback model. This type of model was introduced to the auditory field by Zwicker in the late 1970s, well before the most definitive mechanical measurements were made. The model has been refined ever since, particularly in terms of its anatomical and physiological interpretation.

One highly simplified version of Zwicker's model is shown, along with a few of its predictions, in Figure 2.7. The basilar membrane and the outer hair cells of the organ of Corti are considered to form a positive feedback loop, with the membrane serving to stimulate the hair cells, and the hair cells, in turn, acting to enhance or amplify the movements of the membrane (Zwicker 1979, 1986; Patuzzi et al. 1989).

As far as the ability to explain compression is concerned, the key feature of Zwicker's model is the inclusion of a saturating nonlinearity in the feedback loop. This limits the amount of feedback that can be provided as the level of stimulation increases, and hence limits the amount of amplification that can be achieved by the feedback; the stronger the input to the loop, the stronger the limitation at the output of the nonlinearity and the weaker the feedback becomes relative to the input. The precise form of the nonlinearity is not essential to the overall performance of the model, and even a straightforward clipping device can provide a reasonably good fit to most experimental data (cf. Zwicker 1986). In order to make the model more realistic from a physiological point of view, however, several researchers have used either a first- or second-order Boltzman function as the basis of the nonlinearity (cf. Patuzzi et al. 1989; Nobili and Mammano 1996; Cooper 1998). The reason that this is viewed as physiologically realistic is that the most likely correlate of the nonlinearity in the real cochlea is the displacement-

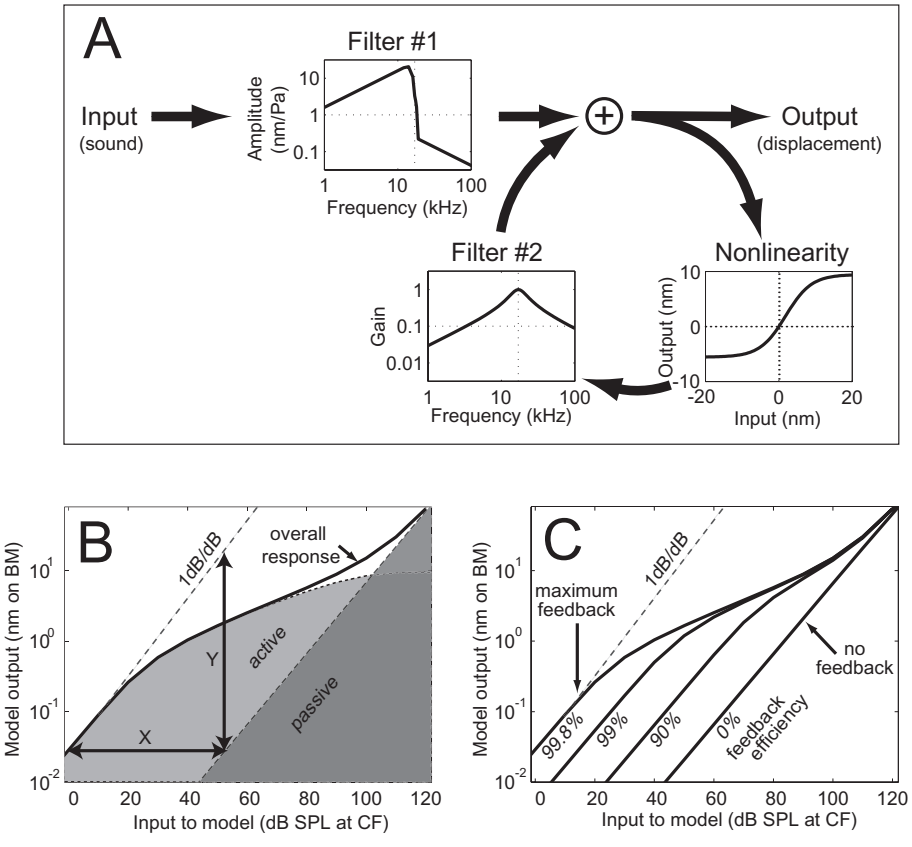


FIGURE 2.7. Modeling compression as the result of a nonlinear, positive feedback system. *A*: positive feedback system containing 2 filters and a saturating nonlinearity. This type of system can replicate many of the response features observed in cochlear mechanics and is thought to be realistic from a physiological point of view (see text). The particular configuration of the model shown here was developed to mimic BM motion in the 17-kHz region of the guinea pig cochlea (Adapted with permission from Cooper 1998. Copyright © 1998 The Physiological Society. [London]) *B*: separation of a model input-output function (bold line) into its active and passive components (i.e., into its feedback and nonfeedback components, respectively; cf. dashed lines). X and Y: alternative measurements of the maximal gain of the feedback loop. *C*: effects of varying the efficiency of the feedback loop in the model (cf. Yates 1990). The maximum feedback efficiency of 99.8% was selected to match the data from Cooper (1998). Dashed lines: growth rate of 1 dB/dB.

to-current conversion that occurs via the transducer channels in the hair cells (cf. Section 1.2; also see Kros 1996; Patuzzi 1996 for reviews).

The two filters of the model illustrated in Figure 2.7*A* serve very different purposes in a signal-processing sense, but they may both be manifesta-

tions of a single filter in the real cochlea (cf. Zwicker 1986). Filter number 1 is intended to represent the effects of the distributed, hydromechanical filtering that occurs during the delivery of a sound to a particular cross section of the cochlear partition. This filter is commonly considered to take the form of a transmission line, with the cochlear fluid acting to transport energy from one cross section of the partition to another (Zwislocki 1948; Peterson and Bogert 1950; Zwicker 1986; see Hubbard and Mountain 1993; de Boer 1996 for reviews). Filter number 2 is intended to represent the effects of the "local" mechanical filtering at the site of an individual cross section (the site of the cochlea that is under investigation, where the local hair cells must have their primary effect). In terms of the compression that we are interested in, the primary function of filter number 2 is to limit the extent of the feedback in the frequency domain. In other words, filter number 2 acts to tune the compression. Because the feedback loop itself tends to emphasize the effects of filter number 2, even a second-order filter such as the one illustrated in Figure 2.7 can limit the extent of the compression to a one-octave-wide band of frequencies. In order to mimic the more broadly tuned compression that is evident in the apical turns of the cochlea, this filter has to be made less selective than it is in the basal regions.

The type of model outlined above allows the input-output functions observed in the real cochlea to be interpreted in a very simple manner. The initial region of almost linear growth in these functions at low intensities is seen to reflect the behavior of a system working in its maximally active state (that is, with its feedback loop providing the maximum amount of amplification that it can; cf. Fig. 2.7*B*,*C*). The compressive regions of the input-output functions then illustrate the system in a series of progressively less active states, as the saturating nonlinearity becomes more and more effective at limiting the amount of feedback. (Note that the amount of feedback in the model never actually decreases in absolute terms; it simply becomes smaller relative to the input.) The model even predicts the eventual demise of the compression (i.e., the return of the growth rate toward 1 dB/dB) at high stimulus levels. This is not a result of the feedback loop at all: it is a simple, passive consequence of increasing the input to the system. The segregation of a CF input-output function into its active and passive components is illustrated in Figure 2.7*B*. CF input-output functions observed under less favorable conditions, intended to replicate those observed in functionally compromised cochleae, are shown in Figure 2.7*C*. Both the degree (i.e., the slope) and the dynamic range of the compression exhibited in these curves decrease with decreases in the efficiency (or gain) of the feedback loop. The similarity between these curves and those observed in the real cochleae (e.g., under conditions of two-tone suppression, as illustrated in Fig. 2.6) is striking.

The one thing that the simple model in Figure 2.7 cannot explain about the input-output functions observed in the real cochlea is the presence of

expansion as well as, or as opposed to, compression for certain frequencies of stimulation. As hinted in Section 2.1.2, there are at least two ways of modifying the model in order to explain such phenomena. One way is simply to change the sign of the feedback loop. That is, to make the positive feedback into negative feedback. This suggestion has been put forward by at least two groups of researchers who studied the apical turns of the cochlea (Khanna and Hao 2000; Zinn et al. 2000), but it is inconsistent with the results observed in the more basal regions of the cochlea. Another, somewhat less radical method is to allow the phase of the input and feedback pathways to vary with respect to one another (cf. Mountain et al. 1983). Phase changes (with frequency) certainly occur in both the active and passive pathways of the real cochlea, as evidenced by the frequency- and/or intensity-dependent phase responses of the basilar membrane in healthy and damaged cochleae (see Robles and Ruggero 2001 for a review). The rates at which these phase changes occur is highest for frequencies just above the CF of a particular site, and so this is where the opportunity for the feedback to become reversed should also be at its highest. Close examination of one of the earliest versions of the feedback model (Zwicker 1986; cf. Zwicker and Peisl 1990), which incorporates the saturating feedback into an electrical transmission line, shows that such effects can indeed result from this type of model.

2.3 Consequences

Perhaps the most obvious consequence of the compression that occurs in the mechanics of the cochlea is the increase in dynamic range that it can offer to subsequent stages in the peripheral auditory system. This topic is expanded in Section 3. The combined frequency and intensity dependencies of the mechanical compression have other consequences as well. These are slightly less obvious than the dynamic range extension, but at least two of them may still be important from the perspective of sound encoding in the auditory periphery. The first is that any amplitude modulation occurring in a sound that undergoes compression will be distorted by the compression. This observation is implicit in our earlier discussion of the temporal features of compression (cf. Section 2.1.3) but has only recently been confirmed by direct measurements at the level of the basilar membrane (Rhode and Recio 2001a,b).

The second consequence of compression is that as the level of a sound changes, so does the spatial pattern of excitation that it produces within the cochlea. This occurs because both the degree (i.e., the slope) and the dynamic range of the compression seen at any one site in the cochlea vary with frequency (cf. Fig. 2.2). The most straightforward consequence of this is a phenomenon known as the half-octave shift, whereby the most effective, or best frequency (BF) at each site shifts downward by around one-half of an octave in most regions of the cochlea with increasing intensity.

By definition,[2] the BF that is observed at low intensities is equal to the CF of the site being studied. Because the CF responses undergo much more compression than the below CF responses, however, the BF at high intensities can be much lower than the CF (particularly clear examples of this are shown in Figs. 2.2*C*,*D*, 2.3*A*, and 2.4*A*). Another way to think about the half-octave shift is to consider the cochlea as having two sets of tuning properties, one passive and one active (cf. filters 1 and 2 in Fig. 2.7). The passive tuning of each site, which will dominate in its responses to high-level tones, is centered around half an octave lower than the active tuning, which dominates the responses of the system to low-level stimuli (cf. Fig. 2.7*B*). The idea that the half-octave shift originated in a spatial shift in the pattern of the excitation within the cochlea has been around for many years (cf. Lonsbury-Martin and Meikle 1978; Davis 1983; Johnstone et al. 1986; Zwicker 1986) but has only recently been confirmed by direct, multiple-point observations (Rhode and Recio 2000). Another recent investigation of the phenomenon actually refuted the existence of a shift in the spatial patterns of basilar membrane excitation (Russell and Nilsen 1997), but the data in that report are substantially noisier than those in Rhode and Recio's (2000) report.

3. Compression in the Auditory Nerve

The response properties of single auditory nerve fibers have been studied extensively over the last four decades and have formed the basis of numerous hypotheses regarding intensity coding in the auditory periphery (see Viemeister 1988; Ruggero 1992 for reviews). The major advantage that the auditory nerve studies have over the direct mechanical studies is that observations can be made from fibers that innervate hair cells at every location along the length of the cochlear partition. Each inner hair cell synapses with many auditory nerve fibers, but each fiber innervates just one inner hair cell (Spoendlin 1967; Liberman 1982). Each nerve fiber therefore provides very selective information about the site that it innervates. Due to the tonotopy of the cochlea (cf. Section 2.1), fibers innervating the basal turns of the cochlea have high CFs and therefore inform us about the processing of relatively high-frequency stimuli, whereas fibers innervating more apical sites have lower CFs and inform us about the processing of lower frequency sounds (Kiang et al. 1965; Liberman 1982). At least in the basal turn of the

[2] The terms CF and BF have not always been used in the sense that they are here. The two terms were originally used synonymously by different groups of researchers working on the auditory nerve (e.g., Kiang et al. 1965; Rose et al. 1967), but it was soon realized that the BF of a nerve fiber (that is, the stimulus frequency that is most effective at driving the fiber at a particular SPL) could deviate considerably from its CF (e.g., see Geisler et al. 1974).

cochlea, which is the only place in which direct comparisons have been performed, the tuning characteristics of each nerve fiber are almost identical to the mechanical tuning characteristics observed at the level of the basilar membrane (Narayan et al. 1998).

The major limitation of the auditory nerve studies is the amount of data that can be collected from each nerve fiber. Single-fiber responses occur in the form of action potentials, or spikes, which have a characteristic "all-or-none" waveform; at any one point in time, each nerve fiber either can fire an action potential, or it cannot (i.e., it never fires one-half of an action potential). Auditory nerve fiber responses are also stochastic: it is only the probability of action potential discharge that is modulated by the presence or absence of a sound. The maximum rate at which each fiber can discharge action potentials is limited (by refractoriness) to around 1,000 spikes per second (sp/s) in very short bursts and (by adaptation) to around 100 sp/s over longer periods (e.g., greater than 100 ms; cf. Fig. 2.11). As a consequence, in order to build a clear picture of the response characteristics of a single fiber in an experiment, the responses have to be averaged over many tens or even hundreds of presentations of a single stimulus. The amount of time necessary to complete this process can often be longer than the period over which electrical contact with the nerve fiber can be maintained, and so the amount of data that can be collected in each experiment is limited.

3.1 Observations

Auditory nerve fibers respond to sound in two ways: they synchronize their discharge patterns to the individual cycles of a sound wave, especially at low frequencies, and they vary their average discharge rates. Even in the absence of intentional acoustic stimulation, the fibers can discharge action potentials spontaneously. The spontaneous discharges occur at completely random times, and individual fibers differ in their average rates of spontaneous activity. There is a bimodal distribution of spontaneous discharge rates across the entire population of auditory nerve fibers, with between 10 and 40% of fibers having spontaneous rates less than approximately 18 sp/s and between 60 and 90% having rates greater than approximately 18 sp/s (Kiang et al. 1965; Kim et al. 1990). The lower-rate subpopulation of fibers can be conveniently divided into two further subgroups using a second average-rate criterion of around 0.5 sp/s (Liberman 1978), yielding a total of three subpopulations with low, medium, and high spontaneous rates. These subpopulations may have quite distinct underlying morphologies (see Liberman 1980), but they are usually considered as part of a continuum in a physiological sense. Most importantly, there is a strong correlation between the spontaneous rate of a fiber and its threshold to acoustic stimulation: low, medium, and high spontaneous rate fibers have high, medium, and low acoustic thresholds, respectively (Liberman 1978; Winter et al. 1990; Yates 1991). This correlation exists regardless of exactly how the

thresholds of the fibers are defined (i.e., using synchrony or average discharge rates as a metric).

There are various ways to quantify the degree to which the fibers synchronize with a sound (see Rose et al. 1967; Goldberg and Brown 1969; Kim et al. 1990), but almost all show the fine-timing patterns of the individual fibers to have a very limited dynamic range (Johnson 1980; cf. Viemeister 1988). The fibers begin to synchronize at SPLs as much as 20 dB below those needed to increase the average rate of discharge, and the degree of their synchrony rises to a maximum over a dynamic range of 20–30 dB (Johnson 1980). Despite this limited dynamic range, the stochastic nature of the responses of an auditory nerve fiber has some interesting consequences in terms of compression. Rather than synchronizing their discharges perfectly to one particular point during each cycle of the waveform of a loud sound, auditory nerve fibers appear to distribute their responses in a graded manner throughout the most effective half-cycle of the stimulus (cf. Rose et al. 1967; Anderson et al. 1971). At least during the most effective half-cycle and at frequencies above 300–500 Hz, the discharge probability of each fiber appears to be scaled in close proportion to the pressure of the (filtered) stimulus. This result is thought to be a consequence of an automatic gain control at the synapse between the nerve fiber and the hair cell that it innervates (cf. Schroeder and Hall 1974). When combined with the increases in the average discharge rate that occur at higher SPLs (cf. Section 3.1.1), this behavior can preserve information about the stimulus waveform over a very wide dynamic range. Response synchronization is not a global phenomenon, however; the maximum degree of synchronization that can be observed decreases rapidly once the frequency of a sound increases above around 1 kHz, and synchronization becomes insignificant above around 4 kHz (Johnson 1980; Palmer and Russell 1986; Joris and Yin 1992). To find evidence of compression across a wider range of frequencies then, we must focus on the average rate of discharge of the fibers.

3.1.1 Input-Output Functions

Auditory nerve fibers generally increase their average rate of discharge as the intensity of a sound is increased (cf. Fig. 2.8*A*). Each fiber has a distinct rate threshold above which its average discharge rate increases slowly at fist, then more rapidly, and finally more slowly again toward high intensities. At the CF of each fiber, the growth characteristics observed at high levels are quite diverse, as illustrated in Figure 2.8*A*. These characteristics have been used to distinguish three subclasses of auditory nerve fiber (cf. Fig. 2.8*A*). Fibers with the lowest rate thresholds (and the highest spontaneous rates) tend to increase their rates of discharge over a very narrow dynamic range before "saturating" at levels around 20–30 dB above their thresholds. Fibers with slightly higher rate thresholds (and, typically, medium spontaneous rates) tend to increase their discharge rate quite

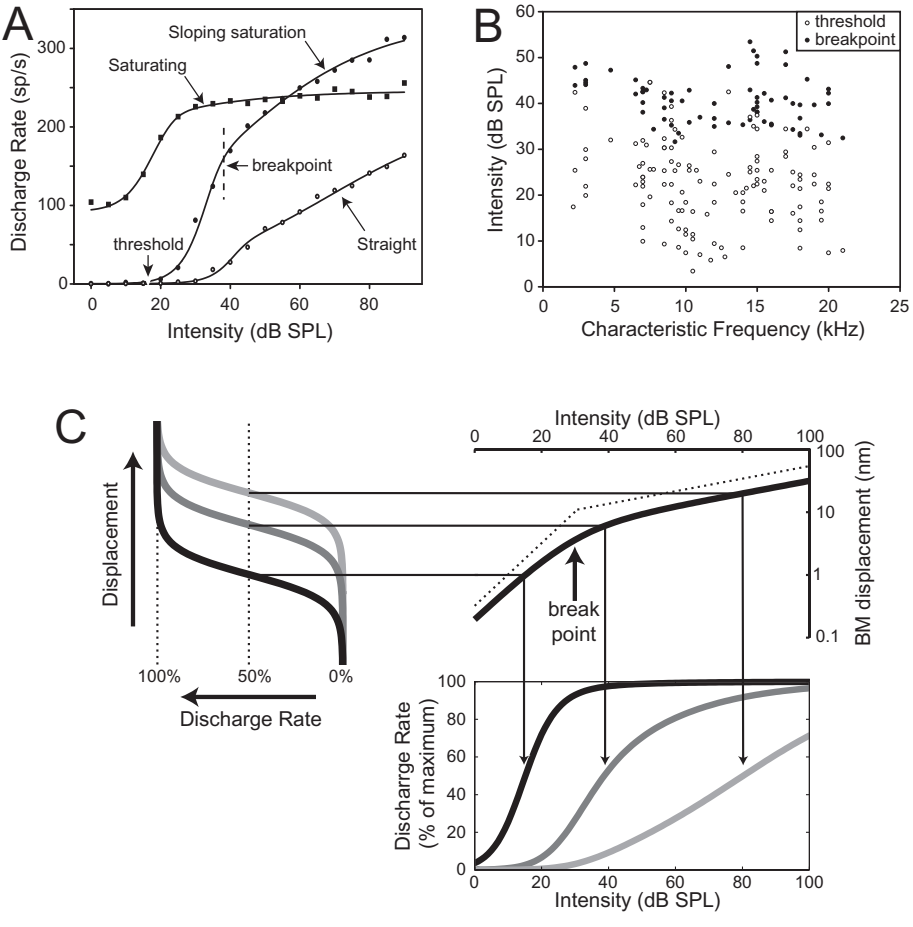


FIGURE 2.8. *A*: rate versus level functions for 3 categories of auditory nerve fiber. Saturating fibers tend to have the lowest thresholds and highest spontaneous rates; sloping saturation fibers tend to have slightly higher thresholds and lower spontaneous rates and exhibit a well-defined breakpoint in their CF rate versus level functions; fibers with straight rate versus level functions have the highest thresholds and lowest spontaneous rates of all. (Guinea pig data reprinted with permission from Müller et al. 1991. Copyright © 1991 Elsevier Science.) *B*: distribution of rate thresholds and compression thresholds (or breakpoints) across a population of auditory nerve fibers in the guinea pig. (Data reprinted with permission from Müller and Robertson 1991a. Copyright © 1991 Elsevier Science.) At any 1 frequency, the compression thresholds (solid circles) are distributed across a much narrower range of intensities than the rate thresholds, suggesting that the origin of the compression is mechanical as opposed to neural (see text). *C*: modeling the 3 categories of auditory nerve rate versus level function by biasing 3 neural sensitivity functions (*top left*) around different set points on a hypothetical BM input-output function (*top right*). The breakpoint of this function is defined by the intersection of its low- and high-level asymptotic slopes of 1.0 and 0.2 dB/dB, respectively (dashed lines). The model is based on the reports of Sachs et al. (1989) and Yates et al. (1990). See text for description.

rapidly over the first 10–15 dB of their dynamic range, but then exhibit a more-or-less clear breakpoint partway up their rate versus level function. Above this breakpoint, these fibers increase their rates of discharge much more slowly with increasing sound pressure. Such fibers have therefore been classified as "sloping-saturation" fibers (Sachs and Abbas 1974; Palmer and Evans 1980; Winter et al. 1990). The third group of fibers exhibits an average discharge rate that grows almost linearly over a very large dynamic range from their relatively high rate thresholds (Winter et al. 1990). These fibers are classified as having "straight" rate versus level functions, and they tend to have the lowest spontaneous rates of all.

The classification of auditory nerve fibers according to the form of their rate versus level functions is somewhat arbitrary because the three groups of fiber have always been thought of as subsets of a continuum (cf. Sachs and Abbas 1974; Evans and Palmer 1980; Sachs et al. 1989; Winter et al. 1990). Indeed, rate versus level functions from all three subclasses can be described quite accurately by a single equation in which the most important parameter is the relative threshold of the fiber (Sachs et al. 1989; Yates 1990; Yates et al. 1990; Müüller et al. 1991). The logic behind this equation is illustrated in Figure 2.8*C*, which shows how a simple saturating input-output function (the sigmoidal curve shown three times in Fig. 2.8*C*, *top left*) can be transformed into any one of the three rate versus level function categories (Fig. 2.8*C*, *bottom right*) simply by preprocessing the driving stimulus (sound pressure) using a second nonlinear function (the compressive nonlinearity of the basilar membrane, as shown in Fig. 2.8*C*, *top right*). According to this model, the most sensitive, high spontaneous rate auditory nerve fibers have saturating rate versus level functions because almost their entire dynamic range fits into the low-level, linear region of the input-output function of the basilar membrane. The slightly less sensitive, medium spontaneous rate fibers exhibit "sloping saturation" because the upper end of their dynamic range coincides with the compressive region of the basilar membrane input-output function (so it gets "stretched out" across a wider range of input SPL). And the most insensitive, low spontaneous rate fibers appear to have "straight" rate versus level functions because almost the whole of their dynamic range coincides with this compressive region (cf. Sachs et al. 1989; Yates et al. 1990).

According to the logic of the model described above, auditory nerve fibers that exhibit either sloping saturation or straight rate versus level functions should provide an independent means to assess the nature of the mechanical preprocessing that occurs within any region of the cochlea. In particular, it should be possible to assess both the strength of the mechanical compression and the dynamic range over which the compression operates (or at least the intensity at which it begins) by modeling the CF rate versus level functions in these fibers. In practice, however, the limitations on the amount of data that can be acquired from each nerve fiber prevent the simultaneous determination of both of these parameters. The only way

to evaluate either parameter accurately is to assume prior knowledge of the other parameter. And because the most reliable prior information that we have from the direct mechanical studies is the slope of the compression above the "breakpoint" of the basilar membrane, it is the position of this breakpoint that has been the focus of most attention in the neural studies. Using their own modeling techniques, both Sachs et al. (1989) and Müller and Robertson (1991a) have argued that mechanical compression must begin at levels around 30 dB higher than the thresholds of the most sensitive auditory nerve fibers. When comparisons are made across fibers within tightly defined frequency limits in individual experiments, the evidence in favor of a common mechanical origin for the neural breakpoints appears quite strong (cf. Sachs et al. 1989; Müller and Robertson 1991a). When the data are pooled across experiments, the evidence tends to become somewhat blurred, but there is still a tendency for the estimated mechanical breakpoints to cluster within a much more limited range of levels than the rate thresholds, as is shown in Figure 2.8*B*. It should be noted that these findings are based on highly selected data sets, however, and even those who argue that the evidence is strong admit that there are exceptions to the rule (cf. Müller and Robertson 1991a). Others have been even more skeptical and have argued that the breakpoints evident in the sloping saturation rate versus level functions do not reflect features of a mechanical origin (Palmer and Evans 1980). In more recent studies, then, investigators have sought a more robust method for estimating the mechanical compression from the neural data.

3.1.2 Quantitative Measurement of Compression

An improved method for estimating the mechanical compression from neural data was developed by Yates et al. in 1990. As illustrated in Figure 2.9*A*, this technique involves using the rate versus level functions for low-frequency (well below CF) tones to normalize the rate versus level functions observed at other frequencies (including the CF) in individual nerve fibers. The technique is known as the derived input-output function technique, and it relies on an assumption that the relatively low-frequency tones evoke linear mechanical responses at the site innervated by each fiber. This assumption is undoubtedly valid in the basal regions of the cochlea, although it is clearly invalid in the apex (cf. Fig. 2.2). The assumption is also consistent with an observation that almost all auditory nerve fibers have very simple, saturating-type rate versus level functions at low frequencies (Yates et al. 1990). This observation is easy to understand in terms of the model in Figure 2.8*C*; when the nonlinear input-output function of the basilar membrane is replaced by a completely linear function (as would be observed well below the CF of the fiber), the rate versus level functions for all categories of nerve fiber (those with either high, medium, or low thresholds) will display the same "saturating" characteristics. Given the form of

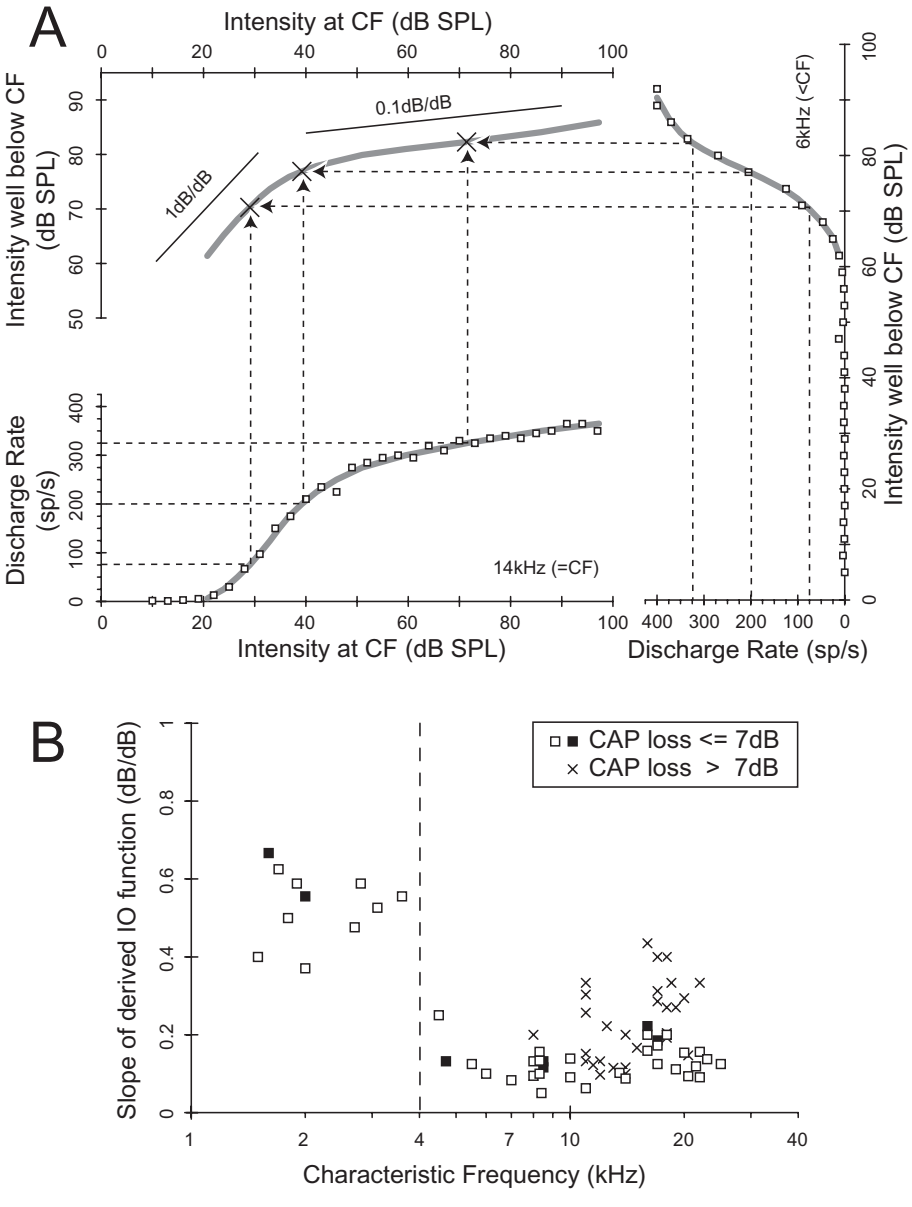


FIGURE 2.9. Input-output functions derived from the responses of individual auditory nerve fibers. A: calculation of a derived input-output function (*top left*) from 2 rate versus level functions (*bottom* and *right*) from a single nerve fiber in the guinea pig cochlea. The derived input-output function plots the stimulus intensities that are needed to evoke various rates of discharge at the CF of the fiber (or any other frequency) against the intensities that are needed to evoke the same rates of discharge well below the CF of the fiber. sp/s, Spikes/second. B: variations in the high-intensity slopes of the input-output (IO) functions derived from a population of guinea pig auditory nerve fibers. Data from fibers recorded under optimal conditions are indicated by compound action potential (CAP) threshold losses of 7 dB or less. Data from fibers recorded under mildly pathological conditions are indicated by CAP threshold losses of 7 dB or more. (Reprinted with permission from Cooper and Yates 1994. Copyright © 1994 Elsevier Science).

the rate versus level functions of each fiber for a low-frequency stimulus, it should be possible to work backward, starting from the rate versus level functions of the fiber for other frequencies, to deduce the form of the mechanical nonlinearities that operate at other frequencies (e.g., at the CF). The entire procedure is illustrated in Figure 2.9*A*.

One of the main advantages that the derived input-output function technique has over other techniques (e.g., rate versus level function modeling) is that it can reveal both the apparent threshold for compression (i.e., the intensity at which the mechanical breakpoint is presumed to occur) and the degree (i.e., the slope) of the compression above the breakpoint. When applied to fibers within very limited CF ranges, the derived input-output technique reveals that the breakpoints in the CF rate versus level functions of the fibers are very tightly clustered (Yates et al. 1990). This is consistent with the hypothesis that the breakpoints originate in the mechanics of the cochlea. When applied across broader ranges of CF, however, and particularly when applied across multiple experiments, the technique reveals more variation in the breakpoint intensities (Cooper, unpublished data). The reasons for this are not entirely clear at present, but other findings suggest that factors other than just the mechanics of the basilar membrane may be contributing to the form of the rate versus level functions observed at different frequencies. One of these findings is that the input-output functions that can be derived from the neural data under conditions of forward masking show small shifts in the breakpoint intensity (Yates et al. 1990). At the time that these observations were made, it was not clear that such effects could not be caused by the basilar membrane. However, subsequent direct observations have shown that the basilar membrane is highly resistant to forward masking (Cooper, unpublished data).

Regardless of the exact origins of the compression that is apparent in the auditory nerve, the derived input-output function technique can readily be applied to fibers that innervate almost any region of the cochlea (other than the extreme apex; see above). It can therefore be used to quantify the compression in these regions and to compare this compression to that seen in the direct mechanical observations (cf. Section 2.2). Figure 2.9*B* illustrates the results of this procedure: the high-level slopes of the input-output functions derived from a population of guinea pig auditory nerve fibers are shown as a function of the CFs of the fibers. Fibers with CFs in excess of around 4 kHz are seen to reflect compression of around 0.1 dB/dB at their CFs, whereas fibers with lower CFs reflect less compression (slopes of around 0.5 dB/dB at high intensities). The slopes of the derived input-output functions in the high CF fibers are slightly lower (i.e., closer to zero) than those evident in the best direct mechanical observations made to date (most modern mechanical studies reveal compression of around 0.2–0.3 dB/dB for CF tones at high stimulus levels; cf. Figs. 2.2–2.4). Whether this reflects a slight mismatch between the physiological conditions of the preparations used in the mechanical and neural studies is not clear. The slopes derived

from neural experiments performed under suboptimal conditions (e.g., as indicated by the crosses in Fig. 2.9*B*) certainly provide a better match to those observed directly in the mechanical experiments. However, it is also possible that something other than just the mechanical preprocessing might be contributing toward the compression that can be seen in the auditory nerve (see comments on forward masking above).

The across-fiber stability of the derived input-output function slopes shown in Figure 2.9*B*, particularly for fibers with CFs in excess of 4 kHz, suggests that mechanical compression operates in a similar manner across a wide range of cochlear locations (i.e., across a wide range of CFs). However, there is a clear discontinuity in the derived input-output function slopes for fibers with CFs above and below 4 kHz. The discontinuity is large enough to suggest that there are real differences between the mechanical preprocessing of the stimuli in the apical and basal turns of the cochlea. One obvious possibility is that the 4-kHz site of the cochlea is the place where the tuning of the mechanical compression becomes broader, such that the assumptions of the derived input-output technique become invalid. In retrospect, this possibility seems quite likely, because the rate versus level functions of low CF fibers often exhibit clear breakpoints even for tones well below the CFs of the fibers (cf. Figs. 2 and 3 in Cooper and Yates 1994). If this is true, then the low-frequency data in Figure 2.9*B* can only provide a lower limit on the strength of the mechanical compression that occurs at the CF in the apical half of the cochlea (i.e., the apical mechanics may actually provide more compression than the 0.5 dB/dB derived input-output function slopes imply).

It is noteworthy that the derived input-output function slopes illustrated in Figure 2.9*B* extend to CFs only as low as 1.5 kHz. The reason for this is that the rate versus level functions of fibers with CFs below this frequency tend to be quite complex and can vary quite markedly and abruptly from frequency to frequency (especially at frequencies well below the CF). It is therefore impossible to know which, if any, of the rate versus level functions in the low CF fibers are likely to reflect linear (or nearly linear) preprocessing. The direct mechanical observations that have been made in the apical turns of the cochlea, showing that compressive nonlinearity can be observed even at frequencies well below the CF of a location (cf. Figs. 2.2–2.4), provide additional reasons not to rely on derived input-output functions in very low CF fibers.

Another way to assess the amount of compression occurring in the cochlea is to study two-tone suppression in the auditory nerve. As outlined in Sections 2.1.3 and 2.2, the mechanical responses of the cochlea to one tone can be suppressed almost instantaneously by the simultaneous presentation of a second tone, leading to clear reductions in both the degree (i.e., the slope) and the dynamic range of the mechanical compression of the first tone (the probe; cf. Fig. 2.6). Neural two-tone suppression shares many of the characteristics of mechanical two-tone suppression (cf. Ruggero et al. 1992), including the fact that only tones falling within a

restricted range of frequencies centered about the CF of a fiber can be suppressed (Sachs and Kiang 1968; Javel et al. 1978; Fahey and Allen 1985). In other words, two-tone suppression only appears to affect probe tones that undergo amplification and compression on the basilar membrane. In the auditory nerve, the reductions in the responses to the probe tone are seen as almost parallel, horizontal shifts[3] of the rate versus probe level function of each fiber, as shown in Figure 2.10*A*. These shifts can be quantified and compared across fibers with different CFs, leading to an independent means of assessing the compression that occurs in different parts of the cochlea.

The results of one of the most careful and comprehensive studies of two-tone suppression in the auditory nerve are shown in Figure 2.10*B*,*C*. Figure 2.10*B* shows the rates of suppression observed in seven fibers covering a wide range of CFs (data from each fiber are joined with lines and the individual CFs are indicated with arrows; cf. Delgutte 1990). The suppression rates within each fiber decrease with increasing suppressor-tone frequency in much the same way that they do at the individual locations studied in cochlear mechanics (cf. Ruggero et al. 1992; Rhode and Cooper 1993; Cooper 1996; Cooper and Rhode 1996b). However, in the neural data, the overall rates of growth of suppression tend to increase with increasing fiber CF. This suggests that the higher CF fibers are more susceptible to two-tone suppression. Given the close correspondence between compression and suppressibility, both in the mechanics of the cochlea and (by implication) within individual nerve fibers, this once again suggests that there is more compression in the higher frequency regions of the cochlea than in the lower frequency regions.

The maximum rates of growth of suppression observed in each of a large population of auditory nerve fibers are shown in Figure 2.10*C* (Delgutte 1990). These confirm the trend described above, although there is some question as to whether the data show a systematic trend across CFs (as shown by the regression line) or fall into two subgroups with low-to-medium and high CFs, respectively. By analogy with the data in Figure 2.9*B*, it is certainly tempting to split the data in Figure 2.10*C* into separate groups with CFs above and below 4 kHz, respectively, and to interpret them in the same way as before: either the strength of the compression that the CF probe tones are undergoing is decreasing with decreasing CF or the compression is becoming more widely distributed across frequency as the CFs decrease (such that the below CF suppressor tones for low CF fibers undergo compression as well as the probe tones).

Other methods of assessing the amount of compression occurring in the cochlea from the discharge patterns of auditory nerve fibers rely even more

[3] Nonparallel shifts are observed if the suppressor tone begins to excite the fibers in its own right (cf. Costalupes et al. 1987) as well as in low and medium spontaneous rate fibers (as would be expected if the suppression has a mechanical origin; cf. Sokolowski et al. 1989), but these are beyond the scope of the present review.

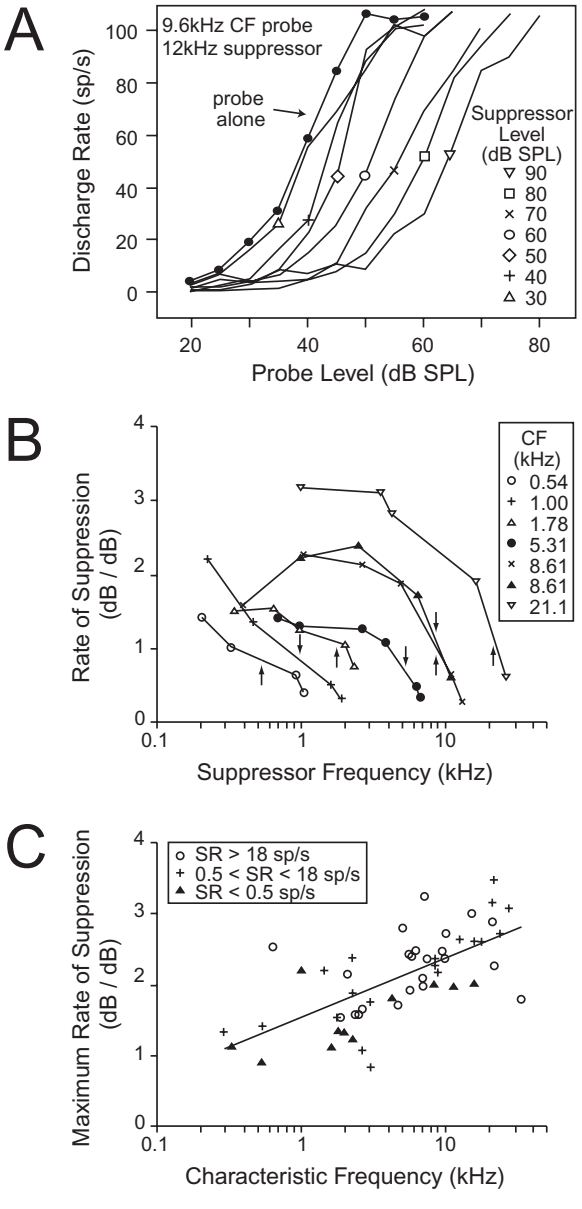


FIGURE 2.10. Two-tone suppression in auditory nerve fibers. *A*: rate versus level function shifts caused by 2-tone suppression. (Cat data reprinted with permission from Javel et al. 1978. Copyright © 1978 Acoustical Society of America.) *B*: rates of growth of suppression in 7 cat auditory nerve fibers. Arrows: CF of each fiber. Within individual fibers, the suppression rates tended to decrease with increasing suppressor frequency. Across fibers, however, suppression rates tended to increase with increasing CF. (Reprinted with permission from Delgutte 1990. Copyright © 1990 Elsevier Science.) *C*: maximum rates of growth of suppression across a population of cat auditory nerve fibers. SR, spontaneous rate. See text for description. (Reprinted with permission from Delgutte 1990. Copyright © 1990 Elsevier Science.)

heavily on the association between amplification and compression in the mechanics of the cochlea. These include the use of ototoxic drugs that preferentially target outer hair cells (e.g., salicylate; cf. Evans and Borerwe 1982; Stypulkowski 1990) or their function (e.g., furosemide; Sewell 1984); the use of direct electrical stimulation to bias the outer hair cell feedback loop away from its point of maximum efficiency (Teas et al. 1970); and stimulation of the medial olivocochlear efferent system (Guinan and Gifford 1988; Guinan and Stankovic 1996; see Guinan 1996 for a review). The results of all these manipulations include the effects of decreasing the amount of compression seen in the mechanics of the cochlea (cf. Section 2.2.3), but they are influenced by other mechanisms as well. Perhaps the simplest results to interpret, and certainly some of the most comprehensively studied effects, are those involving the loop diuretic drug furosemide and those involving the olivocochlear efferents. Each of these evokes much larger shifts in the CF rate thresholds of high-frequency auditory nerve fibers than it does in lower frequency fibers. Particularly in the case of the furosemide experiments, this is difficult to account for by mechanisms other than a decrease in the efficiency of the feedback loop between the outer hair cells and the basilar membrane, with the loop being more efficient to start with in the basal turns of the cochlea than it is in the apical turns.

3.1.3 Dynamic Aspects of Compression

Unlike the situation in the mechanics of the cochlea, the amount of compression that is reflected in the responses of an auditory nerve fiber can depend quite strongly on time. The temporal structure of the response of a typical auditory nerve fiber to a short-tone burst is illustrated in the peristimulus time histograms in Figure 2.11*A*,*B*. Toward the onset of the stimulus, the fiber responds with relatively high discharge rates (or probabilities; cf. Section 3.1). Within a few milliseconds of the stimulus onset, however, the discharge rates decrease substantially through a process known as rapid adaptation (Westerman and Smith 1984, 1985; Yates et al. 1985; Lütkenhoner and Smith 1986). The discharge rates continue to decline for many tens or even hundreds of milliseconds before they reach an approximately steady state via short-term adaptation (Smith and Zwislocki 1975; Westerman and Smith 1984). There are changes over even longer time scales than this, but these are very rarely studied and highly unlikely to affect compression (cf. Young and Sachs 1973; Javel 1996). The dynamics of the neural responses have been studied quite extensively, and they vary significantly across fibers with different CFs and different spontaneous rates (Rhode and Smith 1985; Westerman and Smith 1985; Müller and Robertson 1991b).

The dynamics of the responses of an auditory nerve fiber affect its ability to encode stimuli across a range of intensities. This is illustrated in Figure 2.11*C*–*E*, which shows rate versus level functions for the three categories of auditory nerve fiber described earlier. Figure 2.11*C* shows the responses

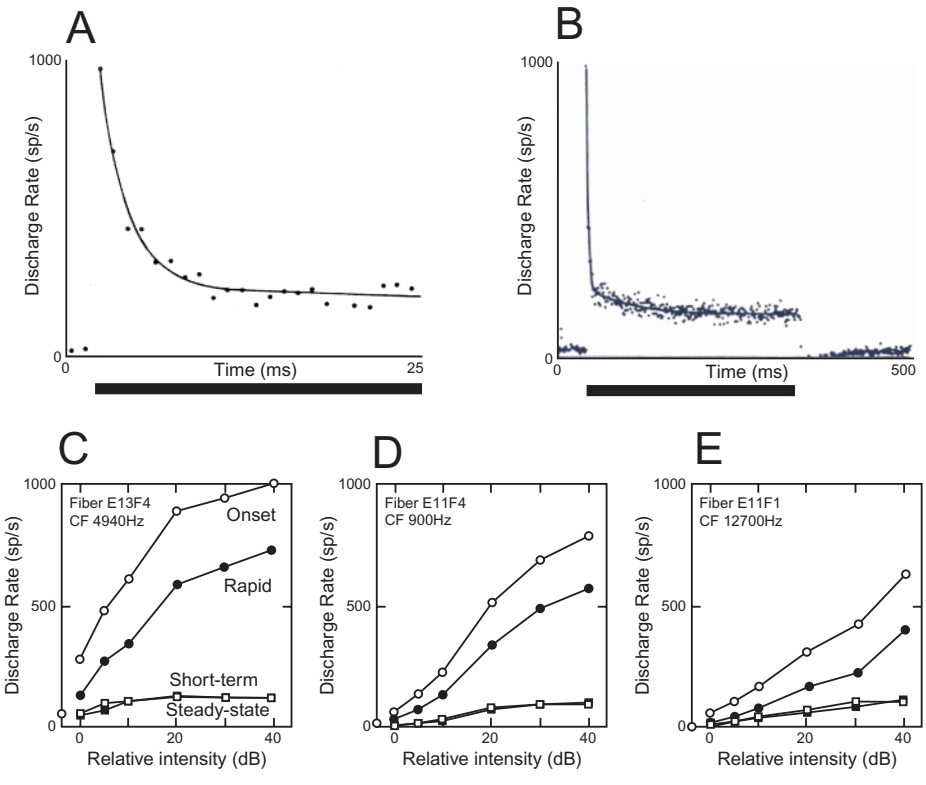


FIGURE 2.11. Compression dynamics in the auditory nerve. *A* and *B*: peristimulus time histogram of tone-evoked responses in a single auditory nerve fiber illustrated on 2 time scales. Solid lines: 3-component model fit to the sound-evoked data. The model sums a rapidly adapting component (exponentially decreasing with a time constant of a few milliseconds) with a short-term adaptation component (exponentially decreasing with a time constant of a few tens of milliseconds) and a steady-state discharge rate. Thick bars, time course of the stimulus. *C–E*: rate versus level functions derived from different temporal components of the tone-evoked responses in three auditory nerve fibers. The abscissas give the stimulus levels relative to the rate threshold of each fiber as determined by audiovisual cues (0 dB on each abscissa is 3 dB above rate threshold). Onset rates (open circles) represent the sums of the rapid (solid circles), short-term (open squares), steady-state-driven (solid squares), and spontaneous (open circles on ordinates) activities. See text for description. (Reprinted with permission from Westerman and Smith 1984. Copyright © 1984 Elsevier Science.)

of a high spontaneous rate fiber with a saturating rate versus level function (see Fig. 2.11*C*, solid squares, which depict the steady-state rate versus level function of the fiber). A mathematical analysis of the responses of the fiber allows its overall discharge rates (or probabilities) to be split into their constituent parts (rapid- and short-term adapting components as well as spon-

taneous and steady-state-driven activities). The onset rate of the fiber, which is shown by the open circles in Figure 2.11*C*, and represents the sum of all of the above components, is seen to have a much wider operating range than the steady-state responses (or the time-averaged responses, which are not shown in Fig. 2.11). The reason for this is that the component of the responses that is attributable to rapid adaptation (as shown by the solid circles in Fig. 2.11*C–E*) has a wider dynamic range than any of the other components (Westerman and Smith 1984). This situation is replicated, to varying degrees, in medium spontaneous rate, sloping saturation fibers (e.g., Fig. 2.11*D*) and in low spontaneous rate, straight rate versus level function fibers (e.g., Fig. 2.11*E*). The degree of dynamic range extension that is offered by the rapidly adapting component of the responses of a fiber appears to be related to the strength of the rapid adaptation, which varies directly with the spontaneous rate of the fiber and with CF (Rhode and Smith 1985; Westerman and Smith 1985; Müller and Robertson 1991b). The dynamic range extensions associated with rapid adaptation seem to play an important role in the encoding of amplitude-modulated stimuli (Yates 1987; Joris and Yin 1992; Cooper et al. 1993). Both low-, medium-, and high-threshold auditory nerve fibers retain an ability to encode dynamic stimuli (via their rapidly adapting response components) well into the intensity range covered by compression in the mechanics of the cochlea.

4. Summary

The peripheral auditory system uses compression to squeeze as much information as possible into the limited dynamic ranges of the neurons that connect it to the central nervous system. The compression is brought about in several ways, each one being important in its own right. At high sound levels, the middle ear muscles can effectively turn down the volume of the sounds that are transmitted to the cochlea. The compression afforded by the middle ear mechanisms is limited in magnitude, however, and is effective only at low frequencies. A much greater degree of compression is offered in the mechanics of the cochlea. The compression here extends from moderately low sound levels all the way through to the most intense sounds that are encountered in everyday life. At each location within the cochlea, only certain frequencies of sound are subjected to compression: high-frequency sounds are compressed in the basal turns of the cochlea and low-frequency sounds are compressed in the apical turns. The degree (i.e., the slope), the extent (i.e., the dynamic range), and the sharpness of the tuning of the mechanical compression all vary from one location in the cochlea to another. The compression is strongest, most wide ranging, and most sharply tuned in the basal turns of the cochlea. The compression appears to result from the mechanical actions of outer hair cells, which are thought to amplify the sound-evoked vibrations of the cochlear partition

in an intensity-dependent manner. The outer hair cells appear to form one part of a positive feedback loop that operates like an instantaneous automatic gain control (i.e., it offers almost distortionless compression, but operates on a cycle-by-cycle basis). The nonlinear nature of the mechano-electrical transduction mechanisms in the hair cells appears to be the dominant source of nonlinearity in the mechanics of the cochlea. The main purpose of mechanical compression is undoubtedly to extend the dynamic ranges of the inner hair cells and the primary afferent neurons that innervate them. One of the side effects of having tuned mechanical compression, however, is that the frequencies evoking the largest responses at each site in the cochlea can change with stimulus level (and they usually do). This is equivalent to saying that the spatial patterns of the cochlear excitation caused by a particular stimulus will change with stimulus level. The synaptic connections between the hair cells and the auditory nerve introduce another form of compression to the auditory periphery. The synapses provide a more classical form of automatic gain control, which is thought to be particularly important in the encoding of dynamic stimuli over wide intensity ranges. Future research on peripheral compression will undoubtedly be aimed at understanding each of these processes in more detail, both mechanistically and functionally.

Acknowledgments. I am indebted to Sid Bacon, Marcus Müller, Fred Nuttall, Alberto Recio, Bill Rhode, and Mario Ruggero, all of whom made valuable contributions toward this review. Support was provided by a Royal Society University Research Fellowship.

References

Anderson DJ, Rose JE, Hind JE, Brugge JF (1971) Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: frequency and intensity effects. J Acoust Soc Am 49:1131–1139.

Ashmore JF (1987) A fast motile response in guinea-pig outer hair cells: the cellular basis of the cochlear amplifier. J Physiol (Lond) 388:323–347.

Borg E, Counter SA (1989) The middle-ear muscles. Sci Am 261:74–80.

Borg E, Zakrisson JE (1975) The activity of the stapedius muscle in man during vocalization. Acta Otolaryngol 79:325–333.

Borg E, Counter SA, Rösler, G (1984) Theories of middle-ear muscle function. In: Silman S (ed) The Acoustic Reflex. Orlando, FL: Academic Press, chapt 3, pp. 63–99.

Brownell WE, Bader CR, Bertrand D, de Ribaupierre Y (1985) Evoked mechanical responses of isolated cochlear outer hair cells. Science 227:194–196.

Cooper NP (1996) Two-tone suppression in cochlear mechanics. J Acoust Soc Am 99:3087–3098.

Cooper NP (1998) Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea. J Physiol (Lond) 509:277–288.

Cooper NP (1999a) Vibration of reflective beads placed on the basilar membrane in the basal turn of the cochlea. J Acoust Soc Am 106:L59–L64.
Cooper NP (1999b) An improved heterodyne laser interferometer for use in studies of cochlear mechanics. J Neurosci Methods 88:93–102.
Cooper NP (2000) Basilar membrane vibrations in the basal turn of the gerbil cochlea. Assoc Res Otolaryngol Abstr 23:205.
Cooper NP, Dong W (2001) Sound-evoked shifts in the baseline position of the apical cochlear partition. Assoc Res Otolaryngol Abstr 24:228.
Cooper NP, Guinan JJ Jr (2002) Fast and slow olivocochlear efferent effects on basilar membrane motion involve different mechanisms. Assoc Res Otolaryngol Abstr 25:82.
Cooper NP, Rhode WS (1992) Basilar membrane mechanics in the hook region of cat and guinea-pig cochleae: sharp tuning and nonlinearity in the absence of baseline position shifts. Hear Res 63:163–190.
Cooper NP, Rhode WS (1995) Nonlinear mechanics at the apex of the guinea-pig cochlea. Hear Res 82:225–243.
Cooper NP, Rhode WS (1996a) Fast travelling waves, slow travelling waves, and their interactions in experimental studies of apical cochlear mechanics. Aud Neurosci 2:289–299.
Cooper NP, Rhode WS (1996b) Two-tone suppression in apical cochlear mechanics. Aud Neurosci 3:123–134.
Cooper NP, Rhode WS (1997) Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea. J Neurophysiol 78: 261–270.
Cooper NP, Yates GK (1994) Nonlinear input-output functions derived from the responses of guinea-pig cochlear nerve fibres: variations with characteristic frequency. Hear Res 78:221–234.
Cooper NP, Robertson D, Yates GK (1993) Cochlear nerve fiber responses to amplitude-modulated stimuli: variations with spontaneous rate and other response characteristics. J Neurophysiol 70:370–386.
Costalupes JA, Rich NC, Ruggero MA (1987) Effects of excitatory and non-excitatory suppressor tones on two-tone rate suppression in auditory nerve fibers. Hear Res 26:155–164.
Dallos P (1985) Response characteristics of mammalian cochlear hair cells. J Neurosci 5:1591–1608.
Davis H (1983) An active process in cochlear mechanics. Hear Res 9:79–90.
de Boer E (1991) Auditory physics: physical principles in hearing theory III. Phys Rep 203:125–231.
de Boer E (1996) Mechanics of the cochlea: modeling efforts. In: Dallos P, Popper AN, Fay RR (eds) The Cochlea. New York: Springer-Verlag, pp. 258–317.
de Boer E, Nuttall AL (1997) The mechanical waveform of the basilar membrane. I. Frequency modulations ("glides") in impulse responses and cross-correlation functions. J Acoust Soc Am 101:3583–3592.
de Boer E, Nuttall AL (2000) The mechanical waveform of the basilar membrane. III. Intensity effects. J Acoust Soc Am 107:1497–1507.
Delgutte B (1990) Two-tone rate suppression in auditory-nerve fibers: dependence on suppressor frequency and level. Hear Res 49:225–246.

Dolan DF, Guo MH, Nuttall AL (1997) Frequency-dependent enhancement of basilar membrane velocity during olivocochlear bundle stimulation. J Acoust Soc Am 102:3587–3596.
Evans EF, Borerwe TA (1982) Ototoxic effects of salicylates on the responses of single cochlear nerve fibres and on cochlear potentials. Br J Audiol 16:101–108.
Evans EF, Palmer AR (1980) Relationship between the dynamic range of cochlear nerve fibres and their spontaneous activity. Exp Brain Res 40:115–118.
Fahey PF, Allen JB (1985) Nonlinear phenomena as observed in the ear canal and at the auditory nerve. J Acoust Soc Am 77:599–612.
Fernàndez C (1952) Dimensions of the cochlea (guinea-pig). J Acoust Soc Am 24: 519–523.
Geisler CD, Nuttall AL (1997) Two-tone suppression of basilar membrane vibrations in the base of the guinea pig cochlea using "low-side" suppressors. J Acoust Soc Am 102:430–440.
Geisler CD, Rhode WS, Kennedy DT (1974) Responses to tonal stimuli of single auditory nerve fibers and their relationship to basilar membrane motion in the squirrel monkey. J Neurophysiol 37:1156–1172.
Goldberg J, Brown P (1969) Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. J Neurophysiol 32:613–616.
Guinan JJ Jr (1996) Physiology of olivocochlear efferents. In: Dallos P, Popper AN, Fay RR (eds) The Cochlea. New York: Springer-Verlag, pp. 435–502.
Guinan JJ Jr, Gifford ML (1988) Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. III. Tuning curves and thresholds at CF. Hear Res 37:29–45.
Guinan JJ Jr, Peake WT (1967) Middle-ear characteristics of anesthetized cats. J Acoust Soc Am 41:1237–1261.
Guinan JJ Jr, Stankovic KM (1996) Medial efferent inhibition produces the largest equivalent attenuations at moderate to high sound levels in cat auditory-nerve fibers. J Acoust Soc Am 100:1680–1690.
Holley M (1996) Outer hair cell motility. In: Dallos P, Popper AN, Fay RR (eds) The Cochlea. New York: Springer-Verlag, pp. 386–434.
Hubbard A, Mountain D (1993) Analysis and synthesis of cochlear mechanical function using models. In: Hawkins H, McMullen T, Popper AN, Fay RR (eds) Auditory Computation. New York: Springer-Verlag, pp. 63–120.
Hudspeth AJ, Corey DP (1977) Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli. Proc Natl Acad Sci USA 74:2407–2411.
Javel E (1996) Long-term adaptation in cat auditory-nerve fiber responses. J Acoust Soc Am 99:1040–1052.
Javel E, Geisler CD, Ravindran A (1978) Two-tone suppression in auditory nerve of the cat: rate-intensity and temporal analyses. J Acoust Soc Am 63:1093–1104.
Johnson DH (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. J Acoust Soc Am 68:1115–1122.
Johnstone BM, Patuzzi RB, Yates GK (1986) Basilar membrane measurements and the travelling wave. Hear Res 22:147–153.
Joris PX, Yin TC (1992) Responses to amplitude-modulated tones in the auditory nerve of the cat. J Acoust Soc Am 91:215–232.

