HEARING

An Introduction to Psychological and Physiological Acoustics

Stanley A. Gelfand

Audiology and Speech Pathology Service
Veterans Administration Medical Center
East Orange, New Jersey
and
Department of Neurosciences
New Jersey Medical School,
College of Medicine and Dentistry of New Jersey
Newark, New Jersey

MARCEL DEKKER, INC. New York and Basel
BUTTERWORTHS London

Library of Congress Cataloging in Publication Data

Gelfand, Stanley A [date]
Hearing, an introduction to psychological and physiological acoustics.

Includes indexes.
1. Hearing. 2. Psychoacoustics. I. Title. [DNLM: 1. Hearing—Physiology. 2. Psychoacoustics. WV 270 G316h]
QP461.G28 612'.85 80-27882
ISBN 0-8247-1189-0

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Current printing (last digit)
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

TO JANICE AND MICHAEL

Foreword

The field of hearing science is of a highly interdisciplinary nature. In order to have a good, broad understanding of the field it is necessary to have a more than superficial knowledge of (in alphabetical order) acoustics; audiology; auditory anatomy, physiology, and neurophysiology; communication theory; data processing; electroacoustics; experimental psychology; instrumentation; signal processing; statistics; and — of course — psychoacoustics. Hearing scientists with a broader vision also know something about acoustic phonetics, auditory development, cognition, speech perception, and related topics; since, after all, the ear is used largely for speech communication and the ear plays a crucial role in the early development of speech and language.

Since Leibnitz was the last man to know everything, it is not surprising that researchers in the field of hearing science have become progressively more specialized over the years. The process of knowing more and more about less and less represents progress of a sort; a dirac pulse of knowledge after all is no mean achievement. However, there are many dangers to overspecialization.

It is important for researchers to keep abreast of major new developments in related (and often seemingly unrelated) areas; and it is equally important for students to be exposed to a broad view of the field before the process of in-depth specialization takes hold. How can this be done? A good textbook providing a broad, unified treatment is a useful start.

Unfortunately, most textbooks in hearing science typically provide extensive coverage in the author's specific area of expertise and only scant coverage elsewhere. Another interestingly common type of book is that of an anthology of chapters by a diverse group of experts. Some of these anthologies provide excellent coverage of a range of topics but there are inevitable gaps. Further, the quality and level of coverage often fluctuate widely (if not ludicrously) between chapters. It takes a smart student to know which chapters to avoid.

One of the strengths of *Hearing: An Introduction to Psychological and Physiological Acoustics* is that it provides broad coverage at a consistent level in terms that students of diverse backgrounds can understand. For the advanced undergraduate student or the postgraduate student in such disciplines as audiology and psychology, this book provides both an introduction and a broad overview of the field of hearing science that is decidedly lacking in contemporary treatments of the subject. It should be an extremely useful guide to these students, as well as to those researchers who wish to refresh their knowledge of the field beyond their areas of specialization. Herein lies another strength of the book in that it is up to date with respect to new developments in all aspects of the field of hearing science—and there have been many of late.

Dr. Gelfand appears to have taken Disraeli's dictum to heart: Wishing to read a good book (on hearing science), he wrote one.

HARRY LEVITT

Professor, Ph.D. Program in Speech and Hearing Sciences
Graduate School and
University Center of the
City University of New York

Preface

This book is concerned with the physiology and psychophysics of audition. Its intent is to introduce the new student to the sciences of hearing and to rekindle the interests of the experienced reader.

The hearing science student is often faced with the frustrating dilemma imposed by textbooks that are either too basic or too advanced. This problem is complicated by the widely divergent backgrounds of those who become interested in hearing. My intent was to provide a volume sufficiently detailed to be useful as a core text, yet not so dependent upon such prior knowledge that it would be beyond the grasp of most students. In order to accomplish this goal, I have tried to minimize the use of mathematical expressions (a difficult chore within the confines of the topic) and have placed qualitative concepts dealing with basic acoustics along with the text material rather than in a separate chapter. This assumes some prior knowledge of elementary physics, but the concepts needed are reiterated with the subject matter so that their application is facilitated.

In addition, the depth and breadth of topic coverage between chapters (and in some cases between sections of the same chapter) was varied based upon reactions to the material by a variety of students. This was done to maximize what the student may come away with as opposed to imposing an arbitrary standard of sophistication across the board. Of course, my own interests and biases have tainted the material by omission and commission in spite of honest attempts to keep these repressed.

Thus, I freely admit to a purposefully irregular level of topic coverage for didactic purposes and apologize for the places where personal bias crept in.

It is doubtful that an introductory masters level or undergraduate course could cover all of the material in a single semester. It is more likely that this volume might be used as a core text for a two-term sequence dealing with psychological and physiological acoustics, along with appropriately selected readings from the research literature and state-of-the-art books. Suggested readings are provided in context throughout the text. However, most of the material has been used in a one semester "scan course." In either case, the student should find a sufficient number of citations for each topic to provide a firm foundation for further study.

Finally, the reader (particularly the beginning student) is reminded that the author of a textbook is merely one who has put material together. Bias finds its way into even descriptive text, and new findings occasionally disprove the "laws" of the past. This book is meant to be a first step; it is by no means the final word.

This work represents the contributions of many others. They share the responsibility for its fine points. The flaws are my own.

Special thanks are expressed to Dr. Maurice H. Miller for his suggestion that I prepare this book and for his constant assistance and confidence. I would also like to thank the reviewers of the manuscript for providing valuable comments and suggestions. Inestimable gratitude is expressed to my dear friend and colleague, Dr. Shlomo Silman, whose advice and confidence provided me with the motivation and drive to complete this work. My sincerest gratitude also goes to Dr. Harry Levitt, who so graciously agreed to write the foreword and who taught me to love hearing science.

The following individuals provided assistance and support in ways too varied and numerous to detail here: Marilyn Agin, Dr. Barbara Bohne, Dr. Joseph Danto, Jack Ficarra, Renee Kaufman, Harry Kubasek, Dr. Janice Leeds, Dr. David Lim, John Lutolf, Adrienne Rubinstein, Stanley Schwartz, Harris and Helen Topel, and especially Neil Piper.

Finally, my greatest appreciation is expressed to my wife, Janice, whose love and confidence kept me going, whose knowledge was invaluable, and who so willingly gave up weekends, holidays, and evenings so that I could write.

STANLEY A. GELFAND

1
Anatomy

GROSS ANATOMY AND OVERVIEW

The auditory system comprises the ears and their connections to and within the central nervous system. From the standpoint of physical layout, the auditory system may be divided into the outer, middle, and inner ears; the auditory nerve; and the central auditory pathways.

This section provides a very brief and simplified overview of the auditory system, in the hope that a brief glance at the forest will help the student avoid being blinded by the trees.

The major divisions of the ear are shown in Fig. 1.1, and their relative positions within the head are given in Fig. 1.2. The outer ear is made up of the pinna (auricle) and ear canal (external auditory meatus). The eardrum (tympanic membrane) separates the outer and middle ears, and is generally considered to be part of the latter. The middle ear also includes the tympanic (middle ear) cavity; the ossicular chain with its associated muscles, tendons, and ligaments; and the Eustachian (auditory) tube. The inner ear begins at the oval window. It includes the sensory organs of hearing (the cochlea) and of balance (the semicircular canals, utricle, and saccule). While the balance system is certainly important, the concern here is hearing, and accordingly the balance apparatus is mentioned only in so far as it is directly associated with the auditory system.

The inner ear, beyond the oval window, is composed of the vestibule, the cochlea, and the vestibular apparatus. A membranous duct is continuous throughout these. In the cochlea, it separates the perilymph-filled

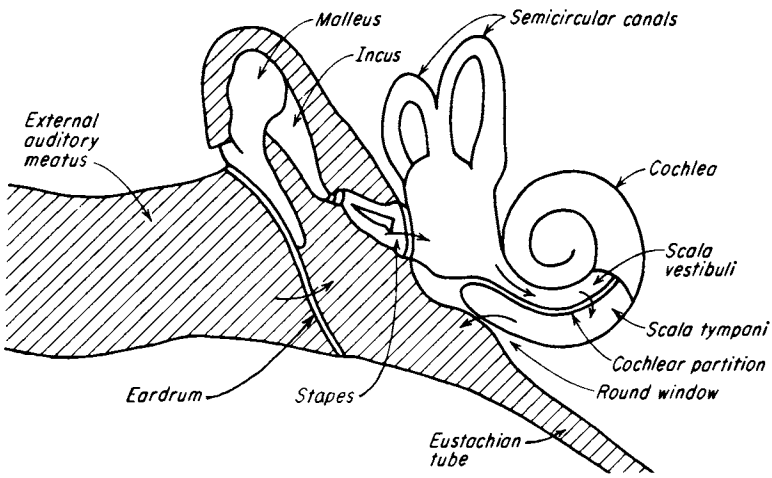

Figure 1.1 Cross section of the human ear. (From *Experiments in Hearing* by G. von Bekesy [1]. Copyright 1960 by McGraw-Hill. Used with permission of McGraw-Hill Book Company.)

scala vestibuli and scala tympani above and below from the endolymph-filled scala media between them. The scala media contains the organ of Corti, whose hair cells are the sensory receptors for hearing. When stimulated, the hair cells initiate activity in the auditory nerve fibers with which they are in contact. The auditory nerve leaves the inner ear through the internal auditory canal (internal auditory meatus), enters the brain at the angle of the pons and cerebellum, and terminates in the brainstem at the cochlear nuclei. We are now in the central auditory system.

TEMPORAL BONE

The ear is contained within the temporal bone. Knowledge of the major landmarks of this bone is thus important in understanding the anatomy and spatial orientation of the ear. The right and left temporal bones are two of the thirty-two bones that make up the skull. Eight of these bones (including the two temporal bones) contribute to the cranium, and the remainder form the facial skeleton. Figure 1.3 gives a lateral (side) view of the skull, emphasizing the temporal bone. The temporal bone forms the inferior portion of the side of the skull. It is bordered by the mandible, zygomatic, parietal, sphenoid, and occipital bones. The temporal bone itself is generally divided into five anatomical divisions: the squamous, mastoid, petrous, and tympanic portions, and the anteroinferiorly protruding styloid process (Fig. 1.4).

The squamous portion is the fan-shaped part of the temporal bone. It is quite thin, often to the point of being translucent. Its inferior surface forms the roof and part of the posterior wall of the ear canal. The zygomatic process protrudes forward from the squamous portion to meet the zygomatic bone. The fan-shaped squamous plate is also in contact with the sphenoid bone anteriorly and with the parietal bone superiorly and posteriorly. The mandible attaches to the temporal bone just anterior to the ear canal, near the base of the zygomatic process, forming the temporomandibular joint.

The mastoid portion lies behind and below the squamous, and forms the posterior aspect of the temporal bone. The mastoid portion attaches

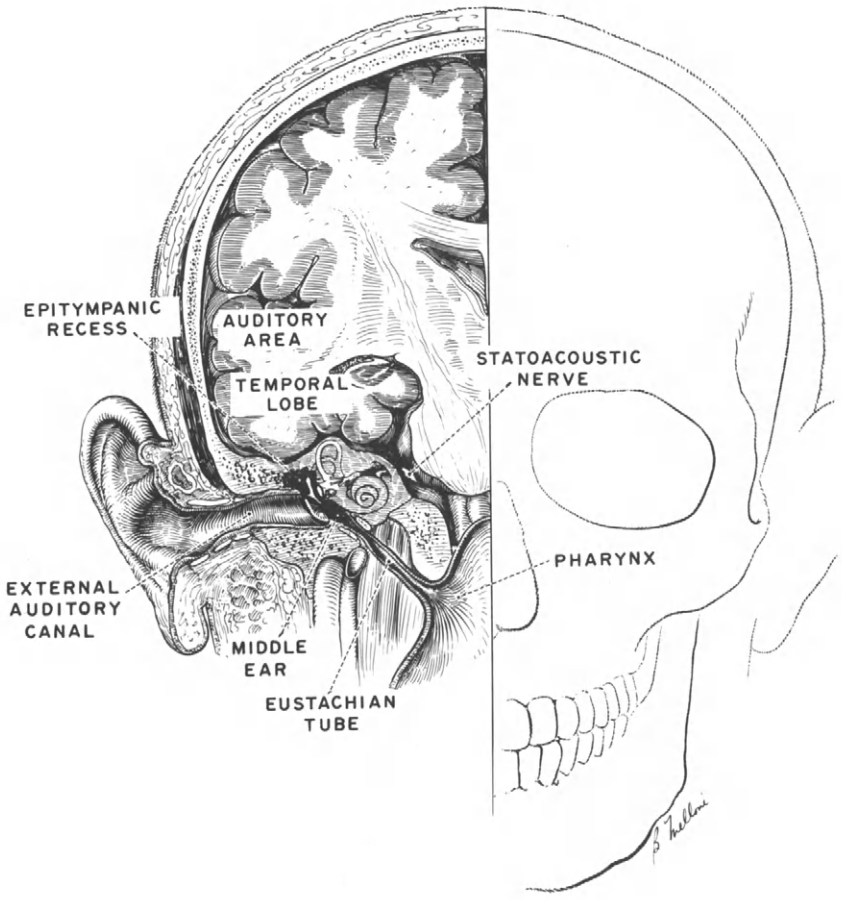

Figure 1.2 Structures of the ear relative to the head. (Courtesy of Abbott Laboratories.)

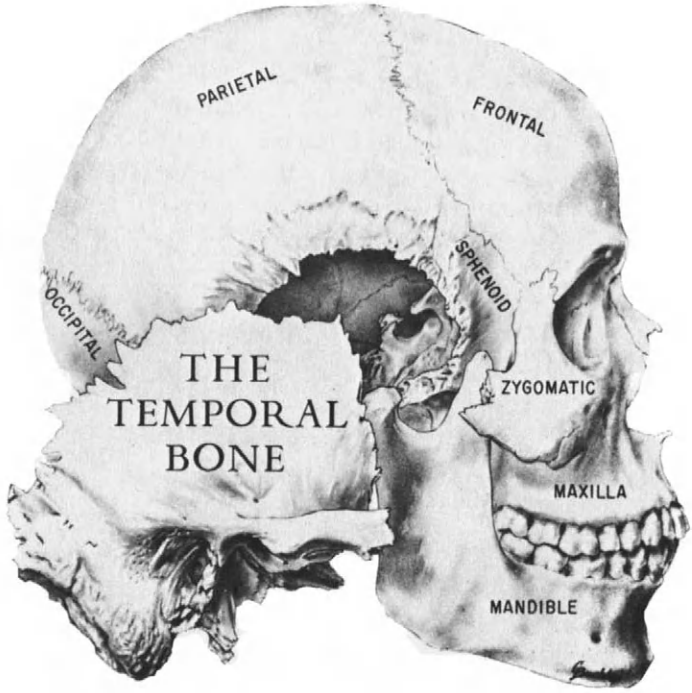

Figure 1.3 Lateral view of the skull emphasizing the position of the temporal bone. [From Anson and Donaldson [2], *The Surgical Anatomy of the Temporal Bone and Ear*, © 1967, W. B. Saunders (with permission).]

to the parietal bone superiorly and to the occipital bone posteriorly. It projects downward to form the mastoid process, which appears as a somewhat cone-shaped extension below the base of the skull. The mastoid process contains interconnecting air cells of variable size, shape, and number. Continuous with these air cells is a cavity known as the tympanic antrum, which lies anterosuperior to the mastoid process. The antrum also connects with the epitympanic recess (attic) of the middle ear via the aditus ad antrum. The antrum is bordered inferiorly by the mastoid process, superiorly by a thin bony plate called the tegmen tympanum, medially by the wall of the lateral semicircular canal, and laterally by the squamous.

The tympanic portion forms the floor and the anterior and inferoposterior walls of the ear canal. It is bordered by the squamous portion above, the petrous portion below, and the mastoid process posteriorly. The lower part of the tympanic portion partially covers the styloid process, a thin, cylinder-like anteroinferior projection from the base of the

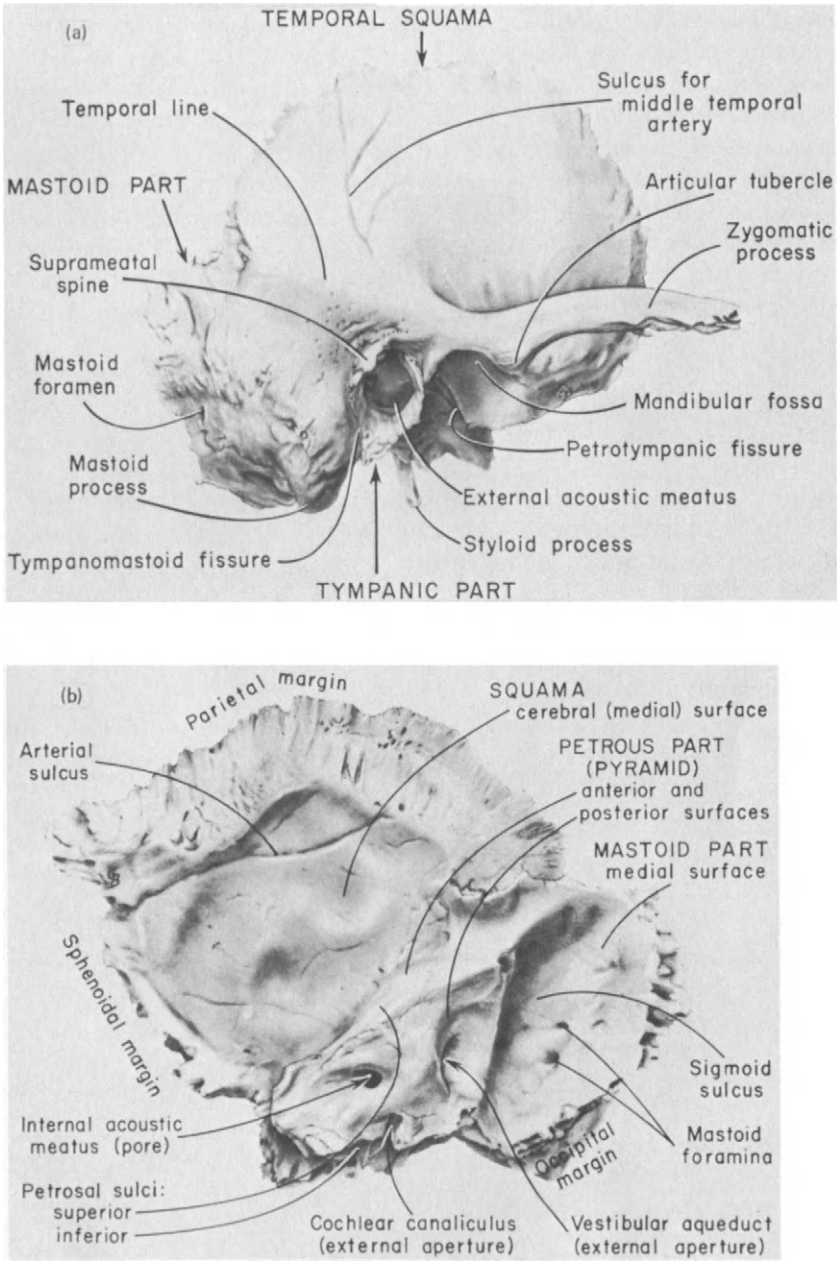

Figure 1.4 Lateral (a) and medial (b) aspects of the temporal bone. [From Anson and Donaldson [2], *The Surgical Anatomy of the Temporal Bone and Ear*, © 1967, W. B. Saunders (with permission).]

temporal bone. The styloid process, which varies in length from as little as 5 mm to as much as ~50 mm, is generally considered to be a separate portion of the temporal bone. While it does not contribute to the hearing mechanism, the styloid process is important as a connecting point for several muscles involved in speech production.

The petrous portion houses the sensory organs of hearing and balance, and contains the internal auditory canal. It is medially directed and is fused at its base to the tympanic and squamous portions. The mastoid lies posterior to the petrous portion, and in fact develops from it postnatally. The details of the petrous portion are equivalent to those of the inner ear, discussed below.

OUTER AND MIDDLE EAR

Pinna

The pinna (auricle) is the external appendage of the ear. It is an irregularly shaped ovoid of highly variable size, which folds over the side of the head posteriorly, superiorly, and inferiorly. It is basically composed of skin-covered elastic cartilage, although it contains some grossly undifferentiated muscles which are of a completely vestigial nature in man. The pinna has a number of extrinsic muscles as well, which are also essentially vestigial in humans.

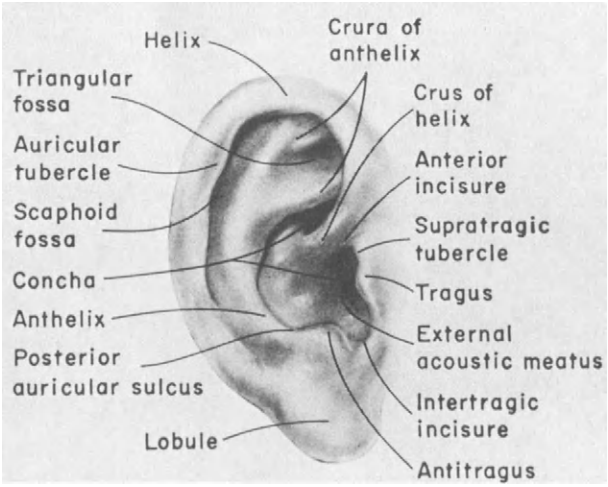

Figure 1.5 Landmarks of the pinna. [From Anson and Donaldson [2], *The Surgical Anatomy of the Temporal Bone and Ear,* © 1967, W. B. Saunders (with permission).]

The landmarks of the pinna are shown in Fig. 1.5. Most of its perimeter is demarcated by a ridgelike rim called the helix. If we first follow the helix posteriorly from the top of the ear, we see that it curves widely back and down to end in the earlobe (lobule) at the bottom of the pinna. Unlike the rest of the pinna, the lobule does not have any cartilage. Starting again from the apex of the helix, we see that it proceeds anteriorly and downward, and then turns posteriorly in a rather sharp angle to form the crus (limb) of the helix — an almost horizontal shelf at about the center of the pinna. The scaphoid fossa is a depression lying between the posterior portion of the helix posteriorly and a structure called the antihelix anteriorly.

The antihelix is a ridge which runs essentially parallel to the posterior helix. Its upper end bifurcates to form two crura, a rather wide superoposterior crus and a narrower anterior crus which ends under the angle where the helix curves backward. A triangular depression is thus formed by the two crura of the antihelix and the anterior part of the helix, and is called the triangular fossa. From the crura, the antihelix curves downward and then forward, and ends in a moundlike widening, the antitragus. Opposite and anterior to the antitragus is a backward-folding ridge called the tragus. The inferoanterior acute angle formed by the tragus and antitragus is called the intertragal incisure. The tragus, the antitragus, and the crus of the helix border a relatively large and cup-shaped depression called the concha. Sebaceous glands are present in the skin of the concha, as well as in the ear canal. At the bottom of the concha, protected by the tragus, is the entrance to the ear canal.

Ear Canal

The ear canal (external auditory meatus) leads from the concha to the eardrum, and varies in both size and shape. The outer portion of the canal, about one-third of its length, is cartilagenous; the remaining two-thirds is bony. The canal is by no means straight; rather it is quite irregular in its course. It takes on a somewhat S-shaped form medially. It curves first anterosuperiorly, then posterosuperiorly, and finally anteroinferiorly. It is for this reason that the pinna must be pulled up and back in order for one to see the eardrum.

The ear canal has a diameter of about 0.7 cm at its entrance, with an average horizontal diameter of 0.65 cm and a mean vertical diameter of 0.9 cm [3]. As would be expected from its irregular course, the length of the canal is also not uniform. Instead, it is approximately 2.5 cm long posterosuperiorly and 3.1 cm long inferoanteriorly [4]. Also contributing to the greater length of the lower part of the canal is the oblique orientation of the eardrum as it sits in its annulus at the end of the canal.

The canal is lined with tight-fitting skin that is thicker in the cartilagenous segment than in the bony part. Ceruminous (wax) and sebaceous (oil) glands are plentiful in the cartilagenous segment, and are also found on the posterior and superior walls of the bony canal. The wax and oil lubricate the canal and help to keep it free of debris and foreign objects. Tiny hairs similarly contribute to the protection of the ear from invasion.