Khanna SM, Hao LF (2000) Amplification in the apical turn of the cochlea with negative feedback. Hear Res 149:55–76.
Khanna SM, Ulfendahl M, Steele CR (1998) Vibration of reflective beads placed on the basilar membrane. Hear Res 116:71–85.
Kiang N, Watanabe T, Thomas C, Clark L (1965) Discharge patterns of single fibers in the cat's auditory nerve. In: Research Monograph number 35. Cambridge, MA, MIT Press.
Kim DO, Chang SO, Sirianni JG (1990) A population study of auditory-nerve fibers in unanesthetized decerebrate cats: response to pure tones. J Acoust Soc Am 87: 1648–1655.
Kros C (1996) Physiology of mammalian cochlear hair cells. In: Dallos P, Popper AN, Fay RR (eds) The Cochlea. New York: Springer-Verlag, pp. 318–385.
Liberman MC (1978) Auditory-nerve response from cats raised in a low-noise chamber. J Acoust Soc Am 63:442–455.
Liberman MC (1980) Morphological differences among radial afferent fibers in the cat cochlea: an electron-microscopic study of serial sections. Hear Res 3:45–63.
Liberman MC (1982) The cochlear frequency map for the cat: labeling auditory-nerve fibers of known characteristic frequency. J Acoust Soc Am 72:1441–1449.
Lim DJ (1980) Cochlear anatomy related to cochlear micromechanics. A review. J Acoust Soc Am 67:1686–1695.
Lonsbury-Martin BL, Meikle MB (1978) Neural correlates of auditory fatigue: frequency-dependent changes in activity of single cochlear nerve fibers. J Neurophysiol 41:987–1006.
Lütkenhoner B, Smith RL (1986) Rapid adaptation of auditory-nerve fibers: fine structure at high stimulus intensities. Hear Res 24:289–294.
Mountain D, Hubbard A, McMullen T (1983) Electromechanical processes in the cochlea. In: de Boer E, Viergever M (eds) Mechanics of Hearing. Delft, The Netherlands: Delft University Press, pp. 11–18.
Müller M, Robertson D (1991a) Shapes of rate-versus-level functions of primary auditory nerve fibres: test of the basilar membrane mechanical hypothesis. Hear Res 57:71–78.
Müller M, Robertson D (1991b) Relationship between tone burst discharge pattern and spontaneous firing rate of auditory nerve fibres in the guinea pig. Hear Res 57:63–70.
Müller M, Robertson D, Yates GK (1991) Rate-versus-level functions of primary auditory nerve fibres: evidence for square law behaviour of all fibre categories in the guinea pig. Hear Res 55:50–56.
Murugasu E, Russell IJ (1995) Salicylate ototoxicity: the effects on basilar membrane displacement, cochlear microphonics, and neural responses in the basal turn of the guinea pig cochlea. Aud Neurosci 1:139–150.
Murugasu E, Russell IJ (1996) The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea. J Neurosci 16: 325–332.
Narayan SS, Temchin AN, Recio A, Ruggero MA (1998) Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae. Science 282:1882–1884.
Nobili R, Mammano F (1996) Biophysics of the cochlea. II: stationary nonlinear phenomenology. J Acoust Soc Am 99:2244–2255.
Nuttall AL, Dolan DF (1993) Two-tone suppression of inner hair cell and basilar membrane responses in the guinea pig. J Acoust Soc Am 93:390–400.

Nuttall AL, Dolan DF (1996) Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig. J Acoust Soc Am 99:1556–1565.
Nuttall AL, Kong WJ, Ren TY, Dolan DF (1995) Basilar membrane motion and position changes induced by direct current stimulation. In: Flock Å, Ottoson D, Ulfendahl M (eds) Active Hearing. Great Yarmouth, UK: Pergamon/Elsevier, pp. 283–294.
Palmer AR, Evans EF (1980) Cochlear fibre rate-intensity functions: no evidence for basilar membrane nonlinearities. Hear Res 2:319–326.
Palmer AR, Russell IJ (1986) Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. Hear Res 24:1–15.
Pang XD, Guinan JJ Jr (1997) Effects of stapedius-muscle contractions on the masking of auditory-nerve responses. J Acoust Soc Am 102:3576–3586.
Pang XD, Peake WT (1986) How do contractions of the stapedius muscle alter the acoustic properties of the middle ear? In: Allen J, Hall J, Hubbard A, Neely S, Tubis A (eds) Peripheral Auditory Mechanisms. New York: Springer-Verlag, pp. 36–43.
Patuzzi RB (1996) Cochlear micromechanics and macromechanics. In: Dallos P, Popper AN, Fay RR (eds) The Cochlea. New York: Springer-Verlag, pp. 186–257.
Patuzzi RB, Robertson D (1988) Tuning in the mammalian cochlea. Physiol Rev 68:1009–1082.
Patuzzi RB, Sellick PM, Johnstone BM (1984) The modulation of the sensitivity of the mammalian cochlea by low frequency tones. III. Basilar membrane motion. Hear Res 13:19–27.
Patuzzi RB, Yates GK, Johnstone BM (1989) Outer hair cell receptor current and sensorineural hearing loss. Hear Res 42:47–72.
Peterson LC, Bogert BP (1950) A dynamical theory of the cochlea. J Acoust Soc Am 22:369–381.
Recio A, Rhode WS (2000) Basilar membrane responses to broadband stimuli. J Acoust Soc Am 108:2281–2298.
Recio A, Rich NC, Narayan SS, Ruggero MA (1998) Basilar-membrane responses to clicks at the base of the chinchilla cochlea. J Acoust Soc Am 103:1972–1989.
Ren T, Nuttall AL (2001) Basilar membrane vibration in the basal turn of the sensitive gerbil cochlea. Hear Res 151:48–60.
Rhode WS, Cooper NP (1993) Two-tone suppression and distortion production on the basilar membrane in the hook region of cat and guinea pig cochleae. Hear Res 66:31–45.
Rhode WS, Cooper NP (1996) Nonlinear mechanics in the apical turn of the chinchilla cochlea *in vivo*. Aud Neurosci 3:101–121.
Rhode WS, Recio A (2000) Study of mechanical motions in the basal region of the chinchilla cochlea. J Acoust Soc Am 107:3317–3332.
Rhode WS, Recio A (2001a) Multicomponent stimulus interactions observed in basilar-membrane vibration in the basal region of the chinchilla cochlea. J Acoust Soc Am 110:3140–3154.
Rhode WS, Recio A (2001b) Basilar-membrane response to multicomponent stimuli in chinchilla. J Acoust Soc Am 110:981–994.
Rhode WS, Smith PH (1985) Characteristics of tone-pip response patterns in relationship to spontaneous rate in cat auditory nerve fibers. Hear Res 18:159–168.
Robles L, Ruggero MA (2001) Mechanics of the mammalian cochlea. Physiol Rev 81:1305–1352.

Robles L, Rhode WS, Geisler CD (1976) Transient response of the basilar membrane measured in squirrel monkeys using the Mossbauer effect. J Acoust Soc Am 59:926–939.
Robles L, Ruggero MA, Rich NC (1991) Two-tone distortion in the basilar membrane of the cochlea. Nature 349:413–414.
Robles L, Ruggero MA, Rich NC (1997) Two-tone distortion on the basilar membrane of the chinchilla cochlea. J Neurophysiol 77:2385–2399.
Rose JE, Brugge JF, Anderson DJ, Hind JE (1967) Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J Neurophysiol 30:769–793.
Rosowski J (1994) Outer and middle ears. In: Fay RR, Popper AN (eds) Comparative Hearing: Mammals. New York: Springer-Verlag, pp. 172–247.
Ruggero MA (1992) Physiology and coding of sound in the auditory nerve. In: Popper AN, Fay RR (eds) The Mammalian Auditory Pathway: Neurophysiology. New York: Springer-Verlag, pp. 34–93.
Ruggero MA, Rich NC (1991) Furosemide alters organ of Corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane. J Neurosci 11: 1057–1067.
Ruggero MA, Robles L, Rich NC (1992) Two-tone suppression in the basilar membrane of the cochlea: mechanical basis of auditory-nerve rate suppression. J Neurophysiol 68:1087–1099.
Ruggero MA, Rich NC, Recio A (1996a) The effect of intense acoustic stimulation on basilar-membrane vibrations. Aud Neurosci 2:329–345.
Ruggero MA, Rich NC, Robles L, Recio A (1996b) The effects of acoustic trauma, other cochlear injury and death on basilar-mem-brane responses to sound. In: Axelsson A, Horchgrevink H, Hamernik R, Hellstrom P-A, Henderson D, Salvi R (eds) Scientific Basis of Noise-Induced Hearing Loss. New York: Thieme Medical, pp. 23–35.
Ruggero MA, Rich NC, Recio A, Narayan SS, Robles L (1997) Basilar-membrane responses to tones at the base of the chinchilla cochlea. J Acoust Soc Am 101:2151–2163.
Russell IJ, Murugasu E (1997) Medial efferent inhibition suppresses basilar membrane responses to near characteristic frequency tones of moderate to high intensities. J Acoust Soc Am 102:1734–1738.
Russell IJ, Nilsen KE (1997) The location of the cochlear amplifier: spatial representation of a single tone on the guinea pig basilar membrane. Proc Natl Acad Sci USA 94:2660–2664.
Russell IJ, Sellick PM (1978) Intracellular studies of hair cells in the mammalian cochlea. J Physiol (Lond) 284:261–290.
Sachs MB, Abbas PJ (1974) Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. J Acoust Soc Am 56:1835–1847.
Sachs MB, Kiang NY (1968) Two-tone inhibition in auditory-nerve fibers. J Acoust Soc Am 43:1120–1128.
Sachs MB, Winslow RL, Sokolowski BH (1989) A computational model for rate-level functions from cat auditory-nerve fibers. Hear Res 41:61–69.
Schroeder MR, Hall JL (1974) Model for mechanical to neural transduction in the auditory receptor. J Acoust Soc Am 55:1055–1060.
Sellick PM, Yates GK, Patuzzi RB (1983) The influence of Mossbauer source size and position on phase and amplitude measurements of the guinea pig basilar membrane. Hear Res 10:101–108.

Sewell WF (1984) The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats. Hear Res 14:305–314.
Siegel JH (1992) Spontaneous synaptic potentials from afferent terminals in the guinea pig cochlea. Hear Res 59:85–92.
Slepecky N (1996) Structure of the mammalian cochlea. In: Dallos P, Popper AN, Fay RR (eds) The Cochlea. New York: Springer-Verlag, pp. 44–129.
Smith RL, Zwislocki JJ (1975) Short-term adaptation and incremental responses of single auditory-nerve fibers. Biol Cybern 17:169–182.
Sokolowski BH, Sachs MB, Goldstein JL (1989) Auditory nerve rate-level functions for two-tone stimuli: possible relation to basilar membrane nonlinearity. Hear Res 41:115–123.
Spoendlin H (1967) The innervation of the organ of Corti. J Laryngol Otol 81: 717–738.
Stypulkowski PH (1990) Mechanisms of salicylate ototoxicity. Hear Res 46:113–145.
Teas DC, Konishi T, Wernick JS (1970) Effects of electrical current applied to cochlear partition on discharges in individual auditory-nerve fibers. II. Interaction of electrical polarization and acoustic stimulation. J Acoust Soc Am 47: 1527–1537.
ter Kuile E (1900) Die Übertragung der Energie von der Grundmembran auf die Haarzellen. Pflügers Arch ges Physiol 79:146–157.
Ulfendahl M, Khanna SM, Flock A (1991) Effects of opening and resealing the cochlea on the mechanical response in the isolated temporal bone preparation. Hear Res 57:31–37.
Ulfendahl M, Khanna SM, Fridberger A, Flock A, Flock B, Jager W (1996) Mechanical response characteristics of the hearing organ in the low-frequency regions of the cochlea. J Neurophysiol 76:3850–3862.
Viemeister NF (1988) Intensity coding and the dynamic range problem. Hear Res 34:267–274.
von Békésy G (1960) Experiments in Hearing. New York: McGraw-Hill.
Westerman LA, Smith RL (1984) Rapid and short-term adaptation in auditory nerve responses. Hear Res 15:249–260.
Westerman LA, Smith RL (1985) Rapid adaptation depends on the characteristic frequency of auditory nerve fibers. Hear Res 17:197–198.
Winter IM, Robertson D, Yates GK (1990) Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibres. Hear Res 45:191–202.
Yates GK (1987) Dynamic effects in the input/output relationship of auditory nerve. Hear Res 27:221–230.
Yates GK (1990) Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions. Hear Res 50:145–162.
Yates GK (1991) Auditory-nerve spontaneous rates vary predictably with threshold. Hear Res 57:57–62.
Yates GK, Robertson D, Johnstone BM (1985) Very rapid adaptation in the guinea pig auditory nerve. Hear Res 17:1–12.
Yates GK, Winter IM, Robertson D (1990) Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range. Hear Res 45:203–219.
Young E, Sachs MB (1973) Recovery from sound exposure in auditory nerve fibers. J Acoust Soc Am 54:1535–1543.

Zheng J, Shen W, He DZ, Long KB, Madison LD, Dallos P (2000) Prestin is the motor protein of cochlear outer hair cells. Nature 405:149–155.
Zinn C, Maier H, Zenner H, Gummer AW (2000) Evidence for active, nonlinear, negative feedback in the vibration response of the apical region of the in-vivo guinea-pig cochlea. Hear Res 142:159–183.
Zwicker E (1979) A model describing nonlinearities in hearing by active processes with saturation at 40 dB. Biol Cybern 35:243–250.
Zwicker E (1986) A hardware cochlear nonlinear preprocessing model with active feedback. J Acoust Soc Am 80:146–153.
Zwicker E, Peisl W (1990) Cochlear preprocessing in analog models, in digital models and in human inner ear. Hear Res 44:209–216.
Zwislocki JJ (1948) Theorie der Schneckenmechanik: qualitative und quantitative analyse. Acta Oto-Laryngol Suppl 72:1–76.

3
Psychophysical Manifestations of Compression: Normal-Hearing Listeners

Andrew J. Oxenham and Sid P. Bacon

1. Introduction

1.1 Psychophysics: Concepts and Methodology

Psychophysics historically represents the attempt to establish a quantitative relationship between physically measurable quantities (such as the intensity and frequency of a sound) and the sensations elicited by those quantities. In hearing, such sensations include loudness and pitch. As our physiological understanding has advanced, the line of division between physical, or "external," variables and psychological, or "internal," variables has shifted. Cooper (Chapter 2) presented detailed measurements of basilar membrane motion in various animals in response to sound, and numerous studies have investigated the response properties of the auditory nerve fibers that contact the sensory cells situated along the basilar membrane (e.g., Kiang et al. 1965; Anderson et al. 1970; Sachs and Young 1980). In this way, variables such as basilar membrane motion (internal in the sense that they are found only within the perceiving organism) become externally observable and hence part of the physical description of the sound stimulus and response.

If we assume that physiological studies in other mammals provide a reasonable model of human auditory processing, the task of the psychoacoustician then expands to accounting for how the known physical transformations of sound energy, such as the mechanical vibrations along the basilar membrane, translate into the truly internal realm of auditory sensation and perception. This chapter addresses two basic questions. The first question is: Can we devise a behavioral measure of basilar membrane response properties, in particular the compressive input-output function found in response to tones with frequencies around the characteristic frequency (CF) of the measurement place? This is an important question for many reasons. First, because direct physiological measurements in the human cochlea are as yet impossible, behavioral measures may provide an important window on these peripheral processes, thereby providing data to

compare to those of other species. Second, measures of cochlear function in individuals may be of benefit in the clinic when diagnosing hearing disorders and prescribing the most appropriate type of hearing aid. The second question this chapter addresses is: What are the perceptual consequences of basilar membrane compression both in the laboratory and in everyday listening situations? Put another way, what aspects of auditory perception can be explained in terms of the response properties of the basilar membrane?

There are a number of ways of approaching the measurement of internal sensation. Classical approaches include magnitude estimation, which involves asking subjects to report their internal sensations by assigning numbers to stimuli based on the magnitude of the perceptual dimension of interest. For instance, sounds of different intensities might be presented, with subjects being instructed to give a number corresponding to the loudness of each sound, ignoring all other aspects of the sound. Such data can be used to derive a function relating physical magnitude (e.g., intensity) to perceptual magnitude (e.g., loudness) (Stevens 1957). Another way of deriving such scales is by ratio estimation, whereby subjects are presented with two stimuli and are asked to give a numerical description of their relationship to each other (e.g., "twice as loud"). Despite the seeming directness of such approaches, they are often plagued by procedural difficulties and pitfalls. Various types of response biases and sequential effects can influence the data to the extent of substantially altering the apparent psychophysical functions (Warren 1970). Overall, the methods and the underlying assumptions of magnitude and ratio estimation remain highly contentious (Poulton 1977).

Another approach to estimating internal representations involves the use of masking. The term masking refers to the reduction in the audibility of one sound in the presence of another sound. It is a phenomenon that we encounter in everyday situations, for instance, when we struggle to maintain a conversation in a noisy restaurant or bar. But with carefully controlled stimuli and experimental conditions, masking can be used to tell us something about how the auditory system processes sound. The underlying and well-founded assumption of much of the work described here is that the masking produced by a sound reflects the way that sound is represented in the auditory periphery. For instance, the fact that a high-level tone produces more masking above its own frequency than below is thought to be a reflection of the similar asymmetric pattern of activity found on the basilar membrane in response to tones.

Masking experiments are generally concerned with measuring the masked threshold of one sound in the presence of another. The concept of threshold is an important one in psychophysics and is worth elaborating on. Possibly the most fundamental of auditory thresholds is the absolute threshold or threshold in quiet. This is the level above which a sound is audible in the absence of any interfering sounds. Although the word threshold

implies something rigid or steplike, this turns out not to be the case. There is no level above which a given sound is always heard and below which it is never heard. Instead, there is always a range of levels over which the likelihood of a sound being detected is greater than zero but less than 100%. Threshold is therefore usually defined as the level at which a sound is detected with a selected probability.

There is an extensive literature on how best to describe the detection process in mathematical terms (e.g., Green and Swets 1966; Macmillan and Creelman 1991). Although the details of these approaches need not concern us here, the results of such research have many practical implications for the measurement of thresholds. For instance, experimenters are unlikely simply to present a sound to subjects and then ask the subjects whether they heard it. At levels near threshold, the answers of the subjects would depend not only on whether the sound had elicited a sensation but also on how willing the subjects were to respond positively when they were uncertain. Factors that influence the response but are unrelated to the stimulus or its sensation are collectively referred to as bias. In probing the auditory system through behavioral techniques, it is important to reduce bias as much as possible. The most common way of measuring thresholds is to use a forced-choice procedure. This involves presenting two or more intervals in succession, only one of which contains the sound to be detected or signal. Subjects are therefore not asked whether they heard the signal but rather which interval contained the signal. This simple manipulation already greatly reduces the scope for bias.

The most straightforward but rather time-consuming way to measure the detectability of a signal is to present it repeatedly at a number of different levels and measure the percentage of times the signal was correctly identified. For instance, in a two-interval forced-choice task, chance corresponds to 50% correct, whereas perfect performance corresponds to 100% correct. Thus there will be a range of levels over which the percent correct score will increase, usually monotonically, from 50 to 100%. A hypothetical example of such data is shown in Figure 3.1. The resulting curve is known as a psychometric function. The threshold may, for instance, be defined as the point at which the curve crosses the 75% correct mark.

The vast majority of psychoacoustic masking studies do not measure the entire psychometric function but instead use a more rapid measurement technique to estimate one point on the psychometric function, which is then defined as the threshold. By far the most popular technique involves an adaptive procedure in which the signal level in one trial is determined by the subject's responses in the previous trial or trials (Levitt 1971). An experimental run usually starts with the signal well above threshold. The level of the signal is then adjusted throughout the run according to a set of rules. The rules determine what point of the psychometric function the experiment will track. The presentation level of the signal therefore adapts to the performance of the subject. In this way, the detectability of a signal can be

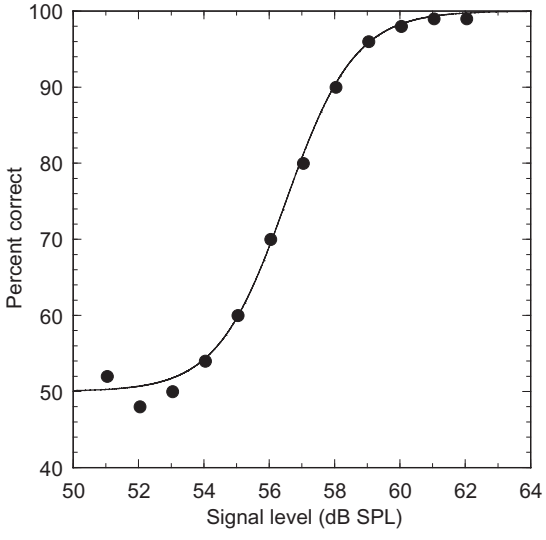


FIGURE 3.1. Schematic diagram of a psychometric function. The percentage of correct responses in identifying the signal generally increases monotonically with increasing signal level. Threshold may be arbitrarily defined as the point at which the curve intersects the 75% correct line. SPL, sound pressure level.

measured and compared across different conditions with relative speed and accuracy.

1.2 Behavioral Implications of Auditory Compression

The human auditory system has a tremendous dynamic range (see Fig. 1.1 in Bacon, Chapter 1). Young people with normal hearing can detect sound pressures as low as 0.00002 Pa in the midfrequency (1–3 kHz) range and generally experience pain only when the sound pressure exceeds that value by a factor of about 1,000,000. Consider the specifications of such a mechanical sensing system, scaled up a little: at one end, vibrations of only 1 mm should be detectable, whereas at the other end, vibrations of 1 km should still be within the operating range of the system! The visual system also has an impressive dynamic range in coping with light intensities. However, while our eyes take several minutes to adjust from a very bright environment to a dark one, our ears can recover to maximum sensitivity from all but the loudest sounds in considerably less than a second. The large range of audible and nondamaging sound pressures is one reason why sound is generally referred to in decibel (dB) units, so that the range from just audible to painful spans about 120 dB at medium frequencies.

The compressive input-output function of the basilar membrane plays a significant role in the ability of the auditory system to cope with the large

dynamic range (see Bacon, Chapter 1). The rapid recovery after loud stimulation is due in part to the fact that the compression on the basilar membrane seems to act nearly instantaneously (Cooper, Chapter 2). Other aspects of basilar membrane nonlinearity, such as the exquisite frequency tuning at low levels, are also of great importance in hearing. Still others, such as spontaneous otoacoustic emissions, or sounds produced by the ear (Zurek 1981), and distortion-product emissions are probably by-products of the nonlinear, active mechanism. Early pitch theories postulated a role for distortion products in strengthening or even generating the fundamental frequency of harmonic complex tones from sources such as the speaking or singing voice and musical instruments (Helmholtz 1885/1954). However, later research showed that distortion products were neither necessary nor sufficient to account for most pitch percepts (Schouten 1940). Nevertheless, the measurement in the ear canal of distortion products generated in the cochlea has been shown to be of benefit in screening for cochlear hearing loss (e.g., Gorga et al. 2000) and is becoming part of the basic battery of clinical tests.

Although changes in frequency tuning with level and distortion products are clearly integral parts of cochlear nonlinearity, the remainder of this chapter concentrates on the perceptual consequences and measurement of the compressive input-output function of the basilar membrane, with particular emphasis on its effect on temporal processing. Recent studies have made it clear that a wide range of seemingly disparate aspects of temporal processing can be understood within the context of basilar membrane compression. As Bacon and Oxenham (Chapter 4) make clear, many of the differences in perception between normal-hearing people and individuals with cochlear hearing loss can be accounted for in terms of a loss of or reduction in basilar membrane compression. Many studies up to 1997 pertaining to the perceptual consequences of cochlear nonlinearity were reviewed in an article by Moore and Oxenham (1998). This chapter outlines some of that material but concentrates on findings that have emerged since that time.

2. Growth of Masking as a Measure of Cochlear Compression

2.1 Growth of Simultaneous Masking

Some of the earliest published data on psychophysical masking already indicated significant nonlinearities in auditory processing. For narrowband noise or sinusoidal maskers with sinusoidal signals, Wegel and Lane (1924) and Egan and Hake (1950) showed that when the masker and signal were close in frequency, the signal intensity at threshold was roughly proportional to the masker intensity. In other words, the signal level at threshold

grew linearly with masker level such that a 10-dB increase in masker level resulted in about a 10-dB increase in the threshold signal level. The situation was very different when the signal frequency was well above that of the masker. In that case, the threshold signal level tended to grow more rapidly than the masker level such that a 10-dB increase in masker level led to a 20-dB increase in signal level at threshold. An example of such data is shown in Figure 3.2 for three masker levels [45, 65, and 85 dB sound pressure level (SPL)]. The masker was a 500-Hz-wide band of noise centered at 2,400 Hz; the signal was a 10-ms sinusoidal signal, gated with 5-ms ramps (no steady state) and temporally centered within the 400-ms masker. As seen in Figure 3.2, the shape of the masking pattern changes considerably with level, becoming increasingly asymmetric with increasing masker level. This nonlinear behavior and the resulting high-level asymmetry is often referred to as the upward spread of masking.

A masking pattern has long been thought of as reflecting the internal excitation produced by the masker (Fletcher and Munson 1937; Zwicker 1970; Florentine and Buus 1981). According to this concept, the pattern traced out by the signal level at threshold in Figure 3.2 is in some way equivalent to the excitation produced by the masker in the auditory periphery: a narrow masking pattern, such as that found with a masker level of 45 dB SPL, suggests that only a small region of the basilar membrane is stimu-

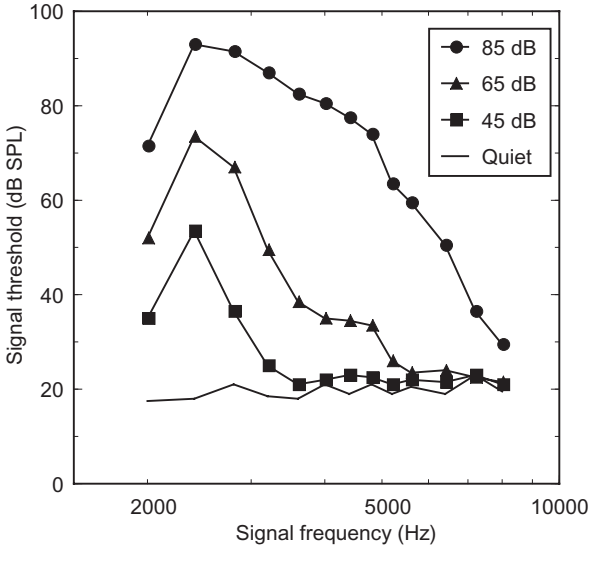


FIGURE 3.2. An example of a masking pattern from a single subject (Oxenham, unpublished data). Signal thresholds are plotted as a function of signal frequency in the presence of a narrowband noise masker centered at 2.4 kHz. The masker level was 45, 65, or 85 dB SPL.

lated by the masker, whereas a broad masking pattern (e.g., the 85-dB curve) suggests that a large portion of the basilar membrane is responding to the masker. The interpretation of masking patterns as a direct reflection of peripheral excitation requires at least two main assumptions. The first assumption is that masking is purely excitatory. That is, the signal is rendered undetectable because the response to it is "drowned out" or "swamped" by the response to the masker. The second assumption is that the peripheral response to the signal is linearly related to signal level so that a 10-dB increase in signal level at threshold can be interpreted as a 10-dB increase in masker excitation at the place with a CF corresponding to the signal frequency. Following these assumptions, the nonlinear growth of masking for signals higher in frequency than the masker must be explained as follows: the peripheral response to the masker grows linearly at places along the basilar membrane with CFs around the masker frequency but grows expansively at places with CFs well above the masker frequency. It turns out, however, that the two assumptions and the resulting conclusion are in all likelihood false. With the benefit of modern physiological findings, an alternative explanation of the upward spread of masking, which is directly related to the frequency-selective and compressive properties of the basilar membrane can be given. This explanation also provides the possibility of deriving a behavioral measure of basilar membrane nonlinearity in humans.

Figure 3.3 shows a schematic diagram of a basilar membrane input-output function in response to a tone at the CF of the point of measure-

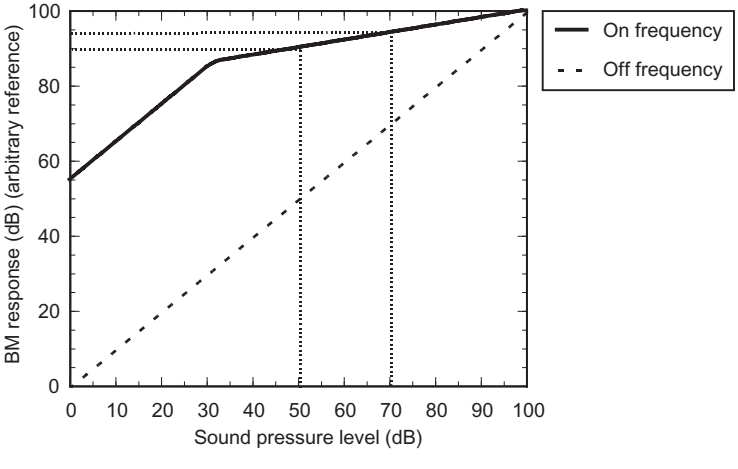


FIGURE 3.3. Schematic diagram of basilar membrane (BM) response to on-frequency and off-frequency tones. The vertical and horizontal dotted lines illustrate how a large change in signal level can result in only a small change in basilar membrane response.

ment (on-frequency tone; solid curve) and to a tone well below CF (off-frequency tone; dashed line). The important characteristics are that the response to the on-frequency tone is nonlinear and compressive for much of the input level range and that the response to the off-frequency tone is linear throughout (see Fig. 2.2 in Cooper, Chapter 2, for examples of basilar membrane input-output functions). When relating these effects to masking, we assume that the signal is detected at a place along the basilar membrane with a CF corresponding to the signal frequency. We also assume that the signal threshold corresponds to a fixed signal-to-masker ratio in terms of basilar membrane response and, for illustrative purposes, that the threshold signal-to-masker ratio is 0 dB so that the signal is at threshold when the masker and signal elicit the same overall response at the place with a CF corresponding to the signal frequency.

Consider a signal masked by an on-frequency masking tone. The basilar membrane response to both the masker and the signal follows the solid curve in Figure 3.3. According to our assumption of a 0-dB signal-to-masker ratio, the masker level and signal level at threshold will always be the same. Therefore, despite the nonlinear input-output function, the growth of masking with an on-frequency masker, measured psychoacoustically, will appear linear: a 10-dB increase in masker level will lead to a 10-dB increase in the signal level at threshold. The situation is different for an off-frequency masking tone. Here the basilar membrane response to the masker follows the dashed line in Figure 3.3, whereas the response to the signal follows the solid curve in Figure 3.3. Consider for example a 50-dB SPL signal, which produces a basilar membrane response of 90 dB in our arbitrary units. The arbitrary units were selected to be identical to those of the off-frequency masker level and so, following the dashed response line of the masker, the masker level at threshold is 90 dB SPL. If the signal level is increased by 20 dB to 70 dB SPL, the basilar membrane response increases to 94 dB. To mask this signal, the masker must have a level of 94 dB SPL, an increase of only 4 dB. Put another way, every 1-dB increase in off-frequency masker level should lead to a 5-dB increase in signal level at threshold, a 5 : 1 growth rate.

Given our simplifying assumptions, the relationship between basilar membrane compression and the predicted masking growth rate is straightforward: the compression slope (or exponent if linear units are used) is the reciprocal of the masking growth rate. In other words, a masking growth rate (p) is predicted from an on-frequency basilar membrane compression slope of 1/p. The nonlinear properties of the basilar membrane therefore provide a qualitative explanation for the nonlinear growth in the upward spread of masking. One potential problem with this approach, however, is that essentially all studies involving off-frequency growth of simultaneous masking (i.e., where the masker and signal are presented at the same time) have shown growth rates of around 2 : 1, depending somewhat on the signal frequency (Bacon et al. 1999) and on the nature of the masker (van der

Heijden and Kohlrausch 1995), but not exceeding 3:1 (Schöne 1977, 1979). This is considerably less than would be expected based on the most recent basilar membrane measurements in other mammals, which have yielded compression slopes of around 0.2, corresponding to a predicted masking growth rate of 5:1. What explains this discrepancy?

2.2 Suppression and the Growth of Forward Masking

One important difference between most physiological and behavioral studies lies in the nature of the stimuli used. In physiological studies, the response to one tone at a time is studied when tracking the input-output function, whereas in behavioral studies, at least two stimuli are present: the masker and the signal. It turns out that this may be a crucial difference between the two paradigms when estimating basilar membrane input-output functions. It has been known for some time that the presence of one sound can influence the physiological response to another in a nonlinear way. One class of these phenomena is known as "two-tone suppression": here the response to one tone can be reduced by the addition of a second, suppressor tone (Cooper, Chapter 2). This is an inherently nonlinear phenomenon: in a linear system, additional sound power at a different frequency will lead to an increase or no change in response but never to a decrease. This effect was first reported in studies of auditory nerve responses (Sachs and Kiang 1968), and qualitatively consistent findings were reported not long afterward in psychoacoustic studies (Houtgast 1972; Shannon 1976). It has since become clear that at least a part of the effect observed in auditory nerve responses can be observed in the motion of the basilar membrane (Ruggero et al. 1992b). Most important from our perspective is that the introduction of a suppressor tone not only reduces the response to the signal tone but also linearizes its input-output function (Ruggero et al. 1992b). Because of suppression, therefore, all growth-of-masking studies using simultaneous masking can be regarded as measuring the (more linear) response of a tone plus suppressor rather than the desired (more compressive) response to the tone alone. This may be one reason why behavioral measures using growth of masking in simultaneous masking have not found the very steep masking slopes predicted by basilar membrane data.

Another reason relates to a phenomenon known as off-frequency listening (Patterson 1976; Johnson-Davies and Patterson 1979; Moore et al. 1984). In physiological measurements of basilar membrane motion, the place of measurement remains fixed and is determined by the positioning of the measuring apparatus. In psychoacoustical measures, we cannot be certain either that the signal is detected via only one location along the basilar membrane or that the place remains constant as the level of the stimuli is changed. For instance, the place that responds best to the signal in isolation may not offer the best signal-to-masker ratio when the masker is added.

Another issue relating to the basalward shift in the peak of the traveling wave at high levels is discussed later in this section.

One way to overcome the problem of suppression is to present the masker and signal at different times, using a paradigm known as forward masking, so that they do not physically overlap. In forward masking, the signal is presented shortly after the offset of the masker. The gap between the end of the masker and the beginning of the signal need not be very long, because suppression seems to begin and end almost instantaneously with the onset and offset of the suppressor (Sachs and Kiang 1968; Ruggero et al. 1992a). A popular way to limit off-frequency listening is to present a noise masker spectrally shaped to limit the detectable spread of the excitation of the signal (O'Loughlin and Moore 1981). Both these techniques were employed by Oxenham and Plack (1997) in an attempt to provide a behavioral estimate of basilar membrane nonlinearity in humans. They measured the forward-masking level necessary to mask a very brief 6-kHz signal as a function of signal level for masker frequencies of 6 kHz (on frequency) and 3 kHz (off frequency). The growth function from the off-frequency masker provided an estimate of the basilar membrane input-output function, as described above. The on-frequency masker was used to test for nonlinearities in forward masking per se.

The mean results from three normal-hearing listeners are shown in Figure 3.4. In contrast to many earlier studies, the masker level is shown on the ordinate and the signal level is on the abscissa. This is because the masker level, and not the signal level, was the dependent variable in this experiment. This form of plotting also has the advantage that the slope of the masking function can be directly interpreted as the compression slope without requiring that the reciprocal be taken. The circles show the data from the on-frequency condition. As seen, the masker level at threshold grows approximately linearly with signal level. This is as expected from the schematic diagram in Figure 3.3: if the masker and signal levels are roughly equal, they will fall within the same part of the basilar membrane input-output function and will therefore be processed similarly. However, it also implies that the forward-masking process is linear. This was not clear at the time; in fact, it was generally believed that forward masking was a highly nonlinear phenomenon. This issue is addressed more thoroughly in Section 2.3. For now, it is sufficient to note that because the on-frequency growth is linear, no further transformation of the off-frequency data is necessary to interpret them in terms of basilar membrane nonlinearity.

The form of the data from the off-frequency masker (Fig. 3.4, squares) is very different from that of the on-frequency masker. In contrast to the linear growth of the on-frequency masker, the off-frequency masker level grows very slowly with signal level, especially at medium levels. If we assume that the basilar membrane response can be described according to the schematic diagram in Figure 3.3, then the data from the off-frequency

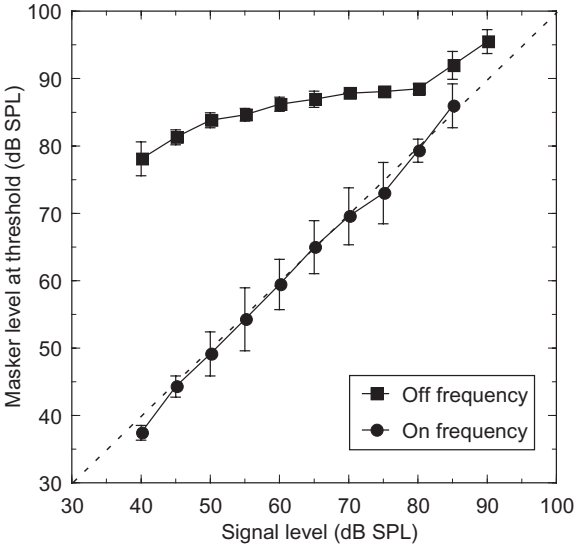


FIGURE 3.4. Mean data from a study by Oxenham and Plack (1997). The masker levels at threshold are plotted as a function of the signal level for a masker at the signal frequency or 1 octave below the signal frequency. Dashed line: linear growth. Values are means ± SE.

condition provide a direct estimate of basilar membrane compression. In particular, the slope of the masking function reflects the slope of the basilar membrane input-output function.

We can now compare the slope derived from Oxenham and Plack's (1997) normal-hearing listeners to the physiological measurements of basilar membrane compression. At signal levels between 50 and 80 dB SPL, the slope of the mean data is 0.16. This is very close to midlevel estimates from cochleae in very good condition (Cooper, Chapter 2) and corresponds to a compression ratio of about 6:1. At lower and higher levels, the growth becomes more linear. More linear growth at low levels is also found in physiological studies. Earlier studies of basilar membrane response also found more linear growth at high levels. However, those findings have been challenged by more recent data that suggest that the compression continues up to levels of 100 dB SPL and beyond (Cooper, Chapter 2). Although the difference between the behavioral and the more recent physiological data at high levels seems puzzling, a simple explanation may resolve the apparent discrepancy. As mentioned above, in psychoacoustic measurements, the place (or places) along the basilar membrane that are responsible for the detection of the signal cannot be known with certainty, although the presence of the noise band should restrict the region of usable information to a large extent.

It is known from physiological studies that as the level of a tone increases, the place of maximum excitation along the basilar membrane shifts toward the base of the cochlea. At very high levels, the peak excitation occurs at a place with a CF well above the frequency of the tone, which therefore responds linearly to the tone. In other words, although the response at CF remains compressive, places more basal along the basilar membrane respond linearly throughout and eventually "catch up" with and overtake the response at the CF in what is referred to as the basalward shift of the traveling wave (Rhode and Recio 2000). Oxenham and Plack (1997) suggested that detection of a signal is more likely to be based on the place of the maximum excitation than on the nominal CF. If so, the psychoacoustical measures would track the peak excitation point, the response of which becomes linear at high levels, whereas the physiological measures track a fixed point on the basilar membrane, where the response remains compressive. This may account for the apparent discrepancy between physiological and behavioral data at high levels. However, results from a recent study by Nelson et al. (2001), described in more detail below, cast some doubt on whether this interpretation is the whole story.

In summary, the data of Oxenham and Plack (1997) suggest that it is possible to estimate basilar membrane compression behaviorally and that the compression observed in humans is similar to that found in other mammals. The difference between this growth-of-masking slope and those found in earlier studies is probably due primarily to the use of forward masking and secondarily to the use of a noise band to reduce off-frequency listening. Although this work showed the possibility of estimating human basilar membrane compression, some aspects of the paradigm made it difficult to extend the results. For instance, very short signal durations are required in order to be able to mask high-level signals. This makes the paradigm difficult to employ at signal frequencies lower than about 4 kHz because the spectral spread of the signal begins to exceed the bandwidth of basilar membrane tuning at the signal frequency, (something known as "spectral splatter"; see Leshowitz and Wightman 1972). Which in turn smears the temporal representation of the stimulus and also makes it difficult to rule out detection of the signal at places with CFs remote from the signal frequency. Also, the use of an additional noise to limit off-frequency listening could present difficulties when transferring the paradigm to hearing-impaired listeners because the noise may produce some direct masking of its own. In Sections 2.3 and 3, extensions of this work, which in different ways overcome the problems mentioned, are discussed.

2.3 Further Estimates of Basilar Membrane Response Using Forward Masking

The two problems with Oxenham and Plack's (1997) measure, namely short signal durations and off-frequency noise, are overcome in a technique pro-

posed by Nelson et al. (2001). They kept a 1-kHz signal fixed at a very low level (10 dB above absolute threshold for each subject) and measured the forward-masking level required to mask the signal at different masker-signal gap durations for a number of different masker frequencies. By assuming that the response to maskers well below the signal frequency was linear, they were able to derive the form of compression at other frequencies by comparing how the masker level changed as a function of gap duration. For instance, if a 10-ms increase in gap duration for the off-frequency masker required a 2-dB increase in the off-frequency masker but a 10-dB increase in the on-frequency masker, the compression slope at the signal frequency could be approximated as 0.2 between the two on-frequency masker levels. Because the signal was always presented at a very low level, off-frequency noise was not required to mask the spread of excitation, and the signal duration could be longer while still keeping the signal masked.

The basic results were similar to those found earlier: maximum compression for on-frequency tones was found at medium levels and more linear responses were observed at low and high levels. Compression exponents reached a minimum of between 0.1 and 0.2, which is consistent both with the earlier behavioral estimate and with physiological data. The earlier results were also extended in this study by systematically examining the responses to a number of masker frequencies ranging from an octave below the signal frequency ($0.5f_s$) to just above the signal frequency ($1.2f_s$). The results suggested that the response to tones starts to become compressive only for frequencies above about $0.7f_s$ (or half an octave below CF). At frequencies above the signal, the response again becomes rapidly linear above about $1.05f_s$. Both these findings are broadly consistent with the relevant physiological data. Although the study of Nelson et al. (2001) was limited to a signal frequency of 1 kHz, the technique could easily be extended to cover a wide range of frequencies.

Another important aspect of the paradigm introduced by Nelson et al. (2001) also relates to the low and constant signal level used. The signal acts as the probe for estimating basilar membrane excitation. Because it remains fixed at a low level, the point along the basilar membrane that is being probed also remains fixed, in possible contrast to the technique of Oxenham and Plack (1997). According to the arguments stated above relating to the basalward shift of the traveling wave, the measure used by Nelson et al. (2001) should continue to show compression even at very high levels. However, the results from most of their subjects do not fulfill this prediction; instead, most of the curves become linear at levels above about 80 dB SPL. It is unclear what accounts for the apparent discrepancy between the physiological and behavioral data, whether it is due to procedural differences or species differences or whether adult human listeners with nominally normal hearing already suffer from sufficient hearing loss to resemble the slightly damaged cochleae of earlier physiological preparations (for a

discussion of possible explanations of a similar change in slope at high levels in simultaneous masking, see Bacon et al. 1999).

3. Behavioral Estimates of Basilar Membrane Compression as a Function of CF

A question that remains basically unanswered in the physiological literature is how basilar membrane nonlinearity changes with CF. Most data come from the basal turn of the cochlea, corresponding to CFs above about 8 kHz; less is known about basilar membrane mechanics at the lower frequency apical region of the cochlea. As discussed by Cooper (Chapter 2), measuring from lower CFs is beset by technical difficulties. In contrast, there are no such difficulties associated with carrying out behavioral experiments at low signal frequencies, with the constraint that stimulus durations and onset-offset ramps must be sufficiently long to avoid problems associated with "spectral splatter" (e.g., Leshowitz and Wightman 1972).

The physiological data that do exist, either from direct basilar membrane measurements (Rhode and Cooper 1996) or from indirect estimates from auditory nerve responses (Cooper and Yates 1994), suggest that nonlinearity and compression are reduced at low CFs (Cooper, Chapter 2). This is somewhat consistent with psychoacoustic data showing, for instance, that auditory filter shapes change less with level at low than at high frequencies (Glasberg and Moore 1990). Hicks and Bacon (1999) carried out a systematic study of how nonlinearity changes with frequency using three different tests, including the growth of on- and off-frequency forward masking. They measured in the more traditional manner of keeping the masker level fixed and adaptively varying the signal level. Plotting signal level as a function of masker level, they found that the growth-of-masking slope for the off-frequency masker increased with increasing signal frequency from 375 to 3,000 Hz, consistent with the hypothesis that compression becomes more pronounced at higher frequencies. Similarly, they found that changes in auditory filter shape with level and estimates of suppression also increased with increasing signal frequency, in line with the idea of increased nonlinearity at higher CFs. A similar conclusion was drawn by Bacon et al. (1999), who measured the growth of off-frequency simultaneous masking for signal frequencies from 400 to 5,000 Hz: they found that the slope of the masking function was shallower at the two lowest signal frequencies (400 and 750 Hz) than at the other frequencies between 1,944 and 5,000 Hz, again corresponding to reduced compression at the lower frequencies.

Using a variant of the so-called pulsation-threshold method (Thurlow and Rawlings 1959; Houtgast 1972), Plack and Oxenham (2000) estimated basilar membrane compression with off-frequency maskers for signal frequencies ranging from 250 to 8,000 Hz. In a pattern of alternating tones, one

tone (the signal) can be made to sound continuous if the other tone (the masker) is sufficiently high in level and close to the signal in frequency. The pulsation threshold is different from a traditional masked threshold in that it relies on a subjective judgment: a listener is asked to judge whether or not the signal sounded continuous to them. Note that what we refer to as the masker does not actually mask the signal. Instead, it masks the intervals between repetitions of the signal, allowing the auditory system to "fill in the gaps." Despite the subjective nature of the judgment, the responses can be relatively consistent. The main conclusion from this study was that nonlinearity seemed reduced at the lowest frequencies of 250 and 500 Hz and remained constant between 1,000 and 8,000 Hz.

In summary, most studies agree that the nonlinearities seem to be reduced at frequencies below about 1,000 Hz, which is at least qualitatively consistent with the available physiological data. However, at high frequencies, the absolute values of estimated compression from the studies discussed above (Bacon et al. 1999; Hicks and Bacon 1999; Plack and Oxenham 2000) are considerably less than those found in the studies of Oxenham and Plack (1997) and Nelson et al. (2001). For the Bacon et al. (1999) study, this may be due to the use of simultaneous masking for reasons relating to suppression (see above). For the other two studies, the difference may lie in the fact that neither study employed band-limited noise to reduce off-frequency listening (see Nelson et al. 2001).

Although the psychophysical and physiological data seem to fit reasonably well, a caveat is in order (Plack and Oxenham 2000). All techniques involving off-frequency growth of masking take the linear response to tones well below CF as a fundamental assumption. The same is true for the indirect estimate of basilar membrane nonlinearity using auditory nerve responses (Cooper and Yates 1994) and for other psychophysical measures such as auditory filter shape and suppression (Hicks and Bacon 1999). If, for instance, nonlinear compression at a particular place along the basilar membrane were applied to all incoming frequencies, measures using the growth of masking would not detect the compression. This is because both the on-frequency and off-frequency maskers would be equally compressed and so no difference in slope would ensue; everything would appear to be linear. There are some physiological data suggesting that this may be the case. Rhode and Cooper (1996) found some compression at a place along the chinchilla basilar membrane with a low CF but found that the compression was not limited to frequencies around the CF; instead, the response to all frequencies was approximately equally compressive. Thus, a more conservative conclusion from the growth-of-masking studies is that they show that the compression at low CFs, if present, is not as frequency selective as that found at high CFs. Some psychophysical evidence using harmonic tone complex maskers (Oxenham and Dau 2001b) seems to support this view and is discussed further in Section 4.5. Other very recent psychophysical estimates, using various aspects of forward masking without the assumption

of a linear off-frequency response, also indicate continued strong compression down to very low CFs (Lopez-Poveda et al. 2003; Plack and O'Hanlon 2003).

4. Temporal Processing

Apart from the importance of basilar membrane nonlinearity in establishing the wide dynamic range of hearing and in accounting for the upward spread of masking, recent research has revealed many aspects of temporal processing that are probably influenced by basilar membrane compression. All sounds in our everyday environment change over time, and our ability to follow rapid changes in sound patterns is crucial to our ability to understand speech. The limits on our ability to follow rapid changes in sound level are tested in measures of temporal resolution. At the other end of the range, our ability to combine information over time is tested in measures of temporal integration. The possible influences of basilar membrane compression on both types of measures are explored here. First, a conceptual model of temporal processing, known as the temporal window model, is described.

4.1 The Temporal Window Model

The temporal window model has been used in the past to conceptualize and quantify psychoacoustic performance in tasks involving temporal processing, such as gap detection, temporal integration, and forward and backward masking. Figure 3.5 illustrates the basic outline of the model. First, sounds are filtered and subjected to a static nonlinearity. For simplicity and in the absence of accurate nonlinear human cochlear filter models, the filtering and nonlinear stages have been treated separately in most modeling studies so far (e.g., Oxenham and Moore 1994; Plack and Oxenham 1998). However, the combination of these two stages is designed to represent the nonlinear filtering of the cochlea. In the next stage, the stimuli are rectified and squared to simulate some aspects of hair cell processing and are then passed through the temporal window (or temporal integrator). The temporal window has the effect of smoothing rapid fluctuations in the temporal envelope of sounds such that slow changes are passed essentially unchanged, but rapid changes are attenuated. This stage has no direct known physiological correlate but could be thought of as representing higher level neurons (perhaps in the cortex?) with relatively long time constants. Defining the shape of the temporal window has been the topic of some previous studies (e.g., Moore et al. 1988). Usually, as with fitting auditory filter shapes (e.g., Patterson and Nimmo-Smith 1980), a simple mathematical shape with relatively few free parameters is assumed and the parameters are then adjusted to best describe the data. So far, studies have

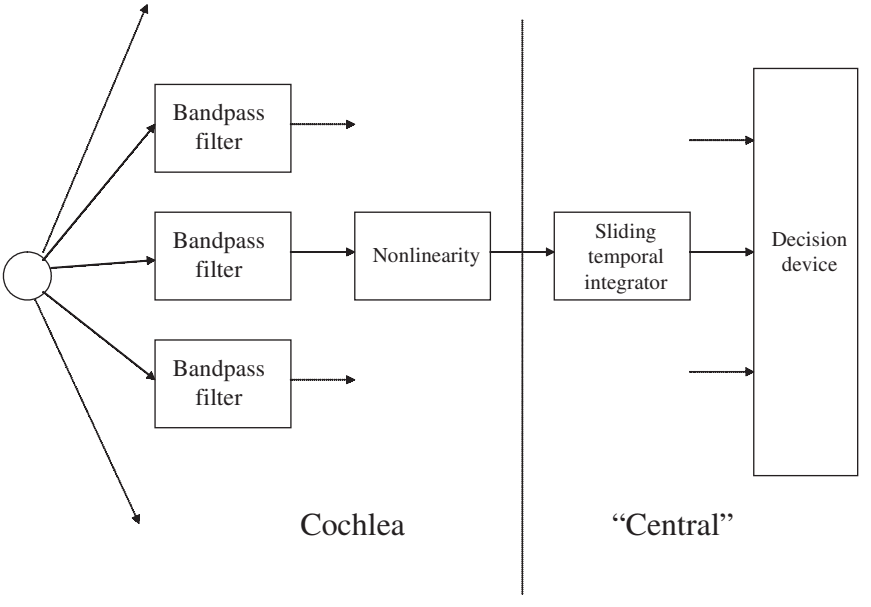


FIGURE 3.5. Block diagram of the temporal window model showing the four basic stages of processing: filtering, nonlinearity, temporal integration, and decision process.

generally assumed a single or double exponential, or rounded exponential (roex), function to describe each side of the window.

The details of the decision device, which follows the temporal window in the model, depend to some extent on the task. For many detection tasks, a computationally convenient decision method is to compare the output of the temporal window in response to the masker and signal to the output in response to the masker alone. The maximum signal-to-masker ratio at the output of the window then determines the predicted threshold. The criterion signal-to-masker ratio is usually treated as a free parameter but is held constant for any given data set.

Figure 3.6 shows a schematic diagram of the temporal envelope of a forward masker followed by a brief signal. Figure 3.6*A* shows the stimulus envelopes before processing by the model, Figure 3.6*B* shows the temporal window through which the stimuli are passed, and Figure 3.6*C* shows the output of the temporal window in response to the masker and signal (solid line) and the response to the masker alone (dashed line). The point in time at which the ratio of these two curves is greatest is the point that is assumed to determine the signal threshold. This ratio, which defines the threshold and is a reflection of the detection efficiency, is assumed to remain constant across conditions.

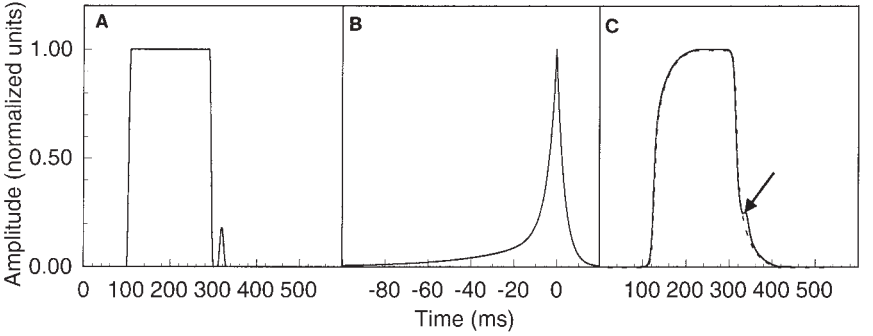


FIGURE 3.6. Schematic diagram showing the process of forward masking according to the temporal window model. *A*: envelopes of a masker followed by the signal. *B*: temporal window. The asymmetry in the window implies that stimuli before the peak of the window have a greater effect than stimuli after. The illustrated window is the time-reversed impulse response. *C*: effect of passing the stimuli in *A* through the window in *B*. Solid line: response to masker and signal; dashed line: response to the masker alone; arrow: point at which the ratio is largest. Often, the maximum ratio of the two curves is used as the decision criterion.

4.2 Additivity of Nonsimultaneous Masking

The question of how masking effects combine has interested psychoacoustic researchers for some time (Green 1967; Wilson and Carhart 1971; Penner and Shiffrin 1980; Humes and Jesteadt 1989). In our everyday acoustic environment, it is rare to have only one interfering sound present at a time. Thus, in order for measurements in the laboratory to find general applicability, we need to know how the masking effects of different sounds interact. When the maskers and signal overlap in time (additivity of simultaneous masking), the variety of potential detection cues as well as the difficulty in providing a functional definition of one, two, or many maskers makes a general theory of masking additivity difficult to formulate (Moore 1985; Oxenham and Moore 1995). In the case of nonoverlapping maskers and signal, the situation is somewhat simplified. Aside from forward masking, discussed above, there is also an effect known as backward masking, where a masker that follows a signal in time can increase its threshold. One way of viewing both forward and backward masking is to assume that they are the result of some "sluggishness" in the auditory system, which can be modeled as a sliding temporal integrator, or temporal window, as described in Section 4.1.

Early studies had found that the effect of combining a forward and a backward masker was greater than would be predicted by a simple energy summation of the two maskers (Pollack 1964; Elliott 1969; Patterson 1971; Wilson and Carhart 1971). Robinson and Pollack (1973) suggested that this effect might be explained in terms of a sliding temporal integrator, as illustrated in Figure 3.7. In the presence of only a forward masker, the best signal-to-masker ratio might occur after the center of the integrator had passed the center of the signal. Graphically, this is when the integrator (marked by the solid bar) is centered to the right of the signal (Fig. 3.7*A*). In the presence of only a backward masker, the opposite would hold: the best signal-to-noise ratio would occur when the integration window was centered to the left of the signal (Fig. 3.7*B*). When both maskers are combined, the optimal position for the window is restricted on both sides, and assuming a symmetric integration window, the integrator should be centered on the signal (Fig. 3.7C). This yields a nonoptimal signal-to-noise ratio for each masker alone and consequently leads to a higher than predicted increase in threshold when the two maskers are combined.