Eardrum

The canal terminates at the eardrum (tympanic membrane), which tilts laterally at the top so as to sit in its annulus at an angle of about 55° to the ear canal (see Fig. 1.1). The membrane is quite thin and translucent, with an average thickness of approximately 0.074 mm [4]. It is elliptical in shape, with a vertical diameter of about 0.9–1.0 cm and a horizontal cross section of approximately 0.8–0.9 cm. The eardrum is concave outward, and the peak of this broad cone is known as the umbo. This inward displacement is associated with the drum's attachment to the manubrium of the malleus, the tip of which corresponds to the umbo (Fig. 1.6). In contact with the drum, the malleus continues upward in a direction corresponding to the one-o'clock position in the right ear and the eleven-o'clock position in the left. The malleal prominence of the malleus is formed by the lateral process of the malleus, from which run the malleal folds which divide the drum into the pars flaccida (Shrapnell's membrane) above and the pars tensa below.

The eardrum is made up of four layers. The outermost layer is continuous with the skin of the ear canal, and the most medial layer is continuous with the mucous membrane of the middle ear. The pars flaccida is composed solely of these two layers. The pars tensa has two additional layers: a layer of radial fibers just medial to the skin layer, and a layer of nonradial fibers between the radial and mucous membrane layers.

Tympanic Cavity

The middle ear cavity (tympanum) may be thought of schematically as a six-sided box. The lateral wall is the eardrum, and opposite to it the promontory of the basal cochlear turn forms the medial wall. Figure 1.7 shows such a conceptualization of the right middle ear, looking in toward the medial wall as though the eardrum (lateral wall) were removed. The roof is formed by the tegmen tympani, which separates the middle ear from the middle cranial fossa above. The floor of the tympanum separates it from the jugular bulb. In the anterior wall is the opening to the Eustachian tube, and above it the canal for the tensor tympani muscle. The canal of the internal carotid artery lies behind the anterior wall, posteroinferior to the tubal opening. The posterior wall contains

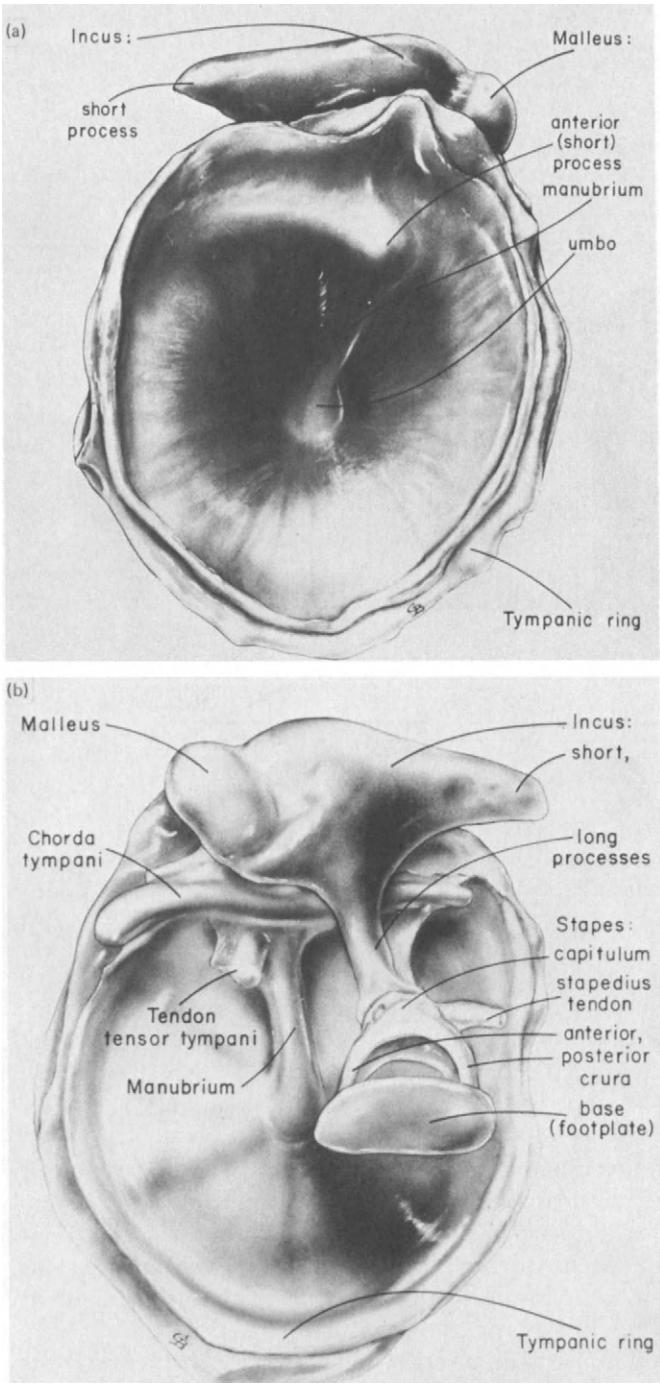

Figure 1.6 Lateral (a) to medial (b) aspects of the tympanic membrane and its connections to the ossicular chain. [From Anson and Donaldson [2], *The Surgical Anatomy of the Temporal Bone and Ear*, © 1967, W. B. Saunders (with permission).]

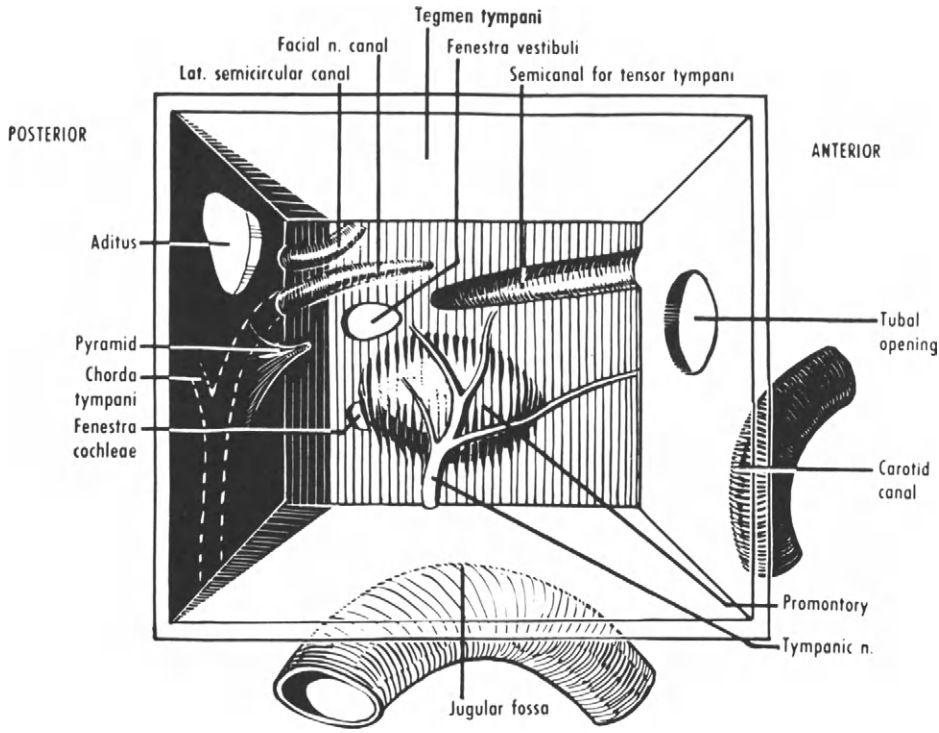

Figure 1.7 Schematic representation of the middle ear. [From Gardner, Gray, and O'Rahilly [5], *Anatomy*, 4th ed., © 1975, W. B. Saunders (with permission).]

the aditus ad antrum through which the upper portion of the middle ear (epitympanic recess) communicates with the mastoid antrum. The posterior wall also contains the fossa incudis, a recess that receives the short process of the incus; and the pyramidal eminence which houses the stapedial muscle. The stapedial tendon exits from the pyramidal prominence at its apex.

Returning to the medial wall, we see that the oval window is located posterosuperiorly to the promontory, while the round window is posteroinferior to the latter. Superior to the oval window lies the facial canal prominence, with the cochleariform process on its anterior aspect. The tendon of the tensor tympani muscle bends around the cochleariform process to proceed laterally to the malleus.

The Eustachian (auditory) tube serves to equalize the air pressure on both sides of the eardrum, and to allow for drainage of the middle ear by serving as a portal to the nasopharynx. It is about 37 mm in length, and

courses medially, down, and forward to exit into the nasopharynx via a prominence called the torus tubarius. The lateral third of the tube is bony, while the remainder is elastic cartilage. The two sections of the tube form an angle of about 160° extending down to the nasopharynx. The meeting of the bony and cartilaginous portions is called the isthmus, and at this point the lumen of the tube may be as little as 1.0–1.5 mm in diameter compared to about 3.0–6.0 mm at the normally open port into the middle ear. The cartilaginous part is normally closed, and opens in response to swallowing, yawning, sneezing, or shouting. This reflexive opening of the Eustachian tube is caused by action of the tensor palatini muscle, which uncurls the normally closed hook-shaped cartilages that make up this part of the tube.

Ossicular Chain

Sound energy impinging upon the eardrum is conducted to the inner ear by way of the malleus, incus, and stapes (the ossicular chain), which are the smallest bones in the body. The ossicles are shown in place in Fig. 1.6 and separately in Fig. 1.8. They are suspended in the middle ear by a series of ligaments, by the tendons of the two intratympanic muscles, and by the attachments of the malleus to the eardrum and of the stapes to the oval window.

The malleus is commonly called the hammer, although it more closely resembles a mace. It is the largest of the ossicles, being about 8.0–9.0 mm long and weighing approximately 25 mg. The head of the malleus is located in the epitympanic space, to which it is connected by its superior ligament. Laterally, the manubrium (handle) is embedded between the mucous membrane and fibrous layers of the eardrum. The anterior process of the malleus projects anteriorly from the top of the manubrium just below the neck. It attaches to the tympanic notch by its anterior ligament, which forms the axis of mallear movement. The malleus is connected to the tensor tympani muscle via a tendon which inserts at the top of the manubrium.

The incus bears a closer resemblance to a tooth with two roots than to the more commonly associated anvil. It weighs approximately 30 mg, and has a length of about 5.0 mm along its short process and about 7.0 mm along its long process. Its body is connected to the posteromedial aspect of the mallear head within the epitympanic recess. The connection is by a saddle joint, which was originally thought to move by a cog-like mechanism when the malleus was displaced inward [6]. However, subsequent research has demonstrated that these two bones move as a unit rather than relative to one another [3]. The short process of the

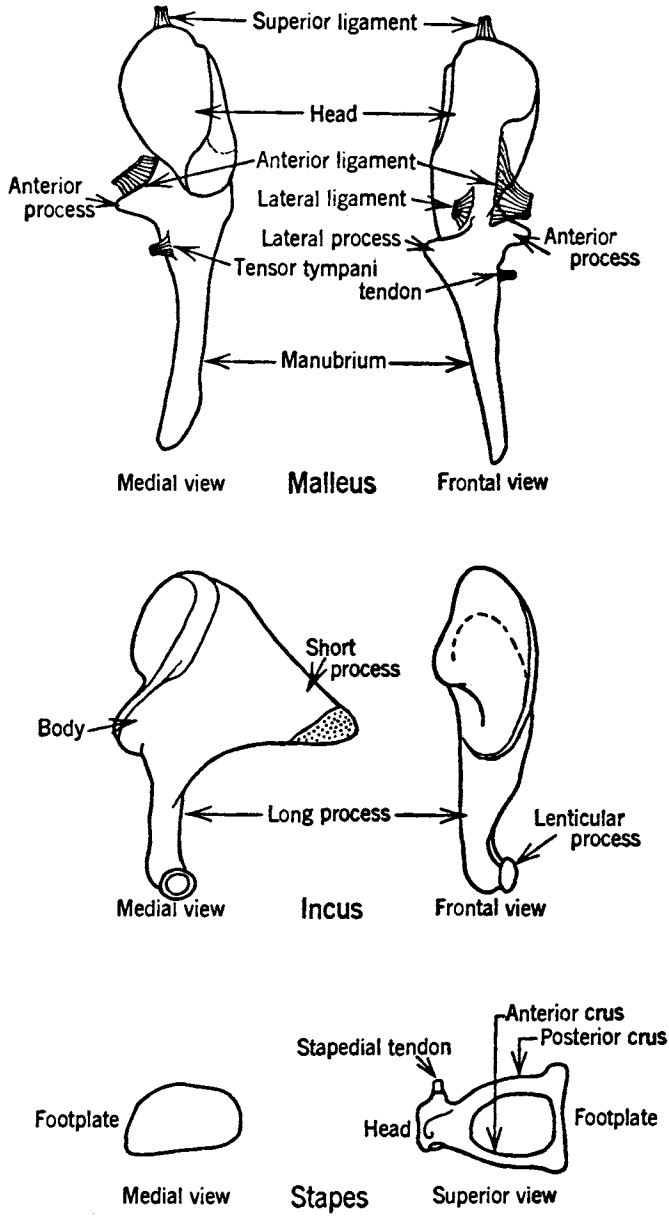

Figure 1.8 Individual ossicles. [From Ernest Glen Wever and Merle Lawrence, *Physiological Acoustics* (copyright 1954 by Princeton University Press): Fig. 1, p. 13. Reprinted by permission of Princeton University Press.]

incus connects via its posterior ligament to the fossa incudis on the posterior wall of the tympanic cavity. Its long process runs inferiorly, parallel to the manubrium. The end of the long process then bends medially to articulate with the head of the stapes in a true ball-and-socket joint.

The stapes (stirrup) is the smallest of the ossicles. It is about 3.5 mm high, and the footplate is about 3.0 mm long by 1.4 mm wide. It weighs on the order of 3.0–4.0 mg. The head of the stapes connects to the footplate via two crura. The anterior crus is straighter, thinner, and shorter than the posterior crus. The footplate, which encases the very fine stapedial membrane, is attached to the oval window by the annular ligament. The stapedius tendon inserts on the posterior surface of the neck of the stapes, and connects the bone to the stapedius muscle.

Intratympanic Muscles

The middle ear contains two muscles, the tensor tympani and the stapedius. The stapedius muscle is the smallest muscle in the body, with an average length of 6.3 mm and a mean cross-sectional area of 4.9 mm^2 [3]. The muscle is completely encased within the pyramidal eminence on the posterior wall of the tympanic cavity, and takes origin from the walls of its own canal. Its tendon exits through the apex of the pyramid and courses horizontally to insert on the posterior aspect of the neck of the stapes. Contraction of the stapedius muscle thus pulls the stapes posteriorly. The stapedius is innervated by the stapedial branch of the seventh cranial (facial) nerve.

The tensor tympani muscle has an average length of 25 mm and a mean cross-sectional area of approximately 5.85 mm^2 [3]. The tensor tympani occupies an osseous semicanal on the anterior wall of the tympanum, just superior to the Eustachian tube, from which it is separated by a thin bony shelf. The muscle takes origin from the cartilage of the auditory tube, from the walls of its own canal, and from the part of the sphenoid bone adjacent to the canal. Emerging from the canal, the tendon of the tensor tympani hooks around the cochleariform process, and inserts on the top of the manubrium of the malleus. Contraction of the tensor tympani thus pulls the malleus anteromedially — at a right angle to the uninterrupted motion of the ossicles. The tensor tympani muscle is innervated by the tensor tympani branch of the otic ganglion of the fifth cranial (trigeminal) nerve.

Both intratympanic muscles are completely encased within bony canals, and attach to the ossicular chain by way of their respective tendons. Bekesy [7] pointed out that this situation reduces the effects that muscular contractions might have upon the transmission of sound through the middle ear system. Contraction of either muscle increases

the stiffness of the ossicular chain as well as of the eardrum. The stapedius muscle pulls posteriorly whereas the tensor tympani pulls anteromedially, so that they might initially be thought to be antagonists. However, the effect of these muscles is to lessen the amount of energy conducted by the ossicular chain, and they thus function as synergists with respect to hearing.

Of particular interest is the acoustic reflex — the response of the intratympanic muscles to intense sounds. It is generally accepted that the acoustic reflex *in man* is due mainly, if not exclusively, to contraction of the stapedius; the tensor responds only to extremely intense sounds, as part of a startle response.

The acoustic reflex arc has been described in detail [8,9]. The afferent (sensory) branch of the arc is the auditory nerve, terminating in the ventral cochlear nucleus, which communicates with the superior olivary complex on both sides (via the trapezoid body). These bilateral connections permit a reflex response in both ears even when only one ear is stimulated. The efferent (motor) pathway of the arc proceeds from the medial accessory nucleus of the superior olivary complex on each side to the nuclei of the facial nerve for the stapedius reflex, and to the nuclei of the trigeminal nerve to cause tensor tympani contraction.

INNER EAR

Osseous Labyrinth

The inner ear structures are contained within a system of spaces and canals, the osseous labyrinth (Fig. 1.9), in the petrous portion of the tem-

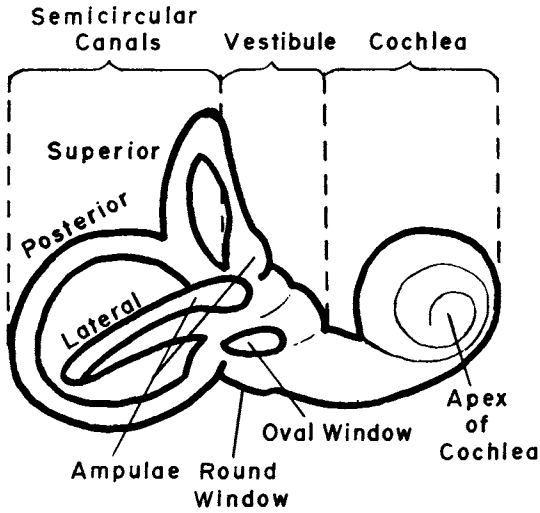

Figure 1.9
The osseous labyrinth.

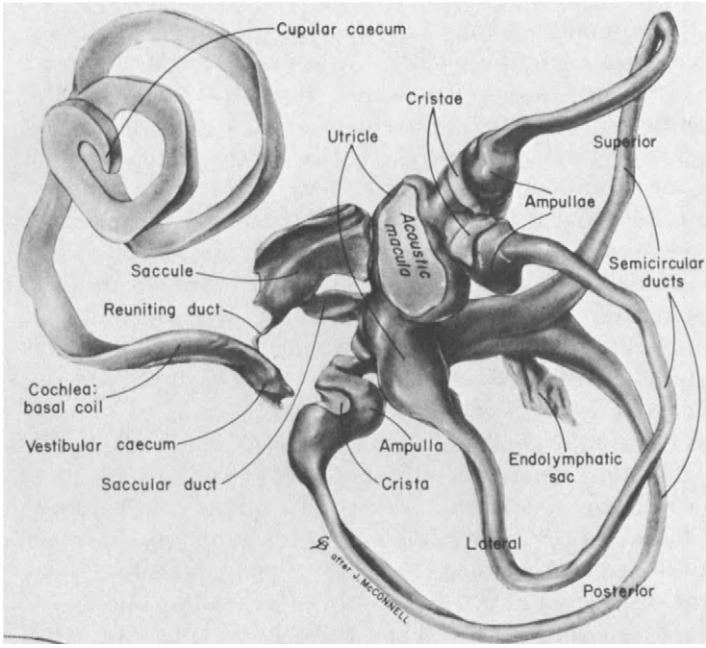

Figure 1.10 The membranous labyrinth. [From Donaldson and Miller [4], "Anatomy of the ear," in *Otolaryngology*, Vol. 1 (Paparella and Shumrick, Eds.), © 1973, W. B. Saunders (with permission).]

poral bone. These spaces and canals are grossly divided into three sections: the vestibule, the cochlea, and the semicircular canals. The oval window accepts the stapedial footplate and opens medially into the vestibule, which is about 4.0 mm in diameter and somewhat ovoid in shape. The snail-shaped cochlea lies anterior and slightly inferior to the vestibule, and is approximately 5.0 mm high and 9.0 mm in diameter at its base. Posterior to the vestibule are the three semicircular canals, lying at right angles to one another, each about 1.0 mm in diameter.

Membranous Labyrinth and Inner Ear Fluids

The general shape of the bony labyrinth is followed by the enclosed membranous labyrinth (Fig. 1.10), which contains the end organs of hearing and balance. The space between the bony walls of the osseous labyrinth and the membranous labyrinth is filled with perilymphatic fluid (perilymph). The membranous labyrinth itself is mostly filled with endolymphatic fluid (endolymph), but the cochlear hair cells are actually bathed in a third fluid, cortilymph, which will be discussed below.

Lawrence [10] enumerated four functions of the inner ear fluids. First, they deliver nutrients to the inner-ear cells which are not in direct contact with the blood (and also remove waste). Second, they provide the chemical environment needed for the transfer of energy from a vibratory stimulus to a neural signal. Third, the fluids are the medium which carries the vibratory stimulus from the footplate to the sensory structures along the cochlear partition. Finally, the distribution of pressure in the inner ear system is controlled by the cochlear fluids, although this last function is still debated. The inner ear fluids are themselves quite interesting, although their study is difficult because they are found in such small amounts. For example, Lawrence reported that the total volumes of these fluids in man are only 78.3 mm^3 of perilymph and the miniscule amount of 2.76 mm^3 of endolymph.

Perilymph is chemically similar to other extracellular fluids. Smith et al. [11] found that guinea pig perilymph has a very high concentration of sodium and a very low concentration of potassium, as does cerebrospinal fluid and blood serum. Interestingly, they found that endolymph has just the opposite concentrations: It is high in potassium but low in sodium. It has the distinction of being the only extracellular fluid in the body which is high in potassium, with the possible exception of some parotid gland secretions [12]. The relative sodium and potassium concentrations in guinea pigs also appear to hold true for man [12,13].

Perilymph appears to be a filtrate formed under capillary pressure from the vessels of the spiral ligament above the insertion of Reissner's membrane in the scala vestibuli [10]. It is probably resorbed by the spiral ligament within the scala tympani near the basilar membrane [14]. Although the stria vascularis plays an important part in the production of endolymph, the precise origin of this fluid is not fully understood. The endolymph seems to flow slowly toward the endolymphatic duct and sac (see below), since occlusion of this duct results in a buildup of endolymph in lab animals [15]. In addition, an exchange of sodium and potassium ions radially between the endolymph and the stria vascularis, and across Reissner's membrane, maintains the appropriate ion balance in the endolymph. This exchange, along with the resting potentials associated with these ions, probably makes possible the energy conversion process in the organ of Corti.

Since endolymph contains a high concentration of potassium similar to that of intracellular fluids, it is a poor environment for the function of the hair cells and the unmyelinated fibers of the auditory nerve within the organ of Corti. The tunnel of Corti and other spaces of the organ are filled with a third fluid, cortilymph, which is high in sodium, and which is effectively isolated from the endolymph by the reticular lamina

[16,17]. The blood vessels beneath the basilar membrane appear to be the source of cortilymph. Although both corilymph and perilymph are high in sodium, it is unlikely that they are the same: Perilymph is toxic to the hair cells, and the two fluids have different sources.

Returning to the structures of the inner ear, the vestibule contains two balance organs which are concerned with linear acceleration and gravity effects. These organs are the utricle and saccule. The semicircular canals, located behind the vestibule, widen anteriorly into five saclike structures, which open into the somewhat elongated utricle. These widenings are the ampullae, and they contain the sensory receptors for rotational acceleration. (The interested reader is referred to any of the fine books in the References section for a detailed discussion of the balance system.) The most important connection between the areas of hearing and balance is the ductus reuniens, which joins the membranous labyrinth between the cochlea and the utricle.

The cochlear aqueduct leads from the vicinity of the round window in the scala tympani to the subarachnoid space medial to the dura of the cranium. Although the aqueduct leads from the perilymph-filled scala to the cerebrospinal-fluid-filled subarachnoid space, it is barely if at all patent in man. Thus, it is doubtful that there is any real interchange between these two fluid systems. The endolymphatic duct leads from the membranous labyrinth within the vestibule to the endolymphatic sac. The sac is located partially between the layers of the dura in the posterior cranial fossa and partly in a niche in the posterior aspect of the petrous portion of the temporal bone. Both the cochlear aqueduct and the endolymphatic duct and sac have been implicated in the regulation of hydraulic pressure in the inner ear.