Although this explanation is appealing, a later test of the hypothesis found that this could not be the whole story. Penner (1980; see also Penner and Shiffrin 1980) tested combinations not only of forward and backward maskers but also of two nonoverlapping forward maskers or two backward maskers. By combining equally effective pairs of maskers, Penner (1980) was able to compare the amount of masking obtained with that predicted by a simple energy summation. In many cases, signal thresholds in the presence of two equally effective maskers exceeded the threshold in the pres-

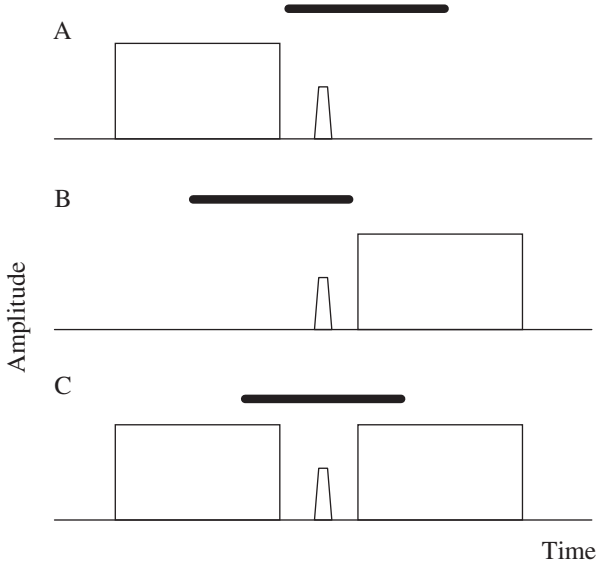


FIGURE 3.7. Schematic diagram illustrating the concept of the sliding temporal integration period proposed by Robinson and Pollack (1973). When only one masker is present, the integration period (solid bar) can be shifted to improve the signal-to-masker ratio. When both maskers are present, the benefit of shifting the integration period is reduced.

ence of a single masker by 10 dB or more. This is considerably more than the 3-dB increase expected on the basis of energy summation. The important point, however, was that the increase in masking was essentially the same whether the two maskers formed a pair of forward and backward maskers or whether both were forward maskers. This is not the result predicted by Robinson and Pollack's (1973) shifting window hypothesis: in the presence of two forward maskers, the integration window could be as optimally positioned as in the presence of a single masker, and so only the increase in masking due to the energy summation of the two maskers would be expected.

Penner (1980) proposed an alternative hypothesis, which accounted in principle for the nonlinear additivity of masking in all the cases discussed so far. She assumed that the stimuli were subjected to an instantaneous compressive nonlinearity before being combined within a linear temporal integrator. It is fairly straightforward to see how this scheme predicts nonlinear masking additivity. Assume for now that the nonlinearity is a simple power law such that the stimulus intensities are transformed according to the equation

$$y = kI^p \tag{4.2.1}$$

where y is the internal response, I is the stimulus intensity, k is a scaling constant, and p is an exponent with a value between 0 and 1. Consider two maskers that in isolation produce a signal threshold intensity of I_1, leading to an internal representation of kI_1^p. When the two equally effective maskers are combined within the linear temporal window, by definition the output of the window will be twice the output due to a single masker. This means that the output due to the signal will have to double in order to maintain a threshold signal-to-noise ratio at the output of the window. If we define the threshold signal intensity in the presence of two maskers as I_2, this results in the equation

$$I_2^{\,p} = 2I_1^{\,p} \tag{4.2.2}$$

Raising both sides to the power of $1/p$ gives

$$I_2 = 2^{1/p} I_1 \tag{4.2.3}$$

If the stimuli are processed linearly with respect to intensity ($p = 1$), then the result is simple energy summation: the intensity of I_2 must be twice that of I_1, an increase of 3 dB. On the other hand, if p is less than 1, implying compression, the value of $2^{1/p}$ becomes greater than 2, and so the increase in masking produced by the two maskers increases accordingly. Equation 4.2.3 can be transformed into dB terms, giving

$$L_2 = L_1 + 10\log_{10}(2)/p \tag{4.2.4}$$

where L is the sound level $[10\log_{10}(I/I_0)]$. In other words, the effect of combining two equally effective maskers is $1/p$ times greater than that predicted by energy summation.

In the works by Penner (1980) and Penner and Shiffrin (1980), a nonlinearity that is more compressive than a power law was proposed. However, this was based on experiments using a broadband click signal and broadband noise maskers. It is not clear which part of the signal spectrum was detected by listeners or whether the detection cues remained the same at different masker levels. Studies since then have concluded that a power law nonlinearity is sufficient to account for the available data, with the following modification: a constant is added, which has the effect of an additional masker to incorporate absolute threshold, thereby ensuring that maskers never predict thresholds lower than the absolute threshold (Humes et al. 1988; Humes and Jesteadt 1989; Oxenham and Moore 1994, 1995). This also has the effect of predicting less than full additivity at low signal levels, in line with the data.

In a study comparing the additivity of forward and backward masking in normal-hearing and hearing-impaired subjects, Oxenham and Moore (1995) found that the data pooled across conditions from three normal-hearing listeners were well described using a compressive exponent (p) of about 0.2. This value may be of significance because it is very similar to physiological estimates of basilar membrane compression (Cooper,

Chapter 2) and to compression estimates from the growth-of-masking studies discussed above. However, it should be noted that this exponent applies to the intensity of a sound and not to its amplitude or pressure. Sound intensity is proportional to the square of sound pressure. The response of the basilar membrane is always quoted in terms of its amplitude or velocity, which relates to sound pressure, not intensity. Therefore, if the exponent derived from the additivity of masking is related to basilar membrane compression, we must postulate a square law nonlinearity after the basilar membrane transformation. Further indications for such a square law nonlinearity are presented in Sections 5 and 6 and by Bacon and Oxenham (Chapter 4).

4.3 Forward Masking

Forward masking has been used to illustrate many points in this chapter, but understanding forward masking is of considerable interest in its own right. Forward masking is a popular measure of temporal resolution or the ability to follow rapid changes in sound level. The ability to recover quickly from stimulation is important in speech perception where relatively large changes in level can occur over very brief time periods.

There are at least two ways of viewing forward masking. The first view is that forward masking is due to neural adaptation, perhaps in the auditory nerve (Smith 1977, 1979). According to this view, the neural response adapts to the masker so that the response to a signal following the masker is reduced. The second view is that forward masking is due to some form of neural integration or persistence occurring at a higher stage in the auditory system, possibly in the cortex. According to this view, the neural response persists after the physical end of the masker and "swamps" the response to the signal. The temporal window model falls into this second category. There are many parallels between this adaptation/persistence debate and the debate between suppression and excitation as possible mechanisms of simultaneous masking (Delgutte 1990).

In many cases the predictions derived from an adaptation or integration explanation of forward masking are identical, and it has so far proved difficult to distinguish between them experimentally using behavioral techniques (Oxenham 2001). However, there is some physiological evidence that the adaptation observed in the auditory nerve cannot account for forward masking. Relkin and Turner (1988) measured the responses in individual auditory nerve fibers to a "signal" tone in isolation or after an intense "masker" tone. In contrast to previous studies, they measured not only the response to the signal but also the response in the absence of the signal. In agreement with previous studies, they found a reduction in the response to the signal when it was preceded by the masker. However, they found a similar reduction in the spontaneous activity after the masker. Thus, the reduced response to the signal remained detectable against the reduced

background activity. Using signal detection analysis (e.g., Green and Swets 1966), they concluded that the amount of masking observed in individual auditory nerve fibers was much less than that observed behaviorally. It remains possible, however, that adaptation at stages higher than the auditory nerve is responsible for forward masking. For the purposes of exploring the effects of peripheral compression on forward masking, we will use the temporal window model, which assumes that forward masking is due to neural persistence. One advantage of this approach is that the same model can be applied to many other aspects of temporal processing such as backward masking (Elliott 1962; Penner 1974), gap detection (e.g., Plomp 1964; Penner 1977; Moore et al. 1989), and increment and decrement detection (e.g., Plack and Moore 1991; Moore et al. 1993; Oxenham 1997), whereas neural adaptation in isolation can only be used to account for forward masking. Sections 4.3.1 to 4.3.3 review various aspects of forward masking and show how they may be influenced by basilar membrane compression.

4.3.1 Growth and Decay of Forward Masking

One interesting aspect of forward masking is that signal thresholds decay to very near absolute threshold within 100–200 ms no matter what the masker level (except very high-level maskers that cause temporary or permanent hearing loss). An example of the decay of forward masking as the gap between the masker and signal is increased is shown in Figure 3.8*A*, taken from a study by Jesteadt et al. (1982). Forward-masking decay curves

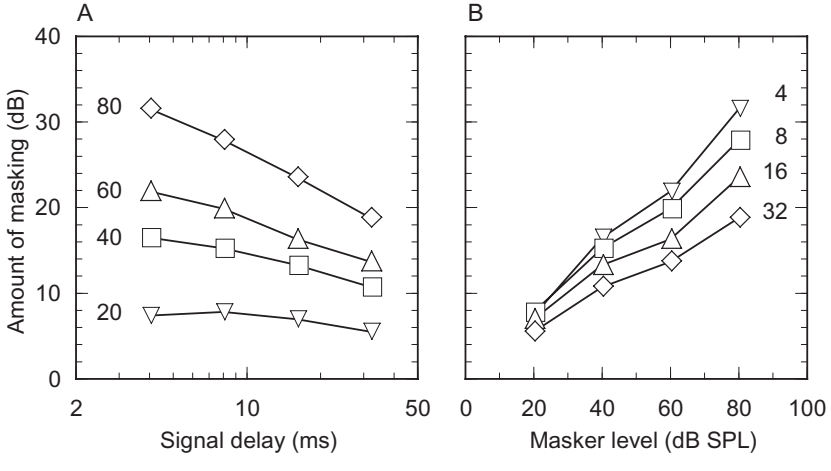


FIGURE 3.8. Forward-masking data using a 1-kHz masker and signal from a study by Jesteadt et al. (1982). *A*: how signal thresholds decay as the gap between the masker and signal is increased for several different masker levels. *B*: same data but now showing how signal level at threshold increases as a function of masker level for several different signal delays.

that become steeper with increasing level imply that the rate of growth of masking depends on the gap between the masker and signal. This can be seen in Figure 3.8*B*. Here, the data in Figure 3.8*A* are simply replotted with the independent variable and the parameter exchanged. This general pattern of results is seen for on-frequency forward masking and for forward masking with a broadband noise masker at all signal frequencies (e.g., Jesteadt et al. 1982; Moore and Glasberg 1983). The situation for forward maskers much lower in frequency than the signal (off-frequency masking) has been examined much less frequently. However, the existing data seem to indicate a rather different pattern from that seen in on-frequency forward masking. Instead of the growth-of-masking slopes being generally nonlinear and becoming shallower with increasing masker-signal gap (e.g., Fig. 3.8*B*), the slopes seem to be approximately linear and independent of the masker-signal delay (Fig. 5 in Kidd and Feth 1981; Fig. 3 in Oxenham and Plack 2000). It turns out that the nonlinear gap-dependent slopes in on-frequency conditions and the linear gap-independent slopes in off-frequency conditions can both be explained within a single unified framework if the effects of basilar membrane compression are considered, as explained below (Oxenham and Moore 1995; Plack and Oxenham 1998; Oxenham and Plack 2000).

Consider the case of on-frequency forward masking, keeping in mind the basilar membrane input-output function shown as the solid curve in Figure 3.3. When the gap between the masker and signal is very short, the signal threshold will be high and the masker and signal levels will be similar to each other. Because of their similarity in level, the masker and signal will be in about the same operating region of the basilar membrane input-output curve. Thus, assuming a constant signal-to-masker ratio at threshold, any change in masker level will lead to a similar change in signal threshold, implying linear growth of masking. Linear growth of forward masking is indeed what is found for very short signals and gaps (Oxenham and Plack 1997). As the gap between the masker and signal is lengthened, the signal level at threshold decreases. In many cases, the masker will be in the midlevel compressive region of the basilar membrane function, whereas the signal will be in the low-level more linear region. Here the growth of masking will no longer be linear; a given change in masker level will require a smaller change in signal level because the signal is compressed less than the masker. This leads to shallow growth-of-masking slopes similar to those often reported in the forward-masking literature (Plomp 1964; Jesteadt et al. 1982; Moore and Glasberg 1983).

Overall, therefore, the change in the rate of masking growth as the gap between the masker and signal increases might be due to the changes in basilar membrane compression with level. Plack and Oxenham (1998) tested this idea directly by collecting data and simulating the results using the temporal window model described above. The basilar membrane nonlinearity assumed in the model was derived by fitting two straight lines to

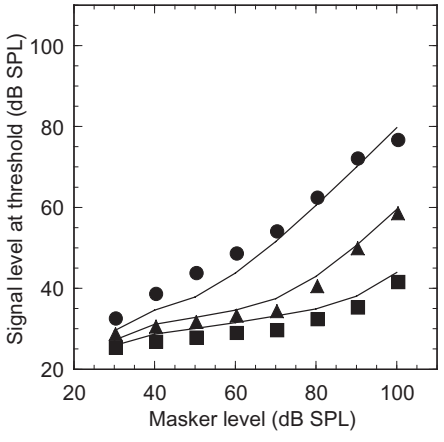


FIGURE 3.9. Growth of forward masking. Data and predictions from Plack and Oxenham (1998). Signal thresholds are plotted as a function of masker level for masker-signal gaps of 2 ms (circles), 20 ms (triangles), and 40 ms (squares). Solid lines: predictions for each condition. Both the masker and the signal were 6-kHz sinusoids.

the data of Oxenham and Plack (1997). The resulting nonlinearity had an input-output slope of 0.78 for levels below 35 dB SPL and 0.16 above that level. The window shape was taken from an earlier study (Oxenham and Moore 1994) so that the only free parameter was the threshold signal-to-masker ratio. Even with this very crude approximation to the shape of the basilar membrane nonlinearity, the model was able to account for the data reasonably well. Figure 3.9 shows a portion of those data together with the model predictions. Despite some systematic deviations between the data and the predictions, the model is able to capture the main features including the steeper growth of masking at the shorter gaps and the change in the slope when the signal level moves into the more compressive region of the basilar membrane nonlinearity.

The nearly linear growth of masking found for maskers well below the signal in frequency (Oxenham and Plack 2000) can be accounted for within the same framework. Again considering Figure 3.3, it can be seen that if the masker is processed linearly (dashed line), then the slope of the growth of masking will be the same regardless of whether the masker is at a low or a high level; for a given range of signal levels, the slope is independent of masker level and hence is independent of the gap between the masker and signal. For low-level signals, both the signal and the masker are processed approximately linearly, leading to linear growth of masking.

One assumption of the analysis in terms of basilar membrane compression is that forward masking itself is a linear process. This assumption is built into the linear temporal window approach but was far from obvious.

In fact, the many nonlinear aspects of forward masking suggested quite the opposite. Even the first analyses of forward and backward masking in terms of a temporal window postulated that the window itself was nonlinear in that it changed shape with level (Moore et al. 1988; Plack and Moore 1990). Only after the form of the basilar membrane nonlinearity was incorporated into such models did it become clear that temporal processing subsequent to basilar membrane compression could be regarded as linear (Oxenham and Moore 1997).

4.3.2 Effects of Masker and Signal Duration in Forward Masking

For a given masker level and masker-signal gap, an increase in the duration of a forward masker usually leads to an increase in the signal threshold, at least for masker durations up to 100–200 ms (e.g., Kidd and Feth 1982; Zwicker 1984; Carlyon 1988). Oxenham and Moore (1994) simulated the data of Zwicker (1984) and showed that the pattern of results could be reasonably well accounted for with the same temporal window that was used to predict the decay of forward masking. A later study showed that the effects of masker duration were influenced more by the signal level at threshold than by the masker level or the masker-signal gap (Oxenham and Plack 2000). This dependency is the expected consequence of basilar membrane nonlinearities; as the masker duration increases, the masker energy falling within the temporal window increases, and the amount by which the signal level needs to be increased to compensate for this will depend only on the compression applied to the signal.

The effects of signal duration on thresholds in forward masking provide an interesting confirmation of a somewhat counterintuitive prediction of the temporal window model. The results also provide another example of how basilar membrane nonlinearity can be used to explain relatively complicated parameter interactions. The study of how signal duration affects thresholds in forward masking has a long and contradictory history. The basic paradigm is shown schematically in Figure 3.10*A*; thresholds are measured as a function of signal duration (d), with the time (t) between the masker offset and the signal offset held constant. In an early study of forward masking using sinusoidal maskers and signals, Zwislocki et al. (1959) found that thresholds were not affected by the duration of the signal. They concluded that only the final portion of the signal determined the threshold. This pattern of results can be explained qualitatively in terms of neural adaptation: the neural response to portions of the signal closer to the masker is more adapted than that to later portions, and so the early part of the signal contributes less to detection than it would in the absence of a forward masker.

Although this makes a plausible case, other data from Zwislocki et al. (1959) and later studies (e.g., Neff 1986) suggest caution in interpreting their results. In particular, their use of sinusoids with long onset and offset ramps

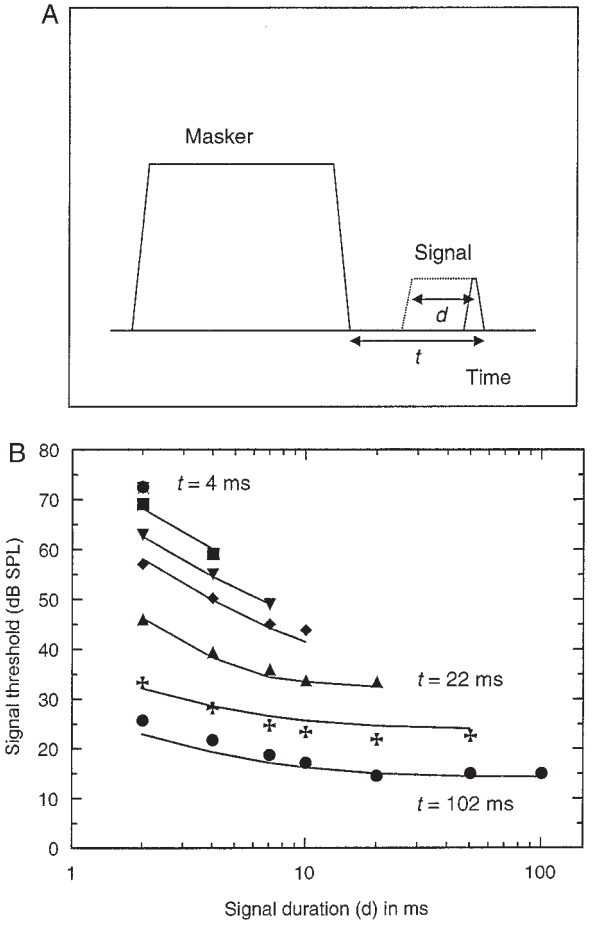


FIGURE 3.10. Temporal integration in the presence of a forward masker. *A*: schematic diagram of the stimuli. The time (*t*) between the masker offset and the signal offset was held constant, and signal thresholds were measured for different signal durations (*d*). *B*: data and predictions, with signal threshold plotted as a function of signal duration. The values of *t* tested (from *top* to *bottom*) were 4, 6, 9, 12, 22, 52, and 102 ms. (Data are from Oxenham 2001.)

made it difficult for subjects to distinguish the masker from the signal even if both were clearly audible. Thus, at short gaps between the masker offset and the signal onset, the threshold estimates of Zwislocki et al. (1959) could have been elevated by so-called confusion effects (Moore and Glasberg 1985; Neff 1985). Confirming these reservations, later studies showed a rather different pattern of results. Fastl (1976a,b, 1979) and Gralla (1992)

both found that signal thresholds decreased with increasing signal duration, much as they do in quiet (e.g., Zwislocki 1960; Watson and Gengel 1969; Gerken et al. 1990). The interesting point is that these later results are in line with the predictions of the temporal window model. The temporal window is thought to be asymmetric (producing more forward masking than backward masking). This means that whether a forward masker is present or not, the signal produces the greatest window output when the window is centered near the end of the signal. Because the window is linear, the presence of the forward masker has no effect on the response of the window to the signal and so the masker has no effect on the integration of signal intensity as the signal duration is increased.

A more recent study (Oxenham 2001) produced results broadly in line with those of Fastl (1976a) and Gralla (1992): signal thresholds decreased with increasing signal duration. The wider range of conditions tested in the recent study also revealed that in some conditions the decrease in signal threshold with increasing duration was actually greater than that observed in quiet. The main results are shown in Figure 3.10*B*. Mean signal thresholds are plotted as a function of signal duration, with the different symbols representing different gaps between the masker and signal offsets. The curves are the predictions of the temporal window model, using the best-fitting temporal window and a basilar membrane nonlinearity, with a linear response at low levels and a compression slope of 0.25 above 35 dB SPL. The value of 0.25 is somewhat higher than that found in some studies (Oxenham and Moore 1995; Oxenham and Plack 1997) but is in line with that found in others (Oxenham and Moore 1994; Oxenham et al. 1997) and is still within the range of values found in physiological studies. Using a more compressive function resulted in somewhat worse fits, but the form of the predictions remained the same.

Important trends in the data include (1) signal thresholds decreasing with increasing signal durations up to 20 ms and remaining roughly constant thereafter; (2) the rate of change in threshold always being at least as great as in quiet for signal durations up to 20 ms; and (3) the rate of change in threshold being substantially greater than that observed in quiet at higher signal levels. All three aspects are well captured by the model predictions. In the model, the similarity between the predictions for conditions in quiet and in the presence of a forward masker, at least at short signal durations, is a consequence of the linear temporal window. The increased effect at high signal levels is due to the compressive nonlinearity of the model. As Penner (1978) showed, a compressive nonlinearity can increase the effect of signal duration on thresholds. This can be understood intuitively as follows. For simplicity, consider a rectangular integration window and assume that a signal is detected if the output of the window exceeds a certain criterion level. If the duration of a signal at threshold is doubled, the output of the window will also double as long as the signal is still shorter than the window duration. In a linear system of energy detection, the signal level would have

to be reduced by 3 dB in order to halve the window output level and return the signal to threshold. However, if the signal is compressed using a power law exponent (p), the level of the signal would need to be reduced by 3/p dB in order to halve the window output level. Further simulations, using the temporal window model and a model of neural adaptation, showed that the main trends could be captured by both classes of model. However, in both cases, the inclusion of a nonlinearity resembling that found on the basilar membrane was essential in generating reasonable predictions (Oxenham 2001). This provides further evidence for the importance of basilar membrane nonlinearity in forward masking.

4.3.3 Forward Masking by Harmonic Tone Complexes

Harmonic tone complexes are a combination of sinusoids, the frequencies of which are all integer multiples of a common fundamental frequency (f_0). The waveform produced by a combination of such sinusoids is periodic, repeating itself once every $1/f_0$ seconds. The shape of the waveform depends on the relative amplitudes and phases of the individual sinusoids. In the case where all the components are consecutive harmonics at the same amplitude, the waveform can range from having a very "peaky" waveform, approaching a click train where all the components share the same cosine starting phase, to one with a very flat temporal envelope where the component phases are selected according to an equation proposed by Schroeder (1970). By this equation, the phase (φ) of component n is selected according to the formula

$$\varphi_n = \pm \pi n(n-1)/N, \qquad (4.3.3.1)$$

where N is the total number of components in the complex. When the positive sign is used, it is known as a Schroeder-positive (m+) complex; when the negative sign is used, it is known as a Schroeder-negative (m–) complex. An example of each type is shown in Figure 3.11*A*,*C*. These stimuli have a number of interesting properties both physically and psychophysically. They can be described as rising (m–) or falling (m+) repeating linear frequency sweeps. Physically, by maintaining a flat temporal envelope, they allow the transmission of signals with very low peak-to-root mean square (rms) ratios. Psychophysically, they have very interesting masking properties that have been used to investigate the phase response of peripheral auditory filtering (Smith et al. 1986; Kohlrausch and Sander 1995; Lentz and Leek 2001; Oxenham and Dau 2001b). Most important from our point of view is that the phase response of the basilar membrane filter seems to interact with the stimuli such that the filtered waveform in response to m+ stimuli is considerably more peaked than the waveform in response to m– stimuli (Recio and Rhode 2000). This is illustrated in Figure 3.11*B*,*D*. Here, the m+

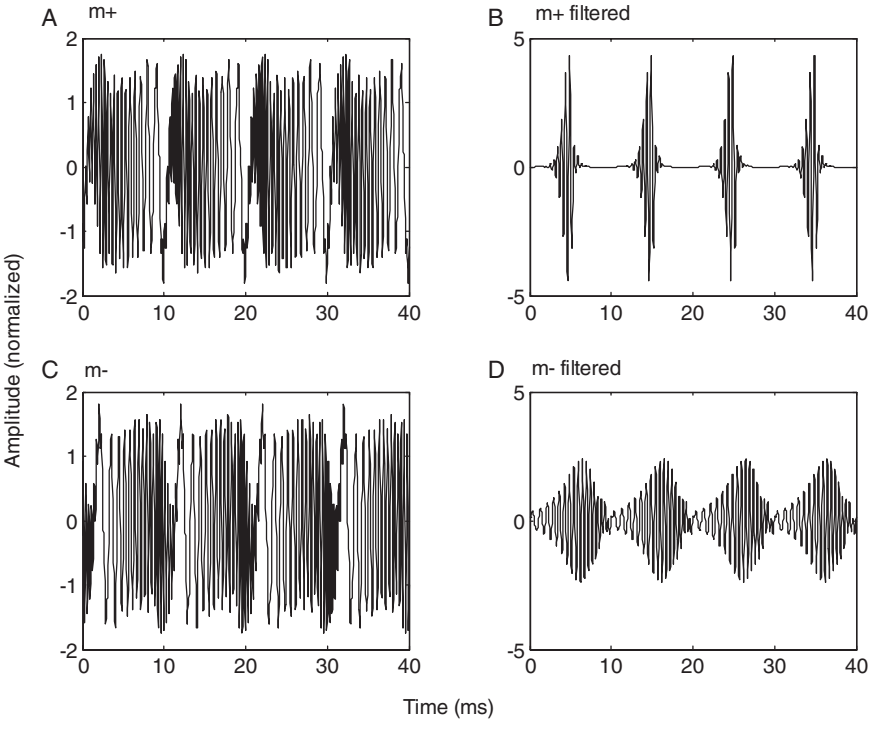


FIGURE 3.11. Schroeder-phase complexes. *A* and *C*: unfiltered versions of Schroeder-positive (m+) and Schroeder-negative (m–) waveforms. The complexes have a fundamental frequency (f_0) of 100 Hz and are comprised of components from 200 to 3,200 Hz. *B* and *D*: same waveforms passed through a filter centered at 2 kHz with a gammachirp magnitude response at 60 dB (Irino and Patterson 1997) and a phase response modified to simulate the phase curvature of the auditory filter (Oxenham and Dau 2001b). All waveforms were normalized to have a root mean square value of 1.

and m– waveforms have been filtered using the magnitude spectrum of a gammachirp filter (Irino and Patterson 1997), designed to simulate the auditory filter, and the phase response with a constant-phase curvature derived from psychophysical estimates of the phase response of the auditory filter (Oxenham and Dau 2001a,b). It is clear that the filtered m+ complex (Fig. 3.11*B*) has the more peaked temporal envelope, with very little energy in the masker valleys. This is because the positive phase curvature (or rate of change in group delay with frequency) of the m+ complex is compensated for by the negative phase curvature of the filter, leading to a filtered stimulus with a near-zero phase curvature (similar to that of a sine- or cosine-phase complex) and a commensurately peaky waveform.

It is straightforward to show that if two stimuli have the same rms value before instantaneous compression, the one with a more peaked, or more modulated, temporal envelope will have a lower rms value after compression. Therefore, if forward masking reflects the temporal integration of a compressed response to the masker, differences in masking effectiveness should be observable between maskers with highly modulated and unmodulated temporal envelopes. Specifically, m+ complexes, eliciting a peaky response, should produce less forward masking than m– complexes, which elicit a less modulated response at the basilar membrane. This hypothesis was tested by Carlyon and Datta (1997).

As predicted, Carlyon and Datta (1997) found that the m+ complexes were less effective forward maskers than were the m– complexes. For the mean data, m+ complexes produced about 10 dB less masking than the m– complexes. The same pattern was found when the signal level was held fixed and the masker level was varied; generally, the m+ complexes had to be presented at a higher level than the m– complexes to produce the same amount of masking. The exception was at low overall levels where both maskers were roughly equally effective. All these results are in qualitative agreement with predictions based on the effects of basilar membrane nonlinearity. At medium and high levels, basilar membrane compression produces a difference in the effectiveness of the two complexes, whereas at low levels, the basilar membrane acts more linearly, and so the difference in effectiveness is reduced. Unfortunately, in this case, qualitative agreement is not translated into quantitative correspondence with basilar membrane compression. When Carlyon and Datta attempted to simulate their results, they found that even raising the stimulus intensity to a power of 0.05 (or 0.1 in terms of amplitude) was not sufficiently compressive to account for the size of their observed effect. Such strong compression is consistent neither with the other behavioral measures of compression nor with the physiological estimates of basilar membrane compression. Other results that question whether basilar membrane nonlinearity can fully account for forward-masking differences using Schroeder complexes were presented by Ewert and Oxenham (2002). They showed that some threshold differences between m+ and m– maskers persisted even when all the masking components were well below the signal in frequency and so should have been processed linearly by the basilar membrane. At the present time, it remains unclear why these particular stimuli should not produce results in agreement with the predictions of basilar membrane compression. Carlyon and Datta (1997) suggested that some other form of (neural) compression may be playing a role. If this is so, however, it is not yet obvious why the effects of the hypothesized neural compression are not observed in other paradigms such as the additivity of forward and backward masking described above.

4.4 Temporal Integration in Simultaneous Masking

It has been known for many years that signal thresholds tend to decrease with increasing signal duration in quiet and in a background of continuous noise, at least up to durations of 200 ms (Hughes 1946; Feldtkeller and Oetinger 1956; Zwislocki 1960). Early model formulations suggested that stimulus intensity was simply integrated and that the integrator had a time constant of about 100–200 ms. Viemeister and Wakefield (1991) took a different approach. They postulated that the auditory system did not require an integrator with a long time constant and that instead thresholds decreased because information was combined over several brief observation periods, termed "multiple looks." The advantage of this approach is that it resolves the apparent discrepancy between the short time constant required to describe measures of temporal resolution, such as gap detection and forward masking, and the longer time constant required for temporal integration (e.g., de Boer 1985). Yet another approach was employed by Dau et al. (1996a,b), who implemented a flexible "template" in their model of auditory masking such that the integration window is matched to the signal characteristics.

Oxenham et al. (1997), following the approach of Viemeister and Wakefield (1991), postulated that detection thresholds within a period of 10–20 ms were mediated by "true" temporal integration, corresponding to a single "look," and that thresholds for signal durations longer than about 20 ms were the result of combining information across such looks. Integration over the period of about 20 ms was called "short-term integration." In this way, the temporal window approach could be extended to cover simultaneous as well as nonsimultaneous masking. One prediction that arises from this approach is that the measured amount of temporal integration (i.e., the rate at which thresholds decrease as a function of signal duration) should be greatest in level regions where basilar membrane compression is greatest (Penner 1978; see also Section 4.3.2). This prediction was confirmed: short-term temporal integration was greatest at medium levels and less at low and high levels, where basilar membrane compression is thought to be reduced. The results, together with model predictions, are shown in Figure 3.12. The predictions are from the temporal window model, assuming the same window as that used to describe forward masking and preceded by a nonlinearity with a compressive slope of 0.25 in the midlevel region and a linear slope of 1 in the low- and high-level regions, as was also used by Oxenham (2001) to simulate temporal integration in forward masking (see Fig. 3.10).

4.5 Simultaneous Masking in Harmonic Tone Complexes

The possible role of basilar membrane compression in forward masking using harmonic tone complexes, in particular with m+ and m– phases, was

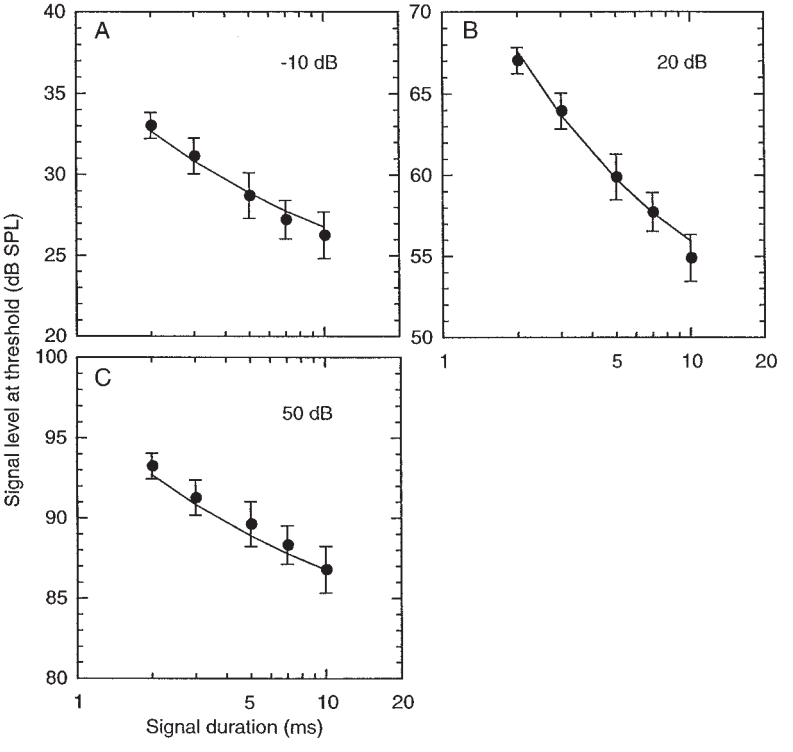


FIGURE 3.12. Mean thresholds for a 6.5-kHz signal as a function of signal duration. *A* – *C*: data from indicated masking noise spectrum levels. Solid lines: predictions of the temporal window model. (Data are from Oxenham et al. 1997.)

discussed above. There we concluded that although the results were qualitatively in agreement with the predictions, quantitative agreement had proved elusive. In simultaneous masking, the situation is more satisfying in that the results from a number of experiments can be understood, quantitatively, in terms of basilar membrane compression. It has been known for some time that m– maskers generally produce higher thresholds for simultaneously presented signals than do m+ maskers (Smith et al. 1986; Kohlrausch and Sander 1995). As mentioned above, this has been ascribed to the interaction between the phase properties of the stimulus and the phase response of the basilar membrane such that the waveform observed on the basilar membrane is more highly modulated, or more peaked, in response to an m+ than to an m– masker. In the original studies, the masking difference was ascribed to the differences in peakiness of the waveforms

and the ability of listeners to use information in the low-level epochs, or "valleys," of the more modulated masker. However, just as peripheral compression affects temporal resolution, it also strongly affects the interactions between a signal and a modulated masker. First, with a compressive nonlinearity, the peaks of a modulated waveform are attenuated more than the valleys, thereby reducing its long-term average level. Similarly, a signal presented in the valleys of the masker will be amplified with respect to the masker peaks, thereby enhancing its representation. In other words, peripheral compression would be expected to enhance the difference in masked thresholds between a highly modulated and an unmodulated masker waveform. This expectation was confirmed by Oxenham and Dau (2001a). Using the same parameters as in earlier studies of forward masking (Plack and Oxenham 1998; Oxenham and Plack 2000), they found that the temporal window model predicted masking differences between m+ and m– maskers of around 25 dB, in reasonably good agreement with the data. On the other hand, removing the compressive nonlinearity from the model resulted in predicted threshold differences between the m+ and m– maskers of less than 0.5 dB. This shows that peripheral compression is necessary to account for the observed masking differences, at least within the framework of the temporal window model.

One interesting finding in this context is that large phase effects can be found in masking for signal frequencies as low as 125 Hz (Oxenham and Dau 2001b). If, as implied by the modeling results, large masking differences require compression, then the results suggest that peripheral compression may continue down to places along the basilar membrane with CFs as low as 125 Hz. In contrast to measures of nonlinearity such as suppression and growth of masking, the masking difference between m+ and m– maskers does not require a nonlinearity that is frequency dependent. These results therefore provide indirect support for the idea, outlined in Section 3, that compression is still present in the apical region of the cochlea, corresponding to low CFs, but that apical compression is not frequency selective.

Thresholds for very brief signals depend on what point in the masker period they are presented. A plot of signal threshold as a function of its position within the masker period is known as a masking period pattern (MPP). Evidence that peripheral compression (or its absence) affects MPPs is presented by Bacon and Oxenham (Chapter 4).

5. Loudness Perception and Integration

A study by Viemeister and Bacon (1988) found that the magnitude estimation values of the loudness (L) for a 1-kHz tone could be related to sound pressure (P) by the equation $L = kP^p$, where p had an average value

close to 0.16 and k is simply a proportionality constant. The similarity between this value and the compression found on the basilar membrane led Yates et al. (1990) to suggest that loudness may be based on a simple coding of basilar membrane displacement at the CF. However, in a comprehensive survey of studies of loudness magnitude estimation, Hellman (1991) found that, on average, the exponent relating loudness to sound pressure was closer to 0.6 and that the value derived by Viemeister and Bacon (1988) fell more than three standard deviations away from the mean value across studies. Aside from anything else, this rather large discrepancy highlights some of the pitfalls of using subjective measures such as loudness judgments.

In a more recent study, Schlauch et al. (1998) revisited the topic of relating loudness to basilar membrane response. In particular, they assessed the idea that changes in loudness perception with hearing impairment could be accounted for by the changes in basilar membrane compression associated with cochlear damage (Yates 1990; Oxenham and Moore 1994; Moore and Glasberg 1997). This question is addressed by Bacon and Oxenham (Chapter 4). Here we examine whether the more typical exponent of around 0.6 can somehow be related to basilar membrane motion. In most models of loudness, the intensity of the incoming sound is compressed within each frequency channel (or critical band) in some way to arrive at what is often termed "specific loudness." These values of specific loudness (one for each critical band) are then summed to arrive at an overall loudness (Zwicker and Scharf 1965; Moore and Glasberg 1996; Moore et al. 1997). In order to arrive at an overall growth of loudness with an exponent of 0.6 (or 0.3 in terms of intensity; I is proportional to P^2), a somewhat more compressive function has to be used in each frequency channel. This is because as the sound level increases, the sound stimulates an increasing number of frequency channels (spread of excitation), thereby increasing the overall loudness percept. For instance, in Zwicker and Scharf's (1965) model, intensity is raised to the power 0.23.

Remember that for the additivity of nonsimultaneous masking, it had to be assumed that a quantity proportional to sound intensity, not pressure, was integrated. If we assume the same for loudness, then the exponent of 0.23 is in the range of exponents between 0.2 and 0.25 often associated with basilar membrane compression. In other words, a square law nonlinearity after basilar membrane compression but before integration is required. This idea has some physiological support. A number of studies have concluded that the inner hair cells may produce something akin to a squaring of basilar membrane displacement or velocity (Goodman et al. 1982; Yates 1990). Aside from the additivity of masking results, experiments involving binaural hearing, discussed in Section 6, also provide support for such a square law nonlinearity. In effect, basilar membrane intensity, not displacement or velocity, seems to be the more psychologically relevant quantity.

Most studies of loudness growth have assumed a simple power law as a nonlinearity. More linear growth of loudness is often observed near threshold but that has mostly been accounted for by adding a constant "internal noise," which has the effect of steepening the initial part of the loudness function (e.g., Zwislocki 1965). More recent data have questioned this approach. Using a combination of loudness matches between long and short tones and between single and multiple tones, Florentine et al. (1996) and Buus et al. (1997, 1998) concluded that loudness grows in proportion to stimulus intensity for levels up to about 20 dB SPL and then grows more compressively than usually assumed at medium levels, with an exponent of about 0.2, relative to intensity. The similarity between these values and those found for physiological and other psychophysical estimates of basilar membrane nonlinearity is striking.

6. Effects of Compression on Binaural Masking

When the same signal and masker are presented simultaneously to both ears, the detectability of the signal can be enhanced by altering the interaural properties of either the signal or the masker. For instance, thresholds for a tone in noise can be decreased by 20 dB or more simply by inverting the polarity of the signal in one ear. These types of phenomena rely on binaural interactions within the auditory system because listening through either ear alone will, of course, produce no benefit. The difference in thresholds between diotic (identical stimuli to both ears) and dichotic (e.g., NoSπ, where the noise is diotic and the signal is presented 180° or π radians out of phase) presentations is known as the binaural masking level difference (BMLD).

An extensive literature on BMLDs and related phenomena has developed over the years (for reviews, see Grantham 1995; Stern and Trahiotis 1995). A popular way of explaining BMLDs in noise maskers has been to consider the normalized interaural cross-correlation of the stimuli (Bernstein and Trahiotis 1996). Using this framework, van de Par and Kohlrausch (1998) investigated BMLDs in multiplied noise. Multiplied noise is obtained by multiplying a sinusoid with a low-pass Gaussian noise. The resulting noise has the same center frequency as the sinusoid and a bandwidth twice that of the low-pass noise. Its envelope and fine-structure properties differ from those of Gaussian noise, with one important feature being that it has regular zero crossings in the fine structure, corresponding to the zero crossings of the sinusoid. It also has more frequent dips in the temporal envelope than does Gaussian noise. van de Par and Kohlrausch (1998) found that the BMLDs obtained with multiplied noise were often considerably greater than those obtained in Gaussian noise. At high center frequencies, where subjects are thought to detect a decrease in the envelope cross-correlation (Bernstein and Trahiotis 1992), NoSπ thresholds were

9 dB lower than for Gaussian noise compared to only a 3-dB difference predicted by the envelope cross-correlation (van de Par and Kohlrausch 1995). This discrepancy was resolved if the envelopes of the stimuli were first subjected to an instantaneous compression. The best fit to the data was obtained by raising the amplitude to a power of 0.4, corresponding to raising the intensity to the power 0.2. They explained the importance of compression in terms of increasing the weight of information in the temporal valleys of the masker. In many ways, this is analogous to the situation with the monaural harmonic tone complexes discussed in Section 4.5.

Bernstein et al. (1999) examined NoSπ thresholds in Gaussian noise and so-called low-noise noise (Hartmann and Pumplin 1988; Kohlrausch et al. 1997). In two earlier studies (Eddins and Barber 1998; Hall et al. 1998), it had been shown that NoSπ thresholds in low-noise noise were higher than those in Gaussian noise with the same average intensity. Eddins and Barber (1998) concluded that the results could not be accounted for by considering differences in the interaural cross-correlation. However, Bernstein et al. (1999) showed that this conclusion was valid only for the square law nonlinearity assumed in the earlier work; when a compressive nonlinearity was applied to the envelope, the predictions were found to match the data rather well. Most interesting from our perspective, the best match to the data was obtained when the envelope was raised to a power of 0.46, corresponding to an exponent of 0.23 applied to the intensity.

Both these studies of binaural masking (van de Par and Kohlrausch 1998; Bernstein et al. 1999), using different paradigms with different subjects, resulted in estimates of compression that are very close to those found in the monaural literature described throughout this chapter. Again, assuming a square law nonlinearity following the basilar membrane, the results are very close to estimates of basilar membrane compression from physiological studies (Cooper, Chapter 2). Overall, it appears that the modeling of at least some aspects of binaural hearing requires the inclusion of a peripheral nonlinearity similar to that observed on the basilar membrane.

7. Summary

This chapter has reviewed many behavioral measures of auditory performance, which all seem to be influenced by peripheral compression. Peripheral compression widens the dynamic range of hearing and effectively improves temporal resolution by emphasizing the information within the temporal valleys of sounds.

The models described in this chapter have been extremely simple, incorporating only a single linear filter centered on the signal frequency, followed by a nonlinearity and, finally, a linear temporal integrator. The compression exponents required to best account for the data ranged from 0.16 to 0.25, which is in good agreement with physiological data from other mammals.

These values correspond to a compression ratio of between 4:1 and 6:1. Given the extreme complexity of the auditory system, it is perhaps surprising that such a simple scheme can provide reasonably accurate predictions of performance in many disparate psychoacoustic tasks. One perhaps obvious, but important, lesson to be taken from these studies is that peripheral processing can have profound influences on subsequent stages. For instance, the amount of compression influences the effective time constant of subsequent temporal processing (Penner 1978; Oxenham and Moore 1997), and the incorporation of compression can reverse conclusions on the viability of certain decision criteria (Bernstein et al. 1999). From that point of view, it is imperative that peripheral transformations are accurately modeled before any conclusions can be drawn about subsequent processing.

One future development, which will be important for describing the processing of more complex spectrotemporal stimuli than have been dealt with here, is the inclusion of a more realistic model of cochlear processing. Although it may not be necessary for a psychoacoustic model to incorporate all the micromechanical intricacies of the cochlea, the effects of such processing, such as two-tone suppression, should be reflected in the model. A first step toward this goal was recently taken by Plack et al. (2002). They incorporated a nonlinear model of effective cochlear processing (Lopez-Poveda and Meddis 2001; Meddis et al. 2001) within the temporal window model and were able to simulate many aspects of forward masking and frequency selectivity, including certain suppression effects.

Overall, the perceptual consequences of peripheral compression are profound. This conclusion is supported not only by the influence of compression in normal hearing but also by the effects of its reduction or absence in impaired hearing. How the auditory system functions with reduced or absent peripheral compression is the topic discussed by Bacon and Oxenham (Chapter 4).

Acknowledgments. The authors were supported by National Institute on Deafness and Other Communication Disorders Grants R01-DC-03909 (to A. J. Oxenham) and R01-DC-01376 (to S. P. Bacon). We thank our colleagues, especially Chris Plack and Brian Moore, for the many collaborations that resulted in the work described here. Chris Plack also provided helpful comments on a draft of this chapter. This chapter was written while A. J. Oxenham was a fellow at the Hanse Institute for Advanced Study in Delmenhorst, Germany.

References

Anderson DJ, Rose JE, Hind JE, Brugge JF (1970) Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: frequency and intensity effects. J Acoust Soc Am 49:1131–1139.

Bacon SP, Boden LN, Lee J, Repovsch JL (1999) Growth of simultaneous masking for $f_m < f_s$: effects of overall frequency and level. J Acoust Soc Am 106:341–350.
Bernstein LR, Trahiotis C (1992) Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies. J Acoust Soc Am 91: 306–316.
Bernstein LR, Trahiotis C (1996) The normalized correlation: accounting for binaural detection across center frequency. J Acoust Soc Am 100:3774–3784.
Bernstein LR, van de Par S, Trahiotis C (1999) The normalized interaural correlation: accounting for NoSπ thresholds obtained with Gaussian and "low-noise" masking noise. J Acoust Soc Am 106:870–876.
Buus S, Florentine M, Poulsen T (1997) Temporal integration of loudness, loudness discrimination, and the form of the loudness function. J Acoust Soc Am 101: 669–680.
Buus S, Muesch H, Florentine M (1998) On loudness at threshold. J Acoust Soc Am 104:399–410.
Carlyon RP (1988) The development and decline of forward masking. Hear Res 32: 65–80.
Carlyon RP, Datta AJ (1997) Excitation produced by Schroeder-phase complexes: evidence for fast-acting compression in the auditory system. J Acoust Soc Am 101:3636–3647.
Cooper NP, Yates GK (1994) Nonlinear input-output functions derived from the responses of guinea-pig cochlear nerve fibres: variations with characteristic frequency. Hear Res 78:221–234.
Dau T, Püschel D, Kohlrausch A (1996a) A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure. J Acoust Soc Am 99: 3615–3622.
Dau T, Püschel D, Kohlrausch A (1996b) A quantitative model of the 'effective' signal processing in the auditory system. II. Simulations and measurements. J Acoust Soc Am 99:3623–3631.
de Boer E (1985) Auditory time constants: a paradox? In: Michelsen A (ed) Time Resolution in Auditory Systems. Berlin: Springer-Verlag, pp. 141–158.
Delgutte B (1990) Physiological mechanisms of psychophysical masking: observations from auditory-nerve fibers. J Acoust Soc Am 87:791–809.
Eddins DA, Barber LE (1998) The influence of stimulus envelope and fine structure on the binaural masking level difference. J Acoust Soc Am 103:2578–2589.
Egan JP, Hake HW (1950) On the masking pattern of a simple auditory stimulus. J Acoust Soc Am 22:622–630.
Elliott LL (1962) Backward masking: monotic and dichotic conditions. J Acoust Soc Am 34:1108–1115.
Elliott LL (1969) Masking of tones before, during, and after brief silent periods in noise. J Acoust Soc Am 45:1277–1279.
Ewert S, Oxenham AJ (2002) Effects of masker phase curvature for on- and off-frequency simultaneous and nonsimultaneous maskers. Assoc Res Otolaryngol Abstr 25:72.
Fastl H (1976a) Temporal masking effects. I. Broad band noise masker. Acustica 35:287–302.
Fastl H (1976b) Temporal masking effects. II. Critical band noise masker. Acustica 36:317–331.

Fastl H (1979) Temporal masking effects. III. Pure tone masker. Acustica 43:283–294.
Feldtkeller R, Oetinger R (1956) Die Hörbarkeitsgrenzen von Impulsen verschiedener Dauer. Acustica 6:489–493.
Fletcher H, Munson WA (1937) Relation between loudness and masking. J Acoust Soc Am 9:1–10.
Florentine M, Buus S (1981) An excitation-pattern model for intensity discrimination. J Acoust Soc Am 70:1646–1654.
Florentine M, Buus S, Poulsen T (1996) Temporal integration of loudness as a function of level. J Acoust Soc Am 99:1633–1644.
Gerken GM, Bhat VKH, Hutchinson-Clutter M (1990) Auditory temporal integration and the power function model. J Acoust Soc Am 88:767–778.
Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. Hear Res 47:103–138.
Goodman DA, Smith RL, Chamberlain SC (1982) Intracellular and extracellular responses in the organ of Corti of the gerbil. Hear Res 7:161–179.
Gorga MP, Nelson K, Davis T, Dorn PA, Neely ST (2000) Distortion product otoacoustic emission test performance when both 2f1-f2 and 2f2-f1 are used to predict auditory status. J Acoust Soc Am 107:2128–2135.
Gralla G (1992) Wahrnehmungskriterien bei Simultan-und Nachhörschwellenmessungen (Detection cues in simultaneous and post masking experiments). Acustica 77:243–251.
Grantham DW (1995) Spatial hearing and related phenomena. In: Moore BCJ (ed) Handbook of Perception and Cognition, Vol 6. New York: Academic, pp. 297–345.
Green DM (1967) Additivity of masking. J Acoust Soc Am 41:1517–1525.
Green DM, Swets JA (1966) Signal Detection Theory and Psychophysics. New York: Krieger.
Hall JW, Grose JH, Hartmann WM (1998) The masking-level difference in low-noise noise. J Acoust Soc Am 103:2573–2577.
Hartmann WM, Pumplin J (1988) Noise power fluctuations and the masking of sine signals. J Acoust Soc Am 83:2277–2289.
Hellman RP (1991) Loudness measurement by magnitude scaling: implications for intensity coding. In: Bolanowski SJ, Gescheider GA (eds) Ratio Scaling of Psychological Magnitude: In Honor of the Memory of S. S. Stevens. Hillsdale, NJ: Erlbaum Associates, pp. 215–228.
Helmholtz HLF (1885/1954) On the Sensations of Tone. New York: Dover.
Hicks ML, Bacon SP (1999) Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing. J Acoust Soc Am 105: 326–338.
Houtgast T (1972) Psychophysical evidence for lateral inhibition in hearing. J Acoust Soc Am 51:1885–1894.
Hughes JW (1946) The threshold of audition for short periods of stimulation. Proc R Soc Lond Ser B 133:486–490.
Humes LE, Jesteadt W (1989) Models of the additivity of masking. J Acoust Soc Am 85:1285–1294.
Humes LE, Espinoza-Varas B, Watson CS (1988) Modeling sensorineural hearing loss. I. Model and retrospective evaluation. J Acoust Soc Am 83:188–202.
Irino T, Patterson RD (1997) A time-domain, level-dependent auditory filter: the gammachirp. J Acoust Soc Am 101:412–419.

Jesteadt W, Bacon SP, Lehman JR (1982) Forward masking as a function of frequency, masker level, and signal delay. J Acoust Soc Am 71:950–962.
Johnson-Davies D, Patterson RD (1979) Psychophysical tuning curves: restricting the listening band to the signal region. J Acoust Soc Am 65:765–770.
Kiang NYS, Watanabe T, Thomas EC, Clark LF (1965) Discharge Patterns of Single Fibres in the Cat's Auditory Nerve. Cambridge, MA: MIT Press.
Kidd G, Feth LL (1981) Patterns of residual masking. Hear Res 5:49–67.
Kidd G, Feth LL (1982) Effects of masker duration in pure-tone forward masking. J Acoust Soc Am 72:1384–1386.
Kohlrausch A, Sander A (1995) Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. J Acoust Soc Am 97: 1817–1829.
Kohlrausch A, Fassel R, van der Heijden M, Kortekaas R, van de Par S, Oxenham AJ, Püschel D (1997) Detection of tones in low-noise noise: further evidence for the role of envelope fluctuations. Acta Acustica 83:659–669.
Lentz JJ, Leek MR (2001) Psychophysical estimates of cochlear phase response: masking by harmonic complexes. J Assoc Res Otolaryngol 2:408–422.
Leshowitz B, Wightman FL (1972) On the importance of considering the signal's frequency spectrum: some comments on Macmillan's "Detection and recognition of increments and decrements in auditory intensity" experiment. Percept Psychophys 12:209–210.
Levitt H (1971) Transformed up-down methods in psychoacoustics. J Acoust Soc Am 49:467–477.
Lopez-Poveda EA, Meddis R (2001) A human nonlinear cochlear filterbank. J Acoust Soc Am 110:3107–3118.
Lopez-Poveda EA, Plack CJ, Meddis R (2003) Cochlear nonlinearity between 500 and 8,000 Hz in listeners with normal hearing. J Acoust Soc Am 113:951–960.
Macmillan NA, Creelman CD (1991) Detection Theory: A User's Guide. New York: Cambridge University Press.
Meddis R, O'Mard LP, Lopez-Poveda EA (2001) A computational algorithm for computing nonlinear auditory frequency selectivity. J Acoust Soc Am 109:2852–2861.
Moore BCJ (1985) Additivity of simultaneous masking, revisited. J Acoust Soc Am 78:488–494.
Moore BCJ, Glasberg BR (1983) Growth of forward masking for sinusoidal and noise maskers as a function of signal delay: implications for suppression in noise. J Acoust Soc Am 73:1249–1259.
Moore BCJ, Glasberg BR (1985) The danger of using narrowband noise maskers to measure suppression. J Acoust Soc Am 77:2137–2141.
Moore BCJ, Glasberg BR (1996) A revision of Zwicker's loudness model. Acustica 82:335–345.
Moore BCJ, Glasberg BR (1997) A model of loudness perception applied to cochlear hearing loss. Aud Neurosci 3:289–311.
Moore BCJ, Oxenham AJ (1998) Psychoacoustic consequences of compression in the peripheral auditory system. Psychol Rev 105:108–124.
Moore BCJ, Glasberg BR, Roberts B (1984) Refining the measurement of psychophysical tuning curves. J Acoust Soc Am 76:1057–1066.
Moore BCJ, Glasberg BR, Plack CJ, Biswas AK (1988) The shape of the ear's temporal window. J Acoust Soc Am 83:1102–1116.

Moore BCJ, Glasberg BR, Donaldson E, McPherson T, Plack CJ (1989) Detection of temporal gaps in sinusoids by normally hearing and hearing-impaired subjects. J Acoust Soc Am 85:1266–1275.
Moore BCJ, Peters RW, Glasberg BR (1993) Effects of frequency on the detection of decrements and increments in sinusoids. J Acoust Soc Am 94:3190–3198.
Moore BCJ, Glasberg BR, Baer T (1997) A model for the prediction of thresholds, loudness, and partial loudness. J Aud Eng Soc 45:224–240.
Neff DL (1985) Stimulus parameters governing confusion effects in forward masking. J Acoust Soc Am 78:1966–1976.
Neff DL (1986) Confusion effects with sinusoidal and narrowband-noise forward maskers. J Acoust Soc Am 79:1519–1529.
Nelson DA, Schroder AC, Wojtczak M (2001) A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners. J Acoust Soc Am 110:2045–2064.
O'Loughlin BJ, Moore BCJ (1981) Off-frequency listening: effects on psychoacoustical tuning curves obtained in simultaneous and forward masking. J Acoust Soc Am 69:1119–1125.
Oxenham AJ (1997) Increment and decrement detection in sinusoids as a measure of temporal resolution. J Acoust Soc Am 102:1779–1790.
Oxenham AJ (2001) Forward masking: adaptation or integration? J Acoust Soc Am 109:732–741.
Oxenham AJ, Dau T (2001a) Reconciling frequency selectivity and phase effects in masking. J Acoust Soc Am 110:1525–1538.
Oxenham AJ, Dau T (2001b) Towards a measure of auditory-filter phase response. J Acoust Soc Am 110:3169–3178.
Oxenham AJ, Moore BCJ (1994) Modeling the additivity of nonsimultaneous masking. Hear Res 80:105–118.
Oxenham AJ, Moore BCJ (1995) Additivity of masking in normally hearing and hearing-impaired subjects. J Acoust Soc Am 98:1921–1934.
Oxenham AJ, Moore BCJ (1997) Modeling the effects of peripheral nonlinearity in normal and impaired hearing. In: Jesteadt W (ed) Modeling Sensorineural Hearing Loss. Hillsdale, NJ: Erlbaum Associates, pp. 273–288.
Oxenham AJ, Plack CJ (1997) A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. J Acoust Soc Am 101: 3666–3675.
Oxenham AJ, Plack CJ (2000) Effects of masker frequency and duration in forward masking: further evidence for the influence of peripheral nonlinearity. Hear Res 150:258–266.
Oxenham AJ, Moore BCJ, Vickers DA (1997) Short-term temporal integration: evidence for the influence of peripheral compression. J Acoust Soc Am 101: 3676–3687.
Patterson JH (1971) Additivity of forward and backward masking as a function of signal frequency. J Acoust Soc Am 50:1123–1125.
Patterson RD (1976) Auditory filter shapes derived with noise stimuli. J Acoust Soc Am 59:640–654.
Patterson RD, Nimmo-Smith I (1980) Off-frequency listening and auditory filter asymmetry. J Acoust Soc Am 67:229–245.
Penner MJ (1974) Effect of masker duration and masker level on forward and backward masking. J Acoust Soc Am 56:179–182.