Cochlea

The human cochlea is about 35 mm long, and forms a somewhat cone-shaped spiral with 2¾ turns. It is widest at the base, where the diameter is approximately 9.0 mm, and tapers toward the apex. It is about 5.0 mm high. The modiolus is the core which forms the axis of the spiral (Fig. 1.11). Through it course the auditory nerve and the blood vessels that supply the cochlea.

It is easier to visualize the cochlea by imagining the spiral uncoiled, as in Fig. 1.12. In this figure, the base of the cochlea is shown at the left and the apex at the right. We see three chambers: the scala media, scala vestibuli, and scala tympani. The scala media is self-contained and separates the other two. The scalae vestibuli and tympani, on the other hand, communicate with one another at the apex of the cochlea, through an

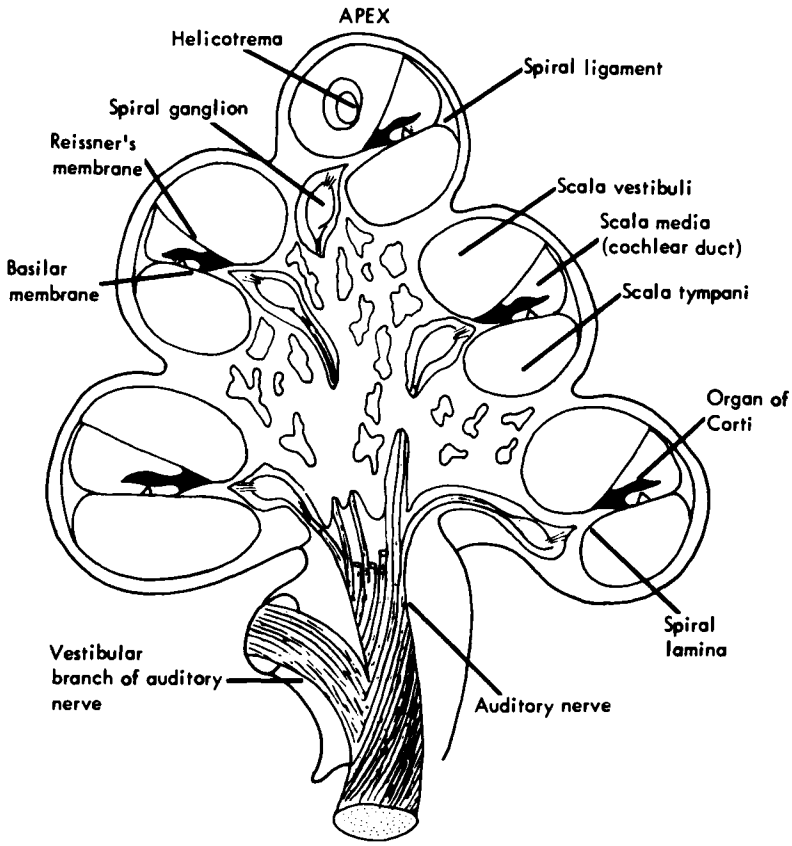

Figure 1.11 Schematic cross section of the cochlea through the modiolus. (From Willard R. Zemlin [18], *Speech and Hearing Science: Anatomy and Physiology*, © 1968, p. 403. Reprinted by permission of Prentice-Hall, Inc., Englewood Cliffs, New Jersey.)

opening called the helicotrema. The scala media is enclosed within the membranous labyrinth and contains endolymph, while the other two contain perilymph. The scala vestibuli is in contact with the stapes at the oval window, while the scala tympani has a membrane-covered contact with the middle ear at the round window. The scala media is separated from the scala vestibuli above by Reissner's membrane, and from the scala tympani below by the basilar membrane. The basilar membrane is approximately 32 mm long, and tapers from about 0.5 mm wide at the apex to about 0.1 mm wide near the stapes [1]. Furthermore, it is thicker at the base than at the apex.

The scala media is connected to the bony shelf on the inner wall of the cochlear spiral — the osseous spiral lamina. The osseous spiral lamina is clearly visible in Fig. 1.13, separating the scala vestibuli above from the scala tympani below. This figure also shows the orientation of the helicotrema at the apical turn, and the relationship between the round window and the scala tympani at the base of the cochlea. The scala media is attached to the outer wall of the cochlea by the spiral ligament, a fibrous connective tissue.

The structures of the scala media are shown schematically in Fig. 1.14. Looking first at the osseous spiral lamina on the inner wall, we see that this bony shelf is actually composed of two plates, separated by a space through which pass fibers of the auditory nerve. These fibers enter via openings called the habenula perforata. Resting on the osseous spiral lamina is a thickened band of periosteum, the limbus. Reissner's membrane extends from the top of the inner aspect of the limbus to the outer wall of the canal. The side of the limbus facing the organ of Corti is concave outward. The tectorial membrane is attached to the limbus at the upper lip of this concave part, forming a space called the internal spiral sulcus. The basilar membrane extends from the lower lip of the limbus to the spiral ligament.

The basilar membrane has two sections. The inner section extends from the osseous spiral lamina to the outer pillars, and is relatively thin. The remainder is thicker and extends to the spiral ligament. These two sections are called the zona arcuata and the zona pectinata, respectively. Sitting on the basilar membrane is the end organ of hearing — the organ of Corti.

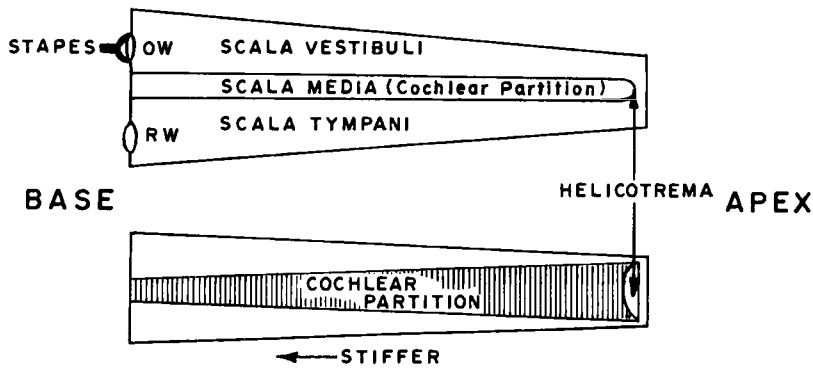

Figure 1.12 Schematic representation of the uncoiled cochlea (OW, oval window; RW, round window).

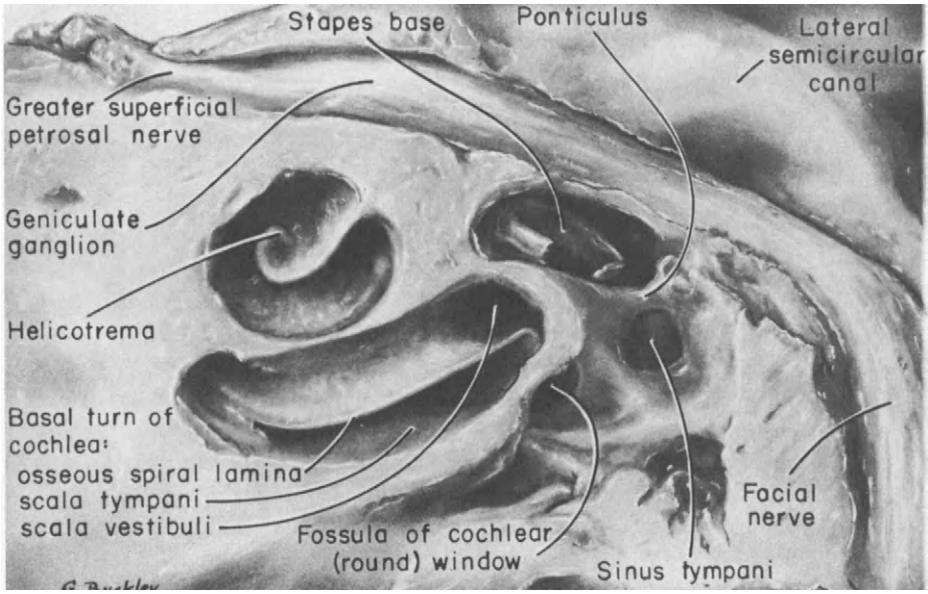

Figure 1.13 The osseous spiral lamina, scala vestibuli, scala tympani, and helicotrema. [From Anson and Donaldson [2], *The Surgical Anatomy of the Temporal Bone and Ear,* © 1967, W. B. Saunders (with permission).]

The organ of Corti runs longitudinally along the basilar membrane. Grossly, it is made up of a single row of inner hair cells (IHCs), three rows of outer hair cells (OHCs) (though as many as four or five rows have been reported in the apical turn), various supporting cells, and the pillar cells forming the tunnel of Corti. This tunnel separates the IHCs and OHCs. Each of the approximately 3500 IHCs is supported by a phalangeal cell which holds the rounded base of the IHC as in a cup. There are about 12,000 OHCs, shaped like test tubes, which are supported by Deiters' cells. Between the inner and outer hair cells (HCs) are the tilted and rather conspicuous pillars (rods) or Corti, which come together at their tops to enclose the triangular tunnel of Corti. Fibers of the eighth cranial (auditory) nerve traverse the tunnel to contact the OHCs. Just lateral to Deiters' cells are several rows of tall, supporting cells called Hensen's cells. Lateral to these are the columnar cells of Claudius, which continue laterally to the spiral ligament and the stria vascularis.

Figure 1.15 gives a closer look at the inner and outer HCs. The upper surface of each HC contains a cuticular plate, which is thick and topped with sterocilia, and an area that is cuticular-free. There are concentrations of mitochondria and Golgi apparatus under the cuticlular-free region, suggesting a high degree of metabolic activity, which may be related to the transduction of mechanical to electrochemical energy. The basal body of a rudimentary kinocilium is found in the noncuticular surface.

The distinctive pattern of the reticular lamina (Fig. 1.16) is formed by the cuticular plates of the HCs along with the upward-extending phalangeal processes of the phalangeal and Deiters' cells. The pillar cells maintain a strong structural attachment between the reticular lamina above and the basilar membrane below, and thus the reticular lamina provides a major source of support for the HCs at their upper surfaces. This relationship is exemplified in Fig. 1.17, showing the OHCs and

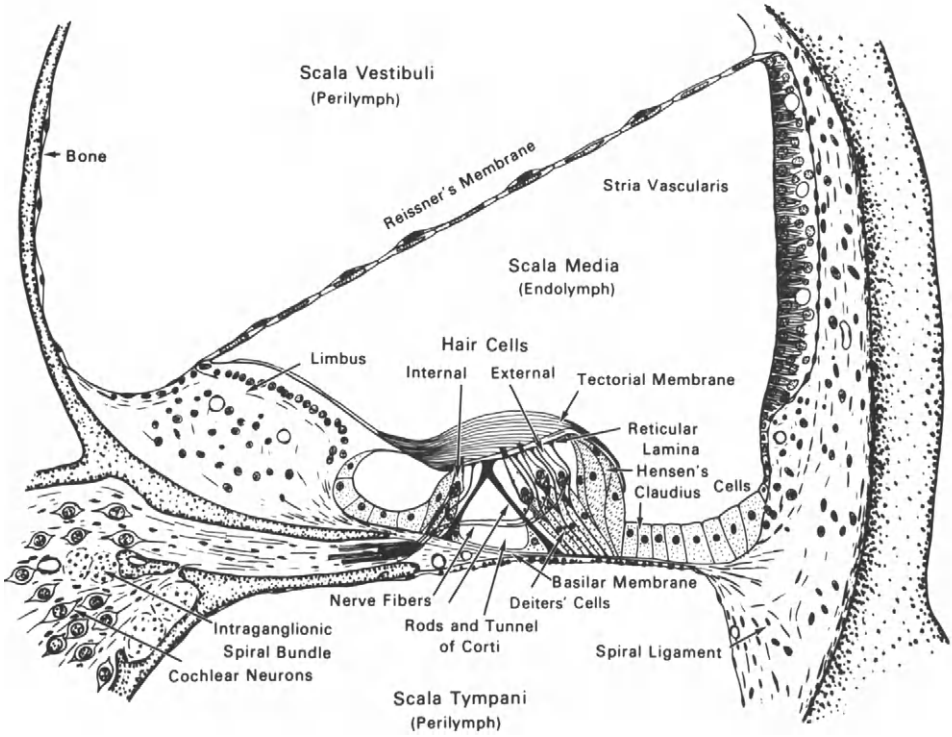

Figure 1.14 Cross section of the organ of Corti. (From Davis [19], with permission.)

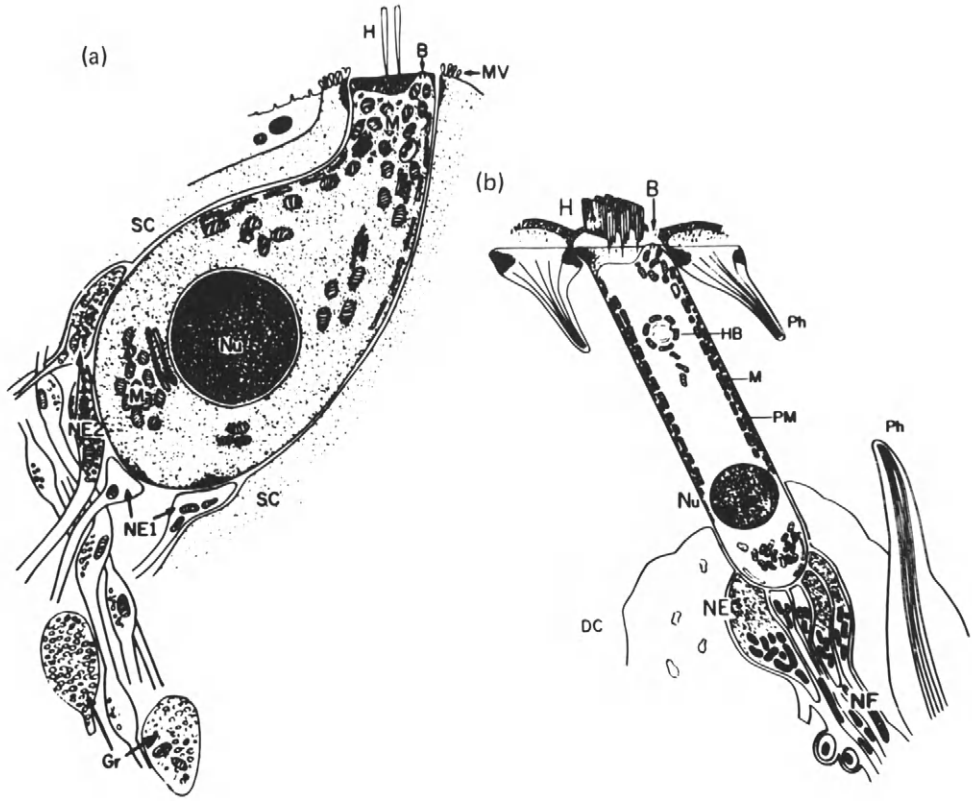

Figure 1.15 Inner (a) and outer (b) hair cells. H, steriocilia; B, basal body of kinocilium; NE1, afferent fiber; NE2, efferent fiber; Ph, phalangeal process. (From J. R. Brodbeck Ed. [20], *Best & Taylor's Physiological Bases of Medical Practice*, 9th ed., © 1973, The Williams & Wilkins Co., Baltimore.)

Deiters' cells in the reticular lamina, as well as the stereocilia of the HCs protruding through the lamina. The reticular lamina effectively isolates the structures and spaces of the organ of Corti from the rest of the endolymphatic duct [23].

There are about 50–70 stereocilia on each IHC. The OHCs each have roughly 40–150 stereocilia arranged in a W-shaped pattern (Fig. 1.18). The base of the W points toward the outer wall of the cochlear duct (i.e., away from the modiolus and toward the stria vascularis). The legs of the "W" form an obtuse angle on the OHCs in the basal turn, and an acute angle on cells in the apical turn.

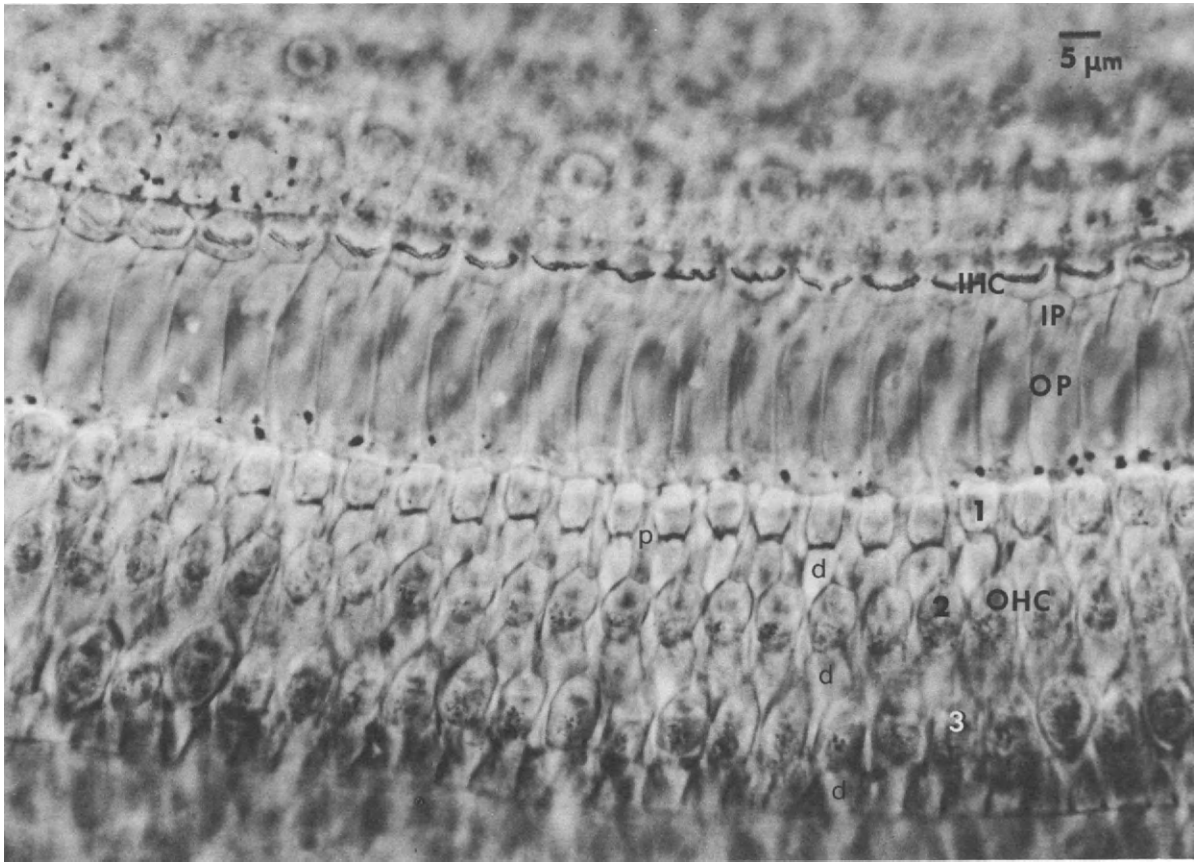

Figure 1.16 Reticular lamina showing a single row of IHCs, inner (IP) and outer (OP) pillars, and three rows (1,2,3) of OHCs. Each hair cell is surrounded by phalangeal processes from pillars (p) or Deiters' cells (d). (Courtesy of Dr. Barbara Bohne, Washington University School of Medicine, St. Louis, Mo.)

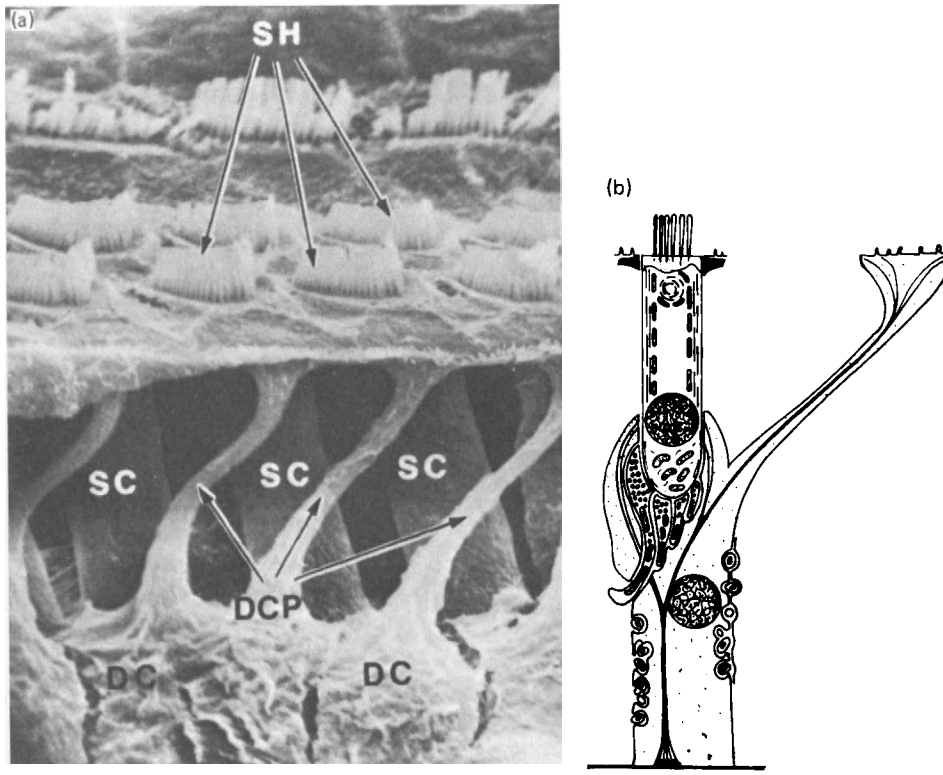

Figure 1.17 Scanning electron micrograph (a) and diagram (b) of Deiters' cells and reticular lamina. (From Bourne and Danieli [21] by permission of Academic Press.)

The tectorial membrane arises from the upper lip of the limbus, courses over the HCs as described below, and is connected by a border net to Hensen's cells lateral to the OHCs. The connection to the limbus permits the tectorial membrane to move relative to the basilar membrane, but it is nevertheless a strong and robust attachment. On the other hand, the attachment of the membrane to the cells of Hensen is quite fragile, to the point that it is often not preserved in specimens prepared for histological study. The tectorial membrane is ribbon-like in appearance and gelatinous in consistency. The attachment of the membrane to the stereocilia of the inner and outer HCs has been a subject of controversy for some time. Scanning electron microscopy has revealed that imprints on the underside of the tectorial membrane correspond to the OHC stereocilia, but have shown no such imprints for the IHC cilia. This suggests that the taller cilia of the OHCs are partially embedded in the tectorial membrane, but that the cilia of the IHCs are not [25].

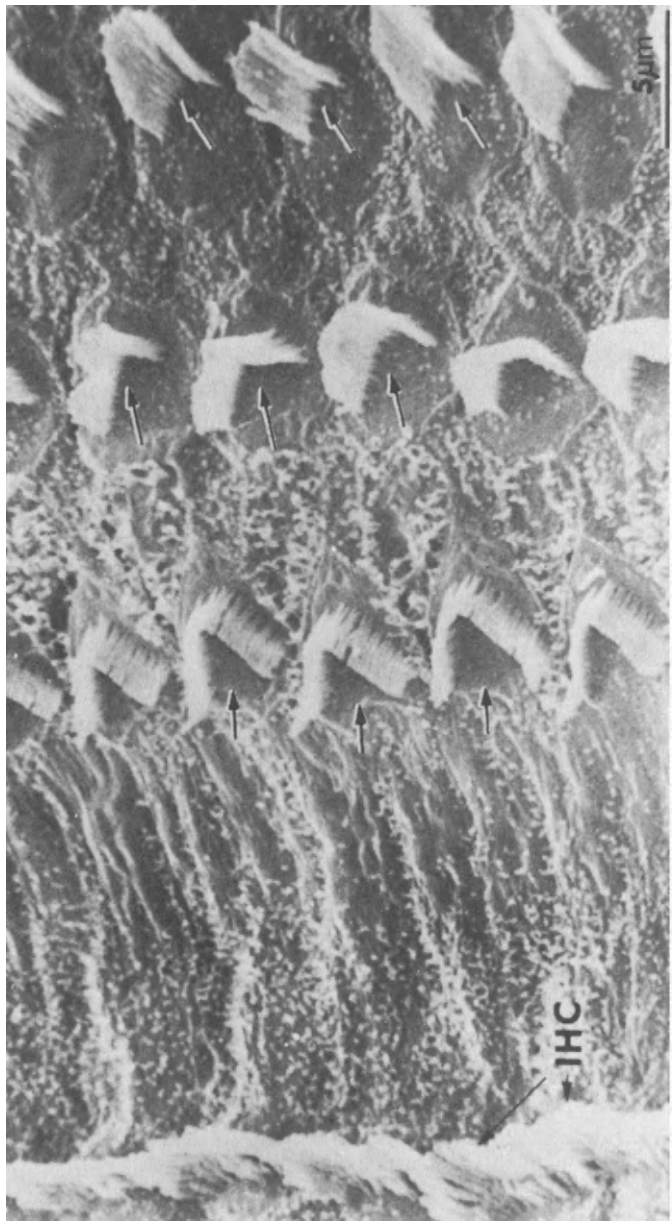

Figure 1.18 Scanning electron micrograph showing the arrangements of the cilia on the inner and outer hair cells. [From D. J Lim [24], Fine morphology of the tectorial membrane: Its relationship to the organ of Corti, *Arch. Otol.* 96, 199–215 (1972) Copyright 1972, American Medical Association.]