Penner MJ (1977) Detection of temporal gaps in noise as a measure of the decay of auditory sensation. J Acoust Soc Am 61:552–557.
Penner MJ (1978) A power law transformation resulting in a class of short-term integrators that produce time-intensity trades for noise bursts. J Acoust Soc Am 63:195–201.
Penner MJ (1980) The coding of intensity and the interaction of forward and backward masking. J Acoust Soc Am 67:608–616.
Penner MJ, Shiffrin RM (1980) Nonlinearities in the coding of intensity within the context of a temporal summation model. J Acoust Soc Am 67:617–627.
Plack CJ, Moore BCJ (1990) Temporal window shape as a function of frequency and level. J Acoust Soc Am 87:2178–2187.
Plack CJ, Moore BCJ (1991) Decrement detection in normal and impaired ears. J Acoust Soc Am 90:3069–3076.
Plack CJ, O'Hanlon CG (2003) Forward masking additivity and auditory compression at low and high frequencies. J Assoc Res Otolaryngol (in press).
Plack CJ, Oxenham AJ (1998) Basilar-membrane nonlinearity and the growth of forward masking. J Acoust Soc Am 103:1598–1608.
Plack CJ, Oxenham AJ (2000) Basilar-membrane nonlinearity estimated by pulsation threshold. J Acoust Soc Am 107:501–507.
Plack CJ, Oxenham AJ, Drga V (2002) Linear and nonlinear processes in temporal masking. Acustica/Acta Acustica 88:348–358.
Plomp R (1964) The rate of decay of auditory sensation. J Acoust Soc Am 36: 277–282.
Pollack I (1964) Interaction of forward and backward masking. J Aud Res 4:63–67.
Poulton EC (1977) Quantitative subjective assessments are almost always biased, sometimes completely misleading. Br J Psychol 68:409–425.
Recio A, Rhode WS (2000) Basilar membrane responses to broadband stimuli. J Acoust Soc Am 108:2281–2298.
Relkin EM, Turner CW (1988) A reexamination of forward masking in the auditory nerve. J Acoust Soc Am 84:584–591.
Rhode WS, Cooper NP (1996) Nonlinear mechanics in the apical turn of the chinchilla cochlea in vivo. Aud Neurosci 3:101–121.
Rhode WS, Recio A (2000) Study of mechanical motions in the basal region of the chinchilla cochlea. J Acoust Soc Am 107:3317–3332.
Robinson CE, Pollack I (1973) Interaction between forward and backward masking: a measure of the integrating period of the auditory system. J Acoust Soc Am 53: 1313–1316.
Ruggero MA, Robles L, Rich NC (1992a) Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression. J Neurophysiol 68:1087–1099.
Ruggero MA, Robles L, Rich NC, Recio A (1992b) Basilar membrane responses to two-tone and broadband stimuli. Philos Trans R Soc Lond B Biol Sci 336:307–315.
Sachs MB, Kiang NYS (1968) Two-tone inhibition in auditory nerve fibers. J Acoust Soc Am 43:1120–1128.
Sachs MB, Young ED (1980) Effects of nonlinearities on speech encoding in the auditory nerve. J Acoust Soc Am 68:858–875.
Schlauch RS, DiGiovanni JJ, Ries DT (1998) Basilar membrane nonlinearity and loudness. J Acoust Soc Am 103:2010–2020.

Schöne P (1977) Nichtlinearitäten im Mithörschwellen-Tonheitsmuster von Sinustönen. Acustica 37:37–44.
Schöne P (1979) Mithörschwellen-Tonmuster maskierender Sinustöne. Acustica 43:197–204.
Schouten JF (1940) The residue and the mechanism of hearing. Proc Kon Akad Wetenschap 43:991–999.
Schroeder MR (1970) Synthesis of low peak-factor signals and binary sequences with low autocorrelation. IEEE Trans Inf Theory 16:85–89.
Shannon RV (1976) Two-tone unmasking and suppression in a forward masking situation. J Acoust Soc Am 59:1460–1470.
Smith BK, Sieben UK, Kohlrausch A, Schroeder MR (1986) Phase effects in masking related to dispersion in the inner ear. J Acoust Soc Am 80:1631–1637.
Smith RL (1977) Short-term adaptation in single auditory nerve fibers: some poststimulatory effects. J Neurophysiol 40:1098–1112.
Smith RL (1979) Adaptation, saturation, and physiological masking in single auditory-nerve fibers. J Acoust Soc Am 65:166–178.
Stern RM, Trahiotis C (1995) Models of binaural interaction. In: Moore BCJ (ed) Handbook of Perception and Cognition. Vol 6. Hearing. New York: Academic, pp. 347–386.
Stevens SS (1957) On the psychophysical law. Psychol Rev 64:153–181.
Thurlow WR, Rawlings IL (1959) Discrimination of number of simultaneous sounding tones. J Acoust Soc Am 31:1332–1336.
van de Par S, Kohlrausch A (1995) Analytical expressions for the envelope correlation of certain narrow-band stimuli. J Acoust Soc Am 98:3157–3169.
van de Par S, Kohlrausch A (1998) Diotic and dichotic detection using multiplied-noise maskers. J Acoust Soc Am 103:2100–2110.
van der Heijden M, Kohlrausch A (1995) The role of envelope fluctuations in spectral masking. J Acoust Soc Am 97:1800–1807.
Viemeister NF, Bacon SP (1988) Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones. J Acoust Soc Am 84:172–178.
Viemeister NF, Wakefield GH (1991) Temporal integration and multiple looks. J Acoust Soc Am 90:858–865.
Warren RM (1970) Elimination of biases in loudness judgements for tones. J Acoust Soc Am 48:1397–1403.
Watson CS, Gengel RW (1969) Signal duration and signal frequency in relation to auditory sensitivity. J Acoust Soc Am 46:989–997.
Wegel RL, Lane CE (1924) The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear. Phys Rev 23:266–285.
Wilson RH, Carhart R (1971) Forward and backward masking: interactions and additivity. J Acoust Soc Am 49:1254–1263.
Yates GK (1990) Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions. Hear Res 50:145–162.
Yates GK, Winter IM, Robertson D (1990) Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range. Hear Res 45:203–220.
Zurek PM (1981) Spontaneous narrowband acoustic signals emitted by human ears. J Acoust Soc Am 69:514–523.
Zwicker E (1970) Masking and psychological excitation as consequences of the ear's frequency analysis. In: Plomp R, Smoorenburg GF (eds) Frequency

Analysis and Periodicity Detection in Hearing. Leiden, The Netherlands: Sijthoff, pp. 376–394.
Zwicker E (1984) Dependence of post-masking on masker duration and its relation to temporal effects in loudness. J Acoust Soc Am 75:219–223.
Zwicker E, Scharf B (1965) A model of loudness summation. Psychol Rev 72:3–26.
Zwislocki JJ (1960) Theory of temporal auditory summation. J Acoust Soc Am 32:1046–1060.
Zwislocki JJ (1965) Analysis of some auditory characteristics. In: Luce RD, Bush RR, Galanter E (eds) Handbook of Mathematical Psychology. New York: John Wiley and Sons, pp. 1–97.
Zwislocki JJ, Pirodda E, Rubin H (1959) On some poststimulatory effects at the threshold of audibility. J Acoust Soc Am 31:9–14.

4
Psychophysical Manifestations of Compression: Hearing-Impaired Listeners

Sid P. Bacon and Andrew J. Oxenham

1. Introduction

This chapter is concerned with how cochlear hearing loss affects various psychophysical measures thought to reflect compression in the auditory periphery and, in particular, at the basilar membrane in the cochlea. As discussed by Oxenham and Bacon (Chapter 3), there are several behavioral measures that appear to be closely linked to basilar membrane compression. Because permanent sensorineural hearing loss usually involves damage to the cochlea and, in particular, to the sensory cells (the outer hair cells) that are responsible for this fast-acting compression (see Cooper, Chapter 2), studies of individuals with sensorineural hearing loss can provide converging evidence for the role of basilar membrane compression in these behavioral measures.

With this in mind, this chapter focuses on many of the topics that were discussed by Oxenham and Bacon (Chapter 3), thus allowing a direct comparison between the psychophysical results obtained in normal-hearing listeners to those obtained in hearing-impaired listeners. For the most part, these hearing-impaired listeners have a permanent sensorineural hearing loss thought to be of cochlear origin. In some cases, however, results from individuals with a temporary loss caused by either intense sound exposure or aspirin ingestion are described; both are known to adversely affect outer hair cells, thus providing a convenient experimental model for examining the role of outer hair cells in auditory perception.

2. Growth of Masking

Masking is often used to measure the frequency-resolving capabilities of the auditory system (Wegel and Lane 1924; Fletcher 1940; Egan and Hake 1950; Small 1959; Houtgast 1973; Patterson 1976; Glasberg and Moore 1990; Rosen et al. 1998; for a review, see Moore 1993). The frequency selectivity or resolution measured psychophysically in humans is similar to the tuning

measured, say, at the level of the auditory nerve or basilar membrane in other mammals (Kiang et al. 1965; Ruggero 1992; Narayan et al. 1998). Generally, comparisons between animal physiological studies and human behavioral studies should be treated with caution. However, a recent study using both otoacoustic emissions and behavioral masking thresholds to independently estimate frequency tuning in humans (Shera et al. 2002) supports the widely held belief that the properties of the basilar membrane are responsible for behavioral auditory frequency selectivity (Moore 1986). This suggests that psychophysical masking can be used to gain some insights into the processing at the basilar membrane.

Most masking studies are concerned with frequency analysis. The emphasis of those studies has been on the frequency or spectral characteristics of the masker and signal. Another aspect, which provided some of the earliest indications of auditory nonlinearity (Wegel and Lane 1924), is how masking properties change with intensity. Inasmuch as masking reflects, in part, the response of the basilar membrane, these psychophysical experiments can provide a noninvasive estimate of how the response of the membrane changes with stimulus level. This section concentrates on how the masking of one pure tone grows or increases as a function of the level of another pure tone (or narrowband noise). It is particularly concerned with the specific condition in which the frequency of the masker is lower than the frequency of the signal. In this situation, it is assumed that the listener detects the signal at a place along the basilar membrane with a characteristic frequency (CF) or best frequency (BF) corresponding (at least approximately) to the frequency of that signal and that the threshold for the signal in the presence of the masker corresponds to some fixed signal-to-masker ratio in terms of an internal response. The growth of masking will then depend on how the response to the signal and the response to the masker grow at the place where the signal is detected. Based on basilar membrane measurements (e.g., Ruggero et al. 1997; see Cooper, Chapter 2), when the masker frequency is much lower than the signal frequency, the growth of response to the masker will be linear and that to the signal will be compressive (see Oxenham and Bacon, Chapter 3, for more details). This is illustrated schematically in Figure 4.1*A*, where the relative response of the basilar membrane at the signal frequency place (i.e., at the CF of the measurement place) is shown as a function of stimulus level separately for a masker (dashed line) and for a signal (solid line). The function for the masker is shifted to the right, to higher stimulus levels, because the masker is less effective than the signal in generating a response at the place corresponding to the signal. To keep the signal-to-masker ratio constant in terms of the internal response, the signal level will need to be increased more than the masker level. This is illustrated in Figure 4.1*A* by comparing the increase in level (x-axis) for both the masker and signal in order to yield a given change in basilar membrane response (y-axis). As shown in Figure 4.1*B*,*C*, the growth-of-masking function for this condition will be either relatively

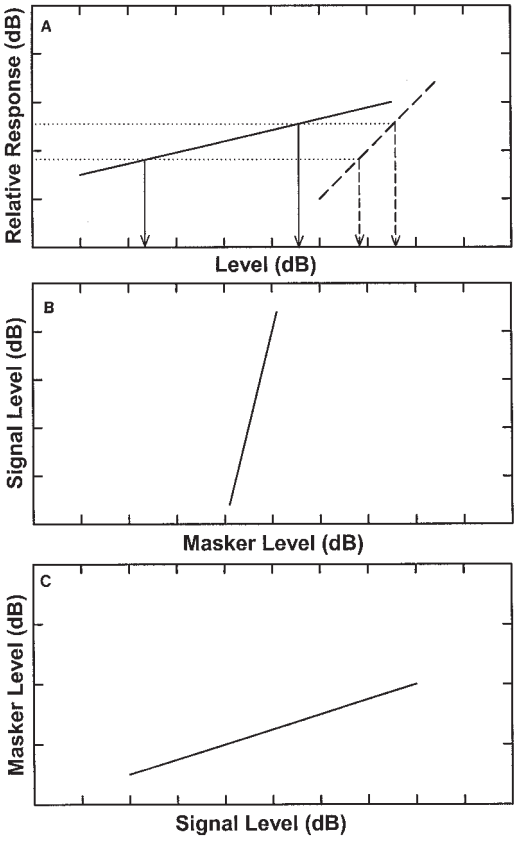


FIGURE 4.1. *A*: schematic representation of the relative response of the basilar membrane at the place corresponding to the signal frequency as a function of the level of the signal (solid line) and of a much lower frequency masker (dashed line). The response to the signal is compressive and the response to the masker is linear. Thus, to elicit a given change in basilar membrane response (dotted lines), the signal level must be changed more than the masker level. *B* and *C*: masking functions that would be expected if masking were determined solely by the basilar membrane and, in particular, if the masked threshold represented a given signal-to-masker ratio in terms of basilar membrane response. *B*: signal level plotted vs. masker level. *C*: masker level plotted vs. signal level.

steep if signal level is plotted as a function of masker level (Fig. 4.1*B*) or relatively shallow if masker level is plotted as a function of signal level (Fig. 4.1*C*). The slopes of the masking functions in Figure 4.1*B*,*C* are inversely related to one another. As discussed by Oxenham and Bacon (Chapter 3), the slope of the masking function provides an estimate of the degree of compression observed at the basilar membrane in response to the signal,

particularly when the signal and masker do not overlap in time. Even when they do overlap, the masking function is probably still influenced by compression, and thus it is worthwhile to consider the effects of hearing loss on the growth of masking obtained in both simultaneous and nonsimultaneous masking.

2.1 Simultaneous Masking

In normal-hearing listeners, when the masker frequency is more than half an octave or so below the signal frequency, the slope of the growth-of-masking function in simultaneous masking ranges anywhere from about 1.5 to 2.5 when signal level (or amount of masking) is plotted as a function of masker level (e.g., Wegel and Lane 1924; Egan and Hake 1950; Schöne 1977; Bacon et al. 1999). According to the arguments outlined above and by Oxenham and Bacon (Chapter 3), this nonlinear growth of masking is thought to reflect the fact that the internal response to the masker (at the place where the signal is detected) will grow linearly, whereas the response to the signal will grow compressively. In individuals with a cochlear hearing loss, the slope of the masking function is consistently less than it is in listeners with normal hearing, with values less than or equal to 1.0 (Smits and Duifhuis 1982; Stelmachowicz et al. 1987; Murnane and Turner 1991; Nelson and Schroder 1997). The slope tends to decrease with increasing hearing loss and, furthermore, appears to depend only on the hearing loss at the signal frequency (Murnane and Turner 1991; Nelson and Schroder 1997). In other words, it appears to depend only on the integrity of the cochlea at the place (the CF) corresponding to the signal frequency. The more gradual slope in hearing-impaired listeners is consistent with a decrease in or loss of basilar membrane compression resulting from cochlear damage. With a complete loss of compression, the internal response should grow linearly for the signal (Cooper, Chapter 2) as it does for the masker, and hence the slope of the masking function should be 1.0. The fact that the slope is often less than 1.0 suggests that the slope of these functions probably does not provide an accurate quantitative estimate of the internal growth of response to the signal. There are at least two possible explanations for this (see Oxenham and Bacon, Chapter 3, for a more detailed discussion). One is related to off-frequency listening (Johnson-Davies and Patterson 1979). If the place along the basilar membrane where the signal was detected were to change as the masker and signal levels increased, to take advantage of a better signal-to-masker ratio, then the amount of masking might increase less with increasing masker level than it would if the signal were detected at only one location. None of the studies of growth of masking in hearing-impaired subjects has used a noise to restrict off-frequency listening.

Another possible explanation for the discrepancy between the measurements of response growth at the basilar membrane and the estimates of

response growth based on simultaneous-masking results has to do with the stimulus presentation. The physiological data are obtained with individual tones presented in isolation, whereas the psychophysical data are obtained with two tones presented simultaneously. It is well known that complex interactions, such as suppression, can exist in the cochlea in response to the simultaneous presentation of more than one sound. These interactions, which are a consequence of the normally functioning outer hair cells, could decrease the slope of the masking function in simultaneous masking (Oxenham and Plack 1997; Bacon et al. 1999; Oxenham and Bacon, Chapter 3). Although this argument pertains particularly to normal-hearing subjects with normally functioning outer hair cells, it could also apply to a lesser extent to hearing-impaired subjects with at least some residual outer hair cell function. To avoid such interactions, the masker and signal can be presented nonsimultaneously so that there is no overlap in time between the two.

2.2 Nonsimultaneous Masking

As discussed by Oxenham and Bacon (Chapter 3), Oxenham and Plack (1997) measured growth-of-masking functions in forward masking for conditions in which the masker frequency was one octave below the signal frequency. They also employed a background noise to limit off-frequency listening. When their results were plotted as masker level versus signal level, the slope of the functions over a range of moderate signal levels was slightly less than 0.2 for their normal-hearing subjects (see Fig. 3.4 in Oxenham and Bacon, Chapter 3). When the data are plotted on these coordinates, the slope of the function provides a direct estimate of the underlying growth of response to the signal at the basilar membrane. Their mean slope value of 0.16 is quantitatively consistent with measures of response growth at the basilar membrane (Cooper, Chapter 2). It corresponds to a compression ratio of about 6:1, meaning that a 6-dB change in the input results in only a 1-dB change in the output. For the three hearing-impaired subjects that Oxenham and Plack (1997) tested, the functions were considerably steeper, as shown in Figure 4.2. For both signal frequencies, the functions have a slope close to 1.0 (i.e., close to the linear growth depicted in Fig. 4.2*A*,*B*, dashed lines). These results are consistent with the expectation that cochlear hearing loss reduces or eliminates basilar membrane compression and thus the compressive growth of response to the signal.

Although the linear slope for the subjects with cochlear hearing loss strongly suggests a loss of compression, it is important to note that most of Oxenham and Plack's (1997) results were restricted to relatively high signal levels where the masking function can be linear even in normal-hearing subjects (see Fig. 3.4 in Oxenham and Bacon, Chapter 3). Their choice of signal levels was an inevitable consequence of testing individuals with a sizeable hearing loss (lower signal levels would not be detected, even

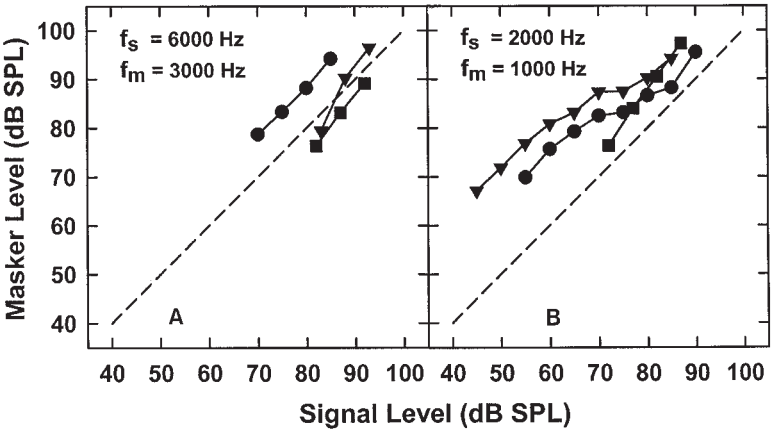


FIGURE 4.2. Growth-of-masking functions obtained in forward masking shown separately for 3 subjects with a permanent cochlear hearing loss. The 4-ms signal was presented shortly after the offset of the 104-ms masker. The frequency of the signal (f_s) was 6,000 (*A*) or 2,000 Hz (*B*). The frequency of the masker (f_m) was 1 octave below the frequency of the signal. Dashed line: linear growth. SPL, sound pressure level. (Data are from Oxenham and Plack 1997.)

without a masker). Individuals with less severe losses could be tested at lower levels, but the less severe their loss, the less likely they are to have sufficient cochlear damage to convincingly alter the slope of the masking function. A somewhat similar problem exists with the new procedure introduced by Nelson et al. (2001). In this procedure, which was described by Oxenham and Bacon (Chapter 3), the signal level is fixed and the masker level necessary to mask the signal is determined as a function of the delay between the offset of the masker and the onset of the signal. The response growth at the signal frequency place is estimated by comparing the results for a masker equal in frequency to the signal (on-frequency condition) with those for a masker well below the signal in frequency (off-frequency condition). Although the signal can be presented at one relatively moderate level (depending, of course, on the magnitude of the hearing loss), the maskers will likely need to be at high levels where the estimated response growth is linear even in normal-hearing subjects (Nelson et al. 2001). Thus, although Nelson et al. estimated the response growth to be compressive in their normal-hearing subjects and linear in their hearing-impaired subjects, it is unclear whether the linear growth in the impaired subjects reflects a true impairment or simply the high masker levels.

One way to avoid the problem of high levels is to test individuals with aspirin-induced hearing loss. When taken in moderate doses, aspirin (acetyl-

salicylic acid) can cause a mild temporary hearing loss by affecting the outer hair cells in the cochlea (Stypulkowski 1990; Shehata et al. 1991; Russell and Schauz 1995; Kakehata and Santos-Sacchi 1996; for a review, see Cazals 2000). Importantly, even small amounts of aspirin-induced hearing loss can have noticeable behavioral consequences (e.g., McFadden et al. 1984; Carlyon and Butt 1993; Hicks and Bacon 1999). This suggests that it might be possible to evaluate the role of basilar membrane compression in the growth of masking over a wide range of signal levels in individuals with temporary cochlear damage.

Hicks and Bacon (1999, 2000) measured the influence of aspirin on masking functions in forward masking. In one study (Hicks and Bacon 1999), two signal frequencies (750 and 3,000 Hz) were used. For each, the masker was lower in frequency than the signal. The results were plotted in terms of the amount of masking versus masker level. Aspirin reduced the slope of the masking function by an amount that increased with increasing temporary hearing loss. Thus the decrease in the amount of basilar membrane compression estimated psychophysically was directly related to the degree of aspirin-induced hearing loss. In their other study (Hicks and Bacon 2000), the 4,000-Hz signal was paired with a pure-tone masker whose frequency ranged from 2,000 to 4,000 Hz. In that study (2000), although not in their previous study (1999), Hicks and Bacon employed a noise to limit off-frequency listening. Their masking functions were plotted in terms of masker level versus signal level (for an example, see Fig. 1.5 in Bacon, Chapter 1). Some of the resulting slope values for each of their five subjects are shown in Figure 4.3 as a function of the frequency ratio between the signal and masker. Before aspirin ingestion (Fig. 4.3, open symbols), the slope of the masking function decreased with increasing frequency ratio, reaching an asymptote at a ratio of 1.8 (also see Nelson et al. 2001). This presumably reflects that at smaller ratios where the masker is much closer in frequency to the signal, the response to the masker at the place corresponding to the signal is at least somewhat compressive, whereas at larger ratios, it is probably linear (Ruggero et al. 1997; Cooper, Chapter 2). At ratios of 1.8 and 2.0, the slope is about 0.2, consistent with the masking results of Oxenham and Plack (1997). After aspirin ingestion (Fig. 4.3, solid symbols), which caused an average hearing loss of 8.2 dB at the 4,000-Hz signal frequency, the slope of the masking function increased, implying reduced basilar membrane compression. At a ratio of 1.2, the average slope value increased from 0.45 to 0.61, whereas at a ratio of 1.8, it increased from 0.25 to 0.43. Because aspirin directly affects the outer hair cells, which are responsible for basilar membrane compression, the results in Figure 4.3 lend strong support to the assertion that the slope of the masking function when the masker frequency is well below the signal frequency reflects basilar membrane compression at the place corresponding to the signal.

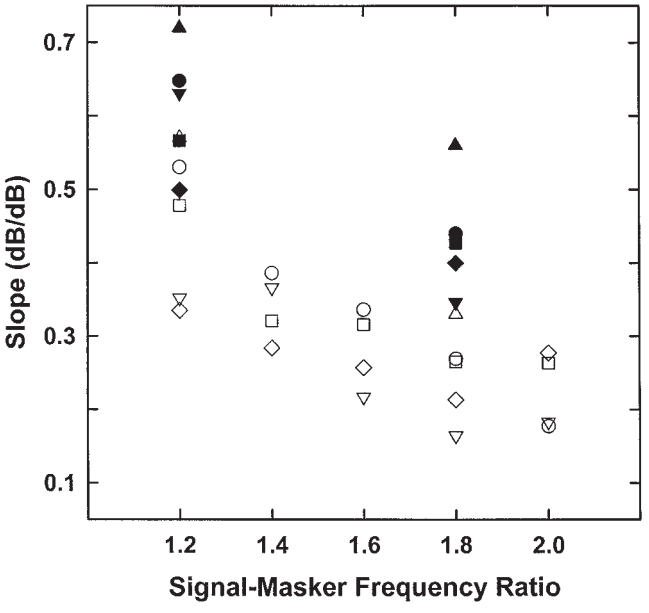


FIGURE 4.3. The slopes of the growth-of-masking functions (masker level vs. signal level) obtained in forward masking as a function of the frequency ratio between the signal and masker. The signal was a 4,000-Hz sinusoid, and the masker was a sinusoid in which the frequency was always less than 4,000 Hz. The different symbols represent the results from different subjects. The functions were obtained before (open symbols) or after (solid symbols) the ingestion of aspirin, which caused a temporary hearing loss of about 8 dB. (Data are from Hicks and Bacon 2000.)

2.3 Upward Spread of Masking and Cochlear Hearing Loss

The masking produced by a pure-tone (or any narrowband) masker is not restricted to frequencies immediately surrounding the masker. Instead, it spreads to frequencies on either side of the masker due to the limited frequency resolution of the ear. At low masker levels, this spread is more or less symmetric. At moderate to high masker levels, however, the spread is considerably greater to frequencies above the masker than to frequencies below. This is the well-known "upward spread of masking." It can be seen, for example, in the classical masking patterns published by Wegel and Lane (1924) and Egan and Hake (1950). An example of such patterns is shown in Figure 3.2 in Oxenham and Bacon (Chapter 3). The degree of asymmetry increases with increasing masker level so that at high masker levels, the

masking patterns are markedly asymmetric. The highly expansive growth of masking at frequencies much higher than the frequency of the masker can be understood in terms of the growth of masking discussed above. That is, it is likely due to the fact that the response to the masker is linear, whereas the response to the signal is compressive at the place along the basilar membrane corresponding to the signal frequency. Thus, a given increase in masker level will require an even greater increase in signal level to maintain a constant signal-to-masker ratio in terms of an internal response and hence a constant signal detectability.

Because individuals with a cochlear hearing loss generally have poorer than normal frequency resolution, the masking tends to spread farther in those listeners than it does in normal-hearing listeners (for a review, see Moore 1998). It is not uncommon for the spread to be especially great toward higher frequencies, indicating an abnormal amount of upward spread of masking. Thus, in everyday listening situations, those hearing-impaired listeners would be especially susceptible to the masking produced by low-frequency environmental noise. However, because compression is reduced or completely eliminated in listeners with cochlear hearing loss, the masking patterns in those listeners should not become increasingly asymmetric with increases in masker level as they do in listeners with normal hearing. Although there are relatively few studies that have measured masking patterns in hearing-impaired subjects at more than one masker level, the limited results that exist show little change in the shape of the pattern with level (Trees and Turner 1986; Alcántara and Moore 2002). Interestingly, the gain of some commercially available hearing aids is intentionally reduced at high overall levels under the (unjustified) assumption that upward spread of masking will become worse with increases in level in individuals with cochlear hearing loss (for a review of amplification, particularly as it pertains to compression, see Levitt, Chapter 5).

3. Temporal Processing

The acoustic spectrum is rarely constant but instead changes considerably over time. A substantial amount of information is conveyed in those time-varying changes. The auditory system must process these changes reasonably accurately in order to extract the necessary information (for a review of temporal processing, see Viemeister and Plack 1993). This section focuses on aspects of temporal processing in which peripheral compression is thought to play an especially important role. It focuses primarily on temporal aspects of masking but also includes a discussion of temporal integration or summation. Of particular interest here is how these measures are affected by cochlear hearing loss and, by implication, a decrease in or loss of basilar membrane compression.

3.1 Forward Masking

Forward masking has been used as a measure of temporal processing for over 50 years (e.g., Lüscher and Zwislocki 1947; Munson and Gardner 1950; Zwislocki et al. 1959; Plomp 1964; Elliott 1971; Widin and Viemeister 1979; Jesteadt et al. 1982; Plack and Oxenham 1998). Despite the sustained interest, an understanding of the physiological processing underlying forward masking has remained elusive. As discussed in some detail by Oxenham and Bacon (Chapter 3), recent work suggests that basilar membrane compression may play an important role in forward masking (Oxenham and Moore 1994, 1995; Plack and Oxenham 1998; Oxenham and Plack 2000; Oxenham 2001). More specifically, the results from a number of forward-masking experiments can be predicted reasonably well by a temporal window model in which a linear temporal integrator or window follows a compressive stage that represents basilar membrane compression (see Fig. 3.5 in Oxenham and Bacon, Chapter 3). The purpose of this section is to describe how well some results from hearing-impaired subjects can be understood in the context of the temporal window model by assuming a decrease in or loss of compression.

A consistent finding in the literature is that the masked threshold (or amount of masking) decreases more slowly as a function of the delay between the offset of the masker and the onset of the signal in hearing-impaired subjects than in normal-hearing subjects (Garner 1947; Elliott 1975; Glasberg et al. 1987; Nelson and Freyman 1987; Oxenham and Moore 1997). This is especially true when the maskers are presented at the same overall level to the two groups of subjects. When the maskers are presented at the same sensation level (i.e., the same number of decibels above the detection threshold for the masker), the differences are markedly reduced, although not completely eliminated (Glasberg et al. 1987; Nelson and Freyman 1987). Some example results from a study by Glasberg et al. (1987) are shown in Figure 4.4. The masker was a band-pass noise centered at the 2,000-Hz signal frequency. The short signal was presented at different times after the onset of the masker; the three leftmost points are for simultaneous masking and the three rightmost points are for forward masking. The decrease in forward-masked threshold is more gradual in the impaired ears (Fig. 4.4, open symbols) than in the normal ears (Fig. 4.4, solid symbols), particularly when the maskers are presented at the same overall levels (compare Fig. 4.4, open and solid circles). This is often described as a longer "recovery" from forward masking.

Another finding in the literature is that the growth of forward masking (signal level as a function of masker level) tends to be steeper in hearing-impaired subjects than in normal-hearing subjects in an on-frequency masking condition (Jesteadt 1980; Kidd et al. 1984; Oxenham and Moore 1995). This is in contrast to the results of off-frequency masking conditions in which the masker is much lower in frequency than the signal. In those

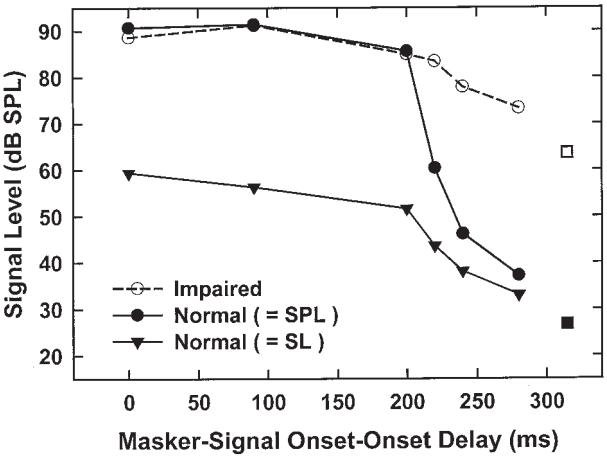


FIGURE 4.4. Signal thresholds as a function of the delay between the onset of a 210-ms narrowband noise masker and the onset of a 10-ms sinusoidal signal. The masker was centered at the 2,000-Hz signal frequency. The data points are the average of 5 unilaterally impaired subjects. In the impaired ear, the masker was presented at 1 relatively high SPL; in the normal ear, the masker was presented either at the same SPL or at the same sensation level (SL) as in the impaired ear. The three leftmost points on each function are for simultaneous masking, whereas the three rightmost points are for forward masking. The unconnected symbols at the far right are the absolute thresholds for the 10-ms signal (solid for the normal ear and open for the impaired ear). (Data are from Glasberg et al. 1987.)

cases, the growth of forward masking tends to be shallower in subjects with cochlear hearing loss. As discussed in Section 2.2, this can be understood in terms of a decrease in or loss of basilar membrane compression.

Both the slower recovery and the steeper growth of masking in on-frequency conditions can be understood in terms of a decrease in compression, as illustrated in Figure 4.5 (also see Oxenham and Moore 1997). In Figure 4.5, the output of a temporal window model is shown for various forward-masking conditions. The compression parameter in the model was adjusted to simulate either normal basilar membrane compression or a complete loss of compression. In the normal case, the slope of the compression function in the model was 0.2 for stimulus levels above 35 dB sound pressure level (SPL) and 1.0 for levels below that. For the impaired case, the slope was 1.0 throughout. In Figure 4.5*A*, recovery functions are shown in response to a masker presented at a level of 80 dB SPL. The recovery function is more rapid with compression (Fig. 4.5*A*, dashed line) than without (Fig. 4.5*A*, solid line). This occurs because as the delay between the offset of the masker and the onset of the signal increases and thus the internal response to the masker at the time of signal presentation decreases, the

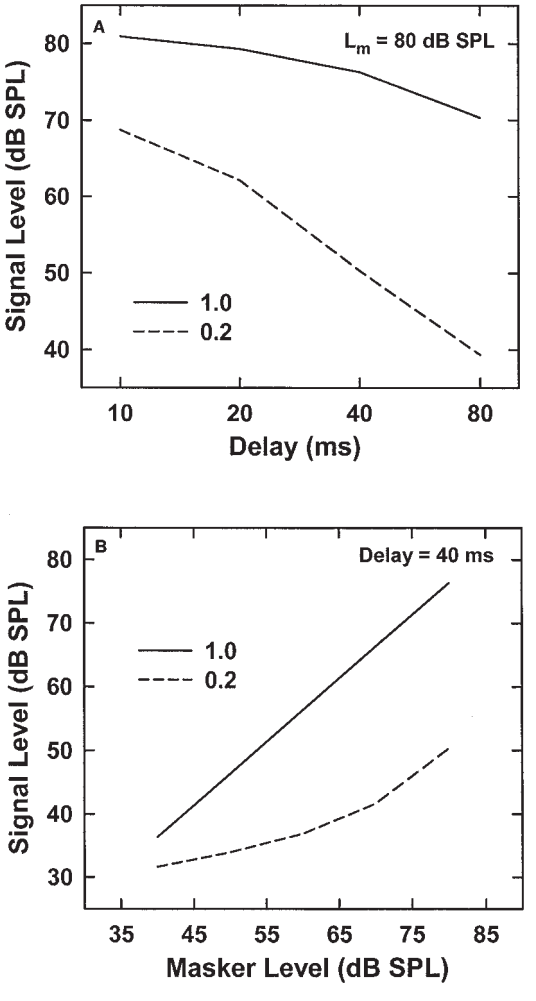


FIGURE 4.5. The output of a temporal window model for various forward-masking conditions, with a 2,000-Hz signal and 2,000-Hz masker. The parameter is the slope of the compression function in the model (1.0 indicates no compression). *A*: signal threshold as a function of the delay between the offset of the masker and the onset of the signal; the masker level (L_m) was 80 dB SPL. *B*: signal threshold as a function of masker level; the delay was 40 ms.

signal level has to be decreased more when its response is compressed than when it is not compressed in order to maintain a constant internal signal-to-masker ratio. In Fig. 4.5*B*, growth-of-masking functions are shown for a masker signal delay of 40 ms. The masking function is more gradual with compression (Fig. 4.5*B*, dashed line) than without (Fig. 4.5*B*, solid line). This arises because over a large range of masker levels, the more intense masker

is compressed, whereas the less intense signal is not. Thus, a given increase in masker level requires a smaller increase in signal level to maintain constant signal detectability. These simulations thus capture the basic trends in the literature with regard to the effects of cochlear hearing loss on forward masking and, in conjunction with the results described by Oxenham and Bacon (Chapter 3), strongly suggest that compression normally plays an important role in forward masking.

3.2 Additivity of Nonsimultaneous Masking

In many real-world environments, the listener is confronted with multiple maskers or background sounds that could interfere with the processing of the sound(s) of interest. It is important to determine how the effects of such maskers combine, not only in terms of gaining a theoretical understanding of the underlying processing but also in terms of being able to predict the audibility of various sounds in complex acoustic environments. As discussed by Oxenham and Bacon (Chapter 3), the situation is somewhat more complicated when the signal and maskers overlap in time (i.e., in simultaneous masking) than when they do not. Thus, as in Oxenham and Bacon (Chapter 3), the focus here is on additivity of nonsimultaneous masking (where both of the maskers and the signal do not overlap in time).

The energy detector model of masking (Green and Swets 1966) predicts that when two equally effective maskers (whether simultaneous with the signal or not) are combined, the amount of masking should increase by 3 dB. In other words, it assumes the linear addition of intensity or an intensity-like quantity. However, in many situations, the increase in masking is considerably more than 3 dB; the additional amount is referred to as "excess" masking. Penner (1980) suggested that the excess masking observed in nonsimultaneous masking could be accounted for by assuming that all stimuli are subjected to a compressive nonlinearity before being combined within a linear temporal integrator (see Oxenham and Bacon, Chapter 3, for more details). Oxenham and Moore (1994) suggested that this compressive nonlinearity might reflect the compression that exists at the basilar membrane in the cochlea. They later argued that if this were indeed the case, then the amount of excess masking should be reduced in individuals with cochlear hearing loss due to their reduced compression (Oxenham and Moore 1995). To test this, they examined the additivity of masking in a nonsimultaneous-masking paradigm in subjects with normal hearing and in subjects with a hearing loss of cochlear origin. Thresholds were measured for a brief 4,000-Hz signal masked by a broadband noise in forward masking, backward masking, and a combined forward and backward masking in which two equally effective maskers were combined. They showed that in normal-hearing subjects, the two maskers combined to produce as much as 15 dB of excess masking and that this could be understood in terms of a power law model of masking with a compressive power

law exponent of about 0.2. This degree of compression is consistent with that observed at the basilar membrane (Cooper, Chapter 2) and that estimated psychophysically using growth of forward masking (see Section 2; also Oxenham and Bacon, Chapter 3).

For the three hearing-impaired subjects tested by Oxenham and Moore (1995), however, the two maskers produced little or no excess masking. Their data were fitted very well with a power law exponent of 1.0, indicating the absence of a compressive nonlinearity. This is, of course, consistent with the fact that cochlear hearing loss reduces or eliminates basilar membrane compression (Cooper, Chapter 2). Furthermore, the fact that two equally effective maskers generally combined to produce a 3-dB increase in masking in the hearing-impaired subjects is consistent with the idea that, after basilar membrane processing, the effective intensity of each masker is combined linearly. This intensity or intensity-like quantity may result from a square law nonlinearity provided by the inner hair cells (Goodman et al. 1982; Yates 1990; see Oxenham and Bacon, Chapter 3). In other words, the inner hair cells may effectively square the basilar membrane vibration amplitude or velocity. Additional support for such a square law nonlinearity was discussed by Oxenham and Bacon (Chapter 3).

3.3 Temporal Effects in Simultaneous Masking

Temporal effects in simultaneous masking have been investigated for several decades (e.g., Samoilova 1959; Elliott 1965; Zwicker 1965a,b; Green 1969; Fastl 1976; Bacon and Viemeister 1985a,b; McFadden and Wright 1990, 1992; Bacon and Liu 2000), either by comparing the amount of masking produced by a brief gated masker with that produced by a continuous masker or by measuring the threshold for a brief signal as a function of its temporal position within a longer duration masker. When a temporal effect exists, the gated masker tends to produce more masking than the continuous masker and the amount of masking is greater for a signal at the beginning than near the temporal center or end of the masker. Zwicker (1965a) dubbed the temporal effect "overshoot" to describe the increase (or overshoot) in masking at the beginning of the masker.

The masker of choice in these studies typically has been either a broadband noise or a pure tone. The magnitude of the effect can be quite large for both, on the order of 10–20 dB or more. The broadband noise maskers overlap the signal spectrally, whereas the pure-tone maskers usually do not. Indeed, the effect is considerably larger when the tonal masker and signal do not overlap (Bacon and Viemeister 1985a). Although there are clear similarities between the temporal effects observed with the two types of masker, there are important differences as well (see Bacon and Healy 2000). In fact, it is unclear whether the same processing underlies the temporal effect with both types of masker.

One explanation, however, that has been proposed to account for the temporal effect with both broadband noise and pure-tone maskers is peripheral adaptation (Green 1969; Bacon and Viemeister 1985a,b; Bacon and Smith 1991). This is based on physiological research (Smith and Zwislocki 1975; Smith 1977, 1979) showing that the response of an auditory nerve fiber to a masker tone declines or adapts over time, whereas the response of that fiber to a brief signal tone does not adapt. Consequently, the neural signal-to-masker ratio increases with time after masker onset. Although peripheral adaptation may indeed play a role in the psychophysical masking experiments, it is clear that it cannot account for the entire temporal effect (e.g., Bacon 1990). More recently, it has been suggested that the outer hair cells may play an important role in the temporal effect in simultaneous masking with both noise and tonal maskers. In particular, it has been suggested that basilar membrane compression may be intimately involved with the temporal effect with broadband noise maskers. It is less clear, however, whether compression is involved in the temporal effect with tonal maskers. Considerable support for the influence of outer hair cells on temporal effects in simultaneous masking comes from research with individuals with either a permanent or temporary cochlear hearing loss. These results are described in Sections 3.3.1 and 3.3.2.

3.3.1 Noise Maskers

Carlyon and Sloan (1987) were the first to measure overshoot in subjects with cochlear hearing loss. They examined overshoot at two signal frequencies in subjects with unilateral impairment and found about 3–5 dB of overshoot in both the normal and impaired ears. They concluded that overshoot was unaffected by cochlear loss. Bacon and Takahashi (1992) noted some procedural problems with the Carlyon and Sloan (1987) study. They presented their own results from a group of hearing-impaired subjects, some of which are shown in Figure 4.6, where the level of the signal at threshold is plotted as a function of the spectrum level of the broadband noise masker. When the signal was in the temporal center of the masker (Fig. 4.6*B*), the results for the hearing-impaired subjects (open symbols, dashed lines) overlapped those for the normal-hearing subjects (solid symbols, solid lines). The only exceptions occurred at the lower masker level where the quiet thresholds for two of the hearing-impaired subjects were higher than the masked thresholds for the other subjects. When the signal was near the beginning of the masker (Fig. 4.6*A*), however, the thresholds for the hearing-impaired subjects were generally lower than those for the normal-hearing subjects. In other words, in this condition, the hearing-impaired subjects performed better than the normal-hearing subjects. The amount of overshoot (the difference between the thresholds at the two signal delays) was thus smaller in the hearing-impaired subjects. In the impaired ears, overshoot was never larger than 5 dB. In the normal ears,

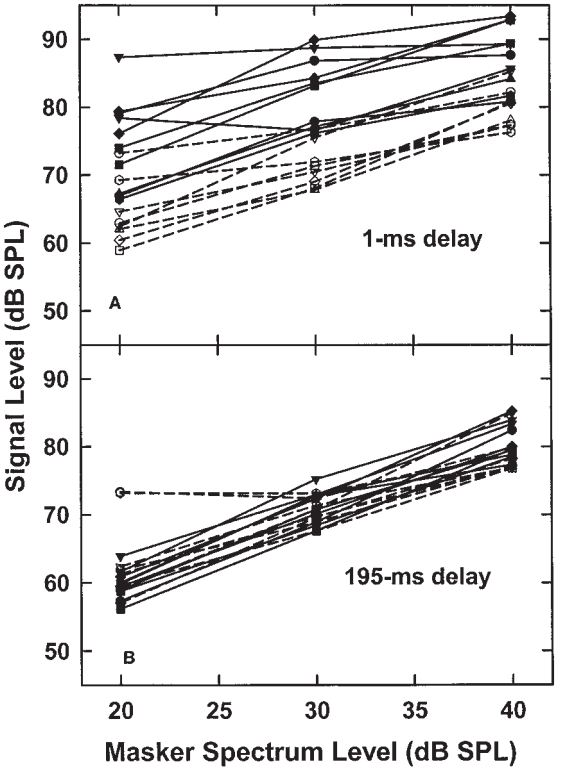


FIGURE 4.6. Thresholds for a 10-ms, 4,000-Hz sinusoidal signal as a function of the spectrum level of a 400-ms broadband noise masker. Results are shown for normal-hearing subjects (solid symbols, solid lines) and hearing-impaired subjects (open symbols, dashed lines). *A*: onset of the signal occurred 1 ms after the onset of the masker. *B*: onset of the signal occurred 195 ms after the onset of the masker. (Data are from Bacon and Takahashi 1992.)

maximum overshoot ranged from 7 to 26 dB. Turner and Doherty (1997) later confirmed the results of Bacon and Takahashi (1992).

The results in subjects with permanent cochlear hearing loss are generally similar to those in subjects with a temporary cochlear loss. Champlin and McFadden (1989) exposed subjects to an intense 2,500-Hz tone set to a level that produced about 15 dB of temporary hearing loss at the 3,550-Hz test frequency [maximal temporary hearing loss tends to occur half an octave above the exposure frequency (e.g., Davis et al. 1950; McFadden 1986)]. They measured overshoot for each subject before and various times after the exposure. Some of their results are shown in Figure 4.7. Immediately after the initial exposure, when the thresholds in quiet (Fig. 4.7, open

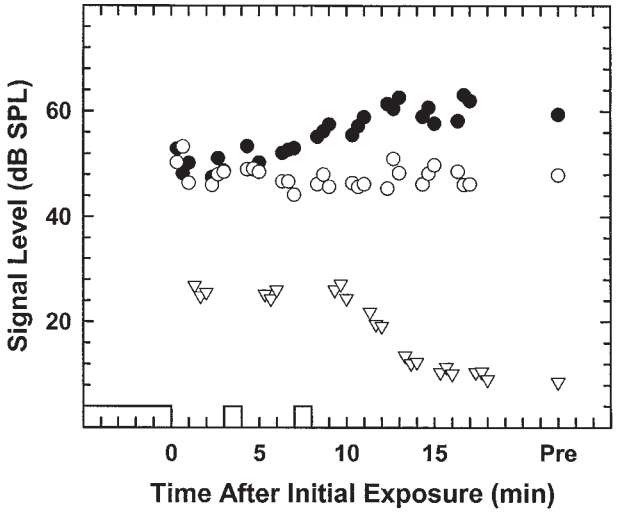


FIGURE 4.7. Thresholds for a 6-ms, 3,550-Hz sinusoidal signal at various times after exposures to an intense 2,500-Hz tone (exposures are indicated by small histogram bars along the *x*-axis). The sinusoidal signal was presented in quiet (open triangles) or near the onset (solid circles) or end (open circles) of a broadband noise masker. Thresholds at the far right were obtained before exposure. (Data are from Champlin and McFadden 1989.)

triangles) were highest (i.e., when the hearing loss was greatest), overshoot was negligible. In other words, there was little or no difference between the threshold for a signal at the beginning of the masker (Fig. 4.7, solid circles) and the threshold for a signal near the end of the masker (Fig. 4.7, open circles). Over time, however, as the hearing loss subsided, overshoot recovered to preexposure values. Notice that the recovery of overshoot was due entirely to an increase in the threshold for a signal near masker onset. Similar results have been observed in subjects with temporary loss due to aspirin ingestion (McFadden and Champlin 1990; Bacon and Hicks 2000). That is, aspirin reduces overshoot by lowering the threshold for a signal near masker onset.

There is thus an apparent paradox: permanent or temporary cochlear insult results in an improvement in the detectability of a brief signal positioned near the beginning of a broadband noise masker. This paradox can be understood, however, in light of the proposed role of basilar membrane compression in overshoot, as first suggested by von Klitzing and Kohlrausch (1994). The basic aspect of their argument is shown in Figure 4.8. The solid line in Figure 4.8*A* shows a basilar membrane input-output function in schematic form. It is linear at low and high levels and compressive with a slope of 0.2 dB/dB at moderate levels (e.g., Johnstone et al. 1986; Ruggero

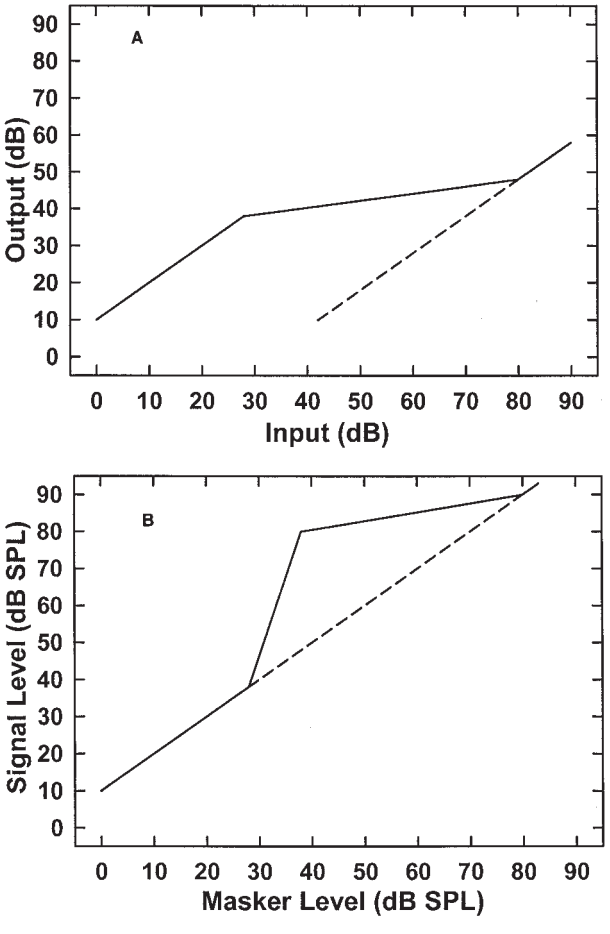


FIGURE 4.8. *A*, solid line: schematic of a basilar membrane input-output function, which is linear at low and high input levels and compressive (slope of 0.2) at moderate input levels; dashed line: linear input-output function. *B*, solid line: growth-of-masking function based on the compressive basilar membrane input-output function, assuming that the signal level must be 10 dB higher than the masker level at the output of the input-output function; dashed line: masking function that would result if the basilar membrane input-output function were linear.

and Rich 1991).[1] The solid line in Figure 4.8*B* shows a hypothetical masking function based on the input-output function in Figure 4.8*A* (the dashed line is discussed later). This function is qualitatively similar to the masking functions obtained in normal-hearing subjects when the signal is near the begin-

[1] Although recent basilar membrane data indicate that the input-output function for a given point along the membrane remains compressive up to extremely high levels

ning of the masker (Bacon 1990). The shape of this masking function can be understood in terms of the basilar membrane input-output function in the following way. Because the signal is brief, its level at threshold is considerably higher than the level of the noise in the auditory filter surrounding the signal. In this example, the internal signal-to-masker ratio at threshold is assumed to be 10 dB. At low masker levels, both the masker and signal will be processed linearly, resulting in a linear masking function. At somewhat higher levels, the masker will be processed linearly, but the moderate-level signal will be compressed; the signal level will have to be increased more than the masker level, resulting in a steep masking function. At even higher levels, the moderate-level masker will be compressed, but the signal will now be processed linearly; a given increase in masker level will result in a small increase in signal level, yielding a shallow masking function. Finally, at very high levels, both the masker and signal will be processed linearly, resulting in a linear masking function.

Now consider what happens with cochlear damage. In this case, the normally compressive portion of the input-output function will be linear or nearly so. A linear input-output function is shown as the dashed line in Figure 4.8*A*. When the signal is at moderate levels in an ear with cochlear damage, it will not be compressed as it is in a normal cochlea, and thus the signal level will not have to be as intense as it normally would in order to be heard. In other words, at threshold, the signal level will be lower in subjects with a cochlear loss than in subjects with normal hearing (compare the dashed line with the solid line in Fig. 4.8*B*). Thus, the reduction in or absence of compression, concomitant with cochlear damage, results in a lower masked threshold for a signal near the onset of a broadband noise masker.

It is important to note that although basilar membrane compression can account for the particularly high thresholds at the onset of the masker in normal-hearing subjects, it does not, by itself, account for the decrease in threshold over time. However, it has been suggested that the decrease in masking over time in normal-hearing subjects reflects a decrease in the degree of basilar membrane compression during the course of masker stimulation, possibly as a result of feedback from the efferent system. This feedback is presumably through the influence of the medial olivocochlear system on the outer hair cells (Turner and Doherty 1997; Bacon and Liu 2000; Strickland 2001). Consistent with this possibility is research with "precursors," which are sounds presented before the onset of the masker and

---

(110 dB SPL) in a healthy cochlea (Ruggero et al. 1997; see Cooper, Chapter 2), it may nevertheless be appropriate to model psychophysical data with a function that becomes linear at high levels. This is because, at high levels, the maximum response to the signal likely will be at a place basal to that corresponding to the signal frequency (Ruggero et al. 1997; Rhode and Recio 2000), where the growth of response will be linear or nearly linear. [An example of this basal shift can be seen in Figure 2.2*C* in Cooper (Chapter 2).] In this case, a given location becomes more responsive to a tone lower in frequency than the CF at high stimulus levels.)

signal. Precursors have been shown to decrease the threshold for a signal at masker onset, but they have no effect on the threshold for a signal presented at least 100–200 ms later in time (e.g., Bacon and Smith 1991). Thus precursors can reduce or eliminate overshoot by lowering the threshold for a signal near masker onset (much like cochlear hearing loss can reduce overshoot). Although most investigators have presented the precursor to the ear receiving the masker and signal (i.e., to the ipsilateral ear), Turner and Doherty (1997) and Bacon and Liu (2000) presented the precursor to the opposite or contralateral ear. One advantage of this approach is that the effect of the precursor cannot be mediated peripherally, such as through adaptation of auditory nerve fibers. The results of both studies indicated that contralateral precursors could reduce overshoot. Furthermore, Bacon and Liu (2000) showed that the effectiveness of ipsilateral and contralateral precursors is affected similarly by changes in their spectral content, implying that the processing underlying the influence of ipsilateral and contralateral precursors is largely the same. The intriguing possibility that the degree of basilar membrane compression is reduced during the course of stimulation as a consequence of an efferent feedback loop (from the medial olivocochlear bundle to the outer hair cells) does not, as yet, have any support from direct mechanical measurements of basilar membrane vibration. It is important to note, however, that (1) stimulation of the efferent system has been shown to result in a more linear basilar membrane input-output function (Murugasu and Russell 1996); (2) basilar membrane responses to broadband noise have not been measured as a function of time after noise onset; and (3) the anesthesia used in the animal experiments designed to measure basilar membrane motion would be expected to reduce the influence of the efferent system (Liberman and Brown 1986).

3.3.2 Tonal Maskers

The temporal effect with tonal maskers depends on the frequency relationship between the masker and signal. It is relatively small for conditions in which the masker frequency is less than or equal to the signal frequency but can be quite large when the masker frequency is greater than the signal frequency (Bacon and Viemeister 1985a,b). A consequence of this is that measures of frequency selectivity, such as psychophysical tuning curves (PTCs), sharpen with time (e.g., Bacon and Viemeister 1985a; Bacon and Moore 1986). There are only a few studies of the effects of cochlear hearing loss on the temporal effect with tonal maskers.

Kimberley et al. (1989) measured PTCs in normal-hearing subjects and in subjects with cochlear hearing loss. The signal was presented at the onset or temporal center of the masker. For the normal-hearing subjects, the PTC sharpened as the signal was moved from the beginning to the temporal center of the masker; the sharpening was primarily on the high-frequency side. For the hearing-impaired subjects, however, the PTC was essentially

independent of the temporal position of the signal. The authors suggested that the sharpening with time was related to the active mechanism in the cochlea (i.e., to the outer hair cells that are responsible for the compressive nonlinearity that is normally observed). Bacon et al. (1988) also observed reduced temporal effects in subjects with cochlear hearing loss. They suggested that the normal temporal effect might be influenced by the cochlear nonlinearity.

Bacon and Hicks (2000) examined the effects of aspirin on the temporal effect with both noise and tonal maskers. As noted in Section 3.3.1, aspirin reduced the effect with noise maskers by reducing the threshold for a signal near masker onset. Aspirin also reduced the temporal effect with tonal maskers but in a different way. In particular, aspirin increased the masked threshold at both delays but more so for a signal in the temporal center of the masker. In general, the reduction in the temporal effect was larger with the noise than with the tonal masker. Because aspirin directly affects the outer hair cells, the results suggest that these cells normally are involved in the temporal effect with both types of masker. For noise maskers, the effects of hearing loss are reasonably well explained by a loss of compression (see Section 3.3.1). For tonal maskers, it is less clear whether the decrease in the temporal effect is the result of a reduction in basilar membrane compression or whether it reflects the modification of some other aspect of the cochlear nonlinearity associated with outer hair cells.