The stria vascularis contains a rich network of capillaries, and is attached to the spiral ligament on the lateral wall of the scala media. The stria vascularis has three layers, composed of characteristic cell types. Its hexagonal, marginal epithelial cells are rich in mitochondria and have several microvilli on the endolymphatic surface. The more irregular intermediate cells have a smaller number of mitochondria, and have projections into the marginal layer. The long, flat basal cells form the deepest layer, in contact with the spiral ligament, and are adjacent to and often surround the capillaries. Considerable metabolic activity is suggested by the generous supply of mitochondria in these cells; and the structure of the stria vascularis is consistent with a secretory or absorptive function. However, it is doubtful that the stria vascularis is actually the blood supply of the organ of Corti, as will be discussed below. It appears that its function is that of an ion transport and control structure, producing the ion concentrations in the endolymph that enable energy transformation to occur; it also possibly controls certain inorganic substances and water for the same purpose [10].

Blood Supply to the Cochlea

The arterial supply to the cochlea, the cochlear artery, is derived from the internal auditory artery, which is a branch of the basilar artery. Venous drainage of the cochlea is into the internal auditory vein. The vascular supply of the cochlea is shown schematically in Fig. 1.19 [10]. The system involves one branch supplying (1) the spiral ligament above the point of attachment of Reissner's membrane, (2) the capillary network within the stria vascularis, (3) the spiral prominence, and (4) the terminal branches that descend to collecting venules; and also a second branch feeding (5) the limbus and (6) a plexus under the basilar membrane.

The cochlear blood supply is intriguing since there is no direct contact between its structures and the vascular system. This is complicated by the fact that in addition to the vascularized stria vascularis and spiral eminence, whose proximity to the endolymph makes them intuitively good candidates for supplying the organ of Corti, there is also a capillary bed under the basilar membrane. To determine the source of nutritive support to the HCs, Lawrence [23] interrupted the blood supply of the vessels entering the spiral ligament in one group of guinea pigs, and the supply to the basilar membrane vessels in a second group. Interruption of the basilar membrane blood supply caused HC degeneration without affecting the stria vascularis or spiral eminence. Interrupting the blood supply of the latter structures resulted in their own destruction but had no such effect upon the organ of Corti. It thus appears that the blood supply of the organ of Corti is the basilar membrane vessels, most likely

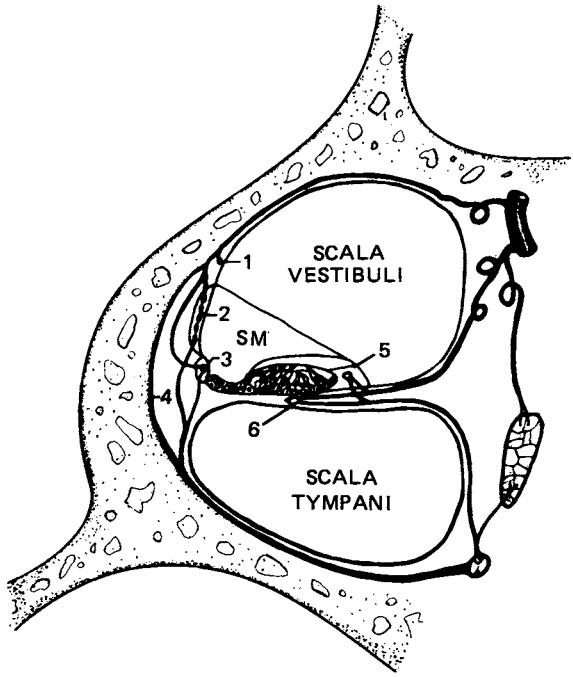

Figure 1.19 Cochlear blood supply (SM, scala media). [From Lawrence [10], "Inner ear physiology," in *Otolaryngology*, Vol. 1 (Paparella and Shumrick, Eds.), © 1973, W. B. Saunders (with permission).]

acting via the cortilymph. On the other hand, the stria vascularis and spiral eminence probably maintain the ionic character of the endolymph.

Innervation of the Cochlea

Rasmussen [26] reported an average of 31,400 neural fibers in the human auditory nerve. These are bipolar neurons (see Chap. 2) whose cell bodies (the spiral ganglia) are in Rosenthal's canal in the modiolus, and whose dendrites course through the habenula perforata to enter the organ of Corti. These neurons are myelinated in the modiolus and osseous spiral lamina, but are unmyelinated within the organ of Corti.

Findings by Fernandez [27] suggested that the innervation of the IHCs is primarily from radial fibers and secondarily from longitudinal (spiral) fibers; whereas spiral fibers are the primary neural supply for the OHCs, with little participation by radial fibers. Radial fibers course directly to the HCs (i.e., essentially perpendicular to the cochlear turns).

Longitudinal (spiral) fibers first radiate out and then turn to follow a course which spirals with the cochlear turns (Fig. 1.20).

The cochlea's sensory innervation is far from equally distributed between the inner and outer HCs. About 95% of the afferent (mainly radial) supply goes to the IHCs, while the remaining 5% consists of the 2500–3000 basilar fibers that cross the tunnel of Corti to innervate the

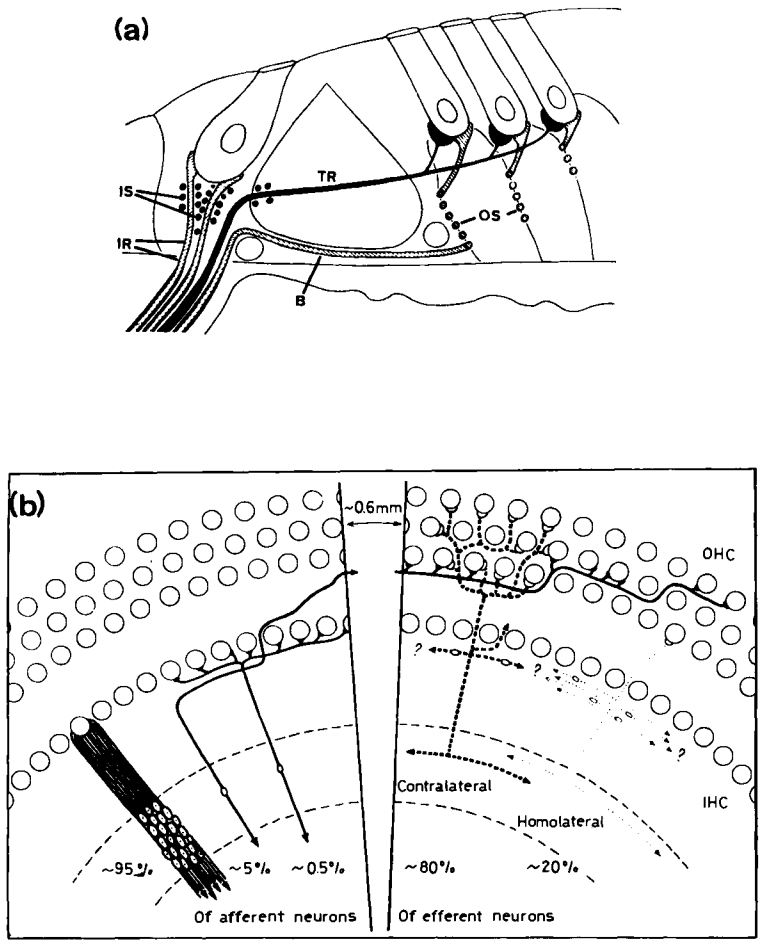

Figure 1.20 Afferent and efferent innervation of the organ of Corti. (a) Afferents gray and efferents black; IS, inner fibers; IR, inner radial fibers; TR, tunnel radial fibers; B, basilar fibers; OS, outer spiral fibers. (b) Arrangement of nerve fibers to the inner and outer hair cells. (From Spoendlin [30], with permission.)

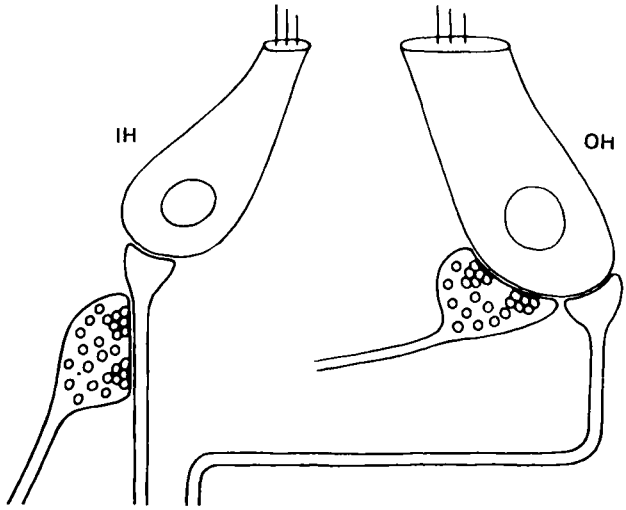

Figure 1.21 Relationship between afferent and efferent fibers and the inner (IH) and outer (OH) cells. (From Spoendlin [30], with permission.)

OHCs* via the outer spiral bundle (Fig. 1.20) [22,28]. Upon entering the organ of Corti, about 20 fibers follow a direct radial course to provide afferent innervation to each IHC; and others turn to form the internal spiral bundle (ISB) which courses beneath the IHCs. The ISB is made up of at least three types of fibers of different diameter, so that it is actually a bundle of a variety of neural elements [29]. Fibers coursing spirally toward the apex give off collaterals that innervate several HCs each, and each HC receives several neurons. Other (basilar) fibers from the ISB course between the inner pillar cells and along the floor of the tunnel of Corti, then up between Deiters' cells to supply the OHCs. These form the outer spiral bundle which follows a route toward the base of the cochlea, with each neuron supplying about 20 OHCs. The 1:10 neuron:HC ratio of the OHCs may thus be viewed as *convergent* upon the nerve, permitting an anatomical basis for spatial summation of stimulation. This cannot be handled by the 20:1 *divergent* relationship of the radial fibers and IHCs [30].

Figures 1.20 and 1.21 show that two different types of fibers have been identified, differing in their endings [32,33]. These are (1) cells with smaller nonvesiculated endings, which are afferent (ascending sen-

*Recent findings by Spoendlin [31] indicate that there are no actual connections to the outer hair cells.

sory) neurons; and (2) cells with larger vesiculated endings, which are efferent (descending) neurons from the olivocochlear (Rasmussen's) bundle. The vesiculated endings contain acetylcholine. Various studies [34,35] found degeneration of the vesiculated cells when parts of the descending olivocochlear pathway were cut. However, there was no degeneration of the nonvesiculated afferent neurons. The endings of the efferent fibers are in direct contact with the OHCs, whereas they terminate on the afferent fibers of the IHCs rather than on the sensory cells themselves (Fig. 1.21). This suggests that the efferents act directly upon the OHCs (*pre*synaptically), but that they act upon the associated afferent fibers of the IHCs (*post*synaptically). While it is generally accepted that the descending pathway has an inhibitory function, the actual mechanism of the efferent control is still unconfirmed [30].