3.4 Simultaneous Masking by Schroeder-Phase Maskers

There has been considerable interest recently in studying the masking produced by so-called Schroeder-phase maskers (Schroeder 1970; Oxenham and Bacon, Chapter 3). Positive-Schroeder (m+) and negative-Schroeder (m–) harmonic complexes have identical, flat amplitude spectra but different phase spectra (for more details, see Section 4.3.3 in Oxenham and Bacon, Chapter 3). The m+ and m– waveforms are relatively flat and are mirror images of one another (as shown in Fig. 3.11*A*,*C* in Oxenham and Bacon, Chapter 3). Nevertheless, the m+ complexes tend to produce less masking than the m– complexes (Kohlrausch and Sander 1995). This difference in masking effectiveness between the m+ and m– maskers is referred to as the Schroeder-phase effect. It has been hypothesized that this phase effect results from the m+ complexes producing a more peaked or modulated response along the basilar membrane coupled with the influence of fast-acting compression (Carlyon and Datta 1997a,b; Summers and Leek 1998). In support of this is recent physiological evidence demonstrating that in the chinchilla, the basilar membrane response to m+ complexes is more peaked than the response to m– complexes (Recio and Rhode 2000; see Fig. 3.11*B*,*D* in Oxenham and Bacon, Chapter 3, for an illustration of the possible influence of auditory filtering on the m+ and m– waveforms). As discussed by Oxenham and Bacon (Chapter 3), a variety of psychophysical

data from normal-hearing subjects is consistent with the possibility that basilar membrane compression influences the Schroeder-phase effect. This section considers the extent to which results from individuals with cochlear hearing loss are also consistent with such an interpretation.

Summers and Leek (1998) examined the Schroeder-phase effect in subjects with normal hearing and in subjects with cochlear hearing loss. Some of their results are shown in Figure 4.9. They varied the level of the masker in order to mask a long-duration signal presented at several fixed levels. For the normal-hearing subjects (Fig. 4.9, solid symbols), the level of the m+ masker was higher than that of the m– masker at all signal levels, although the difference between the two maskers decreased somewhat with increasing signal level. The m+ masker was presumably a less effective masker, in part, because the internal response (after auditory filtering at the basilar membrane) was more modulated than it was for the m– masker. This would occur if the negative phase curvature of the auditory filter more or less offset the positive phase curvature of the m+ masker (see Kohlrausch and Sander 1995; Lentz and Leek 2001; Oxenham and Dau 2001b). Because the

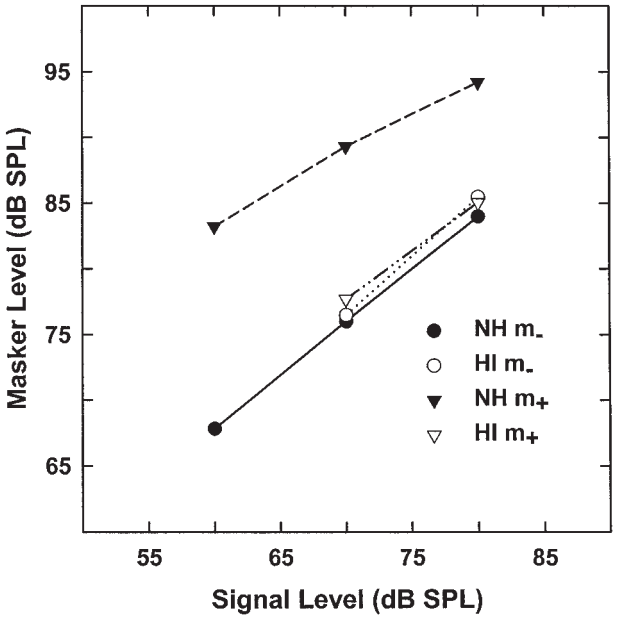


FIGURE 4.9. The overall level of a 460-ms masker required to simultaneously mask a 260-ms, 4,000-Hz sinusoidal signal presented at levels ranging from 60 to 80 dB SPL. The masker was either a positive-phase (m^+) or negative-phase (m^-) Schroeder-phase masker. The data points are the averages of 5 normal-hearing (NH) subjects (solid symbols) or 4 hearing-impaired (HI) subjects (open symbols). (Data are from Summers and Leek 1998.)

signal would be easier to hear during the "valleys" of the internally modulated waveform, the m+ masker had to be at a higher overall level than the m– masker to mask the signal. For the hearing-impaired subjects (Fig. 4.9, open symbols), there was essentially no difference in the level of the masker needed to mask the signal for the two different maskers. In other words, the phase effect was more or less eliminated by the cochlear loss. Because there is no clear evidence that hearing loss alters the phase response of the auditory filter (Nuttall and Dolan 1996; Ruggero et al. 1997), this suggests that the differences in the two maskers at the output of the auditory filter are not themselves sufficient to account for a substantial Schroeder-phase effect. The fact that the phase effect decreased in subjects with cochlear hearing loss led Summers and Leek (1998) to suggest that the Schroeder-phase effect exists only in conditions where basilar membrane compression is operative.

Summers (2000) used Schroeder-phase maskers to measure masking period patterns (MPPs) in subjects with normal hearing and in subjects with cochlear hearing loss. As noted by Oxenham and Bacon (Chapter 3), these types of patterns are obtained by measuring the threshold for a brief signal as a function of its temporal position within the period of a masker. These patterns thus reflect how masking varies within a cycle of a periodic masker. In this experiment, the signal level was fixed at 60 or 80 dB SPL, whereas the masker level was varied to just mask the signal. Some of the results are shown in Figure 4.10. For the normal-hearing subjects (Fig. 4.10*A*), the results clearly depend on signal level. At the lower level (Fig. 4.10*A*, solid symbols), the MPPs are modulated for both maskers but more so for the m+ masker. These results are similar to results observed by others (Kohlrausch and Sander 1995; Carlyon and Datta 1997a). At the higher signal level (Fig. 4.10*A*, open symbols), however, the patterns are flatter for both maskers. For the hearing-impaired subjects (Fig. 4.10*B*), the patterns are more or less flat for both maskers at both signal levels. Taken together, the results from the normal-hearing and hearing-impaired subjects suggest that the Schroeder-phase effect is smallest in conditions where the basilar membrane input-output function is thought to be relatively linear (at high levels in normal-hearing subjects and in subjects with cochlear hearing loss). However, rather than being affected by compression per se, Summers (2000) argued that the results might reflect a modulation in the amount of physiological suppression during each masker period. Suppression is a nonlinear phenomenon that requires normally functioning outer hair cells. In other words, it is another aspect of a nonlinear cochlea (see Cooper, Chapter 2). In this scheme, suppression of the signal by the masker would be greatest during the high-amplitude portions of the (basilar membrane-filtered) peaky m+ masker and weakest during the low-amplitude portions. Because the m– masker would be less peaky after basilar membrane filtering, suppression would not be expected to vary as much with that masker and thus the MPPs would be expected to be flatter. Basilar membrane data

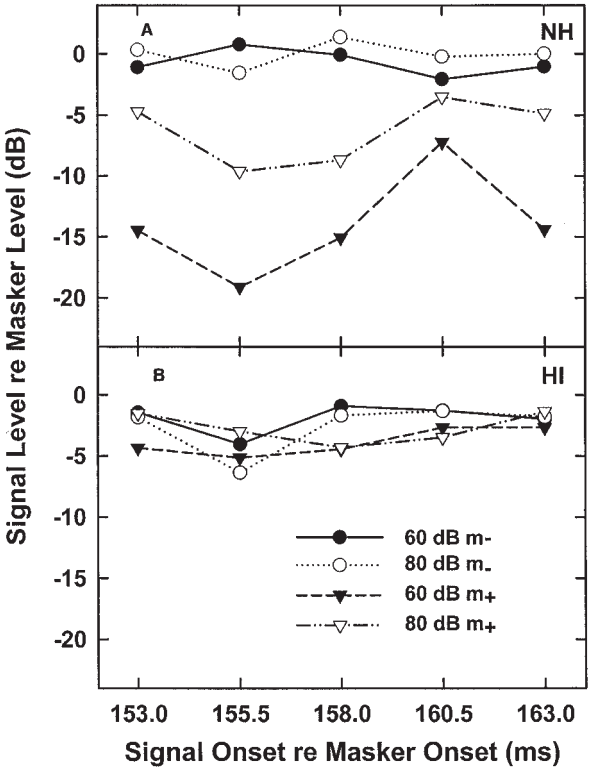


FIGURE 4.10. The level of a 460-ms masker required to simultaneously mask a 5-ms, 4,000-Hz sinusoidal signal presented at different times relative to the onset of the masker. The level of the signal was 60 or 80 dB SPL. The results are plotted in terms of the level of the signal relative to the overall level of the masker. The masker was either a m^+ or m^- Schroeder-phase masker. The data points are the average of 7 NH subjects (*A*) or 7 HI subjects (*B*). (Data are from Summers 2000.)

from Recio and Rhode (2000) are consistent with the possibility that suppression may play a role in the Schroeder-phase effect.

The small or negligible phase effects in subjects with cochlear hearing loss are certainly consistent with the possibility that the outer hair cells play an important role in the Schroeder-phase effect. Whether the influence of the outer hair cells is via basilar membrane compression or suppression is somewhat unclear. However, as discussed by Oxenham and Bacon (Chapter 3), the results from several experiments can be understood in terms of basilar membrane compression. Furthermore, the modeling results of Oxenham and Dau (2001a) indicate that in the context of the temporal window model, compression is necessary to generate model predictions in line with the psychophysical masking results.

3.5 Simultaneous Masking by Sinusoidally Amplitude-Modulated Tonal Maskers

One way in which investigators have examined the ability of the auditory system to follow the short-term fluctuations in the amplitude of a sound is to measure the threshold for a brief signal as a function of its temporal position within a period of a fluctuating masker. As noted in Section 3.4, the resulting masking patterns are referred to as MPPs (e.g., Zwicker 1976a,b,c). Zwicker and Schorn (1982) suggested that it would be more efficient simply to measure two thresholds, one for a signal at the peak and one for a signal at the valley of the masker, rather than to map out the entire MPP. They went on to show that the difference between those two thresholds could be estimated from the difference between the threshold for a long-duration signal in the presence of an unmodulated masker and that obtained in the presence of a modulated masker. Regardless of the experimental approach, the results from a wide range of studies suggest that listeners can indeed take advantage of short-term fluctuations in the amplitude of a masker to improve signal detectability, particularly if those fluctuations are relatively slow and quite pronounced. These studies have either used a broadband noise masker, where the masker completely overlaps the signal in frequency (e.g., Zwicker and Schorn 1982; Bacon and Lee 1997; Bacon et al. 1997; Eddins 2001), or a narrowband (noise or tonal) masker, where the masker and signal may or may not overlap in frequency (Fastl 1975; Zwicker 1976c; Buus 1985; Moore and Glasberg 1987). Although this masking paradigm could be considered a measure of temporal resolution, it is clear that other factors play a role and thus the results from this paradigm should not be taken as a measure solely of temporal resolution. For example, when the masker is a broadband noise, factors related to comodulation masking release (e.g., Hall et al. 1984) and suppression (e.g., Houtgast 1974) may play a role (for a discussion, see Bacon and Lee 1997; Bacon et al. 1997). When the masker is narrowband and the signal is located considerably higher in frequency than the masker, basilar membrane compression may influence the apparent size of the internal fluctuations caused by the fluctuating masker. The remainder of this section considers the role that compression may play, particularly in light of the effects of cochlear hearing loss in various masking conditions.

When the masker is narrowband, listeners are especially able to take advantage of the fluctuations in the masker when the signal is located much higher in frequency than the masker (Fastl 1975; Zwicker 1976c; Buus 1985; Moore and Glasberg 1987; Nelson and Swain 1996; Gregan et al. 1998). Nelson and Schroder (1996) compared the masking produced by a sinusoidally amplitude-modulated (SAM) masker with that produced by a quasi-frequency-modulated (QFM) masker. For both, the carrier was a 500-Hz tone and the modulation rate was 20 Hz. The amplitude spectra were identical, consisting of a 500-Hz carrier plus sidebands at 480 and 520 Hz,

but their waveforms were different. The amplitude of the SAM masker, which was modulated at a depth of 100%, fluctuated considerably, whereas the amplitude of the QFM masker was relatively constant. The signal frequency ranged from 500 to 3,200 Hz. The results from their normal-hearing subjects were similar to those published previously. The highly fluctuating SAM masker produced less masking than the relatively steady QFM masker, particularly for signal frequencies well above the masker carrier frequency. This masking difference was markedly reduced or absent when the signal was in a region of hearing loss in their hearing-impaired subjects (the masker was always in a region of normal hearing). Nelson and Schroder explained their results in terms of the influence of basilar membrane compression on the growth of masking (see Section 2) and, in particular, on what they called a "linearized response growth" model of masking. In essence, they argued that the large masking difference in normal-hearing subjects was due to the steep growth-of-masking function observed in those subjects when the masker frequency is less than the signal frequency. Thus a large instantaneous decrease in level (with the SAM masker) would produce an especially large decrease in masking. They argued further that the reduced difference in masking between the SAM and QFM maskers in their hearing-impaired subjects was due to the relatively shallow masking functions in those subjects subsequent to a decrease in or loss of basilar membrane compression (see Section 2.1).

In a related study from the same laboratory, Wojtczak et al. (2001) measured MPPs for a SAM masker modulated at a rate of about 4 Hz and a depth of 100%. The masker carrier either had the same frequency as the signal or was one octave lower in frequency. Some of their results are shown in Figure 4.11. Figure 4.11*A*,*B* shows the results from one of their normal-hearing subjects, whereas Figure 4.11*C*,*D* shows the results from one of their hearing-impaired subjects. For the normal-hearing subject, the valley of the MPP is considerably deeper and wider when the signal is higher in frequency than the masker (Fig. 4.11*B*) than when it is the same frequency as the masker (Fig. 4.11*A*). For the hearing-impaired subject, however, the shape of the MPP depended considerably less on the frequency relationship between the masker and signal. The two subjects differ most noticeably when the masker is located lower in frequency than the signal. Wojtczak and her colleagues modeled their results with a version of the temporal window model described earlier (Oxenham and Bacon, Chapter 3). For their compression function, they used a formula proposed by Glasberg and Moore (2000), which provides an estimate of the basilar membrane input-output function for various amounts of maximum gain (G_{max}), as shown in Figure 4.12. In Figure 4.12, G_{max} ranges from 20 to 60 dB in 10-dB steps (the dashed line is a linear function with no gain). Note that for a given function, the gain (difference between the output and input in dB) is maximal at low input levels but that it decreases and ultimately reaches zero at high input levels. The predictions of the temporal window

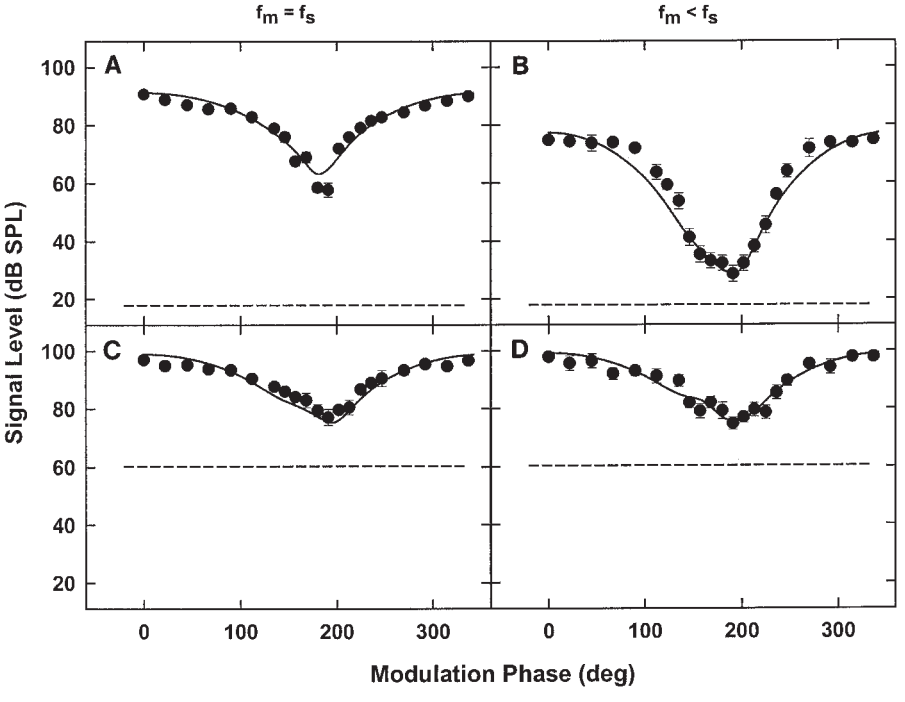


FIGURE 4.11. Thresholds for an 8-ms signal presented at various times during the modulation period of a tonal masker that was sinusoidally amplitude modulated at a rate of about 4 Hz and a depth of 100%. The masker was either the same frequency as the signal (*A* and *C*) or 1 octave lower in frequency than the signal (*B* and *D*). *A* and *B*: results from a NH subject. *C* and *D*: results from a HI subject. The masker level was either 90 (*A* and *B*) or 95 dB SPL (*C* and *D*). Solid lines: best fits to the data using a version of the temporal window model. Dashed lines: absolute thresholds for the 8-ms signal. f_s, signal frequency; f_m, masker frequency. (Data and model are from Wojtczak et al. 2001.)

model are shown as solid lines in Figure 4.11. The differences between the on-frequency (Fig. 4.11*A*) and off-frequency (Fig. 4.11*B*) masking conditions for the normal-hearing subject can be accounted for by the fact that in the off-frequency case, the response to the masker at the place where the signal is detected is linear, whereas the response to the signal is compressed. Thus the internal fluctuations caused by the masker will be "magnified" by the compressive response to the signal (i.e., a given change in the instantaneous masker level will result in an even larger change in signal level). In the on-frequency case, both the masker and signal will be compressed and thus the internal fluctuations will not be magnified by the compression. For the normal-hearing subject whose results are shown in Figure 4.11, the best-fitting G_{max} was 48 dB. For the hearing-impaired subject, on the other hand,

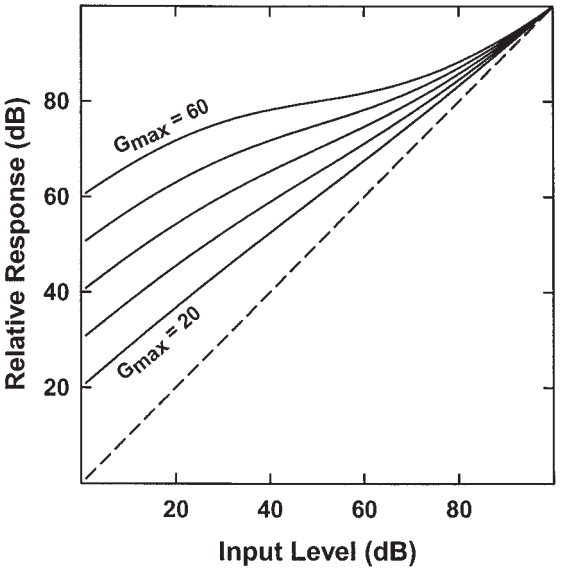


FIGURE 4.12. The compression function suggested by Glasberg and Moore (2000). The parameter is the maximum gain (G_{max}), which varies here from 20 to 60 dB. For a given G_{max}, the gain (difference between the relative response and the input level) decreases with increasing input level. Dashed line: simple linear function with no gain.

the best-fitting G_{max} was 8 dB. Overall, the results for the hearing-impaired subjects could be predicted by assuming little or no basilar membrane compression: G_{max} ranged from 42 to 48 dB in the normal-hearing subjects but only from 0 to 15 dB in the hearing-impaired subjects.

3.6 Temporal Integration in Simultaneous Masking

Temporal integration typically refers to how the threshold for the detection of a signal depends on the duration of that signal. The threshold for a tonal signal in quiet or in broadband noise typically decreases by about 3 dB for every doubling of signal duration over a range of durations from about 20 to 200 ms (Hughes 1946; Garner and Miller 1947; Green et al. 1957; Plomp and Bouman 1959; Zwislocki 1960, 1969). It is still unclear what mechanisms underlie temporal integration, although recent work suggests that basilar membrane compression may influence the results, at least in certain conditions. This possibility is discussed below in the context of the effects of cochlear loss on temporal integration.

Most individuals with cochlear hearing loss exhibit abnormally small amounts of temporal integration (e.g., Wright 1968; Gengel and Watson

1971; Florentine et al. 1988). As noted by Carlyon et al. (1990), this is somewhat surprising given that the integration process, whatever it may be, is thought to be at a central site in the auditory system and thus presumably unaffected by cochlear damage. Some of their results, however, may shed light on that apparent paradox. They measured temporal integration by comparing the threshold for a single 5-ms, 4,000-Hz tone pulse to the threshold for 10 such pulses, with each successive pulse separated by 80 ms. Their hearing-impaired subjects had less integration than their normal-hearing subjects (i.e., the threshold decreased less when going from 1 to 10 pulses than it did for the normal-hearing subjects). The authors also measured psychometric functions for the 1- and 10-pulse conditions. They found that the increase in detectability, in terms of the detectability index (d′), was the same for the 2 groups of subjects when going from 1 to 10 pulses but that the threshold difference in decibels was smaller for the hearing-impaired group due to their steeper psychometric functions. [For a discussion of the detectability index, see Green and Swets (1966); Macmillan and Creelman (1991).] Figure 4.13 shows hypothetical results in schematic form to illustrate how integration in terms of detectability could be normal despite an abnormally small change in threshold. In Figure 4.13, the dashed line represents the psychometric function for the normal-hearing subject, whereas the solid line represents the function for the hearing-impaired subject. A given change in detectability resulting from, say, an increase in the number of pulses or an increase in the duration of a single pulse corresponds to a smaller change in decibels for the hearing-impaired subject. Carlyon and

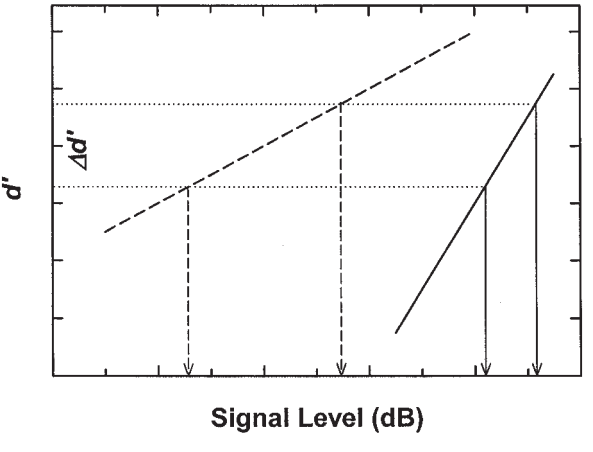


FIGURE 4.13. A schematic representation showing how a given change in detectability (Δd′) can correspond to quite different changes in threshold (in dB) depending on the slope of the psychometric function. Dashed line: hypothetical psychometric function for a NH subject; solid line: one for a HI subject.

his colleagues (1990) noted that the steeper psychometric functions in their hearing-impaired subjects could reflect a reduction in basilar membrane compression. A more linear basilar membrane response could explain the reduced integration in decibels regardless of whether the integration is in terms of a summation of information obtained from multiple looks at the stimulus (Viemeister and Wakefield 1991) or in terms of a summation of neural spikes (Zwislocki 1960; Penner 1972), given that a steeper basilar membrane function will result in a steeper rate-level function in the auditory nerve (see Cooper, Chapter 2). It is worth noting, however, that in order for this explanation to account for reduced integration (in dB) in subjects with cochlear hearing loss, the basilar membrane input-output function in normal-hearing subjects must be at least somewhat compressive at low levels. It is unclear whether this condition is generally met (see Cooper, Chapter 2).

Oxenham et al. (1997) designed an experiment specifically to evaluate the potential role of peripheral compression in temporal integration. They were especially interested in what they called "short-term" integration. This refers to integration for durations less than 20 ms. They thus measured thresholds as a function of signal duration in the presence of a broadband noise presented at levels intended to place the signals at low, moderate, or high levels. As discussed by Oxenham and Bacon (Chapter 3), they found that the slope of the short-term integration function was steeper at moderate signal levels than at low or high levels. They successfully modeled that result as being due to a compression of about 4:1 at moderate levels (the model assumed no compression at the low and high levels). They also measured integration in subjects with cochlear hearing loss. Contrary to the results from normal-hearing subjects, the slope of the short-term integration function did not change with level. Furthermore, that slope value was essentially the same as the slope value obtained at low and high levels in the normal-hearing subjects. Thus the data from the hearing-impaired subjects were fitted well at all levels by assuming no compression.

In summary, basilar membrane compression may play a role in short-term integration (i.e., the integration for durations less than 20 ms), and the decrease or lack of compression may help explain why temporal integration (in dB) is usually less than the 3 dB per doubling that is typically observed in normal-hearing subjects for durations from about 20 to 200 ms.

4. Loudness Recruitment

Most individuals with a cochlear hearing loss have a reduced dynamic range. That is, the range (in dB) between the absolute threshold for a sound and the level at which that sound becomes uncomfortably loud is smaller in someone with a cochlear hearing loss than it is in someone with normal hearing. This is because the absolute threshold is elevated, but the level at

which sounds become uncomfortably loud is often about the same as it is in someone with normal hearing. Thus, for someone with a cochlear hearing loss, there is a range of levels over which the rate of growth of loudness level with increasing sound level is greater than normal. This is known as loudness recruitment.[2]

Loudness recruitment has been the subject of study for decades. Various measures of recruitment have been employed in the audiology clinic as a diagnostic tool for cochlear hearing loss (e.g., Fowler 1928, 1936; Reger 1936; Steinberg and Gardner 1937; Hallpike and Hood 1951; Hood 1977). This section describes some common ways to measure recruitment, the likely cause of recruitment, and some consequences of recruitment for the perception of temporally varying sounds.

4.1 Measuring Loudness Recruitment

Loudness recruitment can perhaps be demonstrated most clearly in subjects with a unilateral cochlear impairment by obtaining loudness matches between the normal and impaired ears in the same subjects (e.g., Fowler 1928; Miskolczy-Fodor 1960; Moore et al. 1985; Zeng and Turner 1991; Moore et al. 1996). One way in which this is done is to present tones alternately at the two ears. The level of the tone is fixed in one ear and is varied in the other until it is judged to be equal in loudness to the one that is fixed in level. An example of results from such a procedure is shown in Figure 4.14, where loudness matches are shown for one unilaterally impaired subject at two different test frequencies. A subject with normal hearing in both ears would provide results along the diagonal (i.e., to be judged equal in loudness, the level would be roughly the same in the two ears). The results from the subject with unilateral impairment obviously do not fall along the diagonal. There is evidence for loudness recruitment at both test frequencies because the slopes of the curves fitted to the data points are greater than one. At 500 Hz (Fig. 4.14*A*), the data points reach the diagonal, indicating complete recruitment. At 2,000 Hz (Fig. 4.14*B*), the data points do not quite reach the diagonal, meaning that the recruitment is incomplete (this is sometimes referred to as partial or under recruitment).

There are other ways in which loudness recruitment can be measured. One is by using a categorical loudness scaling procedure in which the subject is asked to judge the loudness of a sound by classifying it into one

[2] Historically, loudness recruitment has been defined as an abnormally rapid growth of loudness rather than loudness level. Loudness is measured in sones, and it represents the sensory dimension that corresponds most closely to the physical dimension of sound intensity (see Stevens 1975). Loudness level is measured in phons, and it corresponds to the level of an equally loud 1,000-Hz tone (e.g., Fletcher and Munson 1933). As noted by Allen (1997), it is loudness level, not loudness, that grows more rapidly in hearing-impaired listeners than in normal-hearing listeners.

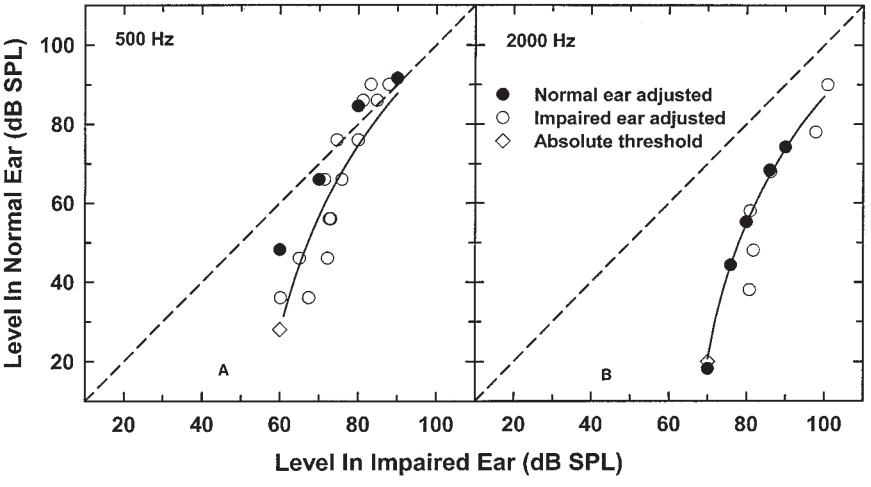


FIGURE 4.14. The levels of a sinusoidal signal in the normal and impaired ears of a subject with a unilateral hearing loss necessary to equate the perceived loudness in the 2 ears. The level of the tone was either fixed in the impaired ear and adjusted in the normal ear (solid circles) or fixed in the normal ear and adjusted in the impaired ear (open circles). Solid line: fit to the data based on a model of loudness, modified to account for the effects of hearing loss; dashed line: where points would lie if equal levels represented equal loudness in the two ears. A: 500-Hz signal; B: 2,000-Hz signal. (Data and model fits are from Moore and Glasberg 1997.)

of several verbal categories ranging, for example, from "cannot hear" to "too loud" (Pasco 1978; Allen et al. 1990; Cox et al. 1997). Other scaling procedures such as magnitude estimation or magnitude production also can be used to measure loudness recruitment (Stevens 1957; Hellman and Meiselman 1990, 1993). With all of these scaling procedures, the results from a subject with a cochlear hearing loss can be compared to those from a subject with normal hearing. An advantage of these procedures is that they can be used to measure recruitment in subjects with bilateral hearing loss.

4.2 Mechanisms Underlying Loudness Recruitment

It is unclear precisely what mechanisms underlie loudness recruitment or, for that matter, the normal perception of loudness. It is generally assumed, however, that the loudness of a sound is related to the total excitation produced by that sound (Zwicker 1960; Moore and Glasberg 1996). Thus the growth of loudness should be related not only to the rate at which the response around CF grows with increases in level but also to the rate at which regions with remote CFs are excited (or "recruited") with increases in level.

It is well known that cochlear hearing loss results in relatively poor frequency selectivity (e.g., Evans 1975; Zwicker and Schorn 1978; Carney and Nelson 1983; Stelmachowicz et al. 1985; Nelson 1991). As a consequence, the spread of excitation in the cochlea is broader than normal. This broad spread of excitation was once considered a likely explanation for loudness recruitment (Kiang et al. 1970; Evans 1975). Moore et al. (1985) tested this by having subjects with unilateral hearing loss match the loudness of a tone that alternated between the normal and impaired ears. Their subjects showed recruitment both without and, importantly, with a noise intended to mask spread of excitation. Thus their results strongly suggest that recruitment is not due to an abnormal spread of excitation (although it may contribute somewhat to it). In support of this conclusion are results from Zeng and Turner (1991), who observed recruitment in unilaterally impaired subjects whose hearing losses would have made the use of spread of excitation unlikely.

From their results, Zeng and Turner (1991) concluded that recruitment was related primarily to the amount of hearing loss at the test frequency. Consistent with this are results showing that, in general, the steepness of loudness growth curves, such as those in Figure 4.14, increases with increasing hearing loss at the test frequency (Miskolczy-Fodor 1960; Glasberg and Moore 1989; Hellman and Meiselman 1990, 1993). This has led many researchers to conclude that recruitment is related to a decrease in or loss of the compression that is normally seen at the basilar membrane (e.g., Moore et al. 1985, 1996; Yates 1990; Zeng and Turner 1991; Schlauch et al. 1998).

The contributions of abnormal spread of excitation and a loss of compression to loudness growth can be evaluated within a model of loudness developed by Moore and Glasberg (1996, 1997). Their model is conceptually similar to previous models (Fletcher and Munson 1937; Zwicker 1958; Zwicker and Scharf 1965; Zwicker et al. 1984; Zwicker and Fastl 1990). Its application to cochlear hearing impairment extends previous extensions of similar models (Florentine and Zwicker 1979; Humes et al. 1992; Florentine et al. 1997) by attempting to separate the effects of inner hair cell loss from the effects of outer hair cell loss. The model assumes that the effect of inner hair cell loss corresponds to an attenuation of the input, whereas the effect of outer hair cell loss corresponds to a loss or reduction of basilar membrane compression and an increase in the spread of excitation. The major sections of the model include the calculation of an excitation pattern, the transformation of excitation to a quantity called specific loudness (a transformation that involves a compressive nonlinearity), and then a summation of the area under the so-called specific loudness pattern, which is the specific loudness as a function of frequency. Their model does a reasonably good job in predicting loudness-matching data in hearing-impaired subjects (e.g., the solid lines in Figure 4.14 represent predictions from their model). According to Moore (1998), the results of their modeling efforts

suggest that loudness recruitment is caused mainly by the steeper input-output function of the basilar membrane. Abnormal spread of excitation appears to play only a minor role.

Using a different approach, Schlauch et al. (1998) evaluated the extent to which the growth of loudness level could be accounted for by the basilar membrane input-output function. They derived basilar membrane input-output functions from loudness-matching data for tones alternating between a normal and a threshold-shifted ear (where the threshold shift was due to either a hearing loss or the presence of a continuous masking noise). They compared the derived input-output functions with those measured directly in a chinchilla (from Ruggero et al. 1997). Their results were broadly consistent with the idea that the growth of loudness largely can be accounted for by the growth of response at the basilar membrane and that loudness recruitment reflects a loss of or decrease in basilar membrane compression.

In a recent study, Buus and Florentine (2002) examined how loudness grows with increasing stimulus level near threshold in individuals with cochlear hearing loss. They found that the growth of loudness near threshold was normal (Buus et al. 1998) but that the loudness at threshold was greater in subjects with cochlear loss. Their results are what would be expected if changes in loudness perception were mediated by changes in basilar membrane compression. Because the normal basilar membrane response growth is linear at low levels, the results in this level region should correspond well with those from hearing-impaired listeners, whose response is presumably linear at all levels. Buus and Florentine (2002) argued that the elevated loudness at threshold was a better definition of loudness recruitment than the classical definition of an abnormally rapid growth of loudness. Regardless of how recruitment is defined, it seems likely that differences in loudness and loudness growth between normal-hearing and hearing-impaired subjects reflect, at least to some extent, differences in the degree of residual basilar membrane compression.

4.3 Influences of Recruitment on the Perception of Temporally Varying Sounds

Although loudness recruitment is typically measured with steady-state sounds, it may impact the perception of temporally varying sounds such as speech or music. In particular, the fluctuations in those sounds may be "magnified" perceptually by the recruitment so that they appear more prominent to individuals with a cochlear hearing loss than they do to individuals with normal hearing. To evaluate this possibility, Moore et al. (1996) asked subjects with a unilateral cochlear hearing loss to adjust the modulation depth of a tone in one ear to match the perceived depth of a tone in the other ear. For the modulation rates tested (4–32 Hz), the modulation

depth had to be greater in the normal ear than in the impaired ear for the depth to be perceived as equal. An example of some of their results is shown in Figure 4.15. The solid line in Figure 4.15 represents a straight-line fit to the data points, whereas the dashed line represents predictions based on loudness-matching functions using steady-state tones. The dot-dashed line indicates where the data points would lie if the modulation depths were perceived equally in the two ears. The matching functions for the modulated tones were predicted reasonably well from the loudness-matching functions for steady-state tones, suggesting a close correspondence between loudness recruitment for steady-state and dynamic sounds. Moore and his colleagues (1996) found that the perceptual magnification caused by loudness recruitment was independent of modulation rate and modulation depth over the ranges they evaluated. Although the range of modulation rates was relatively small, the lack of an effect of modulation rate suggests that the mechanism responsible for loudness recruitment is

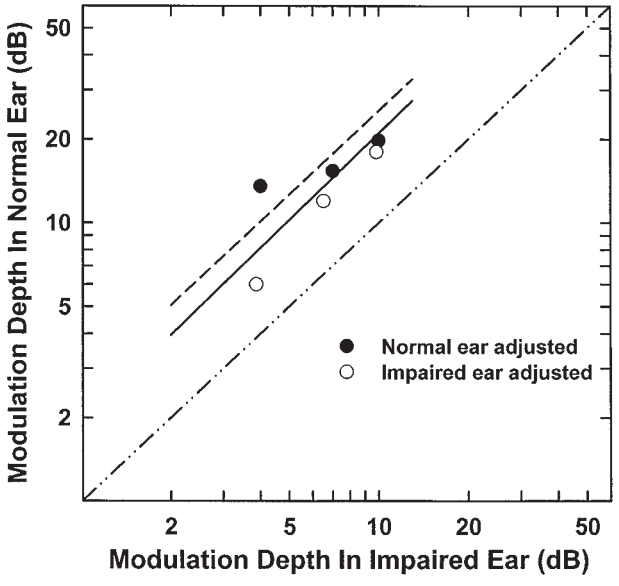


FIGURE 4.15. The modulation depth in the normal and impaired ears of a subject with a unilateral hearing loss necessary to equate the perceived modulation depth in the 2 ears. The stimuli were 1,000-Hz tones modulated sinusoidally on a dB scale at a rate of 8 Hz. The modulation depth was either fixed in the impaired ear and adjusted in the normal ear (solid circles) or fixed in the normal ear and adjusted in the impaired ear (open circles). Solid line: best-fitting line to the data; dashed line: predicted relationship based on loudness matches with steady-state tones; dot-dashed line: where points would lie if the modulation depth were perceived identically in the two ears. (Data and fits are from Moore et al. 1996.)

fast acting. As such, their results are consistent with the possibility that recruitment is due to a loss of basilar membrane compression, which is also fast acting.

Recruitment also can affect the ability of an individual to detect the presence of a temporal gap in an otherwise uninterrupted burst of narrowband noise. Gap detection is a common measure of auditory temporal resolution (see Viemeister and Plack 1993). The subject's task is to determine which one of several sounds contains a temporal gap. The goal is to find the smallest detectable gap (the gap-detection threshold). Gap detection is generally unaffected by cochlear hearing loss when the sound marking the gap is a sinusoid and thus does not contain any inherent amplitude fluctuations (Moore and Glasberg 1988), but it is affected by hearing loss when the sound contains slow fluctuations, like those in a narrow band of noise (Fitzgibbons and Wightman 1982). The task with narrow bands of noise can be especially difficult for any listener because the inherent fluctuations can be mistaken for the experimentally introduced temporal gap. Glasberg and Moore (1992) suggested that the poorer performance by hearing-impaired subjects might reflect the fact that the task is even more difficult for them because the inherent fluctuations are being perceptually enhanced due to loudness recruitment, making them less easily distinguishable from the gap itself. To evaluate this possibility, they measured gap-detection thresholds in narrow bands of noise in conditions where the envelope of the noise was processed by raising it to a power that varied from less than 1.0 (envelope fluctuations reduced) to greater than 1.0 (envelope fluctuations enhanced). When the envelope was raised to a power of 2.0 for the normal-hearing subjects, they performed about as poorly as did the hearing-impaired subjects listening to noise under normal conditions (power of 1.0). Similarly, when the envelope was raised to a power of 0.5 for the hearing-impaired subjects, they performed about as well as the normal-hearing subjects under normal conditions (power of 1.0). These results are consistent with the idea that a loss of basilar membrane compression hampers certain aspects of temporal processing by accentuating the temporal fluctuations in sounds.

5. Summary

It is clear from this chapter and from Oxenham and Bacon (Chapter 3) that there are several psychophysical phenomena that appear closely linked to the compression observed at the basilar membrane. Indeed, under certain circumstances (in nonsimultaneous masking, when the frequency of the masker is well below the frequency of the signal), there is quantitative agreement between the slope of the growth-of-masking function and the slope of the basilar membrane input-output function as measured directly in other mammals. Both indicate a compression ratio of about 5:1, at least

in cases of normal hearing. This quantitative agreement provides strong evidence in favor of a close link between auditory perception and peripheral compression. Further support comes from the fact that temporary or permanent cochlear damage reduces both the physiologically measured and the psychophysically estimated amount of compression. Growth of masking in simultaneous masking also probably reflects the influence of basilar membrane compression. However, it considerably underestimates the degree of that compression, probably as a result of complex interactions (e.g., suppression) between the temporally overlapping masker and signal.

A wide range of temporal-processing tasks appears to be influenced by basilar membrane compression. Support for this comes not only from the modeling results described by Oxenham and Bacon (Chapter 3) but also from the psychophysical results described in this chapter showing the effects of temporary or permanent cochlear hearing loss on behavioral performance. For some of the tasks, the link to basilar membrane compression appears fairly well established, whereas for others, it is weaker. Importantly, the focus on the potential role of compression in temporal processing sheds light on how a peripheral hearing loss, which will most likely affect the outer hair cells in the cochlea and hence basilar membrane compression, could disrupt performance on tasks that are typically thought to reflect primarily central processing.

Loudness recruitment is common in individuals with cochlear hearing loss. Although the understanding of the mechanisms underlying normal loudness perception remains somewhat sketchy, there is growing evidence that loudness recruitment is due primarily to a loss of basilar membrane compression and the concomitant loss of gain at low stimulus levels (see Fig. 1.4 in Bacon, Chapter 1). This explains why the lower end (the absolute or quiet threshold) of the dynamic range is elevated, but the upper end (the uncomfortable loudness level) is usually unaffected. Not only does recruitment reduce the size of the dynamic range of hearing, it also accentuates amplitude fluctuations. This can, for example, adversely affect the ability of listeners to follow the temporal structure or pattern of sound.

Finally, although the focus of this chapter and Chapter 3 by Oxenham and Bacon was on the role of basilar membrane compression in accounting for various psychophysical results, it is worth emphasizing that processing beyond the cochlea is obviously vital to auditory perception. The value of emphasizing the auditory periphery is that it provides insight into the initial auditory processing, which is both obligatory and adversely affected by cochlear hearing loss. This emphasis can thus provide insights into the limits imposed by this processing as well as insights into the potentially debilitating effects of hearing loss. Levitt (Chapter 5) and Zeng (Chapter 6) describe some of the difficulties that cochlear hearing loss presents in terms of rehabilitation, particularly with regard to signal processing strategies in hearing aids (Levitt, Chapter 5) and cochlear implants (Zeng, Chapter 6).

Acknowledgments. The writing of this chapter was partially supported by National Institute on Deafness and Other Communication Disorders Grants DC-01376 (to S. P. Bacon) and DC-03909 (to A. J. Oxenham). We thank Li Liu for her assistance with the figures.

References

Alcántara JI, Moore BCJ (2002) The relative role of beats and combination tones in determining the shapes of masking patterns. II. Hearing-impaired listeners. Hear Res 165:103–116.

Allen JB (1997) Derecruitment by multiband compression in hearing aids. In: Jesteadt W (ed) Modeling Sensorineural Hearing Loss. Mahwah, NJ: Erlbaum Associates, pp. 99–112.

Allen JB, Hall JL, Jeng PS (1990) Loudness growth in ½-octave bands (LGOB)—a procedure for the assessment of loudness. J Acoust Soc Am 88:745–753.

Bacon SP (1990) Effect of masker level on overshoot. J Acoust Soc Am 88:698–702.

Bacon SP, Healy EW (2000) Effects of ipsilateral and contralateral precursors on the temporal effect in simultaneous masking with pure tones. J Acoust Soc Am 107:1589–1597.

Bacon SP, Hicks ML (2000) The influence of aspirin on temporal effects in simultaneous masking with noise and tonal maskers. J Acoust Soc Am 107:2914.

Bacon SP, Lee J (1997) The modulated-unmodulated difference: effects of signal frequency and masker modulation depth. J Acoust Soc Am 101:3617–3624.

Bacon SP, Liu L (2000) Effects of ipsilateral and contralateral precursors on overshoot. J Acoust Soc Am 108:1811–1818.

Bacon SP, Moore BCJ (1986) Temporal effects in simultaneous pure-tone masking: effects of signal frequency, masker/signal frequency ratio, and masker level. Hear Res 23:257–266.

Bacon SP, Smith MA (1991) Spectral, intensive, and temporal factors influencing overshoot. Q J Exp Psychol 43A:373–399.

Bacon SP, Takahashi GA (1992) Overshoot in normal-hearing and hearing-impaired subjects. J Acoust Soc Am 91:2865–2871.

Bacon SP, Viemeister NF (1985a) Simultaneous masking by gated and continuous sinusoidal maskers. J Acoust Soc Am 78:1220–1230.

Bacon SP, Viemeister NF (1985b) The temporal course of simultaneous tone-on-tone masking. J Acoust Soc Am 78:1231–1235.

Bacon SP, Hedrick MS, Grantham DW (1988) Temporal effects in simultaneous pure-tone masking in subjects with high-frequency sensorineural hearing loss. Audiology 27:313–323.

Bacon SP, Lee J, Peterson DN, Rainey D (1997) Masking by modulated and unmodulated noise: effects of bandwidth, modulation rate, signal frequency, and masker level. J Acoust Soc Am 101:1600–1610.

Bacon SP, Boden LN, Lee J, Repovsch JL (1999) Growth of simultaneous masking for $f_m < f_s$: effects of overall frequency and level. J Acoust Soc Am 106:341–350.

Buus S (1985) Release from masking caused by envelope fluctuations. J Acoust Soc Am 78:1958–1965.

Buus S, Florentine M (2002) Growth of loudness in listeners with cochlear hearing losses: recruitment reconsidered. J Assoc Res Otolaryngol 3:120–139.

Buus S, Müsch H, Florentine M (1998) On loudness at threshold. J Acoust Soc Am 104:399–410.
Carlyon RP, Butt M (1993) Effects of aspirin on human auditory filters. Hear Res 66:233–244.
Carlyon RP, Datta AJ (1997a) Excitation produced by Schroeder-phase complexes: evidence for fast-acting compression in the auditory system. J Acoust Soc Am 101:3636–3647.
Carlyon RP, Datta AJ (1997b) Masking period patterns of Schroeder-phase complexes: effects of level, number of components, and phase of flanking components. J Acoust Soc Am 101:3648–3657.
Carlyon RP, Sloan EP (1987) The "overshoot" effect and sensory hearing impairment. J Acoust Soc Am 82:1078–1081.
Carlyon RP, Buus S, Florentine M (1990) Temporal integration of trains of tone pulses by normal and by cochlearly impaired listeners. J Acoust Soc Am 87: 260–268.
Carney AE, Nelson DA (1983) An analysis of psychophysical tuning curves in normal and pathological ears. J Acoust Soc Am 73:268–278.
Cazals Y (2000) Auditory sensorineural alterations induced by salicylate. Prog Neurobiol 62:583–631.
Champlin CA, McFadden D (1989) Reductions in overshoot following intense sound exposures. J Acoust Soc Am 85:2005–2011.
Cox RM, Alexander GC, Taylor IM, Gray GA (1997) The contour test of loudness perception. Ear Hear 18:388–400.
Davis H, Morgan CT, Hawkins JE Jr, Galambos R, Smith FW (1950) Temporary deafness following exposure to loud tones and noise. Acta Otolaryngol Suppl 88:1–57.
Eddins DA (2001) Measurement of auditory temporal processing using modified masking period patterns. J Acoust Soc Am 109:1550–1558.
Egan JP, Hake HW (1950) On the masking pattern of a simple auditory stimulus. J Acoust Soc Am 22:622–630.
Elliott LL (1965) Changes in the simultaneous masked threshold of brief tones. J Acoust Soc Am 38:738–746.
Elliott LL (1971) Backward and forward masking. Audiology 10:65–76.
Elliott LL (1975) Temporal and masking phenomena in persons with sensorineural hearing loss. Audiology 14:336–353.
Evans EF (1975) The sharpening of cochlear frequency selectivity in the normal and abnormal cochlea. Audiology 14:419–442.
Fastl H (1975) Loudness and masking patterns of narrow noise bands. Acustica 33: 266–271.
Fastl H (1976) Temporal masking effects. I. Broad band noise masker. Acustica 35: 287–302.
Fitzgibbons PJ, Wightman FL (1982) Gap detection in normal and hearing-impaired listeners. J Acoust Soc Am 72:761–765.
Fletcher H (1940) Auditory patterns. Rev Mod Phys 12:47–65.
Fletcher H, Munson WA (1933) Loudness, its definition, measurement and calculation. J Acoust Soc Am 5:82–108.
Fletcher H, Munson WA (1937) Relation between loudness and masking. J Acoust Soc Am 9:1–10.
Florentine M, Zwicker E (1979) A model of loudness summation applied to noise-induced hearing loss. Hear Res 1:121–132.

Florentine M, Fastl H, Buus S (1988) Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. J Acoust Soc Am 84:195–203.

Florentine M, Buus S, Hellman RP (1997) A model of loudness summation applied to high-frequency hearing loss. In: Jesteadt W (ed) Modeling Sensorineural Hearing Loss. Mahwah, NJ: Erlbaum Associates, pp. 187–198.

Fowler EP (1928) Marked deafened areas in normal ears. Arch Otolaryngol 8:151–155.

Fowler EP (1936) A method for the early detection of otosclerosis. Arch Otolaryngol 24:731–741.

Garner WR (1947) The effect of frequency spectrum on temporal integration of energy in the ear. J Acoust Soc Am 19:808–815.

Garner WR, Miller GA (1947) The masked threshold of pure tones as a function of duration. J Exp Psychol 37:293–303.

Gengel RW, Watson CS (1971) Temporal integration. I. Clinical implications of a laboratory study. II. Additional data from hearing-impaired subjects. J Speech Hear Disord 36:213–224.

Glasberg BR, Moore BCJ (1989) Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. Scand Audiol Suppl 32:1–25.

Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. Hear Res 47:103–138.

Glasberg BR, Moore BCJ (1992) Effects of envelope fluctuations on gap detection. Hear Res 64:81–92.

Glasberg BR, Moore BCJ (2000) Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. J Acoust Soc Am 108:2318–2328.

Glasberg BR, Moore BCJ, Bacon SP (1987) Gap detection and masking in hearing-impaired and normal-hearing subjects. J Acoust Soc Am 81:1546–1556.

Goodman DA, Smith RL, Chamberlain SC (1982) Intracellular and extracellular responses in the organ of Corti of the gerbil. Hear Res 7:161–169.

Green DM (1969) Masking with continuous and pulsed sinusoids. J Acoust Soc Am 46:939–946.

Green DM, Swets JA (1966) Signal Detection Theory and Psychophysics. New York: John Wiley and Sons.

Green DM, Birdsall TG, Tanner WP Jr (1957) Signal detection as a function of signal intensity and duration. J Acoust Soc Am 29:523–531.

Gregan MJ, Bacon SP, Lee J (1998) Masking by sinusoidally amplitude-modulated tonal maskers. J Acoust Soc Am 103:1012–1021.

Hall JW, Haggard MP, Fernandes MA (1984) Detection in noise by spectro-temporal pattern analysis. J Acoust Soc Am 76:50–56.

Hallpike CS, Hood JD (1951) Some recent work on auditory adaptation and its relationship to the loudness recruitment phenomenon. J Acoust Soc Am 23:270–274.

Hellman RP, Meiselman CH (1990) Loudness relations for individuals and groups in normal and impaired hearing. J Acoust Soc Am 88:2596–2606.

Hellman RP, Meiselman CH (1993) Rate of loudness growth for pure tones in normal and impaired hearing. J Acoust Soc Am 93:966–975.

Hicks ML, Bacon SP (1999) Effects of aspirin on psychophysical measures of frequency selectivity, two-tone suppression, and growth of masking. J Acoust Soc Am 106:1436–1451.
Hicks ML, Bacon SP (2000) The effects of aspirin on a psychophysical estimate of basilar membrane compression. J Acoust Soc Am 107:2914.
Hood JD (1977) Loudness balance procedures for the measurement of recruitment. Audiology 16:215–228.
Houtgast T (1973) Psychophysical experiments on "tuning curves" and "two-tone inhibition." Acustica 29:168–179.
Houtgast T (1974) Lateral Suppression in Hearing. Ph.D. Thesis, Free University of Amsterdam. Amsterdam: Academische Pers. BV.
Hughes JW (1946) The threshold of audition for short periods of stimulation. Philos Trans R Soc Lond B 133:486–490.
Humes LE, Jesteadt W, Lee LW (1992) Modeling the effects of sensorineural hearing loss on auditory perception. In: Cazals Y, Demany L, Horner K (eds) Auditory Physiology and Perception. Oxford, UK: Pergamon Press, pp. 617–624.
Jesteadt W (1980) Effects of masker level and signal delay on forward masking in normal and impaired listeners. J Acoust Soc Am 68:S28.
Jesteadt W, Bacon SP, Lehman JR (1982) Forward masking as a function of frequency, masker level, and signal delay. J Acoust Soc Am 71:950–962.
Johnson-Davies D, Patterson RD (1979) Psychophysical tuning curves: restricting the listening band to the signal region. J Acoust Soc Am 65:765–770.
Johnstone BM, Patuzzi R, Yates GK (1986) Basilar membrane measurements and the travelling wave. Hear Res 22:147–153.
Kakehata S, Santos-Sacchi J (1996) Effects of salicylate and lanthanides on outer hair cell motility and associated gating charge. J Neurosci 16:4881–4889.
Kiang NYS, Wantanabe T, Thomas EC, Clark LF (1965) Discharge Patterns of Single Fibers in the Cat's Auditory Nerve. Cambridge, MA: MIT Press.
Kiang NYS, Moxon EC, Levine RA (1970) Auditory-nerve activity in cats with normal and abnormal cochleas. In: Wolstenholme GEW, Knight JJ (eds) Sensorineural Hearing Loss. London: Churchill, pp. 241–268.
Kidd G Jr, Mason CR, Feth LL (1984) Temporal integration of forward masking in listeners having sensorineural hearing loss. J Acoust Soc Am 75:937–944.
Kimberley BP, Nelson DA, Bacon SP (1989) Temporal overshoot in simultaneous-masked psychophysical tuning curves from normal and hearing-impaired listeners. J Acoust Soc Am 85:1660–1665.
Kohlrausch A, Sander A (1995) Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. J Acoust Soc Am 97:1817–1829.
Lentz JJ, Leek MR (2001) Psychophysical estimates of cochlear phase response: masking by harmonic complexes. J Assoc Res Otolaryngol 2:408–422.
Liberman MC, Brown MC (1986) Physiology and anatomy of single olivocochlear neurons in the cat. Hear Res 24:17–36.
Lüscher E, Zwislocki J (1947) Adaptation of the ear to sound stimuli. J Acoust Soc Am 21:135–139.
Macmillan NA, Creelman CD (1991) Detection Theory: A User's Guide. New York: Cambridge University Press.
McFadden D (1986) The curious half-octave shift: evidence for a basalward migration of the traveling-wave envelope with increasing intensity. In: Salvi RJ,

Henderson D, Hamernik RP, Colletti V (eds) Basic and Applied Aspects of Noise-Induced Hearing Loss. New York: Plenum Press, pp. 295–312.
McFadden D, Champlin CA (1990) Reductions in overshoot during aspirin use. J Acoust Soc Am 87:2634–2642.
McFadden D, Wright BA (1990) Temporal decline of masking and comodulation detection differences. J Acoust Soc Am 88:711–724.
McFadden D, Wright BA (1992) Temporal decline of masking and comodulation masking release. J Acoust Soc Am 92:144–156.
McFadden D, Plattsmier HS, Pasanen EG (1984) Aspirin-induced hearing loss as a model of sensorineural hearing loss. Hear Res 16:251–260.
Miskolczy-Fodor F (1960) Relation between loudness and duration of tonal pulses. III. Response in cases of abnormal loudness function. J Acoust Soc Am 32:486–492.
Moore BCJ (1986) Parallels between frequency selectivity measured psychophysically and in cochlear mechanics. Scand Audiol Suppl 25:139–152.
Moore BCJ (1993) Frequency analysis and pitch perception. In: Yost WA, Popper AN, Fay RR (eds) Human Psychophysics. New York: Springer-Verlag, pp. 56–115.
Moore BCJ (1998) Cochlear Hearing Loss. London: Whurr Publishers.
Moore BCJ, Glasberg BR (1987) Factors affecting thresholds for sinusoidal signals in narrow-band maskers with fluctuating envelopes. J Acoust Soc Am 82:69–79.
Moore BCJ, Glasberg BR (1988) Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears. J Acoust Soc Am 83:1093–1101.
Moore BCJ, Glasberg BR (1996) A revision of Zwicker's loudness model. Acustica 82:335–345.
Moore BCJ, Glasberg BR (1997) A model of loudness perception applied to cochlear hearing loss. Aud Neurosci 3:289–311.
Moore BCJ, Glasberg BR, Hess RF, Birchall JP (1985) Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears. J Acoust Soc Am 77:1505–1513.
Moore BCJ, Wojtczak M, Vickers DA (1996) Effect of loudness recruitment on the perception of amplitude modulation. J Acoust Soc Am 100:481–489.
Munson WA, Gardner MB (1950) Loudness patterns—a new approach. J Acoust Soc Am 22:177–190.
Murnane O, Turner CW (1991) Growth of masking in sensorineural hearing loss. Audiology 30:275–285.
Murugasu E, Russell IJ (1996) The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea. J Neurosci 16: 325–332.
Narayan SS, Temchin AN, Recio A, Ruggero MA (1998) Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae. Science 282:1882–1884.
Nelson DA (1991) High-level psychophysical tuning curves: forward masking in normal-hearing and hearing-impaired listeners. J Speech Hear Res 34:1233–1249.
Nelson DA, Freyman RL (1987) Temporal resolution in sensorineural hearing-impaired listeners. J Acoust Soc Am 81:709–720.
Nelson DA, Schroder AC (1996) Release from upward spread of masking in regions of high-frequency hearing loss. J Acoust Soc Am 100:2266–2277.
Nelson DA, Schroder AC (1997) Linearized response growth inferred from growth-of-masking slopes in ears with cochlear hearing loss. J Acoust Soc Am 101: 2186–2201.

Nelson DA, Swain AC (1996) Temporal resolution within the "upper accessory excitation" of a masker. Acustica 82:328–334.
Nelson DA, Schroder AC, Wojtczak M (2001) A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners. J Acoust Soc Am 110:2045–2064.
Nuttall AL, Dolan DF (1996) Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig. J Acoust Soc Am 99:1556–1565.
Oxenham AJ (2001) Forward masking: adaptation or integration? J Acoust Soc Am 109:732–741.
Oxenham AJ, Dau T (2001a) Reconciling frequency selectivity and phase effects in masking. J Acoust Soc Am 110:1525–1538.
Oxenham AJ, Dau T (2001b) Towards a measure of auditory-filter phase response. J Acoust Soc Am 110:3169–3178.
Oxenham AJ, Moore BCJ (1994) Modeling the additivity of nonsimultaneous masking. Hear Res 80:105–118
Oxenham AJ, Moore BCJ (1995) Additivity of masking in normally hearing and hearing-impaired subjects. J Acoust Soc Am 98:1921–1934.
Oxenham AJ, Moore BCJ (1997) Modeling the effects of peripheral nonlinearity in listeners with normal and impaired hearing. In: Jesteadt W (ed) Modeling Sensorineural Hearing Loss. Mahwah, NJ: Erlbaum Associates, pp. 273–288.
Oxenham AJ, Plack CJ (1997) A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. J Acoust Soc Am 101: 3666–3675.
Oxenham AJ, Plack CJ (2000) Effects of masker frequency and duration in forward masking: further evidence for the influence of peripheral nonlinearity. Hear Res 150:258–266.
Oxenham AJ, Moore BCJ, Vickers DA (1997) Short-term temporal integration: evidence for the influence of peripheral compression. J Acoust Soc Am 101: 3676–3687.
Pascoe DP (1978) An approach to hearing aid selection. Hear Instrum 29:12–16.
Patterson RD (1976) Auditory filter shapes derived with noise stimuli. J Acoust Soc Am 59:640–654.
Penner MJ (1972) Neural or energy summation in a Poisson counting model. J Math Psychol 9:286–293.
Penner MJ (1980) The coding of intensity and the interaction of forward and backward masking. J Acoust Soc Am 67:608–616.
Plack CJ, Oxenham AJ (1998) Basilar-membrane nonlinearity and the growth of forward masking. J Acoust Soc Am 103:1598–1608.
Plomp R (1964) The ear as a frequency analyser. J Acoust Soc Am 36:1628–1636.
Plomp R, Bouman MA (1959) Relation between hearing threshold and duration for tone pulses. J Acoust Soc Am 31:749–758.
Recio A, Rhode WS (2000) Basilar membrane responses to broadband stimuli. J Acoust Soc Am 108:2281–2298.
Reger SN (1936) Differences in loudness response of the normal and hard-of-hearing ear at intensity levels slightly above the threshold. Ann Otol Rhinol Laryngol 45:1029–1039.
Rhode WS, Recio A (2000) Study of mechanical motions in the basal region of the chinchilla cochlea. J Acoust Soc Am 107:3317–3332.

Rosen S, Baker RJ, Darling A (1998) Auditory filter nonlinearity at 2 kHz in normal hearing listeners. J Acoust Soc Am 103:2539–2550.
Ruggero MA (1992) Responses to sound of the basilar membrane of the mammalian cochlea. Curr Opin Neurobiol 2:449–456.
Ruggero MA, Rich NC (1991) Furosemide alters organ of Corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane. J Neurosci 11: 1057–1067.
Ruggero MA, Rich NC, Recio A, Narayan SS, Robles L (1997) Basilar-membrane responses to tones at the base of the chinchilla cochlea. J Acoust Soc Am 101: 2151–2163.
Russell IJ, Schauz C (1995) Salicylate ototoxicity: effects on stiffness and electromotility of outer hair cells isolated from the guinea pig cochlea. Aud Neurosci 1:309–319.
Samoilova IK (1959) Masking of short tone signals as a function of the time interval between masked and masking sounds. Biofizika 4:550–558.
Schlauch RS, DiGiovanni JJ, Ries DT (1998) Basilar membrane nonlinearity and loudness. J Acoust Soc Am 103:2010–2020.
Schöne P (1977) Nichtlinearintäten im mithörschwellen-tonheitsmuster von sinustönen. Acustica 37:37–44.
Schroeder MR (1970) Synthesis of low peak-factor signals and binary sequences with low autocorrelation. IEEE Trans Inf Theory 16:85–89.
Shehata WE, Brownell WE, Dieler R (1991) Effects of salicylate on shape, electromotility and membrane characteristics of isolated outer hair cells from guinea pig cochlea. Acta Otolaryngol 111:707–718.
Shera CA, Guinan JJ Jr, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc Natl Acad Sci USA 99:3318–3323.
Small AM Jr (1959) Pure-tone masking. J Acoust Soc Am 31:1619–1625.
Smith RL (1977) Short-term adaptation in single auditory nerve fibers: some poststimulatory effects. J Neurophysiol 40:1098–1112.
Smith RL (1979) Adaptation, saturation, and physiological masking in single auditory-nerve fibers. J Acoust Soc Am 65:166–178.
Smith RL, Zwislocki JJ (1975) Short-term adaptation and incremental responses in single auditory-nerve fibers. Biol Cybern 17:169–182.
Smits JTS, Duifhuis H (1982) Masking and partial masking in listeners with a high-frequency hearing loss. Audiology 21:310–324.
Steinberg JC, Gardner MB (1937) The dependence of hearing impairment on sound intensity. J Acoust Soc Am 9:11–23.
Stelmachowicz PG, Jesteadt W, Gorga MP, Mott J (1985) Speech perception ability and psychophysical tuning curves in hearing-impaired listeners. J Acoust Soc Am 77:620–627.
Stelmachowicz PG, Lewis DE, Larson LL, Jesteadt W (1987) Growth of masking as a measure of response growth in hearing-impaired listeners. J Acoust Soc Am 81: 1881–1887.
Stevens SS (1957) On the psychophysical law. Psychol Rev 64:153–181.
Stevens SS (1975) Psychophysics. New York: John Wiley and Sons.
Strickland EA (2001) The relationship between frequency selectivity and overshoot. J Acoust Soc Am 109:2062–2073.
Stypulkowski PH (1990) Mechanisms of salicylate ototoxicity. Hear Res 46:113–146.

Summers V (2000) Effects of hearing impairment and presentation level on masking period patterns for Schroeder-phase harmonic complexes. J Acoust Soc Am 108: 2307–2317.
Summers V, Leek MR (1998) Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners. Hear Res 118:139–150.
Trees DE, Turner CW (1986) Spread of masking in normal subjects and in subjects with high-frequency hearing loss. Audiology 25:70–83.
Turner CW, Doherty KA (1997) Temporal masking and the "active process" in normal and hearing-impaired listeners. In: Jesteadt W (ed) Modeling Sensorineural Hearing Loss. Mahwah, NJ: Erlbaum Associates, pp. 387–396.
Viemeister NF, Plack CJ (1993) Time analysis. In: Yost WA, Popper AN, Fay RR (eds) Human Psychophysics. New York: Springer-Verlag, pp. 116–154.
Viemeister NF, Wakefield GH (1991) Temporal integration and multiple looks. J Acoust Soc Am 90:858–865.
von Klitzing R, Kohlrausch A (1994) Effect of masker level on overshoot in running- and frozen-noise maskers. J Acoust Soc Am 95:2192–2201.
Wegel RL, Lane CE (1924) The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear. Phys Rev 23:266–285.
Widin GP, Viemeister NF (1979) Intensive and temporal effects in pure-tone forward masking. J Acoust Soc Am 66:388–395.
Wojtczak M, Schroder AC, Kong YY, Nelson DA (2001) The effect of basilar-membrane nonlinearity on the shapes of masking period patterns in normal and impaired hearing. J Acoust Soc Am 109:1571–1586.
Wright HN (1968) Clinical measurement of temporal auditory summation. J Speech Hear Res 11:109–127.
Yates GK (1990) Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions. Hear Res 50:145–162.
Zeng FG, Turner CW (1991) Binaural loudness matches in unilaterally impaired listeners. Q J Exp Psychol 43A:565–583.
Zwicker E (1958) Über psychologische und methodische Grundlagen der Lautheit. Acustica 8:237–258.
Zwicker E (1960). Ein Verfahren zur Berechnung der Lautstärke. Acustica 10: 304–308.
Zwicker E (1965a) Temporal effects in simultaneous masking by white-noise bursts. J Acoust Soc Am 37:653–663.
Zwicker E (1965b) Temporal effects in simultaneous masking and loudness. J Acoust Soc Am 38:132–141.
Zwicker E (1976a) Psychoacoustic equivalent of period histograms. J Acoust Soc Am 59:166–175.
Zwicker E (1976b) Masking period patterns of harmonic complex tones. J Acoust Soc Am 60:429–439.
Zwicker E (1976c) Mithörschwellen-periodenmuster amplitudenmodulierter töne. Acustica 36:113–120.
Zwicker E, Fastl H (1990) Psychoacoustics—Facts and Models. Berlin: Springer-Verlag.
Zwicker E, Scharf B (1965) A model of loudness summation. Psychol Rev 72:3–26.
Zwicker E, Schorn K (1978) Psychoacoustical tuning curves in audiology. Audiology 17:120–140.

Zwicker E, Schorn K (1982) Temporal resolution in hard-of-hearing patients. Audiology 21:474–492.
Zwicker E, Fastl H, Dallmayr C (1984) BASIC-Program for calculating the loudness of sounds from their 1/3-oct. band spectra according to ISO 522B. Acustica 55:63–67.
Zwislocki J (1960) Theory of temporal auditory summation. J Acoust Soc Am 32: 1046–1060.
Zwislocki JJ (1969) Temporal summation of loudness: an analysis. J Acoust Soc Am 46:431–441.
Zwislocki J, Pirodda E, Rubin H (1959) On some poststimulatory effects at the threshold of audibility. J Acoust Soc Am 31:9–14.

5
Compression Amplification

Harry Levitt

1. Background

The normal ear has a dynamic range of well over 100 dB (see Fig. 1 in Bacon, Chapter 1). This is a remarkably wide range, corresponding to a power ratio of more than 10 billion to 1. Even more remarkable is the neural processing mechanism whereby this extremely wide range of sound levels is encoded by neural elements having a relatively narrow dynamic range (generally on the order of 30 dB). Although the exact mechanism whereby acoustic signals are encoded by the auditory system is not fully understood, it is clear that some form of amplitude compression is taking place.

The available experimental evidence indicates that the outer hair cells play an important role in compressing the amplitude of the encoded acoustic signals. Damage to the outer hair cells can thus reduce or eliminate the compression characteristics of the auditory system (Bacon, Chapter 1; Cooper, Chapter 2; Bacon and Oxenham, Chapter 4; Zeng, Chapter 6). It is important in fitting hearing aids to know the extent to which the compression characteristics of the ear have been damaged and the most effective method of acoustic signal processing to compensate for the reduction in or loss of amplitude compression in auditory signal processing.

The dynamic range of hearing is usually reduced in the case of sensorineural hearing loss (Bacon, Chapter 1; Bacon and Oxenham, Chapter 4). The amount by which the dynamic range is reduced for this type of impairment is roughly equal to the magnitude of the hearing loss. This is because elevation of the hearing threshold is not accompanied by an increase in the loudness discomfort level. Typically, loudness discomfort levels in sensorineural loss are roughly the same as for normal hearing. As a consequence, the dynamic range of the impaired ear (in dB) is approximately equal to the normal dynamic range less the magnitude of the hearing loss.

It should be noted that for a purely conductive hearing loss, the loudness discomfort level is increased by an amount equal to the hearing loss and

that the dynamic range of a conductively impaired ear is essentially the same as that for a normal ear. Most conductive hearing losses can now be treated successfully by medical or surgical means and do not usually require acoustic amplification, although there are cases where medical/surgical intervention is contraindicated or unsuccessful and acoustic amplification is necessary.

The vast majority of candidates for acoustic amplification have a sensorineural hearing loss, and the problem of a significantly reduced dynamic range needs to be addressed in many of these cases. This was not always the case. Before the 1960s, the majority of candidates for acoustic amplification had conductive and/or mild to moderate sensorineural hearing losses. Individuals with severe sensorineural losses were often regarded as "unamplifiable." The change in the demographics of hearing aid candidacy needs to be taken into account in interpreting the results of early studies on acoustic amplification and methods of hearing aid fitting. Most of the subjects in the classic Harvard and British Medical Research Council studies, for example, had conductive hearing losses (Davis et al. 1947; Medical Research Council 1947). Major advances have been made in the medical/surgical treatment of conductive losses since these early studies so that today hearing aids are fitted mostly to individuals with sensorineural rather than conductive losses. Furthermore, advances in hearing aid technology have been such that severe sensorineural losses with narrow dynamic ranges are no longer regarded as unamplifiable.

An important consideration in addressing the problem of reduced dynamic range is that of loudness distortion. The rate at which loudness grows with increasing signal level is the subject of much controversy. A commonly held view is that loudness grows much more rapidly than normal in sensorineural hearing impairment, giving rise to an effect known as loudness recruitment (Bacon and Oxenham, Chapter 4). On the other hand, Neely and Allen (1997) have pointed out that if loudness is specified in linear units (e.g., sones) rather than log units (e.g., loudness level in phons), then the rate of growth of loudness with intensity (also specified in linear units) is actually less than the normal rate of loudness growth over most of the audible range. Another view, held by Buus and Florentine (2002), is that loudness is abnormally high at elevated thresholds but that the rate of loudness growth near threshold is similar in normal and impaired ears.

The loudness of broadband sounds such as speech is affected in a complex way by the reduction in dynamic range in sensorineural hearing loss. Part of the problem is that the dynamic range varies with frequency, usually decreasing with increasing frequency. Loudness summation (the increase in loudness with increasing duration) is also affected by the hearing loss (Florentine et al. 1980). The resulting distortion in loudness relationships among the different frequency components of speech is believed to contribute significantly to the reduction in speech intelligibility (Villchur 1973).

This chapter is concerned with methods of amplification designed to deal with the problem of reduced dynamic range. The primary goal is to make speech more intelligible and to improve the overall quality of amplified sound (for music and other important sounds as well as speech). These techniques may also help restore normal loudness relationships.

2. The Many Forms of Amplitude Compression

Simple linear amplification can compensate for the elevation of the auditory threshold, but it cannot compensate for the reduced dynamic range of hearing. If the weaker sounds of speech are amplified so as to make them audible, the stronger speech sounds become uncomfortably loud. In addition, the abnormal growth of loudness with signal level distorts the loudness relationships among the different sounds of speech. A practical approach to this problem is to use compression amplification in which the gain of the amplifier is reduced for more intense sounds (see Fig. 6 in Bacon, Chapter 1). There are, however, many different forms of compression amplification, each having its own advantages and disadvantages. There are also different schools of thought as to how best to implement compression amplification in a hearing aid.

2.1 Compression Variables

The earliest compression amplifiers used analog components that imposed constraints on the characteristics of the compression circuit. A typical analog compression amplifier would operate in this way. The level of the incoming signal is determined by means of a rectifier with a fixed integration time. When an increase in signal level is detected, the gain of the amplifier is reduced. The change in gain is typically accomplished by discharging a capacitor, the voltage across the capacitor being used to control the gain of the amplifier. A resistance is used to slow the rate of discharge. Because the voltage across the capacitor decays exponentially, the gain of the amplifier also falls exponentially. The half-life of the exponential decay is determined by the product of the resistance and capacitance of the discharging circuit and is known as the resistance-capacitance time constant. The time taken for the circuit to detect the increase in signal level and reduce the gain so that the output signal level falls within a specified amount of its final value is known as the attack time.

A similar sequence of events takes place when the incoming signal is reduced in level. In this case, the time taken for the circuit to detect a decrease in signal level and to increase the gain so that the output level increases to within a specified amount of its final value is known as the release time. As before, the change in gain is exponential. It is common practice for the release time to be longer than the attack time.

Figure 5.1 shows the input and output signals for a compression amplifier of the type described above. Figure 5.1*A* shows the input signal, which is steady in level until time T_1 when its amplitude is suddenly increased by 12 dB. (An increase of only 12 dB is used for purposes of illustration. The standard definitions of attack and release times specify much larger changes in level.) The output signal shows the same 12-dB increase in amplitude initially, but as the gain of the amplifier is reduced, the output signal shows an exponential reduction in amplitude over time. At time T_2, the output signal has fallen to within 26% (i.e., within 2 dB) of its target steady-state value. The time elapsed between T_1 and T_2 is defined as the attack time of the compression amplifier.[1] The difference between the output signal and its target value is known as the overshoot transient.

At time T_3, the input signal is suddenly decreased in amplitude by 12 dB. The output signal initially shows the same 12-dB decrease in amplitude at time T_3, but as the gain of the amplifier is increased, the output signal shows an exponential increase in amplitude over time. At time T_4, the output signal has risen to within 26% of its target value. The time elapsed between T_3 and T_4 is defined as the release time of the compression amplifier. In this case, the difference between the output signal and its target value is known as the undershoot transient.

A well-designed compression amplifier will generate relatively little nonlinear distortion even while the gain is changing. This is difficult to accomplish using analog electronic circuits but can be achieved without much difficulty using digital techniques. Digital compression amplifiers have considerable flexibility in adjusting the rate of change in gain and can synchronize gain changes to occur at specific times relative to the signal waveform so as to minimize any nonlinear distortions. Figure 5.2 shows how distortion of the output waveform can be reduced by synchronizing the change in gain with a zero crossing of the output waveform (i.e., at that instant in time when the value of the waveform is zero). Figure 5.2*A* shows the change in the output waveform for a change in gain occurring at a peak

[1] The exact definitions of attack and release times depend on which hearing aid standard is used. Hearing aid standards have also varied over the years. The current European standard (IEC 118-2) defines release time as the time taken for the output signal to increase to within 2 dB of its final value after a decrease in input level from 80 to 55 dB SPL. The American standard (ANSI S3.22-1996) defines release time in terms of the time taken for the output level to increase to within 4 dB of its final value for a decrease in input level from 90 to 55 dB SPL. The definition of attack time is essentially the same except that in this case, the time taken for a decrease in output level after an increase in input level is used. These definitions are likely to change as different methods of implementing compression using newer digital techniques are introduced. Although the details of these definitions may change over time, the basic concepts will remain the same because the underlying purpose of these standards is to provide uniformity in defining the characteristics of a hearing aid regardless of whether analog or digital technology is used.

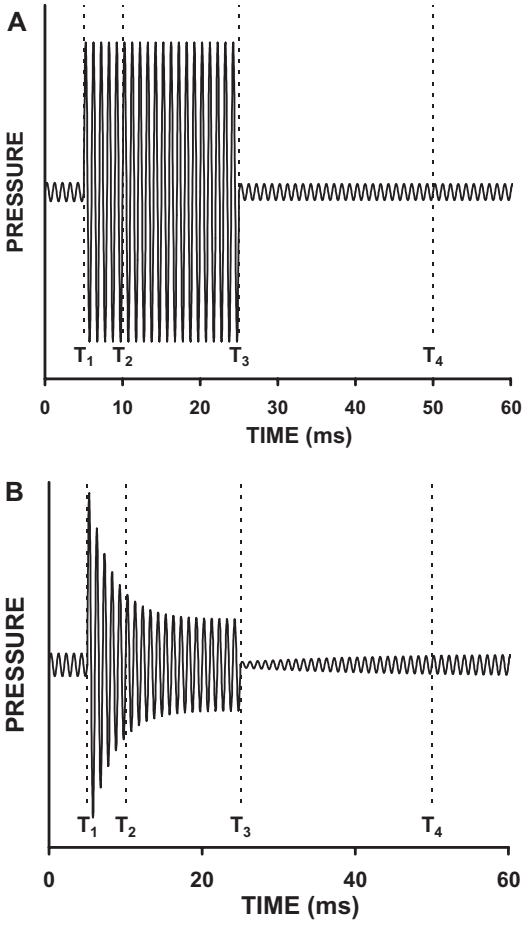


FIGURE 5.1. *A*: input signal to a compression amplifier. *B*: the output signal. At time T_1, the amplitude of the input signal is increased suddenly by 12 dB. The output signal showed the same 12-dB increase in amplitude initially, followed by an exponential decrease in amplitude over time. At time T_2, the amplitude of the output signal has fallen to within 2 dB of its target steady-state value. The time elapsed between T_1 and T_2 is the attack time, and the difference between the output signal and its target value is the overshoot transient. At time T_3, the amplitude of the input signal is decreased suddenly by 12 dB. The output signal shows the same 12-dB decrease in amplitude initially, followed by an exponential increase in amplitude over time. At time T_4, the output signal has increased to within 2 dB of its target steady-state value. The time elapsed between T_3 and T_4 is the release time, and the difference between the output signal and its target value is the undershoot transient.

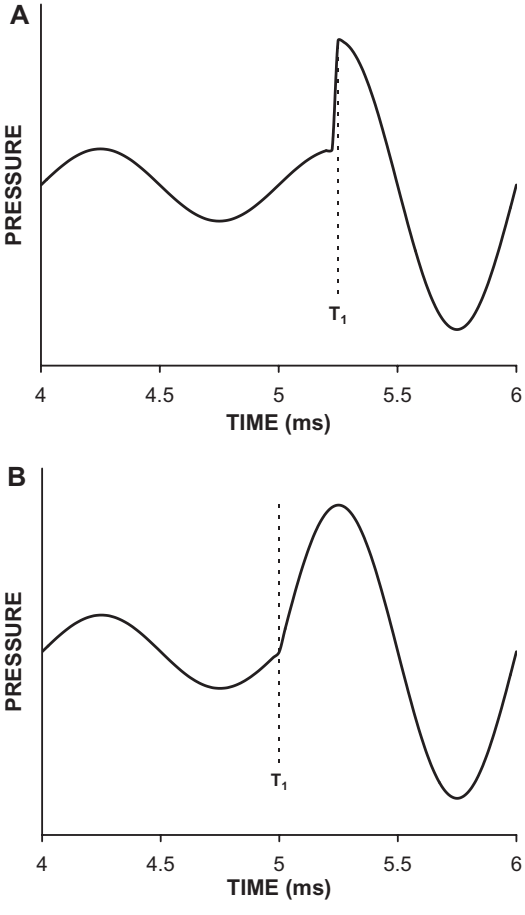


FIGURE 5.2. *A*: how a signal can be distorted by a sudden increase in gain. *B*: how the distortion can be reduced by constraining the change in gain to occur at a zero crossing of the signal.

of the signal waveform. Figure 5.2*B* shows the output waveform for the same change in gain occurring at a zero crossing. The distortion of the output waveform is substantially less in the latter case.

Two other variables are of importance in specifying the characteristics of a compression amplifier. The compression ratio is defined as the change in input level relative to the change in output level. A compression ratio of 2:1, for example, indicates that for every 2-dB increase in input level, the output level is increased by 1 dB. The compression threshold is the signal level at which the amplifier begins to operate as a compression amplifier. Several illustrative input-output curves showing the joint effects of compression threshold and compression ratio are shown in Figure 5.3.

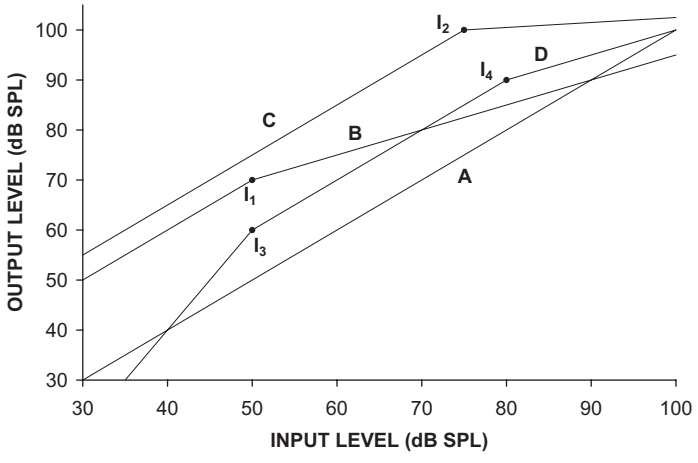


FIGURE 5.3. Input-output curves for several compression amplifiers. Curve A is the input-output curve for a conventional amplifier with 0 dB gain. It has a slope of 1. Curve B corresponds to a compression amplifier with an input compression threshold (I_1) of 50 dB sound pressure level (SPL) and a compression ratio of 2:1. The curve has a slope of 0.5 above I_1. Curve C represents a compression limiter. The unit has an input compression threshold (I_2) of 75 dB SPL and operates as a conventional amplifier with a gain of 25 dB for input levels below I_2. The compression ratio is 10:1 for input levels greater than 75 dB SPL. The curve has a slope of 1 below I_2 and a slope of 0.1 above I_2. Curve D corresponds to a compression amplifier with two compression thresholds, I_3 and I_4. The compression ratio for input levels below I_3 is 0.5:1. The curve has a slope of 2 below I_3, and the signals are expanded rather than compressed at these low levels. The compression ratio is 1:1 between I_3 and I_4, and signals within this range receive conventional amplification with a gain of 10 dB (= output level—input level for this region of the curve). For input levels greater than the second compression threshold (I_4) of 80 dB SPL, the compression ratio is 2:1, and the curve has a slope of 0.5 and is parallel to curve B in this region.

Curve A in Figure 5.3 shows the input-output curve for a conventional amplifier with a gain of 0 dB; i.e., the output level is uniformly equal to the input level. This curve has a slope of 1.

Curve B in Figure 5.3 shows the input-output curve for a compression amplifier with a compression threshold (I_1) at an input level of 50 dB SPL and a compression ratio of 2:1. Below the compression threshold, the amplifier operates as a conventional amplifier with a fixed gain of 20 dB. The slope of this input-output curve is 1 below the compression threshold and 0.5 above this threshold. Note that curve B crosses the 0-dB gain curve (curve A) at an input level of 90 dB sound pressure level (SPL); that is, for input levels above 90 dB SPL, the output level is lower than the input level.

Curve C in Figure 5.3 shows the input-output curve for a compression amplifier with a compression threshold (I_2) of 75 dB SPL and a compres-

sion ratio of 10:1. The gain below the compression threshold is 25 dB. This compression amplifier effectively limits the output level to a maximum value slightly above 100 dB SPL, the output level at the compression threshold (I_2).

Curve D in Figure 5.3 shows the input-output curve for a compression amplifier with two compression thresholds. Below the lower compression threshold (I_3), which corresponds to an input level of 50 dB SPL, the amplifier has a compression ratio of 0.5:1. The slope of the input-output curve is greater than 1 at these low levels and the output level increases at a greater rate than the input level. Expansion rather than compression takes place when the compression ratio is less than 1. In this case, the compression ratio is 0.5:1 and the output level is increased by 2 dB (= 1/0.5 dB) for every 1-dB increase in input level. The input level for the higher compression threshold (I_4) is 80 dB SPL. The amplifier has a compression ratio of 1:1 between the two compression thresholds and operates as a conventional amplifier at these input levels, with an input-output slope of 1. At input levels above 80 dB SPL, the amplifier has a compression ratio of 2:1. The slope of the input-output curve is 0.5 at these high levels, the output level increasing by 0.5 dB for every 1-dB increase in input level.

2.2 *Four Basic Forms of Compression Amplification*

Different combinations of the above four variables (compression ratio, compression threshold, attack time, and release time) result in very different forms of compression amplification. Four basic combinations are as follows:

(1) Automatic gain control (AGC) is obtained by combining a low threshold of compression with a long release time. In this form of compression amplification, long-term fluctuations in signal level are reduced. AGC is the oldest form of compression amplification and was first used in radio to combat fading, that is, to maintain a fixed average loudness despite fluctuations in the strength of the received radio signal.

(2) Compression limiting is obtained by combining a high threshold of compression with a high compression ratio and very short attack and release times. This form of compression imposes a limit on the maximum output level that can be generated by the hearing aid. This method of output limiting produces relatively little distortion compared to the method of peak clipping. The latter method simply clips the peaks in the signal waveform, leaving the waveform with a flat top in place of a peak. Compression limiting, by lowering the gain of the amplifier, retains the peaks in the waveform but at a lower level. Peak clipping was widely used before the development of compression limiting.

(3) Wide dynamic range (WDR) compression is obtained using a low compression threshold. In this form of compression, a wide range of signal levels is compressed. WDR compression can be designed to approximate

the compression characteristics of the normal ear (to the best of our knowledge).

(4) Syllabic compression is obtained by combining short attack and release times with a low compression threshold. This form of compression can be used to adjust the level of individual speech sounds. The term syllabic compression refers to that condition in which weak syllables are amplified by an amount greater than that for the strong syllables. If very short time constants are used, then individual phonemes will differ in gain; compression amplification of this form should properly be called phonemic compression. In accordance with common usage, the term syllabic compression will be used as shorthand for both syllabic and phonemic compression.

The above four forms of compression amplification are not mutually exclusive. It is possible, for example, to have a WDR syllabic compressor as well as a syllabic compressor with compression limiting. There are also many variations of these four basic forms of compression. One such example is a compressor with two thresholds, a low threshold as in WDR compression and a high threshold to provide compression limiting. Another possibility is to have a low compression ratio at low input levels (possibly a compression ratio that expands rather than compresses low level signals), with no compression at intermediate input levels (i.e., compression ratio = 1:1) and a high compression ratio at high input levels, as illustrated by curve D in Figure 5.3.

2.3 Various Implementations of Compression Amplification

There are typically several stages of amplification in a hearing aid, such as a preamplifier that is usually immediately after the microphone, a low-power amplifier containing a volume control and a frequency-shaping network, and a power amplifier that drives the output transducer, that is, the hearing aid receiver. (The latter term, although widely used, is a misnomer because the output transducer delivers rather than receives signals.) If the compression amplifier is placed before the volume control, e.g., immediately after the preamplifier, the hearing aid is said to have input compression. If the compression amplifier is placed after the volume control, for example, at or just before the output power amplifier, the hearing aid is said to have output compression. Output compression is more effective than input compression in controlling the maximum output level of a hearing aid. On the other hand, there are engineering advantages to input compression and many hearing aids are designed with input compression.

Compression amplifiers are often referred to as "nonlinear," whereas conventional hearing aid amplifiers are referred to as "linear." This termi-

nology is misleading. A compression amplifier operates as a nonlinear device only while the gain is changing. Once the gain has changed, a compressor will once again operate as a linear amplifier. The amount of nonlinear distortion produced while the gain is changing is relatively small. It should also be noted that a conventional hearing aid amplifier does not always amplify signals linearly. Some form of output limiting is needed in order to protect the user from excessive amplification. Noncompression hearing aids typically use peak clipping, a highly nonlinear operation, to limit the output level.

To illustrate the point, consider a compression amplifier with an input-output curve corresponding to curve C in Figure 5.3. This amplifier will operate as a linear device for as wide a range of signal levels as a conventional hearing aid amplifier with the same output limiting level. Furthermore, for signals that require limiting, the so-called "nonlinear" compression amplifier will produce substantially less nonlinear distortion than a conventional "linear" amplifier with peak clipping.

Overshoot and undershoot transients as well as nonlinear distortions that are produced as gain is varied dynamically can be reduced substantially, if not eliminated, using a technique known as look-ahead compression (Robinson and Huntington, 1973; Verschuure et al. 1994). This technique requires that the signal be delayed, analyzed in order to determine how to alter gain with minimal distortion (e.g., synchronizing gain changes with zero crossings; see Figure 5.2), and then amplified. Although relatively simple forms of look-ahead compression have been implemented some time ago using analog technology (Shorter et al. 1967), recent advances in digital technology have opened up new possibilities regarding signal-processing techniques of this type.

A powerful form of signal processing that can be implemented using digital technology is that of block processing. This method of signal processing, which can be used to implement look-ahead compression, subdivides the input signal into a sequence of time windows, the waveform in each time window being analyzed and processed as a unit or "block." The processed signals in successive time windows are then concatenated so as to produce a continuous output waveform. Overlapping time windows are used to eliminate discontinuities that would otherwise occur at the juncture of two time windows. Experimental digital hearing aids using block processing have been used for research purposes for more than two decades (Levitt 1982; Levitt et al. 1986). Recent advances in the microminiaturization of digital technology have made it possible for this form of processing to be incorporated in instruments small enough to fit in the ear.

Block processing introduces a delay equal to 1–1.5 time windows. A delay of 1 time window is sufficient to implement look-ahead compression, but the delay should not exceed 10 ms in order to avoid adverse perceptual effects (Stone and Moore 1999). A valuable feature of block processing is that the short-term frequency spectrum (i.e., the spectrum within a finite

time window) can be obtained using the fast Fourier transform. Both the amplitude and phase of the short-term frequency spectrum can thus be modified conveniently and with great precision. For example, frequency shaping can be combined with compression while also using phase information to reduce acoustic feedback (Levitt 1993).

2.4 Multichannel Compression

The various forms of compression described thus far have been limited to a single frequency band. Villchur (1973) has pointed out that if a compression amplifier operates on a single, wide frequency band containing weak high-frequency components as well as strong low-frequency components, then the weak high-frequency components of the sound will be compressed together with the strong low-frequency components. He recommended that a bank of band-pass filters be used with a separate compression amplifier in each band. This approach, known as multichannel or multiband compression, avoids the above problem in that it can provide the appropriate amount of compression for different speech sounds in different frequency regions.

Multichannel compression, however, can introduce significant distortions to the speech spectrum, thereby destroying phonetic cues conveyed by spectral shape. It is possible, however, to reduce the amount of spectral distortion introduced by frequency-dependent compression. This can be done in several different ways.

One approach is to constrain the amount of gain that is applied to adjacent frequency bands so as to avoid large between-band changes in intensity that distort the normal spectral contrasts in speech. Bustamante and Braida (1987a) have developed a method of this type using a principal component analysis in which the pattern of correlations among the band intensities is determined and then constraints are imposed based on these correlations so as to control the between-band variations in gain.

An alternative approach to the problem is to estimate the shape of the spectral envelope over short time intervals (i.e., time windows) using an orthogonal polynomial expansion and then to amplify the signal such that the variation over time of specific terms in the orthogonal polynomial expansion is compressed (Levitt and Neuman 1991). For example, the lowest order term in the expansion corresponds to the short-term power of the signal (within a time window). Compressing the range of variation of this term over time results in a form of compression that is very similar to conventional amplitude compression. The second term in the expansion corresponds to the average slope of the spectral envelope. Compressing the range of variation of this term reduces the variation in spectral balance over time. The third term in the expansion corresponds to the average curvature of the spectral envelope (as approximated by a quadratic function) and so on.

If all the terms in the expansion were compressed by a very large amount, the spectrum of the processed signal would remain constant over time. The average shape of the spectrum over time can also be controlled by adjusting the average value of each term in the expansion. Computation of the orthogonal polynomial expansion requires relatively few arithmetic operations, thereby allowing the technique to be implemented in real time (Levitt 1993).

A method of signal processing that is similar to frequency-dependent amplitude compression is that of altering the frequency-gain characteristic of the hearing aid as a function of signal level. An amplifier of this type, known as the K-amp (Killion 1993), is widely used in modern hearing aids. As in the case of multiband amplitude compression, there are different ways in which the frequency-gain characteristic can be adjusted with level; for example, either the low or high frequencies can be adjusted with increasing signal level. The term BILL is used to describe an amplifier in which low frequencies receive relatively more gain with decreasing signal level (i.e., bass increase at low levels). The term TILL is used for amplifiers in which the high frequencies receive relatively more gain with decreasing level (i.e., treble increase at low levels). For a discussion of this terminology, see Killion et al. (1990).

2.5 Combining Different Forms of Compression

The number of possible forms of compression that can be implemented in a hearing aid is immense. Each of the four basic methods of compression can, in principle, be implemented in each of a large number of frequency bands. In addition, it is possible to combine single-channel and multichannel compression in a useful way, as illustrated in the following example.

Figure 5.4 shows the block diagram of a compression hearing aid that combines two-channel syllabic compression with single-channel AGC and single-channel output limiting. The first stage of amplification consists of a single-channel AGC amplifier with a long release time. Long-term fluctuations in the level of the speech signal are thus reduced. The signal is then filtered into two frequency bands, one for low frequencies and the other for high frequencies. Low compression thresholds with short release times are used in each frequency band so as to increase the relative level of the weaker speech sounds (i.e., syllabic compression). Because low- and high-frequency speech sounds are compressed in separate channels, the problem of a weak high-frequency speech sound being compressed together with a strong low-frequency sound in the same channel is avoided. The outputs of the two frequency channels are then combined and the overall gain is adjusted to a comfortable level. The final stage of amplification consists of a single-channel compression limiter to protect the user from excessive amplification.

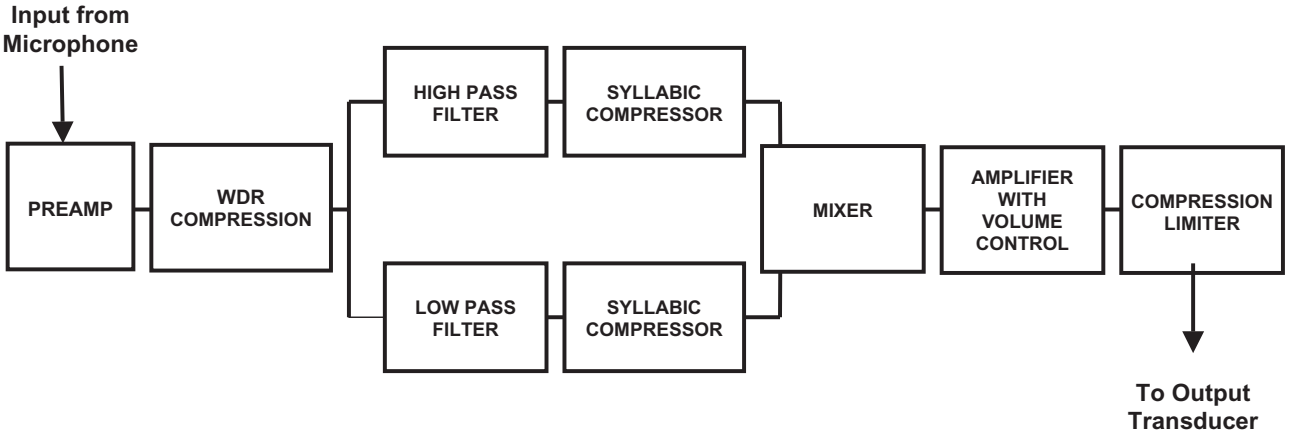


FIGURE 5.4. A compression hearing aid combining automatic gain control (AGC), compression limiting, and 2-channel syllabic compression. The first stage of amplification is a low-noise preamplifier followed by a wide dynamic range (WDR) compressor with long time constants to provide AGC amplification. The signal is then filtered into 2 frequency bands with syllabic compressors in each band. The compressed signals are then added in the mixer, and the overall gain is adjusted by means of a volume control. The last stage is a compression limiter to protect the listener from excessive output levels. This 2-channel compression hearing aid is modeled after the hearing aid evaluated by Moore (1987).

Experimental evaluations of hearing aids embodying this basic design have yielded positive results (Laurence et al. 1983; Moore et al. 1985; Moore 1987; Moore and Glasberg 1986, 1988; Moore et al. 1991, 1992; Humes et al. 1999), and commercial instruments based on this design have been very successful. Many modern hearing aids have a similar design but use more than two channels. Digital technology has greatly facilitated the development of hearing aids with advanced signal-processing capabilities, and instruments with as many as 19 channels are now available.

The hearing aid shown in Figure 5.4 not only serves as an example of a basic design that has been very successful but also illustrates the point that different forms of compression should not be viewed as competing approaches but rather as complimentary approaches to a difficult problem. The issue that needs to be addressed is how best to combine different forms of compression amplification in a hearing aid so as to capture the advantages of compression amplification while minimizing its shortcomings.

3. Benefits and Limitations of Compression Amplification

There are both benefits and limitations to compression amplification. Some of these limitations are quite subtle in that compression amplifiers do not always do what they are supposed to do. Other limitations (or lack of anticipated benefits) involve the way sound is processed by the auditory system and are more difficult to deal with because of our limited understanding of auditory perception, especially in the case of a hearing impairment. These difficulties are compounded when real-world listening conditions are considered (e.g., compression amplification in the presence of background noise). There is also substantial disagreement within the field regarding the merits of the various forms of compression amplification designed to deal with real-world listening conditions.

3.1 Do Compression Amplifiers Do What They Are Supposed To Do?

This question is relatively easy to address because the answer depends on physical measurements and not on our understanding of how sound is processed by the auditory system. Compression amplification was first developed to compensate for fading in radio transmissions. The variations in level of these radio signals were large but relatively slow. As a consequence, AGC amplification (wide dynamic range, long time constants) provided the necessary variations in gain to compensate effectively for the variations in signal level. In this application, the compression amplifier is doing what it is supposed to be doing with good effect.

The use of compression in hearing aids is not as simple. Speech has both long-term and short-term variations in level. Long-term variations in speech level are quite common in everyday communication. The sound level of a nearby speaker, for example, will be higher than that of someone further away. These differences in average speech level can be compensated for quite well by AGC compression. Compression amplifiers designed for this purpose typically accomplish their design requirements satisfactorily.

The situation is very different for the case of compression amplifiers designed to reduce short-term variations in speech level (such as a syllabic compressor). These amplifiers are not only subject to conflicting design requirements (as discussed shortly), but in many cases, they do not do exactly what they are supposed to do. A sudden change in level, for example, can produce overshoot or undershoot transients as illustrated in Figure 5.1. These transients not only affect the subjective quality of the amplified sound, but they can also result in an effective compression ratio that is less than the intended compression ratio.

The reduction in compression ratio results from a complex interaction between the time constants of the compressor and the temporal characteristics of the signal being amplified. The effect is not immediately obvious and is best explained by means of an example. Consider the utterance "We'll buy food for four" as processed by a syllabic compressor with significant overshoot and undershoot. The speech level during the early part of the utterance is relatively steady and fairly intense. The compressor gain will thus have stabilized at a moderately low value shortly before the word "food" is produced. The fricative at the start of "food" is relatively weak and the gain of the amplifier will initially be too low for this sound. The compressor will increase its gain so as to increase the audibility of the weak fricative, but because the sound is of short duration, the gain may not reach its target value before the speech level is increased again as the vowel in "food" is produced.

The increase in speech level during the vowel in "food" will cause the compressor to begin reducing its gain toward a value appropriate for a strong vowel. The gain of the compressor, however, may once again not reach its target value before the speech level is reduced during the weak fricative at the start of the word "for." The sequence of events is then repeated again and again as the level of the speech signal fluctuates at a rate faster than the tracking rate of the compressor. Because the target gains are not reached as the speech fluctuates in level, the effective compression ratio is reduced. The magnitude of the reduction in compression ratio depends in a complex way on the time constants of the compressor and the rate at which the speech signal fluctuates in level (Stone and Moore 1992; Verschuure et al. 1996).

The effective compression ratio is also dependent on the bandwidth of the signal driving the compressor. If a compressor operates on a broadband signal, the amount of compression within a narrow band of that signal will

be less than that for the broadband signal (Verschuure et al. 1996). This observation has important implications for compressors with bandwidths much larger than the analysis bandwidth of the ear and is deserving of further investigation.

One application in which compression amplifiers not only do what they are supposed to do but also do it very well is that of compression limiting. This form of compression is rapidly replacing peak clipping as the preferred method of limiting the output of a hearing aid. Although peak clipping is an effective method of output limiting, it also introduces substantial nonlinear distortion. Compression limiting is equally effective in limiting the output of a hearing aid, but in contrast to peak clipping, it introduces relatively little nonlinear distortion. Furthermore, if the change in gain introduced by the compressor is constrained to occur at a zero crossing of the waveform (see Figure 5.2), the resulting nonlinear distortion will be minimal. Peak clipping, however, does provide more power in a given channel than compression limiting and is better suited for hearing aids in which a very high-power output is the dominant requirement. These hearing aids are designed for people with profound hearing losses who need a substantial power output and for whom most of the nonlinear distortion produced by peak clipping is not audible.

3.2 How Effective Is Compression Amplification in Hearing Aids?

There is no simple answer to this question. The effectiveness of compression amplification in a hearing aid depends on many variables including the nature and severity of the hearing impairment, the listening environment (e.g., in quiet or in noise), and the extent to which the characteristics of the hearing aid have been matched to the needs of the user. These variables interact in a complex way, and in many cases, compression characteristics that work well for one set of conditions are unsuitable for other conditions. Similarly, compression characteristics that are appropriate for one person may be inappropriate for another. Individuals with narrow dynamic ranges are likely to benefit from a greater amount of compression but not in every case (Neuman et al. 1994). Furthermore, for profoundly hearing-impaired individuals, compression amplification may hinder rather than help speech understanding (Boothroyd et al. 1988).

It should also be noted that the theoretical underpinnings of compression amplification have focused on the speech signal, with only a secondary concern for other signals such as music, the user's own voice, or background noise. As a consequence, there have been many unanticipated problems in the development and evaluation of practical compression hearing aids. Post hoc solutions have been the order of the day in finding practical solutions for the real-world use of hearing aids. These solutions are not entirely sat-

isfactory in that they are typically no more than empirical compromises devoid of an underlying theory.

The effect of background noise on compression amplification is the most difficult problem. Consider the case of a continuous low-level background noise. A WDR compressor with a short release time will amplify the noise during breaks or pauses in the speech, thereby making the background noise sound louder than would otherwise be the case. If a longer release time is used, then another problem known as pumping can occur, in which weak sounds (speech or noise) after a strong sound are heard to grow in loudness.

A compression threshold that is slightly above the level of the background noise will avoid the problem of within-pause amplification but at the cost of losing the benefit of compression amplification for low-level speech sounds in quiet. It may be possible to operate efficiently in both quiet and noise by automatically adjusting the compression threshold to an appropriate level for the background noise. This technique represents a practical compromise in which some of the benefits of compression amplification (e.g., increased audibility) are sacrificed in order to deal with another problem (background noise). Because the best compromise between the conflicting requirements varies in the everyday use of a hearing aid, adaptive adjustment of the compression parameters is needed depending on the listening environment.

Different types of background noise can also have very different effects on a compression amplifier. An intense noise of short duration, such as the slamming of a door, can cause an amplifier with a short attack time and long release time to reduce its gain for a relatively long period of time. The perceived effect is that of a dead period in which the hearing aid does not appear to be working. If a long attack time is used, an intense short burst of noise will be amplified before the compression amplifier has time to reduce its gain. The result can be a very unpleasant experience for the hearing aid user and could further damage an already impaired auditory system.

The time constants appropriate for intense sounds of short duration are not the same as those for sounds of long duration. Here again, there are conflicting demands on the design of an effective compression hearing aid, and, as before, a practical compromise is possible by adjusting the compression parameters adaptively depending on the nature of the sounds being amplified. Hearing aids with different release times for sounds of short and long duration have been developed and have proven to be quite effective (Moore and Glasberg 1988; Teder 1991; Killion et al. 1992).

Given an appropriate choice of parameter values for the hearing loss and listening environment, the core question remains: Does compression amplification improve speech intelligibility (as predicted by theoretical analyses)? The answer is “yes,” but under restricted conditions for the simpler

forms of compression and a cautious "possibly" for more complex forms of compression such as multichannel syllabic compression.

AGC will improve the intelligibility of low-level speech in quiet. Manual adjustment of the gain control will achieve the same result, but there are many situations in which the necessary manual adjustments are too frequent to be manageable (Section 6.5.1 in Dillon 2001).

Compression limiting is generally favored over peak clipping except for profound hearing impairments, in which case maximizing output power may be of greater consequence than the substantial amount of nonlinear distortion produced by peak clipping. Larson et al. (2000) in a large-scale clinical study (360 patients) found that compression limiting and WDR compression were judged superior to peak clipping not only in terms of perceived distortion but also in showing significantly higher word recognition scores ($P = 0.002$), more comfortable loudness ($P = 0.003$) and overall liking ($P = 0.001$). No significant differences were observed between compression limiting and WDR compression with respect to these variables.

3.3 Does Single-Channel Syllabic Compression Improve Intelligibility?

An issue of great interest is whether WDR compression with short time constants (i.e., syllabic compression) improves intelligibility. Whereas an AGC amplifier will place much of the speech signal between the threshold of hearing and the loudness discomfort level (the residual hearing area), the dynamic range of speech is such that individual speech sounds of low intensity will still be inaudible. Syllabic compression is designed to improve the audibility of these low-level speech sounds, thereby improving intelligibility.

Research on this topic has yielded mixed results. Whereas some studies have shown improved intelligibility with single-channel compression in quiet (e.g., Dreschler 1988, 1989), other studies have not shown the anticipated improvements, especially in the presence of background noise (Braida et al. 1979; Dreschler et al. 1984; Tyler and Kuk 1989). In some cases, syllabic compression will reduce rather than improve speech intelligibility. Boike and Souza (2000), for example, showed no significant change in speech recognition scores with compression ratios in quiet over the range from 1:1 to 10:1, but there was a significant reduction in speech recognition scores with increasing compression ratio for speech in noise. The reduction in score was substantial for compression ratios greater than 2:1.

The clinical trial recently completed by Larson et al. (2000) sheds some light on these diverse experimental results. In this study, several measures of benefit were obtained for three hearing aid circuits: peak clipping, compression limiting and WDR compression. The results showed that if WDR compression is compared to peak clipping, then WDR compression shows

a significant advantage. If, however, WDR compression is compared to compression limiting, then there is no significant advantage for WDR compression. The choice of the reference condition in evaluating compression amplification is of crucial importance. Differences in the choice of the reference condition account in large measure for the diversity of experimental results that have been reported.

Another important observation in the Larsen et al. (2000) study was the extent of individual differences in comparing the different forms of amplification. In particular, some subjects showed significantly greater benefit with WDR compression, whereas others did not. The importance of taking individual differences into account in the design and fitting of compression hearing aids should not be underestimated.

Laboratory studies involving adjustment to the consonant-vowel (CV) intensity ratio provide some useful insights as to how individual differences can affect studies of this type and why syllabic compression has not, as yet, yielded the anticipated benefits in practice. When the CV intensity ratio is increased by a single, fixed amount so as to improve the audibility of the weak consonants, the improvement in intelligibility is small (Gordon-Salant 1986, 1987; Montgomery and Edge 1988). If, however, individualized adjustments are made to the CV intensity ratio for each subject and for different phonetic environments, then a substantial improvement in intelligibility is possible for speech in quiet (Kennedy et al. 1998).

Adjustment to the CV intensity ratio is possible using WDR compression with an appropriate choice of attack and release times. The real difficulty is finding the right time constants as a function of phonetic context and taking individual subject differences into account.

3.4 Benefits and Limitations of Multichannel Compression

Interest in multichannel compression has been substantial ever since Villchur (1973) pointed out the limitations of single-channel compression and recommended that multichannel compression be used instead. Multichannel compression is now implemented in most of the newly introduced digital hearing aids.

As in the case of single-channel compression, experimental evaluations of multichannel syllabic compression have yielded mixed results. Some studies have shown improved intelligibility (Villchur 1973; Yanick 1973; Yund et al. 1987; Kiessling and Steffens 1991; Moore and Glasberg 1986, 1988; Yund and Buckles 1995a,b), whereas others have not (Abramovitz 1980; Lippmann et al. 1981; Walker et al. 1984; De Gennaro et al. 1986; Bustamante and Braida 1987b).

As before, the diversity of experimental results can be accounted for in large measure by differences in the choice of the reference condition. If the

reference condition is a relatively poor form of amplification (e.g., the subject's own hearing aid that may not have been fitted properly), then positive results for the experimental multichannel hearing aid are to be expected. Not surprisingly, most of the studies showing improved intelligibility for multichannel compression have not used a truly challenging control condition.

One such control condition, for example, would be a single-channel AGC hearing aid with a frequency response that maximizes intelligibility for each subject, using circuit components of the same quality as in the experimental hearing aid. A significant gain in intelligibility compared to a control condition of this kind would be convincing evidence that multichannel compression does in fact improve intelligibility.

The most promising results with multichannel compression that have been obtained thus far have been with two-channel compression. Laurence et al. (1983), Moore et al. (1985, 1992), Moore and Glasberg (1986, 1988), and Moore (1987)) have consistently obtained good results with two-channel compression in both quiet and noise.

In most of these studies, AGC was used in combination with syllabic compression. In the study by Moore and Glasberg (1988), for example, four combinations of AGC and two-channel compression were evaluated. The best results were obtained with syllabic compression in only the high-frequency band of the two-channel compressor combined with wideband two-stage AGC at the input. The two-stage AGC unit used a short release time of 150 ms for transient sounds and a long release time of 5 s to compensate for gradual variations in sound level. The other three experimental conditions consisted of (1) a two-stage AGC alone, (2) a two-stage in combination with syllabic compression in both channels of the two-channel system, and (3) a conventional single-stage AGC with syllabic compression in both channels.

Humes et al. (1999) have also shown two-channel compression to yield higher intelligibility scores in both quiet and noise. The reference condition in this case, however, was a conventional noncompression hearing aid so that the question of whether two-channel syllabic compression provides improved intelligibility in comparison with single-channel compression was not addressed.

Yund and Buckles (1995a) obtained a similar result with eight-channel syllabic compression compared to conventional noncompression amplification. The improvement in intelligibility was small and was only significant at low speech-to-noise ratios. Also, subjects with less residual hearing showed a decreased intelligibility with multichannel compression.

Although the evidence is not overwhelming, there is sufficient evidence to justify the added complexity of two-channel syllabic compression in combination with wideband AGC (compared to simpler forms of single-channel compression amplification). An obvious follow-up question is whether more than two channels of compression amplification, suitably combined

with an appropriate form of AGC, could improve intelligibility even further. Stone et al. (1999) evaluated a four-channel compression amplifier with characteristics similar to those used in the study by Moore and Glasberg (1988). The results did not show a significant advantage for four-channel compression.

In principle, increasing the number of channels has the potential for placing more of the speech signal in the residual hearing area but at the expense of distorting phonetic cues conveyed by spectral contrasts. This loss of information may not be important for a limited amount of compression, but if the amount of compression is substantial, the loss of information is likely to outweigh the benefits of improved audibility provided by compression. Many place-of-articulation cues are conveyed by relatively subtle spectral contrasts. It is thus not surprising to find that studies evaluating multichannel compression with many channels have reported increased place-of-articulation errors (Lippmann et al. 1981; De Gennaro et al. 1986; Lindholm et al. 1988).

The larger the compression ratio, the greater the distortion of spectral contrasts, and it is not surprising that evaluations of multichannel compression using large compression ratios (e.g., greater than 3:1) obtained a decrease in speech intelligibility (De Gennaro et al. 1986; Bustamante and Braida, 1987b; Drullman and Smoorenburg 1997; Boike and Souza 2000; Dillon 2001). Large compression ratios are more likely to be used for listeners with severe hearing losses, and here again, reduced speech intelligibility has been obtained with multichannel compression for listeners with severe hearing losses (Lippmann et al. 1981; Boothroyd et al. 1988). This observation is not necessarily restricted to multichannel compression.

In contrast to the above, Yund and Buckles (1995b) investigated the effect of number of channels on speech intelligibility and found that speech intelligibility increased as the number of channels was increased from 4 to 8 and that intelligibility remained essentially unchanged as the number of channels was increased to 16. It should be noted that 15 of the 16 subjects in these studies had participated in previous research on multichannel compression and were experienced listeners. Yund and Buckles (1995c) showed the importance of long-term learning effects in studies of this type, and this may help account for some of the differences between this study and other investigations of multichannel compression. It should also be noted that nonsense syllables were used as the test material, and as pointed out by Van Tasell and Trine (1996), subjects can learn "the temporal overshoots associated with compression attack time in the stimuli" and that generalizations to conversational speech should be made with caution.

The data obtained by Yund and Buckles (1995a), however, are consistent with the findings noted above regarding the effectiveness of compression amplification in cases of severe hearing loss. In their evaluation of the eight-channel compression amplifier, subjects with more severe losses received a greater amount of compression. These subjects showed a significant reduc-

tion in intelligibility with multichannel compression, whereas those subjects with less severe losses received less compression and showed significant improvements in intelligibility.

3.5 Underlying Issues

Plomp (1988, 1994) has argued that the temporal contrasts in intensity conveyed by the speech envelope are essential for intelligibility and that reduction of these intensity cues, as occurs in syllabic compression, is more damaging than helpful to intelligibility. He also points out that the use of high compression ratios with many channels of syllabic compression could effectively eliminate all of the temporal and spectral contrasts in speech. In support of his argument, Plomp cites the successful application of the modulation transfer function (Steeneken and Houtgast 1980) in predicting the loss in intelligibility of reverberant speech based on the reduced intensity contrasts in the speech envelope caused by the reverberation.

It should be noted, however, that the loss of temporal contrasts in a reverberant environment is not the only distortion of the speech signal resulting from reverberation. The modulation transfer function is an effective indicator of the loss in intelligibility produced by reverberation because in addition to the reduction in temporal contrasts, the reverberation also causes other distortions such as temporal masking that are correlated with the reduction in temporal contrasts and that also contribute to the loss in speech intelligibility.

Villchur (1989), in response to Plomp's position, has argued that the distortion would only be significant for multichannel compression with a large number of frequency bands. Hohmann and Kollmeier (1995) have also pointed out that the nature of the temporal distortions produced by reverberation and by syllabic compression are quite different and that predictions of reduced intelligibility based on the modulation transfer function for substantial amounts of syllabic compression are not supported by their experimental data.

It is possible to separate temporal distortions from spectral distortions by modulating a broadband noise to have the same temporal structure as speech. The signal-correlated noise (SCN) stimuli have minimal spectral information and are ideal for investigating the temporal distortions produced by syllabic compression without confounding by concomitant spectral distortions.

Souza and Turner (1996) found that the temporal distortions introduced by single-channel syllabic distortion with moderate amounts of compression do not have a significant effect on the recognition of the processed SCN stimuli by both normal-hearing and hearing-impaired listeners. The SCN stimuli were speech-modulated versions of /vowel-consonant-vowel/ nonsense disyllables. Van Tasell and Trine (1996) performed a similar study using SCN stimuli for both nonsense disyllables and sentences. Single-

channel syllabic compression had little effect on the recognition of the SCN disyllables, but recognition of the SCN sentences was reduced significantly.

The effect of both conventional noncompression amplification and multichannel syllabic compression on temporal cues was investigated by Souza and Turner (1998) using the same SCN stimuli as in their previous study (Souza and Turner 1996). The data showed that multichannel compression improved recognition of the SCN stimuli (relative to noncompression amplification) at moderate input levels but that recognition showed a significant reduction at high input levels requiring a large amount of compression. Data obtained with conventional speech stimuli were essentially the same as those obtained with similar multichannel syllabic compressors.

These investigations do not support Plomp's argument in that moderate to large amounts of syllabic compression reduced the temporal information available for speech recognition to some extent but not substantially; that is, reasonable amounts of compression do not disrupt the temporal cues for speech. It is nevertheless important to recognize that there is an inherent loss of information when speech signals are compressed. This loss of information may not be important for a limited amount of compression, but if the amount of compression is substantial, the loss of information is likely to outweigh the benefits of improved audibility provided by compression.

The inherent problem in compression amplification with respect to speech intelligibility is finding the best compromise between the improvement in intelligibility provided by increased audibility with the reduction in intelligibility caused by the loss of speech cues resulting from the compression process.

This compromise is likely to be highly subject dependent as demonstrated by Levitt and Neuman (1991) in their investigation of orthogonal polynomial compression. In this method of compression, the proportion of the speech signal that can be squeezed into the residual hearing area can be increased systematically by increasing the number of terms used in the polynomial expansion. A point of diminishing returns is reached, however, when the loss of speech cues resulting from the reduction in spectral-temporal contrasts and the increase in internal masking exceeds the benefits obtained by squeezing more of the speech signal into the residual hearing area.

Levitt and Neuman (1991) found that some subjects obtained higher intelligibility scores with a moderate reduction in spectral balance (i.e., variations in the slope of the speech spectrum were reduced in range), whereas other subjects had higher intelligibility scores with conventional amplitude compression and no compression of spectral shape.

The importance of fitting compression hearing aids so as to best match the subject's residual hearing is of crucial importance. The magnitude and impact of individual differences is greater with more advanced methods of compression and should not be underestimated.

3.6 Other Considerations

The underlying rationale for the development of compression amplification was to improve the intelligibility of speech for people with a reduced dynamic range of hearing. This discussion has therefore focused on the effect of compression amplification on speech intelligibility. However, due consideration needs to be given to criteria other than speech intelligibility. Other criteria such as comfort, sound quality, clarity, ease of listening, pleasantness, noisiness, overall impression, and preference are currently being considered in experimental evaluations of hearing aids.

There are both advantages and disadvantages to the use of subjective criteria such as those listed above. These criteria address important aspects of acoustic amplification that are beyond the scope of more objective methods of evaluation such as speech recognition testing. Subjective ratings and paired comparison judgments are also quicker and easier to obtain than an objective speech test. Experimental comparisons between speech recognition tests and relevant subjective ratings are correlated so that subjective methods of evaluation can be used under conditions that would be impractical for speech-recognition testing (Sullivan et al. 1988; Neuman et al. 1994, 1998; Boike and Souza 2000). These evaluations have also shown that subjective judgments of speech quality are also more sensitive to differences among hearing aids than conventional speech tests.

The primary disadvantage of subjective assessment techniques is their subjective nature. Factors such as variations in criterion, subjective bias, and misinterpretation of instructions are important problems that need to be addressed. It is beyond the scope of this chapter to cover subjective assessment procedures, but it is relevant to note that because of the correlation between speech recognition scores and relevant subjective criteria (e.g., clarity, sound quality, ease of listening, overall impression), many of the findings obtained by means of speech recognition tests regarding the relative merits of the various methods of compression amplification also apply to these subjective criteria.

The correlation between subjective and objective measurements is usually very good when there are large differences among the amplification systems being evaluated but are not as good for small or subtle differences. That is, when the differences are large, objective measures of speech recognition and subjective evaluations will yield essentially similar results. When the differences are small, however, speech recognition testing may not be sensitive enough to detect the differences, whereas subjective judgments may have sufficient sensitivity to reliably identify these differences. It is in this latter area that subjective techniques are being used increasingly.

Compression amplification can also be used for noise reduction and controlling the loudness of the hearing aid user's own voice. Both of these problems can be very disturbing to the hearing aid user.

Amplification of the user's own voice needs to be reduced because the hearing aid microphone is closer to the user's voice than any other sound source in normal conversation and, as a consequence, is likely to be overamplified unless precautions are taken. In addition, sound produced by one's own voice also reaches the ear by means of bone conduction, which is a relatively efficient means of sound transmission. The sum of the two signals, the voice signal received acoustically via the hearing aid and the signal received by means of bone conduction, usually results in the user's own voice sounding excessively loud. The amount of bone conduction is also increased substantially if the ear is occluded.

Multichannel compression has distinct advantages over single-channel compression with respect to the above problem. Much of the annoying loudness is due to overamplification in the low frequencies, which a multichannel system can compress effectively without reduction of important high-frequency cues. As few as two or three frequency bands may be sufficient for this purpose.

Multichannel compression can also be used to reduce the level of background noise. The technique is to reduce the gain in those frequency bands where the noise level exceeds that of the speech. Because the speech signal is completely masked by the noise in these frequency bands, there is no loss of speech information by reducing the gain in this way.

In principle, the greater the number of frequency bands, the greater the nominal reduction in background noise level, but there is bound to be a point of diminishing returns if too many bands are used. At present, it is an open question as to how many frequency bands are needed to reduce background noise levels in the real-world use of hearing aids.

It should be noted that this method of noise reduction does not produce a concomitant improvement in speech intelligibility. This is because the intelligibility of speech is dependent on the speech-to-noise ratio within the critical bands of the ear. In order to improve speech intelligibility in noise, it is necessary to improve the speech-to-noise ratio within these critical bands. Multichannel compression can effectively change the gain within a given frequency band, but it cannot change the speech-to-noise ratio in that band. Reducing the gain of those bands in which speech is completely masked by noise will not improve speech intelligibility unless the noise is so intense that it causes significant upward spread of masking (Fabry et al. 1993). Background noises of this type are seldom encountered in the everyday use of hearing aids.

Another open question is the extent to which normal loudness relationships across frequency in the perception of speech and other complex signals can be restored using multichannel compression. Loudness distortions can, of course, be reduced to some extent by this means, but here again, it is not known how far one can go in restoring normal loudness relationships. It should be borne in mind that reduction in the dynamic range of hearing is not the only factor causing loudness distortion in sensorineural

hearing loss. There are, in addition, differences in loudness summation between normal-hearing and hearing-impaired listeners that need to be taken into account (Florentine et al. 1980). As a consequence, methods of compensating for both frequency-dependent and duration-dependent loudness distortions are likely to be very complex. Also, as noted earlier, there is a diversity of opinion as to how to model loudness growth in the impaired ear.

4. Summary

Compression amplification has many different forms. Some forms of compression amplification are particularly useful under certain conditions, whereas other forms are of limited or no benefit in their intended application. AGC is effective in compensating for long-term changes in signal level and will improve the intelligibility of low-level speech signals in quiet while maintaining intense signals at a comfortable loudness. Compression limiting is effective in limiting the output of a hearing aid and, with the exception of applications requiring very high output power, is generally superior to peak clipping. Single-channel syllabic compression has not yielded the anticipated improvements in intelligibility based on theoretical considerations, although some investigators have reported small improvements in intelligibility for speech in quiet. Laboratory studies involving adjustments to the CV intensity ratio indicate that larger improvements in intelligibility in quiet should be possible provided phonetic context and individual subject differences are taken into account.

From a theoretical perspective, multichannel WDR compression is much more promising than single-channel compression with respect to improving speech intelligibility. Experimental evaluations of multichannel compression amplification, however, have yielded mixed results. The majority of studies show no significant improvement in intelligibility, and several studies show a decrease in intelligibility. Despite the many conflicting research findings, there appears to be an underlying pattern to the results. Small improvements in speech intelligibility can be obtained but only for limited amounts of compression, such as using only two channels with a moderate compression ratio. Increasing the compression ratio above about 2:1 (for speech in noise) reduces intelligibility significantly. Judgments of sound quality show a similar decrement with increasing compression ratio. A large number of channels with unconstrained intensity variations between channels also introduces significant amounts of spectral distortion that are likely to reduce intelligibility.

There is an inherent loss of information when speech signals are compressed, and the optimum compromise has yet to be found between the information gained by increasing the audibility of the speech signal and the loss of important speech cues, such as the distortion of spectral and tem-

poral contrasts resulting from compression amplification. The optimum compromise is likely to be highly subject dependent. The importance and complexity of taking individual differences into account in fitting compression hearing aids should not be underestimated.

A useful feature of multichannel compression amplification is that the level of intense background noise can be reduced by reducing the gain in those frequency regions where the noise level exceeds that of the speech. This technique reduces the loudness and annoyance of background noise but does not improve intelligibility because the speech-to-noise ratio within the critical bands of the ear remains unchanged. Upward spread of masking can, under certain conditions, be reduced with multichannel compression, but the practical benefits in terms of improved intelligibility in the real-world use of hearing aids have yet to be demonstrated. Multichannel compression can also be used to reduce the annoyance of overamplification of the user's own voice.

The optimum number of frequency bands in multichannel compression is an open question. Good results have already been obtained with two-channel WDR compression that also combines the advantages of single-band AGC and compression limiting. More frequency bands are likely to be more effective in reducing the annoyance of spectrally complex background noises, but at some stage, a point of diminishing returns will be reached if too many bands are used.

In conclusion, compression amplification can be viewed as a family of techniques. AGC and compression limiting are the two good family members. They have their limitations, but they do what they were designed to do and they do it well. Single-channel and multichannel WDR syllabic compression are much like two ill-behaved children that need to be carefully controlled. They do not always do what they are instructed to do. The effective compression ratios for spectrally and temporally complex signals, for example, are likely to be less than what they are supposed to be. Furthermore, these techniques do not provide the improvements in intelligibility of which they are believed to be capable. Improvements in speech recognition have been obtained with syllabic compression under controlled laboratory conditions (such as careful adjustment of the CV intensity ratio taking phonetic context and individual subject differences into account), but the performance of this technique in real-world listening conditions leaves much to be desired.

Multichannel syllabic compression has much greater potential than single-channel techniques. Unfortunately, much of it is unrealized, but there is growing evidence that the technique is maturing and fulfilling some of its promise. Relatively good results have been obtained with two-channel compression. Improved intelligibility is also possible if between-band intensity variations are brought under control.

Multichannel compression has demonstrated real-world benefit by reducing the level of loud background noises. It is very revealing that this

application of multichannel compression was not considered in the underlying rationale that led to the development of the technique. The theoretical underpinnings of multichannel compression have focused primarily on the speech signal with limited success. Environmental noise and other important signals need to be taken into account within the framework of a much broader general theory of compression. The improvements that have been obtained thus far in dealing with real-world applications of multichannel compression are based on post hoc compromises between conflicting requirements. A general underlying theory would bring some discipline to this unruly but promising child.

Acknowledgements. Preparation of this chapter was supported by National Institute on Disability and Rehabilitation Research Grant H133E010107. I am also grateful to the editors for their many helpful comments.

References

Abramovitz RS (1980) Frequency shaping and multiband compression in hearing aids. J Commun Disord 13:483–488.

Boike KT, Souza PE (2000) Effect of compression ratio on speech recognition and speech-quality ratings with wide dynamic range compression amplification. J Speech Lang Hear Res 43:456–468.

Boothroyd A, Springer S, Smith L, Shulman J (1988) Amplitude compression and profound hearing loss. J Speech Hear Res 31:362–376.

Braida LD, Durlach NI, Lippmann RP, Hicks BL, Rabinowitz WM, Reed CM (1979) Hearing aids—a review of past research of linear amplification, amplitude compression and frequency lowering. ASHA Monograph No. 19. Rockville MD: American Speech-Language-Hearing Association.

Bustamante DK, Braida LD (1987a) Principal-component amplitude compression for the hearing impaired. J Acoust Soc Am 82:1227–1242.

Bustamante D, Braida L (1987b) Multiband compression limiting for hearing-impaired listeners. J Rehabil Res Dev 24:149–160.

Buus S, Florentine M (2002) Growth of loudness in listeners with cochlear hearing losses: recruitment reconsidered. J Assoc Res Otolaryngol 3:120–139.

Davis H, Stevens SS, Nichols RH Jr, Hudgins CV, Marquis RJ, Peterson GE, Ross DA (1947) Hearing Aids: An Experimental Study of Design Objectives. Cambridge, MA: Harvard University Press.

De Gennaro S, Braida L, Durlach N (1986) Multichannel syllabic compression for severely impaired listeners. J Rehabil Res Dev 23:17–24.

Dillon H (2001) Hearing Aids. New York: Thieme.

Dreschler WA (1988) Dynamic range reduction by peak clipping or compression and its effects on phoneme perception in hearing-impaired listeners. Scand Audiol 17:45–51.

Dreschler WA (1989) Phoneme perception via hearing aids with and without compression and the role of temporal resolution. Audiology 28:49–60.

Dreschler WA, Eberhardt D, Melk PW (1984) The use of single-channel compression for the improvement of speech intelligibility. Scand Audiol 13:231–236.

Drullman R, Smoorenburg G (1997) Audio-visual perception of compressed speech by profoundly hearing-impaired subjects. Audiology 36:165–177.
Fabry D, Leek M, Walden B, Cord M (1993) Do adaptive frequency response (AFR) hearing aids reduce "upward spread" of masking? J Rehabil Res Dev 30:318–325.
Florentine M, Buus S, Scharf B, Zwicker E (1980) Frequency selectivity in normally-hearing and hearing-impaired observers. Audiology 36:165–177.
Gordon-Salant S (1986) Recognition of natural and time/intensity altered CVs by young and elderly hearing-impaired subjects. J Acoust Soc Am 80:1599–1607.
Gordon-Salant S (1987) Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects. J Acoust Soc Am 81:1199–1202.
Hohmann V, Kollmeier B (1995) The effect of multichannel dynamic compression on speech intelligibility. J Acoust Soc Am 97:1191–1195.
Humes LE, Christense L, Thomas T, Bess FH, Hedley-Williams A, Bentler R (1999) A comparison of the aided performance and benefit provided by a linear and a two-channel wide dynamic range compression hearing aid. J Speech Lang Hear Res 42:65–79
Kennedy E, Levitt H, Neuman AC (1998) Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners. J Acoust Soc Am 103:1098–1114.
Kiessling J, Steffens T (1991) Clinical evaluation of a programmable three-channel automatic gain control amplification system. Audiology 30:70–81.
Killion MC (1993) The K-Amp hearing aid: an attempt to present high fidelity for the hearing impaired. Am J Audiol 2:52–74.
Killion MC, Staab WJ, Preves DA (1990) Classifying automatic signal processors. Hear Instrum 41:24–26.
Killion MC, Teder H, Johnson A, Hanke S (1992) Variable recovery time circuit for use with wide dynamic range automatic gain control for hearing aids. US Patent No. 5,144,675 (1 September 1992).
Larsen VD, Williams DW, Henderson WG, Luethke LE, Beck LB, Noffsinger D, Wilson RH, Dobie RA, Haskell GB, Bratt GW, Shanks JE, Stelmachowicz P, Studebaker GA, Boysen AE, Donahue A, Canalis R, Fausti SA, Rappaport BZ (2000) Efficacy of 3 commonly used hearing aid circuits: a crossover trial. NIDCD/VA Clinical Trial Group. JAMA 284:1806–1813.
Laurence RF, Moore BCJ, Glasberg BR (1983) A comparison of behind-the-ear high fidelity linear hearing aids and two-channel compression aids, in the laboratory and in everyday life. Br J Audiol 17:31–48.
Levitt H (1982) An array-processor computer hearing aid. ASHA J Am Speech Hear Assoc 24:805.
Levitt H (1993) Digital hearing aids. In: Studebaker GA, Hochberg I (eds) Acoustical Factors Affecting Hearing Aid Performance (2nd ed). Boston: Allyn and Bacon, pp. 317–335.
Levitt H, Neuman AC (1991) Evaluation of orthogonal polynomial compression. J Acoust Soc Am 90:241–252.
Levitt H, Neuman AC, Mills, Schwander T (1986) A digital master hearing aid. J Rehabil Res Dev 23:79–87.
Lindholm J, Dorman M, Taylor B, Hannley M (1988) Stimulus factors influencing the identification of voiced stop consonants by normal-hearing and hearing-impaired adults. J Acoust Soc Am 83:1608–1614.

Lippmann R, Braida L, Durlach N (1981) Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss. J Acoust Soc Am 69:524–534.

Medical Research Council (1947) Hearing Aids and Audiometers. Report of the Committee on Electro-Acoustics. Special Report Series No. 261.London: His Majesties Stationary Office.

Montgomery A, Edge R (1988) Evaluation of two speech enhancement techniques to improve intelligibility for hearing impaired adults. J Speech Hear Res 31: 386–393.

Moore BCJ (1987) design and evaluation of a two-channel compression hearing aid. J Rehabil Res Dev 24:181–192.

Moore BCJ, Glasberg BR (1986) A comparison of two-channel and single-channel compression hearing aids. Audiology 25:210–226.

Moore BCJ, Glasberg BR (1988) A comparison of four methods of implementing automatic gain control (AGC) in hearing aids. Br J Audiol 22:93–104.

Moore BCJ, Laurence RF, Wright D (1985) Improvements in speech intelligibility in quiet and in noise produced by two-channel compression hearing aids. Br J Audiol 19:175–189.

Moore BCJ, Glasberg BR, Stone MA (1991) Optimization of a slow-acting automatic gain control system for use in hearing aids. Br J Audiol 25:171–182.

Moore BCJ, Johnson J, Clark T, Pluvinage V (1992) Evaluation of a dual-channel full dynamic range compression system for people with sensorineural hearing loss. Ear Hear 13:349–370.

Neely ST, Allen JB (1997) Relation between the rate of growth of loudness and the intensity DL. In: Jesteadt W (ed) Modeling Sensorineural Hearing Loss. Hillsdale, NJ: Erlbaum Associates, pp 213–222.

Neuman AC, Bakke MH, Helman S, Levitt H (1994) Effect of compression ratio in a slow-acting compression hearing aid: paired comparison judgments of quality. J Acoust Soc Am 96:1471–1478.

Neuman AC, Bakke MH, Mackersie C, Hellman S, Levitt H (1998) The effect of compression ratio and release time on the categorical rating of sound quality. J Acoust Soc Am 103:2273–2281.

Plomp R (1988) The negative effect of amplitude compression in multichannel hearing aids in light of the modulation-transfer function. J Acoust Soc Am 83: 2322–2327.

Plomp R (1994) Noise, amplification, and compression: considerations of three main issues in hearing aid design. Ear Hear 15:2–12.

Robinson CE, Huntington DA (1973) The intelligibility of speech processed by delayed long-term averaged compression amplification. J Acoust Soc Am 54: 314.

Shorter DEL, Manson WI, Stebbings DW (1967) The dynamic characteristics of limiters for sound programme circuits. BBC Engineering Monograph No. 70. London: British Broadcasting Corporation.

Souza PE, Turner CW (1996) Effect of single-channel compression of temporal speech information. J Speech Lang Hear Res 39:901–911.

Souza PE, Turner CW (1998) Multichannel compression, temporal cues, and audibility. J Speech Lang Hear Res 41:315–326.

Steeneken HJM, Houtgast T (1980) A physical method for measuring speech-transmission quality. J Acoust Soc Am 67:318–326.

Stone MA, Moore BCJ (1992) Syllabic compression: effective compression ratios for signals modulated at different rates. Br J Audiol 26:351–361.
Stone MA, Moore BCJ (1999) Tolerable hearing aid delays. I. Estimation of limits imposed by the auditory path alone using simulated hearing losses. Ear Hear 20: 182–192.
Stone MA, Moore BCJ, Alcantara JI, Glasberg BR (1999) Comparison of different forms of compression using wearable hearing aids. J Acoust Soc Am 106: 3603–3619.
Sullivan JA, Levitt H, Hwang JY, Hennessey AM (1988) An experimental comparison of four hearing aid prescription methods. Ear Hear 9:22–32.
Teder H (1991) Hearing instruments in noise and the syllabic speech-to-noise ratio. Hear Instrum 42:15–18.
Tyler RS, Kuk FK (1989) The effects of "noise suppression" hearing aids on consonant recognition in speech-babble and low-frequency noise. Ear Hear 10:243–249.
Van Tasell DJ, Trine TD (1996) Effects of single-band syllabic amplitude compression on temporal speech information in nonsense syllables and in sentences. J Speech Lang Hear Res 39:912–922.
Verschuure H, Prinsen T, Dreschler W (1994) The effects of syllabic compression and frequency shaping on speech intelligibility in hearing impaired people. Ear Hear 15:13–21.
Verschuure H, Maas AJ, Stikvoort E, de Jong RM, Goedegebure A, Dreschler WA (1996) Compression and its effect on the speech signal. Ear Hear 17:162–175.
Villchur E (1973) Signal processing to improve speech intelligibility in perceptive deafness. J Acoust Soc Am 53:1646–1657.
Villchur E (1989) Comments on "The negative effect of amplitude compression in multichannel hearing aids in light of the modulation-transfer function" [J Acoust Soc Am 83:2322–2327]. J Acoust Soc Am 86:425–427.
Walker G, Byrne D, Dillon H (1984) The effects of multichannel compression/expansion amplification on the intelligibility of nonsense syllables in noise. J Acoust Soc Am 76:746–757.
Yanick P (1973) Improvements in speech discrimination with compression versus linear amplification. J Aud Res 13:333–338.
Yund EW, Buckles KM (1995a) Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids. J Acoust Soc Am 97:1224–1240.
Yund EW, Buckles KM (1995b) Multichannel compression hearing aids: effect of number of channels on speech discrimination in noise. J Acoust Soc Am 97:1206–1223.
Yund EW, Buckles KM (1995c) Discrimination of multichannel-compressed speech in noise: long-term learning in hearing-impaired subjects. Ear Hear 16:417–427.
Yund E, Simon HJ, Efron R (1987) Speech discrimination with an 8-channel compression hearing aid and conventional aids in background of speech-band noise. J Rehabil Res Dev 24:161–180.

6
Compression and Cochlear Implants

Fan-Gang Zeng

1. Introduction

Cochlear compression plays an important role in supporting the exquisite sensitivity, fine frequency tuning, and large operating dynamic range of the ear (Bacon, Chapter 1; Cooper, Chapter 2; Oxenham and Bacon, Chapter 3). With cochlear impairment, hearing threshold is elevated, and frequency selectivity and dynamic range are reduced (Bacon and Oxenham, Chapter 4; Levitt, Chapter 5). In cases of cochlear implants, the cochlear compression and other cochlear functions are bypassed altogether; hearing sensation is evoked by direct electric stimulation of the auditory nerve.

The goal of studying compression in cochlear implants is twofold. The first goal is practical. In a cochlear implant, all compression-related functions need to be replaced by a front-end artificial processor. The second goal is theoretical because direct stimulation of the auditory nerve provides a unique opportunity for studying the lack of cochlear compression in auditory perception, thereby complementing the studies in normal-hearing and cochlear-impaired listeners and allowing delineation of peripheral and central contributions to the functions of the overall system.

This chapter first briefly reviews how cochlear implants work (see Section 2). The psychophysical, or more precisely, psychoelectrical capabilities in cochlear implant users is reviewed, with an emphasis on the effect of loss of cochlear compression on perception (see Section 3). Section 4 contrasts the psychoacoustical and psychoelectrical data and discusses their inference on the theoretical models of auditory processing. Section 5 discusses practical issues on compression in cochlear implants in terms of restoring normal loudness growth, increasing electric dynamic range, and improving speech performance in implant users. Section 6 summarizes the current data and discusses future research directions.

2. Review of Cochlear Implants

2.1 Past and Present

When the Italian scientist Alessandro Volta invented the battery more than two centuries ago, one thing he did with his invention was to study how electric stimulation might affect sensations (Volta 1800). While studying the effects of electric stimulation on light, touch, smell, and other sensations, he placed one of two metallic probes in each ear and connected the ends of the probes to a 50-V battery. He observed that, ". . . at the moment when the circuit was completed, I received a shock in the head, and some moments after I began to hear a sound, or rather noise in the ears, which I cannot well define: it was a kind of crackling with shocks, as if some paste or tenacious matter had been boiling . . . The disagreeable sensation, which I believe might be dangerous because of the shock in the brain, prevented me from repeating this experiment . . ." It is believed that that was the first demonstration that electric stimulation can evoke hearing sensation.

At present, more than 50,000 people worldwide, including 10,000 children, have received cochlear implants. The earliest FDA-approved device was the House-3M single-electrode implant, with several hundred users. At present, there are three major cochlear implant companies, including the manufacturers of the Clarion device (Advanced Bionics Corporation, US), the Med-El device (Med-El Corporation, Austria), and the Nucleus device (Cochlear Corporation, Australia). An earlier multielectrode implant (the Ineraid device; Eddington et al. 1978) had a percutaneous plug interface and was ideal for many research purposes. In patients whose auditory nerve was sectioned by tumor removal, an auditory brainstem implant (ABI) has been used to stimulate the cochlear nucleus directly (e.g., Otto et al. 2002). The cochlear implant has evolved from the single-electrode device that was used mostly as an aid for lip reading and sound awareness to modern multielectrode devices that allow an average user to talk on the telephone. Despite the differences in speech processing and electrode design, there appears to be no significant difference in performance among the present cochlear implant users. The audiological criteria for having cochlear implantation has also relaxed, from bilateral total deafness (greater than 110 dB HL hearing loss) to severe hearing loss (greater than 70 dB HL) to the current suprathreshold speech-based criteria (less than 50% open-set sentence recognition with properly fitted hearing aids; NIH Consensus Statement 1995). More importantly, given the appropriate environment, most children who have received cochlear implants have shown language development parallel to that of normal-hearing children (Svirsky et al. 2000).

2.2 Design of a Cochlear Implant

In normal hearing, sound travels from the outer ear through the middle ear to the cochlea where the sound is converted into electric impulses that the brain can understand. Most cases with severe hearing loss involve damage to this sound-to-electric impulse conversion in the cochlea. A cochlear implant bypasses this natural conversion process by directly stimulating the auditory nerve with electric pulses.

Figure 6.1 shows a typical modern cochlear implant. First, a microphone picks up the sound (1) and sends it via a wire (2) to the speech processor (3) that is worn behind the ear or on a belt like a pager for older versions. The speech processor converts the sound into a digital signal according to the individual's degree of hearing loss. The signal travels back to the headpiece (4) that contains a coil transmitting coded radio frequencies through the skin (5). The headpiece is held in place by a magnet attracted to the

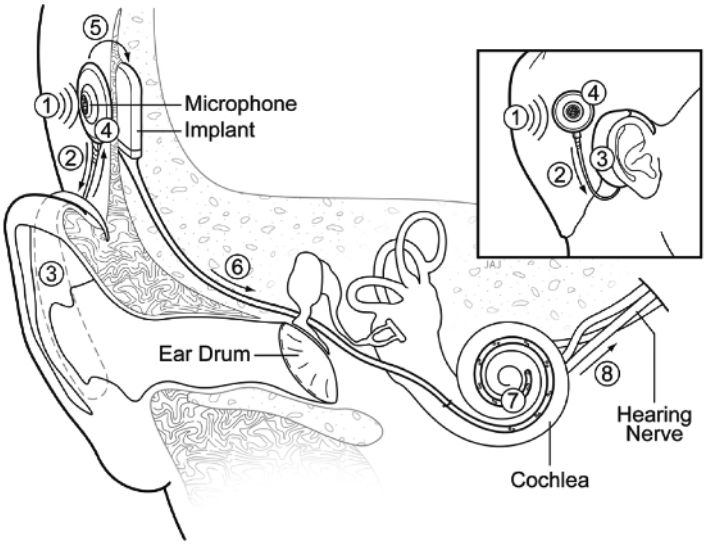


FIGURE 6.1. A typical cochlear implant system showing how it directly converts sound to electric impulses delivered to the auditory nerve. A microphone picks up the sound (1) and sends it via a wire (2) to a behind the ear speech processor (3). The processor converts the sound into a digital signal and sends the processed signal to a headpiece (4). The headpiece is held in place by a magnet attracted to the implant on the other side of the skin. Both the headpiece and the implant contain coils that transmit coded radio frequencies through the skin (5). The implant also contains hermetically sealed electronic circuits that decode the signals, convert them into electric currents, and send them along wires threaded into the cochlea (6). The electrodes at the end of the wire (7) stimulate the auditory nerve (8) connected to the central nervous system where the electrical impulses are interpreted as sound.

implant on the other side of the skin. The implant contains another coil receiving the radio frequency signal and also hermetically sealed electronic circuits. The circuits decode the signals, convert them into electric currents, and send them along wires threaded into the cochlea (6). The electrodes at the end of the wire (7) stimulate the auditory nerve (8) connected to the central nervous system, where the electrical impulses are interpreted as sound.

To a large degree, all modern multielectrode devices attempt to replicate the frequency analysis and amplitude compression mechanisms in acoustic hearing. They all divide a broadband audio signal (between several hundred hertz and 5–10 kHz) into 8–20 narrowband signals via analog or digital filters. However, they differ significantly in the postfiltering processing. One version of processing is to deliver the narrowband analog waveform to a tonotopically appropriate electrode that directly stimulates the auditory nerve. This type of processing has been called compressed analog (CA) or simultaneous analog stimulation (SAS) strategy (Eddington et al. 1978). Another version of processing is to extract the temporal envelope of the narrowband signals via rectification and low-pass filtering and then use the temporal envelope to amplitude modulate a fixed-rate biphasic pulse carrier. To avoid simultaneous stimulation between different electrodes, the pulses of the carrier between electrodes are systematically interleaved so that only one electrode will be stimulated at a time. When the number of analysis bands is the same as the number of electrodes, this nonsimultaneous processing is called continuous interleaved sampling (CIS) strategy (Wilson et al. 1991). However, when only a subset of the bands with maximal activities (e.g., 8 of the 22 electrodes) is stimulated, it is called the n-of-m (i.e., 8-of-20) or SPEAK strategy (McDermott et al. 1992). Recently, several combinations of these strategies have also become commercially available.

In addition to the common frequency analysis, all cochlear implants have adopted an amplitude compression scheme to match the wide acoustic dynamic range to the narrow electric range (see Section 3.1.1). The compression can be achieved with a gain control mechanism in the CA strategy, a logarithmic compression in the CIS strategy, or a power-function compression in the SPEAK strategy. Ideally, this compression scheme in cochlear implants can replicate the cochlear compression in acoustic hearing (see Section 4) and restore normal loudness growth (see Section 5).

2.3 Physiological Responses to Electric Stimulation

The cochlear implant bypasses all cochlear functions, including compression and synaptic transmission, to stimulate the auditory nerve directly. It is both important and interesting to compare the response of the auditory nerve to acoustic and electric stimulation so that we can understand the electrode-nerve interface and better appreciate the role of cochlear compression in perception. In the following paragraphs, the focus is on rate-intensity functions and phase-locking properties of auditory neurons.

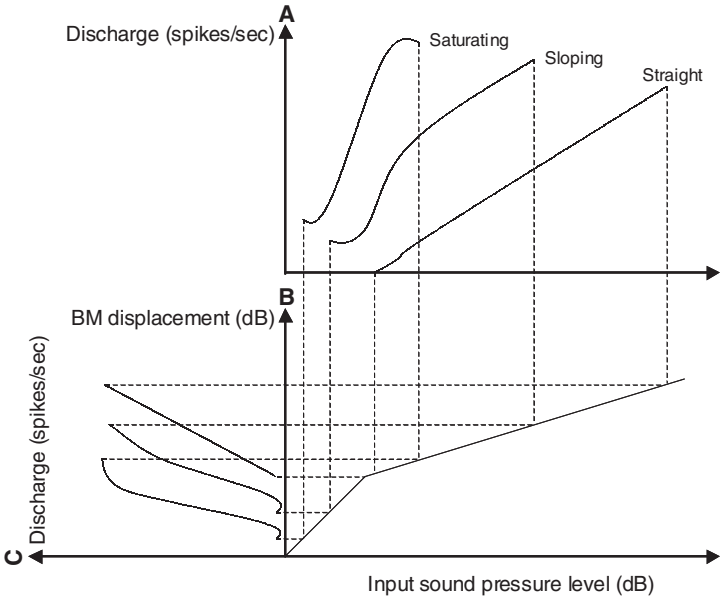


FIGURE 6.2. *A*: 3 typical types of rate-intensity functions for the auditory nerve fibers in acoustic hearing. Note the narrowest dynamic range in the saturating-type fiber and the widest dynamic range in the straight-type fiber. *B*: compressive input-output function of the basilar membrane (BM). *C*: Transformed "rate-intensity" functions in which the *x*-axis is the basilar membrane displacement. Note the uniformly narrow dynamic range in all 3 nerve fibers, suggesting that the acoustically observed different types of rate-intensity functions are mainly due to cochlear compression. (Adapted with permission from Yates et al. 1992. Copyright © 1992 Elsevier Science.)

Figure 6.2*A* shows three typical rate-intensity functions in acoustic hearing from the auditory neurons with high, medium, and low spontaneous activities (e.g., Sachs and Abbas 1974; Liberman 1978; Yates et al. 1990; Cooper, Chapter 2). Generally, the higher the spontaneous activity, the lower the response threshold and the wider the dynamic range. The dynamic range varies from 10 to 20 dB for high spontaneous rate neurons to 50 dB or more for low spontaneous rate neurons. In electric stimulation of a deafened ear, there is no spontaneous activity for the auditory neuron that has lost the dendritic connection to the inner hair cells (Kiang and Moxon 1972; Hartman et al. 1984; van den Honert and Stypulkowski 1984; Parkins and Colombo 1987). Although there are variations in electric thresholds, the dynamic range of the rate-intensity functions in electric stimulation is uniformly narrow and on the order of several decibels (Javel et al. 1987; Miller et al. 1999; Litvak et al. 2001).

Here it is demonstrated that the extremely narrow dynamic range in the rate-intensity function for electrical stimulation is a direct consequence of

the loss of cochlear compression. It has been argued (e.g., Sachs and Abbas 1974; Yates et al. 1992) that the difference in rate-intensity functions for acoustic stimulation reflects the compressive nonlinearity of the basilar membrane (Fig. 6.2*B*) rather than the neural response to the drive of the basilar membrane. For example, the straight rate-intensity function with a large dynamic range becomes surprisingly similar to the saturating function with a narrow dynamic range after taking the cochlear compression into account (Fig. 6.2*C*). Because the direct input to the neuron is the postsynaptic potential of the inner hair cell, we can consider Figure 6.2*C* as the response of the neuron to "electric stimulation" in a normal ear. Clearly, these "electrically stimulated" rate-intensity functions are more uniform and all have narrow dynamic ranges similar to what is observed in direct electric stimulation of the auditory nerve in a deafened ear.

Another example demonstrating the role of cochlear compression at the auditory nerve level is to compare the phase-locking abilities between acoustically and electrically stimulated auditory neurons. Because cochlear compression also contributes to the sharp frequency tuning in acoustic hearing, it must pose a physical limit to the phase-locking abilities of the auditory nerve according to the trade-off between frequency and time resolution. Figure 6.3 shows a synchronization measure as a function of

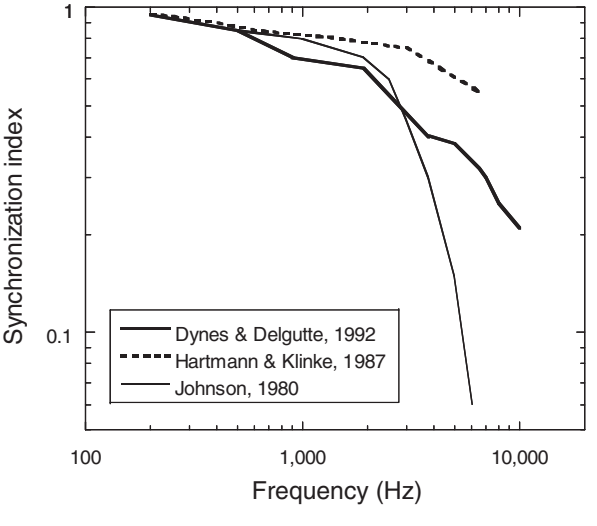


FIGURE 6.3. Synchronization index as a function of stimulus frequency in acoustic and electric stimulation. Note the sharply decreased synchronization in acoustic hearing (thin line) for frequencies above 1 kHz. In contrast, significant synchronization is still present at 6–10 kHz in electric hearing (dashed and thick lines). (Adapted with permission from Dynes and Delgutte 1992. Copyright © 1992 Elsevier Science.)

stimulus frequency with acoustic (Johnson 1980) and electric stimulation (Hartmann and Klinke 1987; Dynes and Delgutte 1992). Although the synchronization of firing decreases sharply for frequencies above 1,000 Hz in acoustic stimulation, a relatively high degree of synchronization is still maintained for frequencies as high as 10,000 Hz in electric stimulation. As seen in Section 4.3, the cochlear compression clearly limits normal-hearing listeners' ability to process temporal information.

2.4 Use of the Cochlear Implant to Probe Normal Auditory Processing

Although the cochlear implant has achieved a high level of medical success as a means of restoring partial hearing to deaf people, its use as a research tool to understand normal auditory processing is still underappreciated. This section shows how to use various patient populations in general and cochlear implant users in particular to understand normal auditory processing mechanisms.

We take an engineering reductionism approach that assumes that the function of the total system can be understood by assessing the function of each component. Figure 6.4 illustrates this approach. In auditory processing, the information is processed by the cochlea, then the auditory nerve, the cochlear nucleus, and, finally, other parts of the central nervous system and the feedback pathway from the central nervous system to the cochlea (the efferents). If we want to understand cochlear function, we can compare the performance between a normal-hearing listener and a cochlear implant user because in the latter case, the cochlea is bypassed and the auditory nerve is directly stimulated. Similarly, if we want to understand auditory nerve function, we can compare the performance between the cochlear implant user and an auditory brainstem implant (ABI) user because in the latter case, both the cochlea and the auditory nerve are bypassed and the cochlear nucleus is directly stimulated (Brackmann et al. 1993). The efferent function can also be inferred by comparing the performance between the ear with intact efferent innervation and the ear with sectioned efferents (Scharf et al. 1997; Zeng et al. 2000).

This approach certainly has its limitations. One limitation is the still poorly understood electrode-to-nerve interface that is likely not equivalent to the natural transducer-to-nerve synapses. Another limitation is the generally degenerated neurons (e.g., cell loss and demyelination) in cochlear implant users. These neural abnormalities may produce additional deficits other than what the lack of the cochlear mechanisms alone would cause. However, as a first approximation, it can be shown that this comparative and reductionism approach can reveal important processing mechanisms in auditory perception. The hope is that with better understanding of the biophysics and physiology of the system, we can delineate the peripheral versus

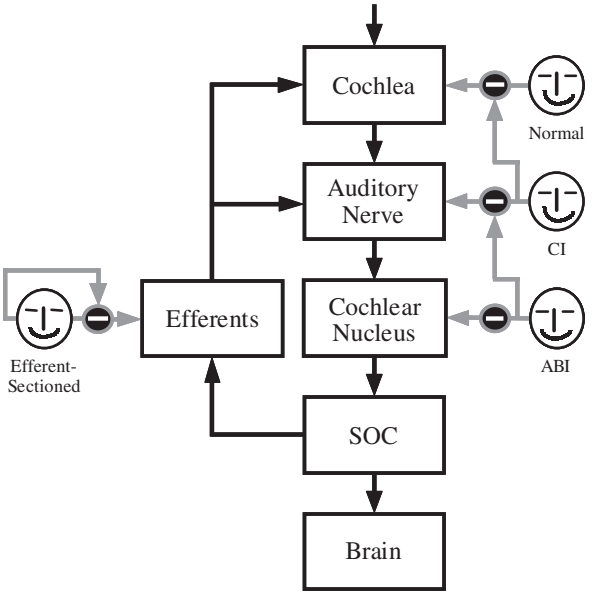


FIGURE 6.4. A reductionism model using hearing-impaired listeners to probe normal auditory functions. The blocks with downward arrows represent the auditory afferent pathway from cochlea to auditory nerve, cochlear nucleus, superior olivary complex (SOC), and the brain. The block with the upward arrow represents the auditory efferent pathway from SOC to the auditory nerve and cochlea. The happy faces represent normal-hearing, cochlear-implant (CI), auditory brainstem implant (ABI), and efferent-sectioned listeners. Because the cochlear implant bypasses the cochlea and directly stimulates the auditory nerve, a comparison between normal and CI listeners' performance will allow us to deduce the role of the cochlea in auditory processing. Similarly, a comparison between CI and ABI listeners will deduce the role of the auditory nerve and that between normal and efferent-sectioned listeners will deduce the role of efferents in auditory processing.

central contribution to the performance of the overall system. As a first step, the focus is on the cochlear implant users' psychoelectrical capabilities (Section 3).

3. Perceptual Consequences of Lost Cochlear Compression

Cochlear compression significantly affects auditory perception. This effect can be direct in terms of intensity processing or indirect in terms of frequency and temporal processing. This section demonstrates both the direct

and indirect effects of cochlear compression on auditory perception by examining perceptual performance in cochlear implant users.

3.1 Intensity Processing

A normal-hearing listener can process sound information with changes in intensity over 12 orders of magnitude, a 120-dB dynamic range in acoustic hearing (Bacon, Chapter 1). In addition, a normal-hearing listener can discriminate up to 200 intensity differences within this 120-dB dynamic range (Rabinowitz et al. 1976; Viemeister and Bacon 1988; Schroder et al. 1994). This remarkable dynamic range and its associated intensity resolution are deeply rooted in cochlear compression. Sections 3.1.1, 3.1.2, and 3.1.3 examine the changes in the dynamic range and intensity resolution caused by the loss of cochlear compression.

3.1.1 Dynamic Range in Electric Hearing

In electric hearing, the absolute threshold does not have the same meaning as in acoustic hearing. The acoustic threshold is limited to the minimal mechanical vibration that can be picked up by the hair cells and the nerve via an active mechanical mechanism (or amplifier). The electric threshold mostly reflects the type of electrode used, the electrode-tissue interface, the distance between the electrode and the nerve, and the degree and pattern of the nerve survival. Figure 6.5 presents dynamic range data as a function of frequency for both sinusoidal (left-slanted hatched areas) and pulsatile (right-slanted hatched areas) stimuli in eight cochlear implant users. The lower boundary of the dynamic range is the electric thresholds, and the upper boundary is the maximum acceptable loudness (Zeng and Shannon 1994, 1999).

Although there are significant individual differences in the absolute value of electric hearing thresholds and maximum loudness levels, the pattern of the data is clear and uniform. For sinusoids, both thresholds and maximum loudness levels increase monotonically as a function of frequency, whereas for pulses, both decrease monotonically as a function of frequency. The 100-Hz sinusoid produced the lowest threshold and the widest dynamic range (mean dynamic range = 30 dB), whereas the 100-Hz pulse produced the highest threshold and the narrowest dynamic range (mean dynamic range = 14 dB). The sinusoidal dynamic range decreases with frequency until 300–500 Hz as a result of a steeper increase in thresholds than in maximum loudness levels. On the other hand, the pulsatile dynamic range increases with frequency as a result of a steeper decrease in thresholds than maximum loudness levels. At 1,000 Hz, there is no statistical difference in dynamic range between sinusoidal (19 dB) and pulsatile (18 dB) stimuli. Both phenomenological (Shannon 1989a; Zeng et al. 1998b) and biophysical (Bruce

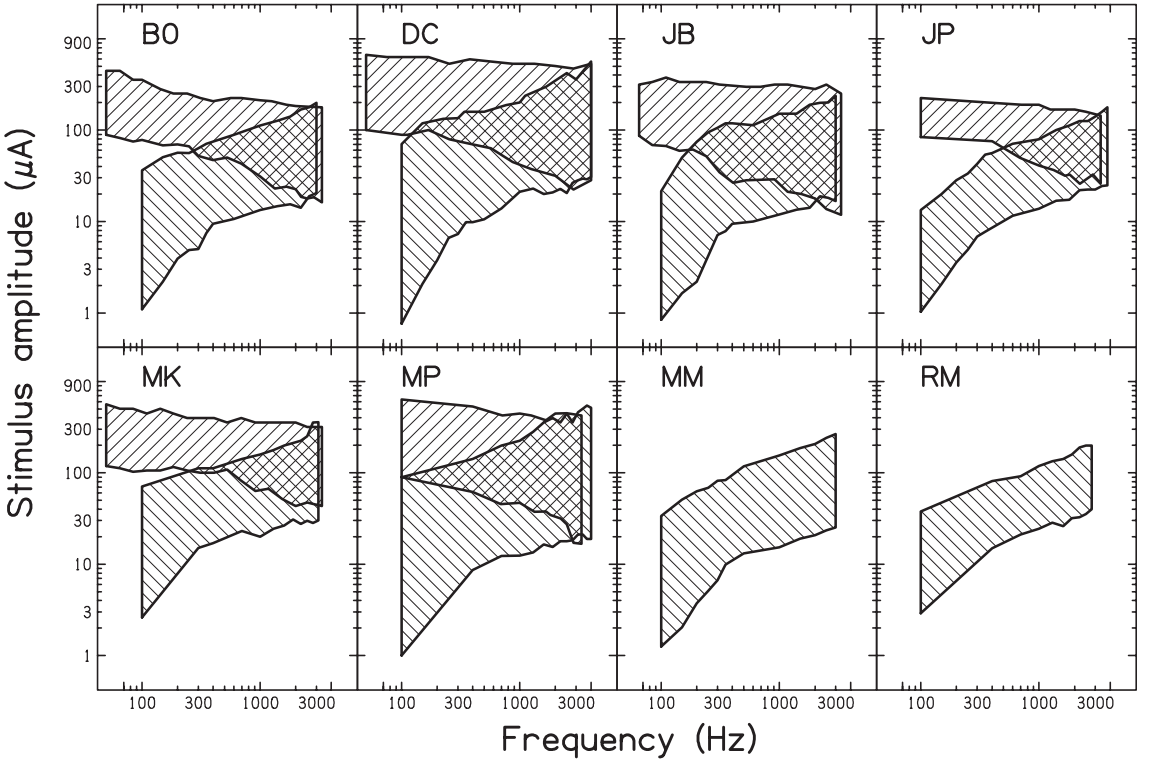


FIGURE 6.5. Dynamic range data as a function of frequency for sinusoidal (left-slanted hatched areas) and pulsatile (right-slanted hatched areas) stimuli in 8 Ineraid CI users. The lower boundary of the dynamic range is the electric threshold, and the upper boundary is the maximum acceptable loudness. All stimuli were 200 ms in duration. The pulsatile stimuli consisted of biphasic pulse trains of 100 ms/phase. The most apical electrode was stimulated in a monopolar mode. In contrast to the 100–120 dB acoustic dynamic range, the electric dynamic range varied between 10 and 30 dB (a ratio of 3 to about 30 between the maximum acceptable loudness and the threshold levels).

et al. 1999c) models have been developed to predict these dynamic range data.

3.1.2 Loudness Growth

How does loudness grow from threshold to maximal comfortable level? It has been well established in acoustic hearing that loudness grows as a power function of sound intensity with an exponent of roughly 0.3 (the famous Stevens' power law, Stevens 1961). In electric hearing, loudness growth has been modeled as either a power function or an exponential function (for a review, see Zeng and Shannon 1992).

There are generally two methods to obtain the loudness function. One method is to use loudness balancing to derive indirectly the loudness function of an unknown stimulus. For example, if we know that loudness (L) grows as a power function (f) of the intensity (I) for an acoustical stimulus [$L = f(I)$] and also know the balance function (g) between acoustic and electric (E) stimuli [$I = g(E)$], then we can easily derive the loudness function for the electric stimulus $\{L = f(I) = f[g(E)]\}$. This method has been used by a number of investigators (Eddington et al. 1978; Zeng and Shannon 1992; Dorman et al. 1993) and has produced the most convincing evidence for an exponential loudness function in electric hearing. Figure 6.6*A* shows a classic power function for loudness growth in the acoustic ear of a subject who had a brain stem implant on the other side. A brainstem rather than a cochlear implant user is used because (1) a complete set of loudness and just noticeable difference (JND) data is available for this subject (Fig. 6.6; see also Fig. 6.8) and (2) there is no difference in loudness growth between brainstem and cochlear implant users, at least for stimulus frequencies higher than 300 Hz (Zeng and Shannon 1994). Figure 6.6*B* shows a linear loudness balance function between linear electric amplitude (A) in microamperes and log acoustic amplitude in decibels ($E = a^*\log A$, where a is a constant). The linear function in these coordinates indicates that loudness grows as an exponential function in electric hearing

$$L = A^{\theta} = \left(10^{E/a}\right)^{\theta} = 10^{E\theta/a} \tag{3.1.2.1}$$

Another method used to obtain loudness growth functions in electric hearing is via the classic magnitude estimation technique in which the subject has to report a number that (s)he believes corresponds to the loudness of the presented sound. Figure 6.7 shows such functions obtained by several investigators using the magnitude estimation method in cochlear implant users. The symbols are actual data, and the dashed lines represent exponential functions best fitted to the data. In some cases, a power function was found to provide as good (Fu and Shannon 1998) or better (Zeng and Shannon 1994) a fit to loudness growth for low-frequency (< 300 Hz) electric stimuli and can be directly compared to the expected 0.3 exponent or slope in normal-hearing listeners. However, it is sometimes

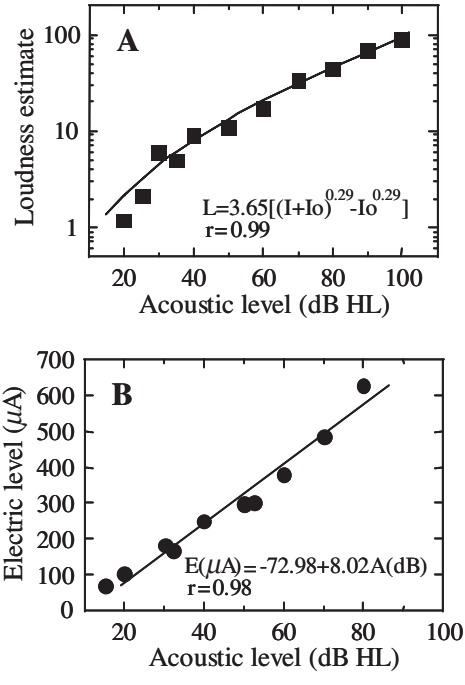


FIGURE 6.6. Loudness function in acoustic hearing (*A*) and loudness balance function between acoustic and electric hearing (*B*) in a listener who had normal hearing on 1 side and an ABI on the other side. The loudness function shows a typical power function with an exponent of 0.29 in acoustic hearing. Note the linear function in *B* in which the *x*-axis is in decibels for acoustic level, whereas the *y*-axis is linear microamperes for electric level, indicating a logarithmic loudness balance function between acoustic and electric hearing. (HL = Hearing Level)

difficult to use the power function to describe adequately the shallower loudness growth ("tail") near the threshold. The exponential loudness growth function can describe a wide range of stimulus conditions over the entire dynamic range (e.g., Chatterjee 1999; Chatterjee et al. 2000; McKay et al. 2001).

3.1.3 JND in Intensity

The intensity JND in electric hearing has been measured extensively (Bilger 1977; Eddington et al. 1978; Shannon 1983; Nelson et al. 1996; Zeng and Shannon 1999). Although the term intensity is used here to follow the general tradition in psychoacoustics (Viemeister and Bacon 1988), the actual unit used in electric hearing is usually current, a measure that is equivalent to pressure in acoustics. Here some interesting effects of the loss of cochlear compression on intensity discrimination in electric hearing are

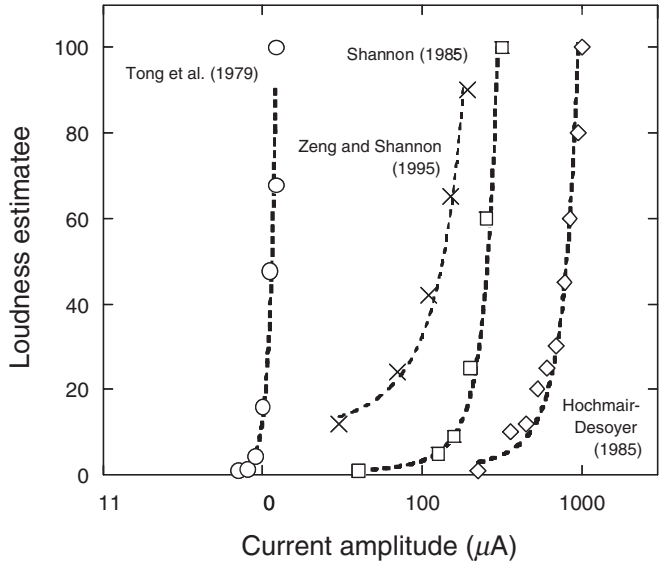


FIGURE 6.7. Loudness functions in electric hearing obtained from several previous studies. Dashed lines: fit to these data using an exponential loudness function in electric hearing. (All except for Zeng and Shannon (1995) are adpated with permission from Zeng and Shannon 1992. Copyright © 1992 Elsevier Science.)

demonstrated. Figure 6.8*A* shows the raw JND data (differences in microamperes as a function of standard level, also in microamperes) obtained in the same auditory brainstem listener whose loudness-matching data were shown in Figure 6.6. The JND in current was relatively constant at 70 μA within the lower half of the dynamic range (250–450 μA) but decreased monotonically to about 30 μA near the maximal comfortable loudness level. Because the overall dynamic range is about 500 μA, the JND size is relatively constant at about 10% of the dynamic range. In other words, the implant subject can resolve about 10 discriminable steps within the entire dynamic range. Similarly, results of 10–20 discriminable steps have been found in most cochlear implant subjects (Nelson et al. 1996; Zeng and Shannon 1999) in contrast to the 50–200 steps found in normal-hearing listeners (Rabinowitz et al. 1976; Viemeister and Bacon 1988; Schroder et al. 1994). If we use the total number of discriminable steps as the measure, then we can conclude that intensity discrimination abilities in cochlear implant users are much poorer than in normal-hearing listeners.

However, if we use a different measure of intensity discrimination, namely, the Weber fraction (the ratio between the absolute difference and the standard), as traditionally used in psychoacoustics, then we will reach a totally different conclusion. Figure 6.8*B* shows Weber fractions for both

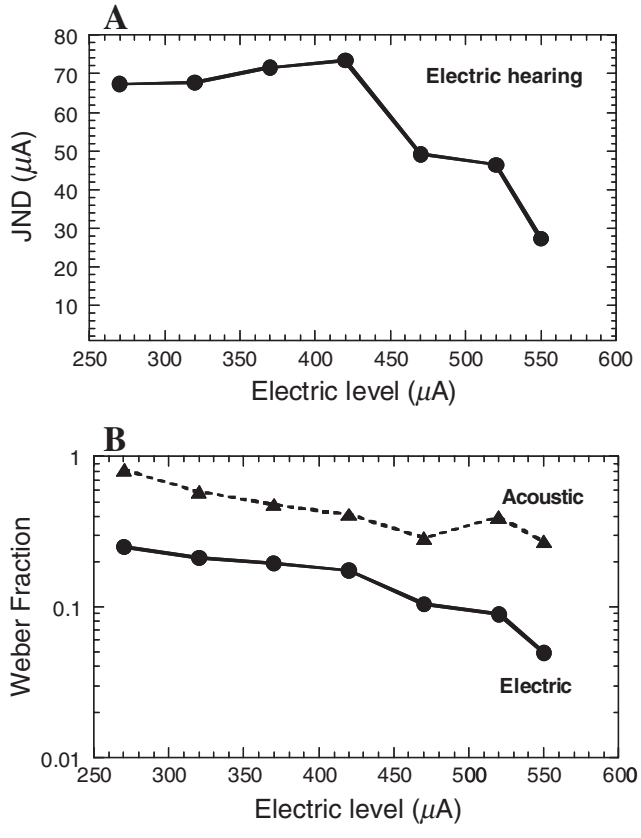


FIGURE 6.8. Intensity discrimination functions between acoustic and electric hearing in a listener who had normal hearing on 1 side and an ABI on the other side. *A*: just noticeable difference (JND) in absolute differences in electric hearing. *B*: JND in relative differences, i.e., the Weber fraction. For comparison, the acoustic Weber fractions measured from the same subject are plotted as a function of the electric currents (*x*-axis) that produced roughly the same loudness as the corresponding acoustic level (estimated from the loudness balance function shown in Fig. 6.6).

acoustic and electric hearing. The electric Weber fraction was obtained by dividing the values on the *y*-axis by their corresponding values on the *x*-axis in Figure 6.8*A*. The acoustic Weber fraction was calculated by converting the decibel difference (x) into the ratio between the pressure difference (ΔP) and the standard pressure (P), namely, $\Delta P/P = [10^{(x/20)}] - 1$. The acoustic Weber fraction is plotted as a function of the electric current that has produced roughly the same loudness as the corresponding acoustic level (estimated from the loudness balance function shown in Figure 6.6).

Figure 6.8*B* shows that the electric Weber fraction is almost an order of magnitude smaller than the acoustic Weber fraction throughout the entire

dynamic range (see also Nelson et al. 1996). As a matter of fact, we have often encountered cases where a cochlear implant user can reliably discriminate a loudness difference between 100 and 101 μA (e.g., Figs. 2 and 3 in Zeng and Shannon 1999). This 1% resolution is equivalent to a 0.08-dB level difference [20 log(1.01)], which is impossible to achieve with acoustic stimulation (the smallest difference is about 0.3–0.5 dB at 90–100 dB standard levels; e.g., Viemeister and Bacon 1988; Zeng et al. 1991). Thus, using the Weber fraction measure, we would reach the opposite conclusion that cochlear implant users have much better intensity discrimination abilities than normal-hearing listeners. The reason for reaching these apparent conflicting conclusions is largely due to the loss of cochlear compression in electric hearing (see Section 4.3). Although the electric JND is about one order of magnitude smaller than the acoustic JND, it does not necessarily result in a greater number of discriminable steps because the electric dynamic range is even more reduced by a factor of 4–5 orders of magnitude (20 vs. 100–120 dB).

3.2 Frequency Processing

3.2.1 Place Pitch

Because the cochlear compression is frequency selective, the sharp tuning that is normally observed in acoustic hearing is due in part to compression (Ruggero 1992). In electric hearing, the hope is to elicit tonotopic pitch perception by systematically stimulating the electrodes from apex to base in the cochlea. Unfortunately, despite equal physical distance between electrodes, pitch sensation evoked by these electrodes does not always conform to an orderly, perceptual relationship. Depending on electrode insertion depth, electrode configuration, nerve status, and other parameters, the pitch evoked by the same electrode array may have a different overall range and have a monotonic or nonmonotonic ("pitch reversal") relationship with the electrode array (Nelson et al. 1995; McKay et al. 1996; Collins et al. 1997; Busby and Clark 2000; Dawson et al. 2000). The spread of excitation as measured by forward-masking techniques has also been obtained in electric hearing (Lim et al. 1989; Cohen et al. 1996; Chatterjee and Shannon 1998). Although there is a greater spread of excitation from apical to basal electrodes, much like there is a greater spread of excitation from low to high frequencies in acoustic hearing (shown in Fig. 3.2 in Oxenham and Bacon, Chapter 3), no level dependence of the excitation pattern as would be observed as a consequence of nonlinear cochlear compression in acoustic hearing is observed in electric hearing.

3.2.2 Temporal Pitch

Because of better synchronization in electric hearing than acoustic hearing (Fig. 6.3), one might think that cochlear implant users can use timing cues

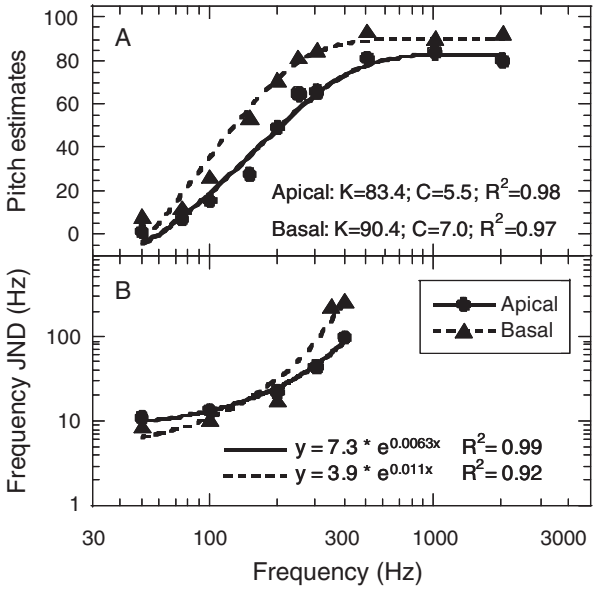


FIGURE 6.9. Pitch magnitude estimates (*A*) and JND (*B*) as a function of frequency in 1 Ineraid cochlear implant subject. Biphasic (100 or 200 ms/phase) pulses of different frequencies were used in the experiments. All stimuli were 300 ms in duration and delivered through a constant current source to the most apical (circles) or basal (triangles) electrode in a monopolar mode. Solid and dashed lines: predicted pitch functions from the frequency discrimination data (modeled as an exponential function; *B*). K, the saturation pitch value; C, constant perceptual difference in pitch (see Zeng 2002 for details).

to elicit more reliable and higher temporal pitch than normal-hearing listeners. Figure 6.9*A* shows pitch estimates as a function of frequency from both apical (circles) and basal (triangles) electrodes in one cochlear implant subject. Surprisingly, the pitch-frequency function saturates at roughly 300 Hz. The 300-Hz boundary has been consistently observed in all previous studies except for a few in which the implant users appeared to be able to use temporally based pitch up to 1 kHz (for a review, see Zeng 2002). Frequency JNDs (Fig. 6.9*B*) have also been measured to be relatively constant at 10–20 Hz for standard frequencies below 100 Hz but to increase drastically above that (see also Bilger 1977; Eddington et al. 1978; Shannon 1983; Townshend et al. 1987; Pijl and Schwarz 1995; McDermott and McKay 1997; Zeng 2002). These frequency JNDs are about one order of magnitude larger than the JNDs obtained with pure tones in normal listeners (Wier et al. 1977) but similar to those obtained with sinusoidally amplitude modulated noise in normal listeners (Formby 1985). These results suggest that the sharp frequency resolution observed in normal-hearing listeners is likely due to the place code rather than the temporal code.

3.3 Temporal Processing

The lack of cochlear compression significantly affects temporal processing in electric hearing. However, these effects are often masked in the literature. For example, gap detection and forward masking are perhaps the most frequently measured temporal processing tasks and have been found to be relatively normal in cochlear implant users (e.g., Bilger 1977; Eddington et al. 1978; Dobie and Dillier 1985; Preece and Tyler 1989; Shannon 1989b). We should be aware that these temporal measures are normal only after they are reorganized by some "normalized" intensity measures between acoustic and electric hearing (e.g., loudness or percent dynamic range). No one has explicitly studied whether the lost compression is the mechanism for the required "normalization" in forward masking in electric stimulation, but the lack of compression has been used to successfully account for the forward-masking data in cochlear-impaired listeners, particularly their slower recovery (Bacon and Oxenham, Chapter 4). Perhaps a temporal window model incorporating a compressive nonlinearity could reconcile the differences in forward masking observed between normal-hearing, cochlear-impaired, and cochlear implant users. Here three additional examples in temporal processing in which the lack of a compressive nonlinearity is clearly required and may not be readily masked by intensity normalization are discussed.

3.3.1 Temporal Integration

The first example is temporal integration. In acoustic hearing, for durations up to 100–200 ms, stimulus level can be traded almost linearly with its duration to achieve detection threshold. In other words, the slope of the temporal integration function [dB vs. log(time)] is about –3 dB per doubling of stimulus duration. It is unclear whether this represents, for example, one "look" at the output of a linear temporal integrator with a time constant of about 100–200 ms or multiple looks at the output of an integrator with a much shorter time constant (Oxenham and Bacon, Chapter 3). In electric hearing, temporal integration has been found to have a much shallower slope, albeit more variable than in acoustic hearing (Eddington et al. 1978; Shannon 1986; Pfingst and Morris 1993; Donaldson et al. 1997). Shannon (1986) found an average slope of –1.1 dB/doubling in 22 cases. Donaldson et al. (1997) found an average slope of –0.42/doubling, with a range from –0.06 to –1.94 dB/doubling, on 21 electrodes from 8 subjects. In addition, Donaldson et al. found that the slope of the temporal integration function tended to decrease with absolute threshold but increase with dynamic range. As seen in Section 4.3, this extremely shallow temporal integration function reflects mostly the lack of cochlear compression in cochlear implant users.

3.3.2 Amplitude Modulation Detection

The temporal modulation transfer function (TMTF) reveals a listener's ability to follow dynamic changes in amplitude and is represented by an amplitude modulation detection threshold [20 log(m)] as a function of modulation frequency (m is the modulation index). Figure 6.10 summarizes modulation detection from four different studies in acoustic and electric hearing. The bottom solid line in Figure 6.10 represents TMTF data obtained using a sinusoid to amplitude modulate a broadband noise carrier (Bacon and Viemeister 1985). The top solid line in Figure 6.10 represents the TMTF data obtained using a sinusoid to amplitude modulate a 5-kHz sinusoidal carrier (Kohlrausch et al. 2000). Amplitude modulation thresholds are also presented for cochlear implant users who detected modulation at a comfortable loudness level with either a pulsatile carrier (Fig. 6.10, open squares) or a sinusoidal carrier (Fig. 6.10, solid circles; from Shannon 1993).

All TMTF data show a low-pass characteristic function. The acoustic data with the noise carrier produce the lowest modulation sensitivity (–26 dB or

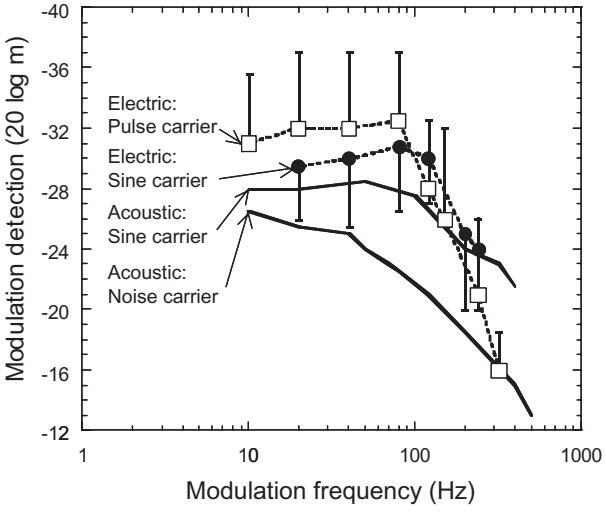


FIGURE 6.10. Temporal modulation transfer functions in acoustic hearing (noise and sinusoidal carriers) and in electric hearing (pulsatile and sinusoidal carriers). Open squares: data obtained using a sinusoid to amplitude modulate a pulse carrier in 7 Nucleus implant users; solid circles: data obtained using a sinusoid to amplitude modulate a sinusoidal carrier in 10 Ineraid implant users (Shannon 1993). Bottom solid line: acoustic temporal modulation transfer function (TMTF) data with the noise carrier (Bacon and Viemeister 1985); top solid line: TMTF with the 5-kHz sinusoidal carrier (averaged from Figs. 2 and 3 in Kohlrausch et al. 2000). Values are means ± SD.

5% modulation at 10 Hz) and 3-dB cutoff frequency (about 60 Hz), whereas the acoustic data with the sinusoidal carrier produce a somewhat higher sensitivity (–28 dB or 4% modulation) and cutoff frequency (about 170 Hz). The degraded performance with the noise carrier has been suggested to reflect the inherent fluctuations of the noise carrier, creating a modulation-masking situation (e.g., Kohlrausch et al. 2000; Moore and Glasberg 2001). Another possibility is that listeners may use an off-frequency listening cue, as evidenced by the presence of level-dependent modulation detection sensitivity with the sinusoidal carrier but the absence of it with the noise carrier. This off-frequency cue, particularly on the high-frequency side of the excitation pattern, tends to enhance the modulation due to the linear growth of excitation at these off frequencies.

The two sets of TMTF data in electric hearing have similar modulation sensitivities of about –30 dB (or 3% modulation) and cutoff frequencies at 150 Hz. These values are significantly better than the values obtained with the noise carrier but similar to those with the sinusoidal carrier in acoustic hearing. At least at an average level, these data suggest that it is more appropriate to use a sinusoid than a noise carrier to simulate modulation detection in electric hearing (which is not the case in the present simulation of cochlear implants; e.g., Shannon et al. 1995). On an individual basis, there is large variability with the modulation detection data in electric hearing, which has been shown to correlate with the productive use of temporal envelope cues in speech recognition (Cazals et al. 1994; Fu 2002). Moreover, the best modulation detection sensitivity reported in normal-hearing listeners was –37 dB (about 1.4% modulation; see Kohlrausch et al. 2000), but it is not unusual to find individual implant subjects who can detect 0.3–1.0% modulations (i.e., 20 log(m) = –50 to –40 dB; see Shannon 1992; Busby et al. 1993; Cazals et al. 1994; Chatterjee and Robert 2001; Fu 2002). This remarkable modulation sensitivity is at least partially due to the lack of cochlear compression and its perceptual consequences in electric stimulation.

3.3.3 Discrimination of Temporal Asymmetry

Patterson (1994a,b) produced “damped” and “ramped” sinusoids with asymmetrical temporal envelopes but identical long-term power spectra. The damped sinusoid is defined by three parameters: the sinusoidal carrier frequency, the exponentially damped temporal envelope (typically defined by the half-life time), and the repetition period of this exponentially damped envelope. The ramped sinusoid is simply the temporal reversal of the damped sinusoid. With an 800-Hz carrier frequency and 50-ms repetition period, Patterson found that normal-hearing listeners could typically discriminate the temporal asymmetry between damped and ramped sounds when the half-life of the envelope is between 1 and 50 ms. Under similar conditions, the cochlear implant users could discriminate between damped

and ramped sounds for half-lives as short as 0.5 ms and as long as 500 ms, an order of magnitude greater than the normal range (Lorenzi et al. 1997). Lorenzi et al. suggested that this superb performance in detecting temporal asymmetry in cochlear implant users is due to the lack of cochlear compression in electric stimulation. They further demonstrated this limiting role of cochlear compression in temporal processing by subjecting the implant listeners to listening to the damped and ramped sounds through their speech processors where there is an artificial amplitude compression circuit. As predicted, the implant users' ability to discriminate temporal asymmetry was greatly reduced to values that were within the normal range (1–24 ms; see Lorenzi et al. 1998).

4. Auditory Processing Revealed by Electric Hearing

Having reviewed selective behavioral responses to electric stimulation of the auditory nerve in cochlear implant users, what these psychoelectrical capabilities may tell us about the normal auditory processing mechanisms is now examined. Following the same outline as in Section 3, the auditory models inferred by electric hearing in intensity, frequency, and temporal processing are examined. Particular attention is paid to those mechanisms related to the lack of cochlear compression in cochlear implant users.

4.1 Intensity Processing

Here two issues in intensity processing are addressed. First, how does the auditory system encode the extremely large dynamic range in intensity? Second, can one derive the intensity-loudness function from the JND data in electric hearing? If so, what does it mean in terms of the role of the internal representation of intensity?

4.1.1 A Compression-Expansion Model for Loudness Coding

It has been known for a long time that loudness grows as a power function of intensity in acoustic hearing (Stevens 1961). Recent data have shown that loudness grows as an exponential function of intensity (or current) in electric hearing (see Section 3.1.2). Zeng and Shannon (1994) explored this difference in loudness growth between acoustic and electric hearing and proposed a general compression-expansion scheme for loudness coding in the auditory system. Figure 6.11 shows that the 100-dB dynamic range is first compressed in the cochlea to produce an output range of roughly 20 dB (Ruggero 1992). The narrowly compressed dynamic range is transmitted through the auditory nerve to the central auditory system and then expanded there to partially restore the large input range.

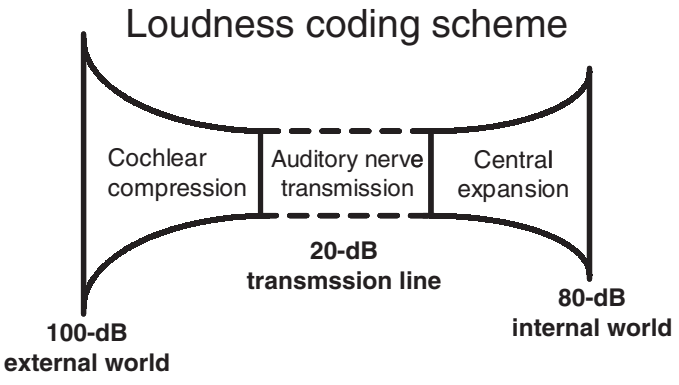


FIGURE 6.11. Compression-expansion model for loudness encoding. To solve the dynamic range problem between the large dynamic range in acoustic environments and the narrow dynamic range in neural transmission, the auditory system first compresses the acoustic dynamic range in the cochlea and then restores it by a central expansion. Similar coding schemes have been used in audio processing.

Mathematically, this compression-expansion coding for loudness can be approximately demonstrated by the following derivations (Fig. 6.12). Assume that N is the nerve output, then the cochlear compression can be approximated by

$$\mathrm{N} = a\log(\mathrm{I}). \tag{4.1.1.1}$$

The loudness is the product of the cochlear compression and the central expansion

$$\exp(\mathrm{N}) = \exp[a\log(\mathrm{I})] = \mathrm{I}^{a}. \tag{4.1.1.2}$$

We have thus obtained a power function describing the loudness and intensity relationship in acoustic hearing.

On the other hand, if the cochlear compression is bypassed as in the case of cochlear implants, then loudness is determined by only the central exponential process

$$\mathrm{L} = \exp(\mathrm{e}). \tag{4.1.1.3}$$

To further demonstrate this compression-expansion model of loudness coding, Zeng et al. (1998a) showed that one could obtain a good match to the power loudness function by exponentially expanding the measured compressive input-output function of the basilar membrane (e.g., Johnstone et al. 1986; Ruggero 1992). Zeng and Shannon (1994) suggested that this expansion occurs at the brainstem level. McGill and Teich (1995) further suggested a branching neural network as the explicit mechanism for expansion. On the other hand, Schlauch et al. (1998) showed that a similarly good

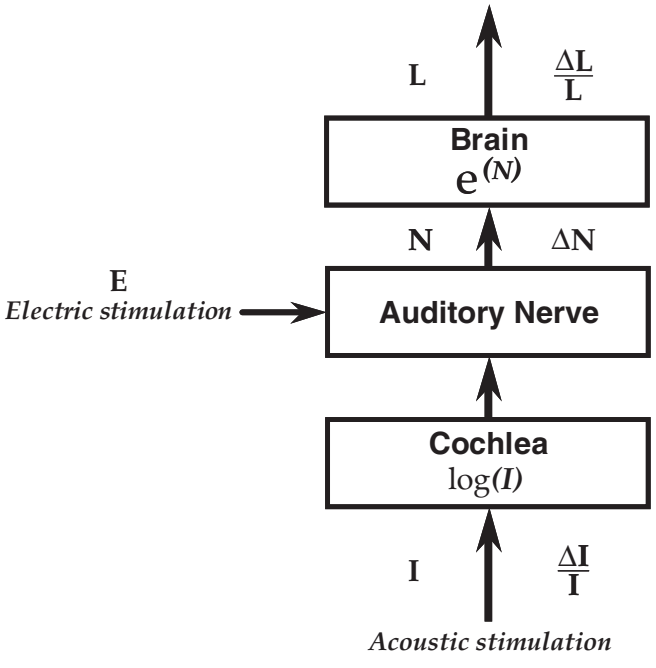


FIGURE 6.12. Unified model for intensity coding. On the left side of the block diagram, the sound intensity (I) is logarithmically compressed to drive the nerve, resulting in a neural discharge or count (N). This neural count is then exponentiated to result in a loudness percept (L). Such a compression-expansion model will give rise to a power loudness function in acoustic hearing. Bypassing the logarithmic compression, the electric stimulation results in the exponential loudness function. On the right side of the block diagram, the logarithmic compression and the exponential expansion transform from a constant relative difference in intensity ($\Delta I/I$) to a constant absolute difference in neural count (ΔN) and again to a constant relative difference in loudness ($\Delta L/L$).

match to the loudness function could be obtained by a presynaptic expansion mechanism: "squaring" the input-output function of the basilar membrane by the inner hair cells (for a further discussion of a square law nonlinearity, see Oxenham and Bacon, Chapter 3; Bacon and Oxenham, Chapter 4). Although the exact sites where expansion occurs are still debatable, note that this compression-expansion encoding scheme has been used frequently in audio engineering applications to overcome the narrow transmission channel problem (Furui 1988). It is not surprising that the sensory system has evolved with a similar mechanism for dealing with the dynamic range problem in the interface between the environment and the organism.

4.1.2 Loudness and the JND in Intensity

Another fundamental issue in psychophysics is whether the input-output function of the sensory system can be derived by its sensitivity measures. In this case, one would like to know whether the loudness function can be obtained by the intensity JND function. Fechner (1966) assumed that a stimulus JND (expressed as a Weber fraction or $\Delta I/I$) represents a constant unit in sensation ($\Delta L = I/I$). He integrated this equation and obtained his famous logarithmic law ($L = \log I$) in psychophysics. Although his logarithmic law was later replaced by a power law, his idea on the JND-loudness relationship still stimulates active research to this date (e.g., Houtsma et al. 1980; Zwislock and Jordan 1986; Viemeister and Bacon 1988; Schlauch et al. 1995; Allen and Neely 1997; Zeng and Shannon 1999; Hellman and Hellman 2001).

Zeng and Shannon (1999) formulated a unified framework that takes into account both the peripheral compression and central expansion and their relationship to the JND-loudness relationship (Figure 6.12). The peripheral compression converts a relative difference in intensity ($\Delta I/I$) into an absolute difference in neural count (ΔN), whereas the central expansion converts it back into a relative difference in sensation magnitude ($\Delta L/L$). The direct relationship between the relative differences in the stimulus and sensation domains has been called Brentano's law or Ekman's law (Ekman 1959; Stevens 1961).

In electric stimulation, neither Weber's law nor the near miss to Weber's law holds; rather, the absolute difference in microamperes or percent dynamic range is constant. Zeng and Shannon (1999) showed that the exponential loudness function can be directly derived by integrating the JND function in cochlear implant users. However, this simple relationship between the JND and loudness functions cannot be obtained in acoustic hearing, possibly due to the internal neural noise (i.e., spontaneous activity), multichannel listening, and compressive nonlinearity present in the normal-hearing listeners.

4.2 Frequency Processing

4.2.1 Pitch Models

Ruggero (1992) showed that cochlear compression and sharp frequency tuning are tightly coupled, and both probably reflect the nonlinear processing of the outer hair cells. With hearing impairment, the cochlea is more linear and both the compression and sharp tuning are reduced or lost (Bacon, Chapter 1; Cooper, Chapter 2; Bacon and Oxenham, Chapter 4). Electric stimulation of the auditory nerve in cochlear implant users reflects the extreme case of the loss of compression and sharp tuning. However, electric stimulation also provides an opportunity to test pitch models, which would be difficult, if not impossible, to do in acoustic stimulation.

One such example is to test the boundary and limit of temporal pitch in hearing. In acoustic stimulation, frequency changes always accompany changes in both the place of excitation and the phase locking of the nerve firing, making it difficult to assess the relative contribution of the place versus timing cues to pitch perception. The boundary of temporal pitch has been suggested to be at 1,500 Hz (Terhardt 1974) or 5,000 Hz, the frequency limit at which the auditory nerve stops to phase-lock to the stimulus (Johnson 1980). Burns and Viemeister (1976) used sinusoidally amplitude-modulated (SAM) noises to overcome this difficulty and found that only temporal cues could encode pitch up to about 500 Hz. However, the weak pitch the SAM noise evokes and the concern about the use of short-term spectral cues have made the results with SAM noise somewhat controversial and inconclusive for the boundary of temporal pitch.

With cochlear implants, the timing cue can be controlled totally independent of the place of stimulation. Surprisingly, virtually all studies of temporal pitch in cochlear implant users have shown that they cannot use temporal cues to form pitch perception for frequencies above 300–500 Hz (see Section 3.2.2). This psychophysical inability to encode temporal pitch above 300–500 Hz is in sharp contrast to the ability of the auditory nerve to phase-lock to electric stimuli up to 10,000 Hz (Fig. 6.3). This drastic difference between the psychophysical and physiological abilities in encoding temporal pitch in cochlear implant users may reflect the fact that the firing pattern due to electric stimulation is too artificially synchronous and the nerve is not totally healthy in the implant ears. However, the evidence from electric stimulation and the SAM noise in acoustic hearing strongly suggest that the brain may not use the timing cue above 300–500 Hz to encode pitch.

4.2.2 Pitch and the Frequency JND

Similar to the simple intensity JND-loudness relationship found in electric hearing, temporal pitch can be easily and reliably derived from frequency JND measures. Figure 6.9 shows that a simple integration of the frequency JND function can predict accurately the saturating temporal pitch function (Zeng 2002). The close couplings between the frequency JND and pitch and between the intensity JND and loudness in electric stimulation show that, without cochlear compression, the relationship between stimulus and sensation can be easily revealed. With the presence of the nonlinear compression, such a relationship is much more difficult to derive and describe in acoustic hearing. To test the generality of Fechner's (1966) classic hypothesis relating stimulus discriminability to sensation magnitude, we need to recognize and isolate the cochlear compression in this relationship.

4.3 Temporal Processing

The significant role of cochlear compression in temporal processing is not always apparent in traditional psychoacoustical studies because most tem-

poral models do not incorporate cochlear compression. By tweaking other parameters such as filter width, postfilter nonlinearity (rectification vs. power law), and decision variables and rules, these temporal models without compression can account for basic temporal phenomena such as temporal integration, gap detection, modulation detection, and forward masking (e.g., Viemeister 1979; Forrest and Green 1987; Moore et al. 1988; Oxenham and Plack 2000). More recently, however, compression has been implemented in the context of a temporal window model (see Fig. 3.5 in Oxenham and Bacon, Chapter 3). This model can account for a wide range of perceptual phenomena, including many aspects of temporal processing, as discussed in detail by Oxenham and Bacon (Chapter 3) and Bacon and Oxenham (Chapter 4). Here how the implant data (see Section 3.3) can reveal the role of compression in temporal processing is discussed.

First, the extremely shallow slope of the temporal integration function as observed in cochlear implant users can be readily modeled as a consequence of the loss of compression in these listeners. Yates et al. (1990) demonstrated that in acoustic hearing, the varieties and dynamic ranges of the rate-intensity functions in the auditory nerve reflect basilar membrane compression (Cooper, Chapter 2). Without the compression, the auditory nerve has a more uniform and steeper rate-intensity function in electric stimulation than acoustic stimulation. The steep rate-intensity function in electric hearing will result in a steep psychometric function as shown by both the experimental data (Donaldson et al. 1997) and a stochastic model of the electrically stimulated auditory nerve (Bruce et al. 1999a,b). Although Donaldson et al. (1997) used a multiple-looks model with a roughly 10-ms window to explain their data, Bruce et al. (1999c) found that the traditional "single-look" 100-ms integrator could also successfully explain the same data, suggesting that the lack of compression in electric hearing rather than the decision variable and rule is the critical factor determining the temporal integration function (see Bacon and Oxenham, Chapter 4, for a discussion of how a decrease in or loss of compression can influence temporal integration in acoustically stimulated hearing-impaired listeners).

Second, the lack of compression and filter ringing in electric stimulation can also account for exquisite sensitivity in intensity discrimination and detection of amplitude modulation and temporal asymmetry in cochlear implant users. A simple and parsimonious model detects the difference between peak and valley intensities. For example, implant listeners can detect amplitude modulations on the order of 1–5%, translating into 2–10% differences between the peak and the valley amplitudes [(minimum/maximum = (1 − m)/(1 + m)]. Similarly, for temporal asymmetry with a 500-ms half-life and 50-ms repetition period, the implant listener can detect about a 7% difference in amplitude between the beginning and the end of the temporal envelope [exp(–0.69*50/500)]. These differences in amplitude are consistent with the intensity discrimination directly mea-

sured in implant listeners at a comfortable loudness (i.e., Weber's fractions in the 1–10% range; see Nelson et al. 1996; Zeng and Shannon 1999; also Fig. 6.8).

Without evoking changes in the internal noise and other central decision variables in electric stimulation, the loss of cochlear compression alone can account for most of the difference in Weber's fraction between acoustic and electric stimulation. Except for the acoustic TMTF data with the sinusoidal carrier, normal-hearing listeners typically can detect intensity differences from 0.5 to 3 dB, translating into about a 6–40% difference in amplitude. This precompression difference, subject to a power function with an exponent of 0.3 (approximating the cochlear compression function; see Cooper, Chapter 2), is reduced to a postcompression difference of 2–11%, similar to the values typically found in cochlear implant users.

5. Practical Issues

Cochlear compression, or the lack of it with electric stimulation of the auditory nerve, significantly affects the design and performance of a cochlear implant. As seen in Section 3.1.1, the direct consequence of the lost compression is the extremely narrow dynamic range of 10–20 dB in cochlear implant users. The foremost important issue in the cochlear implant design is to fit speech sounds into this narrow dynamic range.

5.1 Automatic Gain Control and Instantaneous Compression

A normal-hearing listener can accommodate an extremely wide dynamic range of about 120 dB or an intensity change of over 12 orders of magnitude (Bacon, Chapter 1). The dynamic range of speech varies between 30 and 60 dB depending on the speech material, the acoustic measure [root mean square (rms) or envelope levels], and the definition of the speech dynamic range (for a review, see Zeng et al. 2002). All implants have an automatic gain control that adjusts the microphone sensitivity so that speech sounds, whether soft or loud, near or distant, can be optimally amplified to fit into a fixed 30–60 dB electric range that is compressed further to match the individual user's electric dynamic range. The automatic gain control tends to have a fast attack time (a few milliseconds) and a relatively slow release time (tens to hundreds of milliseconds; Levitt, Chapter 5).

On the other hand, the compression is instantaneous and typically accomplished by a lookup table in a digital implementation of the acoustic-to-electric amplitude-mapping process. Although the goal of this instantaneous compression is necessary because of the apparent mismatch in dynamic range between the environmental sounds and the electric stimu-

lation, it effectively partially recovers the function of the lost cochlear compression in cochlear implant users.

5.2 Restoring Normal Loudness Growth

One goal of the compression in cochlear implants is to restore normal loudness growth. Several researchers have shown a linear loudness-matching function between acoustic amplitude expressed in decibels and electric amplitude expressed in microamperes (Eddington et al. 1978; Zeng and Shannon 1992; Dorman et al. 1993; Fig. 6.6*B*). This linear function can be represented by

$$(20\log A - 20\log Ao)/(20\log Au - 20\log Ao) = a(E - T)/(U - T) + b \tag{5.2.1}$$

where Ao is acoustic threshold, Au is acoustic uncomfortable level, E is linear electric amplitude, T is electric threshold, U is electric uncomfortable level, and a and b are constants.

To restore normal loudness growth, the electric amplitude should then be determined by a logarithmic mapping function; in other words, a logarithmic compression from acoustic amplitude to electric amplitude is needed

$$E = 20\log(A/Ao)/IDR*(U - T)/a - b \tag{5.2.2}$$

where IDR is the input acoustic dynamic range. Applying the boundary conditions where E equals T when A equals Ao and E equals U when A equals Au, we have

$$E = 20\log(A/Ao)/IDR*(U - T) + T \tag{5.2.3}$$

Equation 5.2.3 shows that the ratio between electric dynamic range (U – T) and the IDR is a scaling factor, whereas T can be treated as a DC shift. Figure 6.13 illustrates such a mapping between IDR and electric dynamic range in cochlear implants. The x-axis (i.e., the IDR) determines the range of acoustic input mapped into the electric output range between threshold (T level) and the most comfortable loudness (M level). The speech processor first selects an acoustic level (0 dB on the x-axis) and maps it into an electric level (M level) that evokes the most comfortable loudness. The speech processor then maps either the 10-dB range below the 0-dB acoustic level into the audible electric dynamic range (the rightmost sloping line) or any other acoustic range into the same audible electric dynamic range. Presumably, any acoustic input level that is outside the IDR will be mapped into either a subthreshold electric level (less than T level) or a constant saturating level (greater then M level).

5.3 Speech Recognition

The choice of the acoustic dynamic range, the electric dynamic range, and the conversion from acoustic amplitude to electric amplitude can signifi-

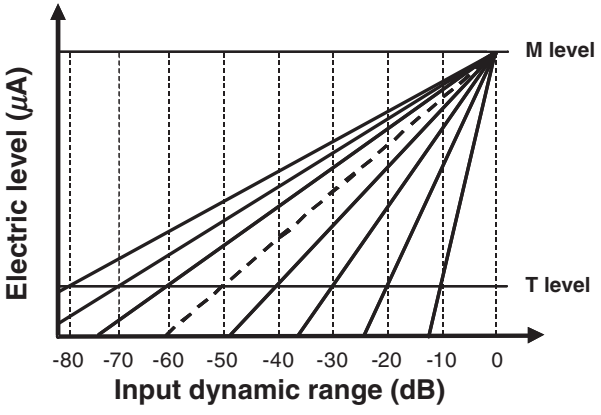


FIGURE 6.13. Logarithmic compression for acoustic-to-electric amplitude conversion in Clarion cochlear implants. The *x*-axis (i.e., the input dynamic range) determines the range of acoustic input mapped into the electric output range between threshold (T level) and the most comfortable loudness (M level). The speech processor first selects an acoustic level (0 dB on the *x*-axis) and maps it into an electric level (M level) that evokes the most comfortable loudness. The speech processor then maps either the 10-dB range below the 0-dB acoustic level into the audible electric dynamic range (the rightmost sloping line) or any other acoustic range into the same audible electric dynamic range. A 60-dB dynamic range map (dashed line) is typically set in the speech processor.

cantly affect speech performance in cochlear implant users. This section will examine each of these three factors in determining the cochlear implant performance and relate it to the psychophysical and theoretical aspects of auditory compression whenever appropriate.

5.3.1 Effect of IDR

Ideally, the IDR would be set to 120 dB so that the acoustic amplitude within this normal range is converted into a current value that evokes sensation between minimal and maximal loudness. Because of the narrow electric dynamic range and the limited discriminable steps within the range (about 10–20 dB dynamic range and 20 discriminable steps; see Nelson et al. 1996; Skinner et al. 1997b; Zeng et al. 1998b; Zeng and Galvin 1999; Zeng et al. 2002), the implant users might not be able to discern enough meaningful variations in sound intensity for the most important speech sound. Traditionally, speech dynamic range has been assumed to be 30 dB (e.g., Dunn and White 1940; ANSI 1969, 1997). Some of the earlier cochlear implants (e.g., Nucleus 22) have also set their IDR to this 30-dB value.

Using the standard logarithmic compression, Zeng et al. (2002) systematically examined the effect of adjusting the IDR on speech performance

in cochlear implant users. For consonants and vowels produced by multiple talkers, speech performance was optimal when the IDR was set at about 50 dB. The performance dropped off when the IDR was either decreased or increased. The optimal performance with a 50-dB IDR was consistent with the acoustically measured dynamic range for the temporal envelope distribution in these speech materials (Cosendai and Pelizzone 2001; Zeng et al. 2002).

5.3.2 Effect of Electric Dynamic Range

Clinicians spend most of the time during the programming of a cochlear implant trying to estimate the electric dynamic range as determined by the threshold and the maximum comfortable level. The threshold is usually called the T level, whereas the maximum comfortable level is called the M level or C level. Of course, we want to make sure that the acoustic information is mapped appropriately within the electric dynamic range; however, it is not clear how important an accurate measurement of this range needs to be to maintain a high level of speech performance in cochlear implant users.

Dawson et al. (1997) simulated errors in estimating the electric dynamic range by quasi-randomly reducing the maximum comfortable level by 20% from the actually measured values. They found a significant effect of this loudness imbalance on speech recognition, particularly in noise. Others (Skinner et al. 1997a; Zeng and Galvin 1999; Loizou et al. 2000a) also found similar effects of reducing the dynamic range on speech recognition, particularly again for vowels and in noise.

5.3.3 Effect of Compression

Several studies have systematically examined the effect of compression on speech performance in cochlear implant users (Fu and Shannon 1998, 2000; Zeng and Galvin 1999; Loizou et al. 2000b). In these studies, the degree of compression was varied from generally little compression to the steepest compression mimicking a step function. Although little or no compression generally produced the worst speech performance, the degree of compression had a relatively small effect on speech performance and depended on speech-processing strategies (Fig. 6.14). For example, Zeng and Galvin (1999) used the most compressive conversion from acoustic amplitude to electric amplitude, a step function that essentially converts variations of acoustic amplitude into a binary representation in electric amplitude. They found in cochlear implant users who used the SPEAK strategy that this extreme form of compression had relatively little effect on phoneme recognition in quiet. On the other hand, Fu and Shannon (1998) and Loizou et al. (2000b) showed in cochlear implant users using the CIS strategy that

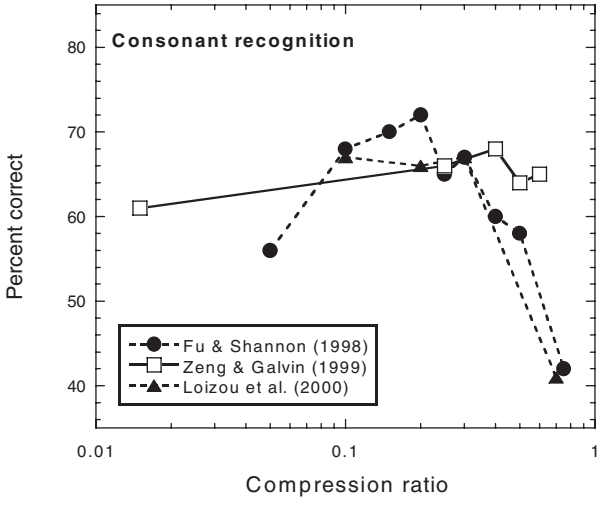


FIGURE 6.14. Consonant recognition as a function of compression ratios in cochlear implant listeners. The compression ratio is defined as the exponent (p) of a power function [$E = Eo + k(A - Ao)^p$], where Eo is the electric threshold, Ao is the minimal acoustic value to be mapped, and k is a constant. A logarithmic compression can be approximated by setting p to 0.3. The steplike compression in Zeng and Galvin (1999) is approximated by a compression ratio of 0.015 (the leftmost square).

the best speech performance was achieved when the compression restored normal loudness in cochlear implant users (with the compression ratio between 0.1 and 0.3). Additional benefits could also be obtained by selectively amplifying the low-value acoustic amplitudes in speech sounds (Geurts and Wouters 1999).

There is no question that compression needs to be restored by cochlear implants, but the question still remains regarding the form of the most effective compression. The preliminary data presented above show only a tip of the iceberg because, in reality, the best compression is not an isolated parameter but rather is one that interacts with other factors such as the listening environment (quiet vs. noise, speech vs. music), the processing strategies (temporal envelope vs. spectral features), and the electrode interactions (single- vs. multiple-electrode stimulation). For example, the popular strategy of extracting and coding the temporal envelope (Wilson et al. 1991) is more sensitive to the degree of compression than strategies of extracting the spectral features (Zeng and Galvin 1999). Ultimately, the compression has to be combined with the electrode-to-neuron interface to restore not only the normal loudness growth but, more importantly, the number of discriminable intensity steps as well as the range and resolution in frequency.

6. Summary

Using the cochlear implant to directly stimulate the auditory nerve has been proven to be an effective means of restoring partial hearing to deaf people. With today's modern multielectrode devices, an average implant user can talk on the telephone. We have begun to explore the utility and potential of the implant in probing basic auditory mechanisms, including cochlear compression. Overall, the basic research in cochlear implants is still in its infancy. It still remains a dream to incorporate compression in the cochlear implant to restore not only the loudness growth but also the number of discriminable steps in intensity, frequency tuning, and natural nerve response. This chapter has shown that the cochlear implant can be used as a powerful research tool to demonstrate the roles of cochlear compression in auditory functions.

(1) Cochlear compression plays an extremely important role in encoding the 120-dB dynamic range in acoustic hearing. Without this compression, the dynamic range is drastically reduced to 10–30 dB in electric hearing. The implant data suggest a general loudness-coding scheme consisting of a peripheral compression and a central expansion.

(2) Comparative studies between acoustic and electric hearing suggest that cochlear compression limits normal-hearing listeners' ability to discriminate differences in intensity and time (e.g., fast amplitude modulations and asymmetrical temporal changes). In this regard, the cochlear implant, sometimes referred to as the bionic ear, truly outperforms its natural counterpart.

(3) Employment of a nonlinear compression, similar to the natural cochlear compression, is important to restore normal loudness growth and improve perceptual performance in cochlear implant listeners. The exact form of compression and its interactions with processing, patient, and environmental factors still remain unclear.

Acknowledgments. I thank Ian Bruce, Gail Donaldson, Christian Lorenzi, and Roy Patterson for discussion on the role of compression in temporal processing, and Ackland Jones, Arthur Popper, and Sid Bacon for helpful comments on an earlier version of this chapter. Preparation of this manuscript was partially supported by National Institute on Deafness and Other Communication Disorders Grant R01-DC-02267-08.

References

Allen JB, Neely ST (1997) Modeling the relation between the intensity just-noticeable difference andloudness for pure tones and wideband noise. J Acoust Soc Am 102:3628–3646.

ANSI (1969) American National Standards Methods for the Calculation of the Articulation Index. ANSI S3.5-1969. New York: American National Standards Institute.

ANSI (1997) American National Standards Methods for the Calculation of the Speech Intelligibility Index. ANSI S3.5-1997. New York: American National Standards Institute.

Bacon SP, Viemeister NF (1985) Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. Audiology 24:117–134.

Bilger RC (1977) Psychoacoustic evaluation of present prostheses. Arch Otol Rhinol Laryngol 86:92–104.

Brackmann DE, Hitselberger WE, Nelson RA, Moore J, Waring MD, Portillo F, Shannon RV, Telischi FF (1993) Auditory brainstem implant. I. Issues in surgical implantation. Otolaryngol Head Neck Surg 108:624–633.

Bruce IC, Irlicht LS, White MW, O'Leary SJ, Dynes S, Javel E, Clark GM (1999a) A stochastic model of the electrically stimulated auditory nerve: pulse-train response. IEEE Trans Biomed Eng 46:630–637.

Bruce IC, White MW, Irlicht LS, O'Leary SJ, Dynes S, Javel E, Clark GM (1999b) A stochastic model of the electrically stimulated auditory nerve: single-pulse response. IEEE Trans Biomed Eng 46:617–629.

Bruce IC, White MW, Irlicht LS, O'Leary SJ, Clark GM (1999c) The effects of stochastic neural activity in a model predicting intensity perception with cochlear implants: low-rate stimulation. IEEE Trans Biomed Eng 46:1393–1404.

Burns EM, Viemeister NF (1976) Nonspectral pitch. J Acoust Soc Am 60:863–869.

Busby PA, Clark GM (2000) Pitch estimation by early-deafened subjects using a multiple-electrode cochlear implant. J Acoust Soc Am 107:547–558.

Busby PA, Tong YC, Clark GM (1993) The perception of temporal modulations by cochlear implant patients. J Acoust Soc Am 94:124–131.

Cazals Y, Pelizzone M, Saudan O, Boex C (1994) Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants. J Acoust Soc Am 96:2048–2054.

Chatterjee M (1999) Effects of stimulation mode on threshold and loudness growth in multielectrode cochlear implants. J Acoust Soc Am 105:850–860.

Chatterjee M, Robert ME (2001) Noise enhances modulation sensitivity in cochlear implant listeners: stochastic resonance in a prosthetic sensory system? J Assoc Res Otolaryngol 2:159–171.

Chatterjee M, Shannon RV (1998) Forward masked excitation patterns in multielectrode electrical stimulation. J Acoust Soc Am 103:2565–2572.

Chatterjee M, Fu QJ, Shannon RV (2000) Effects of phase duration and electrode separation on loudness growth in cochlear implant listeners. J Acoust Soc Am 107:1637–1644.

Cohen LT, Busby PA, Whitford LA, Clark GM (1996) Cochlear implant place psychophysics 1. Pitch estimation with deeply inserted electrodes. Audiol Neurootol 1:265–277.

Collins LM, Zwolan TA, Wakefield GH (1997) Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects. J Acoust Soc Am 101:440–455.

Cosendai G, Pelizzone M (2001) Effects of the acoustical dynamic range on speech recognition with cochlear implants. Audiology 40:272–281.

Dawson PW, Skok M, Clark GM (1997) The effect of loudness imbalance between electrodes in cochlear implant users. Ear Hear 18:156–165.
Dawson PW, McKay CM, Busby PA, Grayden DB, Clark GM (2000) Electrode discrimination and speech perception in young children using cochlear implants. Ear Hear 21:597–607.
Dobie RA, Dillier N (1985) Some aspects of temporal coding for single-channel electrical stimulation of the cochlea. Hear Res 18:41–55.
Donaldson GS, Viemeister NF, Nelson DA (1997) Psychometric functions and temporal integration in electric hearing. J Acoust Soc Am 101:3706–3721.
Dorman MF, Smith L, Parkin JL (1993) Loudness balance between acoustic and electric stimulation by a patient with a multichannel cochlear implant. Ear Hear 14:290–292.
Dunn HK, White SD (1940) Statistical measurements on conversational speech. J Acoust Soc Am 11:278–288.
Dynes SB, Delgutte B (1992) Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea. Hear Res 58:79–90.
Eddington DK, Dobelle WH, Brackmann DE, Mladejovsky MG, Parkin JL (1978) Auditory prostheses research with multiple channel intracochlear stimulation in man. Arch Otol Rhinol Laryngol 87:1–39.
Ekman G (1959) Weber's law and related functions. J Psychol 47:343–352.
Fechner GT (1966) Elements of Psychophysics, Vol. I. Helmut E. Adler (trans). New York: Holt, Rinehart and Winston.
Formby C (1985) Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners. J Acoust Soc Am 78:70–77.
Forrest TG, Green DM (1987) Detection of partially filled gaps in noise and the temporal modulation transfer function. J Acoust Soc Am 82:1933–1943.
Fu QJ (2002) Temporal processing and speech recognition in cochlear implant users. Neuroreport 16:1635–1639.
Fu QJ, Shannon RV (1998) Effects of amplitude non-linearity on phoneme recognition by cochlear implant users and normal-hearing listeners. J Acoust Soc Am 105:2570–2577.
Fu QJ, Shannon RV (2000) Effects of dynamic range and amplitude mapping on phoneme recognition in Nucleus-22 cochlear implant users. Ear Hear 21:227–235.
Furui S (1988) Digital speech processing, synthesis, and recognition. New York: Marcel Dekker, pp. 139–204.
Geurts L, Wouters J (1999) Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants. J Acoust Soc Am 105: 2476–2484.
Hartmann R, Klinke R (1987) Impulse pattern in auditory-nerve fibers to extra- and intra-cochlear electrical stimulation. In: Banfai P (ed) Cochlear Implants: Current Situation, International Cochlear Implant Symposium. Duren, Germany, pp. 73–86.
Hartmann R, Topp G, Klinke R (1984) Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea. Hear Res 13:47–62.
Hellman WS, Hellman RP (2001) Revisiting relations between loudness and intensity discrimination. J Acoust Soc Am 109:2098–2102.

Houtsma AJ, Durlach NI, Braida LD (1980) Intensity perception. XI. Experimental results on the relation of intensity resolution to loudness matching. J Acoust Soc Am 68:807–813.

Javel E, Tong YC, Shepherd RK, Clark GM (1987) Responses of cat auditory nerve fibers to biphasic electrical current pulses. Arch Otol Rhinol Laryngol 96:26–30.

Johnson DH (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. J Acoust Soc Am 68:1115–1122.

Johnstone BM, Patuzzi R, Yates GK (1986) Basilar membrane measurements and the traveling wave. Hear Res 22:147–153.

Kiang NY, Moxon EC (1972) Physiological considerations in artificial stimulation of the inner ear. Arch Otol Rhinol Laryngol 81:714–730.

Kohlrausch A, Fassel R, Dau T (2000) The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. J Acoust Soc Am 108:723–734.

Liberman MC (1978) Auditory-nerve response from cats raised in a low-noise chamber. J Acoust Soc Am 63:442–455.

Lim HH, Tong YC, Clark GM (1989) Forward masking patterns produced by intracochlear electrical stimulation of one and two electrode pairs in the human cochlea. J Acoust Soc Am 86:971–980.

Litvak L, Delgutte B, Eddington D (2001) Auditory nerve fiber responses to electric stimulation: modulated and unmodulated pulse trains. J Acoust Soc Am 110: 368–379.

Loizou PC, Dorman M, Fitzke J (2000a) The effect of reduced dynamic range on speech understanding: implications for patients with cochlear implants. Ear Hear 21:25–31.

Loizou PC, Poroy O, Dorman M (2000b) The effect of parametric variations of cochlear implant processors on speech understanding. J Acoust Soc Am 108: 790–802.

Lorenzi C, Gallego S, Patterson RD (1997) Discrimination of temporal asymmetry in cochlear implantees. J Acoust Soc Am 102:482–485.

Lorenzi C, Gallego S, Patterson RD (1998) Amplitude compression in cochlear implants artificially restricts the perception of temporal asymmetry. Br J Audiol 32:367–374.

McDermott HJ, McKay CM (1997) Musical pitch perception with electrical stimulation of the cochlea. J Acoust Soc Am 101:1622–1631.

McDermott HJ, McKay CM, Vandali AE (1992) A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. J Acoust Soc Am 91:3367–3371.

McGill WJ, Teich MC (1995) Alerting signals and detection in a sensory network. J Math Psychol 39:146–162.

McKay CM, McDermott HJ, Clark GM (1996) The perceptual dimensions of single-electrode and nonsimultaneous dual-electrode stimuli in cochlear implantees. J Acoust Soc Am 99:1079–1090.

McKay CM, Remine MD, McDermott HJ (2001) Loudness summation for pulsatile electrical stimulation of the cochlea: effects of rate, electrode separation, level, and mode of stimulation. J Acoust Soc Am 110:1514–1524.

Miller CA, Abbas PJ, Robinson BK, Rubinstein JT, Matsuoka AJ (1999) Electrically evoked single-fiber action potentials from cat: responses to monopolar, monophasic stimulation. Hear Res 130:197–218.

Moore BC, Glasberg BR (2001) Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners. J Acoust Soc Am 110:1067–1073.

Moore BC, Glasberg BR, Plack CJ, Biswas AK (1988) The shape of the ear's temporal window. J Acoust Soc Am 83:1102–1116.

Nelson DA, Van Tasell DJ, Schroder AC, Soli S, Levine S (1995) Electrode ranking of "place pitch" and speech recognition in electrical hearing. J Acoust Soc Am 98:1987–1999.

Nelson DA, Schmitz JL, Donaldson GS, Viemeister NF, Javel E (1996) Intensity discrimination as a function of stimulus level with electric stimulation. J Acoust Soc Am 100:2393–2414.

NIH Consensus Statement (1995) Cochlear implants in adults and children. Vol. 13, No. 2, pp. 1–30.

Otto SR, Brackmann DE, Hitselberger WE, Shannon RV, Kuchta J (2002) Multichannel auditory brainstem implant: update on performance in 61 patients. J Neurosurg 96:1063–1071.

Oxenham AJ, Plack CJ (2000) Effects of masker frequency and duration in forward masking: further evidence for the influence of peripheral nonlinearity. Hear Res 150:258–266.

Parkins CW, Colombo J (1987) Auditory-nerve single-neuron thresholds to electrical stimulation from scala tympani electrodes. Hear Res 31:267–285.

Patterson RD (1994a) The sound of a sinusoid: spectral models. J Acoust Soc Am 96:1409–1418.

Patterson RD (1994b) The sound of a sinusoid: time-interval models. J Acoust Soc Am 96:1419–1428.

Pfingst BE, Morris DJ (1993) Stimulus features affecting psychophysical detection thresholds for electrical stimulation of the cochlea. II. Frequency and interpulse interval. J Acoust Soc Am 94:1287–1294.

Pijl S, Schwarz DW (1995) Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes. J Acoust Soc Am 98:886–895.

Preece JP, Tyler RS (1989) Temporal-gap detection by cochlear prosthesis users. J Speech Hear Res 32:849–856.

Rabinowitz WM, Lim JS, Braida LD, Durlach NI (1976) Intensity perception. VI. Summary of recent data on deviations from Weber's law for 1000-Hz tone pulses. J Acoust Soc Am 59:1506–1509.

Ruggero MA (1992) Responses to sound of the basilar membrane of the mammalian cochlea. Curr Opin Neurobiol 2:449–456.

Sachs MB, Abbas PJ (1974) Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. J Acoust Soc Am 56:1835–1847.

Scharf B, Magnan J, Chays A (1997) On the role of the olivocochlear bundle in hearing: 16 case studies. Hear Res 103:101–122.

Schlauch RS, Harvey S, Lanthier N (1995) Intensity resolution and loudness in broadband noise. J Acoust Soc Am 98:1895–1902.

Schlauch RS, DiGiovanni JJ, Ries DT (1998) Basilar membrane nonlinearity and loudness. J Acoust Soc Am 103:2010–2020.

Schroder AC, Viemeister NF, Nelson DA (1994) Intensity discrimination in normal-hearing and hearing-impaired listeners. J Acoust Soc Am 96:2683–2693.

Shannon RV (1983) Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. Hear Res 11:157–189.
Shannon RV (1986) Temporal processing in cochlear implants. In: Collins MJ, Glattke TJ, Harker LA (eds) Sensorineural Hearing Loss: Mechanisms, Diagnosis, Treatment. Iowa City, IA: University of Iowa Press, pp. 349–368.
Shannon RV (1989a) A model of threshold for pulsatile electrical stimulation of cochlear implants. Hear Res 40:197–204.
Shannon RV (1989b) Detection of gaps in sinusoids and pulse trains by patients with cochlear implants. J Acoust Soc Am 85:2587–2592.
Shannon RV (1992) Temporal modulation transfer functions in patients with cochlear implants. J Acoust Soc Am 91:2156–2164.
Shannon RV (1993) Psychophysics of electrical stimulation. In: Tyler RS (Ed) Cochlear Implants: Audiological Foundations. San Diego, CA: Singular Publishing Group Inc, pp. 357–388.
Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304.
Skinner MW, Holden LK, Holden TA (1997a) Parameter selection to optimize speech recognition with the Nucleus implant. Otolaryngol Head Neck Surg 117: 188–195.
Skinner MW, Holden LK, Holden TA, Demorest ME, Fourakis MS (1997b) Speech recognition at simulated soft, conversational and raised-to-loud vocal efforts by adults with cochlear implants. J Acoust Soc Am 101:3766–3782.
Stevens S (1961) To honor Fechner and repeal his law. Science 133:80–86.
Svirsky MA, Robbins AM, Kirk KI, Pisoni DB, Miyamoto RT (2000) Language development in profoundly deaf children with cochlear implants. Psychol Sci 11: 153–158.
Terhardt E (1974) Pitch, consonance, and harmony. J Acoust Soc Am 55:1061–1069.
Townshend B, Cotter N, Van Compernolle D, White RL (1987) Pitch perception by cochlear implant subjects. J Acoust Soc Am 82:106–115.
van den Honert C, Stypulkowski PH (1984) Physiological properties of the electrically stimulated auditory nerve. II. Single fiber recordings. Hear Res 14:225–243.
Viemeister NF (1979) Temporal modulation transfer functions based upon modulation thresholds. J Acoust Soc Am 66:1364–1380.
Viemeister NF, Bacon SP (1988) Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones. J Acoust Soc Am 84:172–178.
Volta A (1800) On the electricity excited by mere contact of conducting substances of different kinds. R Soc Philos Trans 90:403–431.
Wier CC, Jesteadt W, Green DM (1977) Frequency discrimination as a function of frequency and sensation level. J Acoust Soc Am 61:178–184.
Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM (1991) Better speech recognition with cochlear implants. Nature 352:236–238.
Yates GK, Winter IM, Robertson D (1990) Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range. Hear Res 45:203–219.
Yates GK, Johnstone BM, Patuzzi RB, Robertson D (1992) Mechanical preprocessing in the mammalian cochlea. Trends Neurosci 15:57–61.
Zeng FG (2002) Temporal pitch in electric hearing. Hear Res 174:101–106.
Zeng FG, Galvin JJ (1999) Amplitude compression and phoneme recognition in cochlear implant listeners. Ear Hear 20:60–74.

Zeng FG, Shannon RV (1992) Loudness balance between acoustically and electrically stimulated ears. Hear Res 60:231–235.
Zeng FG, Shannon RV (1994) Loudness-coding mechanisms inferred from electric stimulation of the human auditory system. Science 264:564–566.
Zeng, FG, Shannon RV (1995) Loudness of simple and complex stimuli in electric hearing. Ann Otol Rhinol Laryngol 104 (Suppl 166): 235–238.
Zeng FG, Shannon RV (1999) Psychophysical laws revealed by electric hearing. Neuroreport 10:1931–1935.
Zeng FG, Turner CW, Relkin EM (1991) Recovery from prior stimulation. II. Effects upon intensity discrimination. Hear Res 55:223–230.
Zeng FG, Shannon RV, Hellman WS (1998a) Physiological processes underlying psychophysical laws. In: Palmer A, Rees A, Summerfield AQ, Meddis R (eds) Psychophysical and Physiological Advances in Hearing. London: Whurr Publishers, pp. 473–481.
Zeng FG, Galvin JJ, Zhang CY (1998b) Encoding loudness by electric stimulation of the auditory nerve. Neuroreport 9:1845–1848.
Zeng FG, Martino KM, Linthicum FH, Soli SD (2000) Auditory perception in vestibular neurectomy subjects. Hear Res 142:102–112.
Zeng FG, Grant G, Niparko J, Galvin J, Shannon R, Opie J, Segel P (2002) Speech dynamic range and its effect on cochlear implant performance. J Acoust Soc Am 111:377–386.
Zwislocki JJ, Jordan HN (1986) On the relations of intensity jnd's to loudness and neural noise. J Acoust Soc Am 79:772–780.

Index