Collaterals of the efferent fibers of the ISB are sent off to the IHCs. Other efferents (tunnel radial fibers) course through the tunnel of Corti to terminate on the OHCs. The density of efferent fibers is substantially greater for the OHCs than for the IHCs. Furthermore, there is greater efferent innervation for the OHCs at the base of the cochlea than at the apex; and this innervation of the OHCs also tapers from the first row through the third.

CENTRAL AUDITORY PATHWAYS

Afferent Pathways

The auditory (or cochlear) nerve appears as a twisted trunk, its core made up of fibers derived from the apex of the cochlea and its outer layers coming from more basal regions. The nerve leaves the inner ear via the internal auditory meatus, and enters the brain at the lateral aspect of the lower pons. Thus, the fibers of the auditory nerve constitute the first-order neurons of the ascending central auditory pathways.

The number of nerve fibers associated with the auditory system increases dramatically at the various way stations from the auditory nerve and the cortex. Chow [36] estimated that these fibers increase in number from roughly 30,000 first-order neurons to approximately 10 million at the auditory cortex in the rhesus monkey.

Upon entering the brainstem, the first-order neurons of the auditory nerve synapse with cell bodies in the dorsal and ventral cochlear nuclei (Fig. 1.22). Specifically, neurons arising from the more basal areas of the cochlea terminate in the dorsal portion of the dorsal cochlear nucleus (DCN); and the ventral cochlear nucleus (VCN) and the ventral portion of the DCN receive neurons originating in the more apical parts

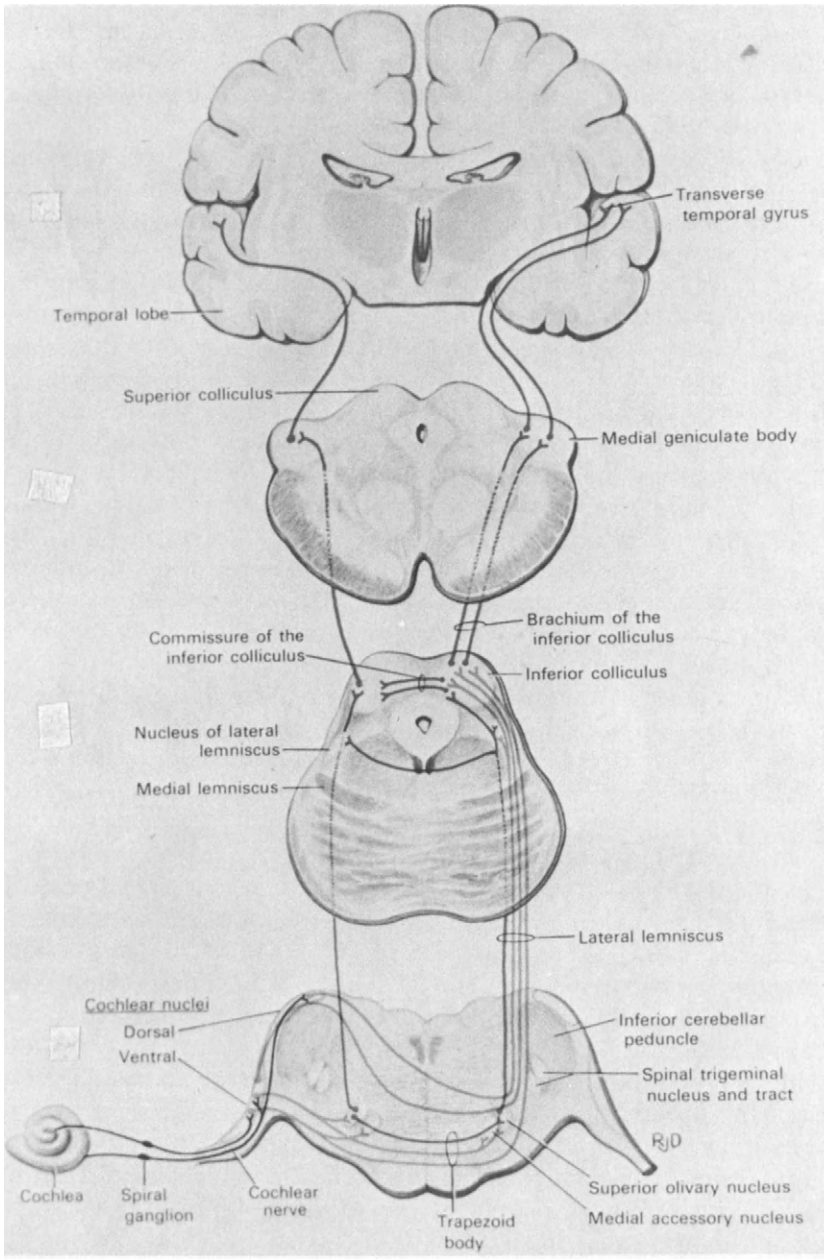

Figure 1.22 Schematic representation of the central auditory pathways. (From M. B. Carpenter [37], *Core Text of Neuroanatomy*, 2nd ed., © 1978, The Williams & Wilkins Co., Baltimore.)

of the cochlea. Several studies [38,39] have shown that degeneration of the cochlear nuclei follows lesions of the cochlea. These place- or frequency-related connections are covered in the discussion of tonotopic organization in Chap. 6.

Second-order neurons arise from the cochlear nuclei. The trapezoid body is formed from the ventral acoustic stria, which arises from the VCN. The fibers of the trapezoid body decussate — cross to the opposite side — to synapse with the nuclei of the contralateral superior olivary complex (SOC) or to ascend in the lateral lemniscus. Other fibers of the trapezoid body terminate at the SOC on the ipsilateral side, and at the trapezoid nuclei. The dorsal part of the VCN gives rise to the intermediate acoustic stria, which contralateralizes to ascend in the lateral lemniscus of the opposite side. The dorsal acoustic stria is made up of fibers from the DCN, which cross to the opposite side and ascend in the contralateral lateral lemniscus.

The superior olivary complex contains a lateral principal nucleus which approximates an S in shape, and a medial accessory nucleus. Fibers from the VCN on both sides mainly synapse with the medial accessory nucleus; the medial accessory nucleus in turn is the source of third-order neurons which ascend via the lateral lemniscus on the same side. Thus, the SOC receives a *bilateral* representation (at least from the VCN), and the lateral lemniscus includes fibers from the SOC as well as from the trapezoid body and acoustic stria. While fibers from the cochlear nuclei of one side ascend directly up the opposite lateral lemniscus, it appears that there are no fibers from the homolateral cochlear nuclei [40].

Also ascending within the lateral lemniscus are fibers arising from the several nuclei of the lateral lemniscus itself. Although the nuclei of the lateral lemniscus have classically been viewed as dorsal and ventral, recent evidence [41] has revealed a third nucleus which appears to be a continuation of the nuclei stemming from the SOC. Furthermore, the ventral nucleus, and to a slightly lesser extent the dorsal nucleus, are highly variable; they are in fact even quite variable between the right and left sides of the same subject. Communication between the lateral lemnisci of the two sides is via the commissural fibers of Probst, which appear to involve, at least primarily, the dorsal nuclei.

The majority of the ascending fibers synapse with the nuclear mass in the inferior colliculus (IC) at the level of the midbrain. There is also communication between the colliculi of the two sides via the commissure of the inferior colliculus. Several fibers may pass the IC and follow a direct course to the medial geniculate body (MGB) of the thalamus. The pathway from the IC to the MGB goes by way of the brachium of the in-

ferior colliculus. This pathway does not appear to contain neuron bodies in the human adult, and is made up of fibers from the IC as well as of the fibers that bypass the colliculus as they ascend to the MGB [42].

The medial geniculate body is the last subcortical way station, and all ascending pathways to the auditory cortex synapse here. The MGB has two main parts, the par principalis and a ventral nucleus. Fibers ascending from the IC terminate primarily in the par principalis.

The auditory (geniculotemporal) radiations project from the MGB to the transverse temporal gyrus on the temporal cortex of the same side. Also referred to as Heschl's gyrus, this cortical area corresponds to areas 41 and 42 of Brodmann's classification system (Fig. 1.23). Area 41 is the primary auditory cortex. Area 41, which consists of the middle portion of the anterior and part of the posterior temporal gyri, is actually obscured from view within the central sulcus. Area 42, which is an auditory association area, is adjacent to area 41 on parts of the posterior transverse and superior temporal gyri.

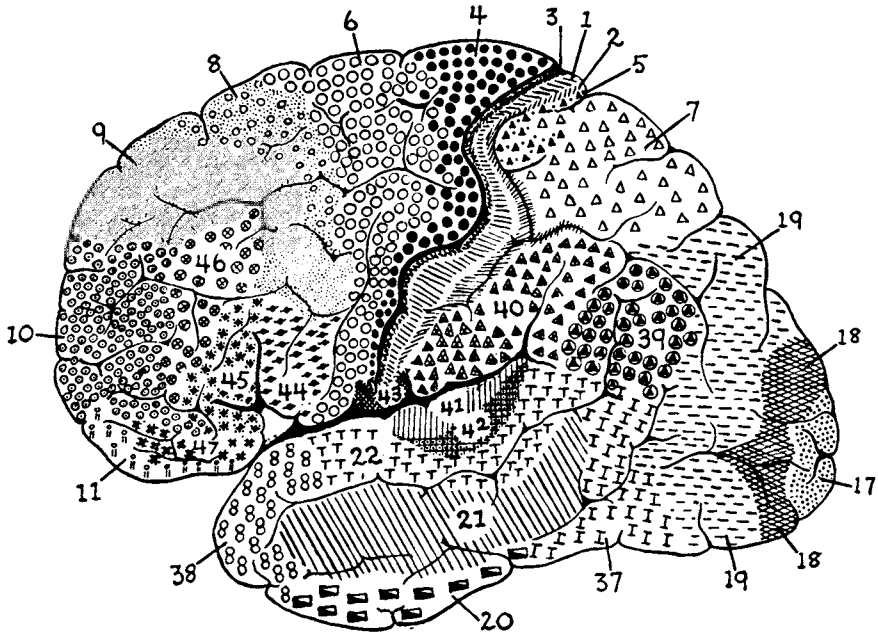

Figure 1.23 Lateral view of the human brain showing Brodmann's areas. (From M. B. Carpenter [43], *Human Neuroanatomy,* 7th ed., © 1976, The Williams & Wilkins Co., Baltimore.)

Experiments on monkeys [44,45] indicate that the primary auditory cortex communicates with the auditory association cortex by way of short neural fibers. The auditory association area in turn gives rise to generous projections to other parts of the brain, such as the temporoparietal cortex. There appears to be interhemispheric communication between the auditory association cortices of each side via the corpus callosum but the interhemispheric connections between the primary auditory areas on both sides are not as homotopically related. Interhemispheric projection between the auditory areas of the two temporal lobes appears to be greater for the medial than for the lateral parts of these areas.

Efferent Pathways

As described above, descending efferent fibers enter the inner ear and make contact with the OHCs directly, and with the IHCs indirectly via synapses with the associated afferent fibers. These efferent fibers of the auditory nerve represent the terminal portion of the descending olivocochlear tract, which was first described by Rasmussen [46,47] and which is often referred to as Rasmussen's bundle. The olivocochlear bundle (Fig. 1.24) is composed of both crossed and uncrossed pathways. Rasmussen [48] described the parent portion of the efferent pathway as being composed of about 500 fibers, although the number of fine collateral fibers entering the organ of Corti is far greater. The crossed pathway contributes roughly four times as many fibers as does the uncrossed bundle.

The crossed efferent bundle in the cat originates in the dorsomedial periolivary nucleus of the medial superior olive, though there are some contributions from the ventral and medial nuclei of the trapezoid body [49] These fibers cross under the fourth ventricle to exit from the brainstem on the opposite side in association with the vestibular nerve. The uncrossed efferents arise from the lateral periolivary nucleus, join the bundle from the opposite side within the brainstem, and enter the ipsilateral inner ear with the vestibular nerve. The efferent fibers are distributed to the inner and outer HCs in the manner described above.

In addition to the crossed and uncrossed olivocochlear pathways, Rasmussen [48,50,51] has also demonstrated efferent connections to the cochlear nuclei from the inferior colliculus, the S-shaped segment of the superior olivary complex, the nuclei of the lateral lemniscus, and possibly the cerebellum. Descending fibers may also provide cortical control over lower centers (corticofugal pathways), although these are beyond the scope and intent of this text.

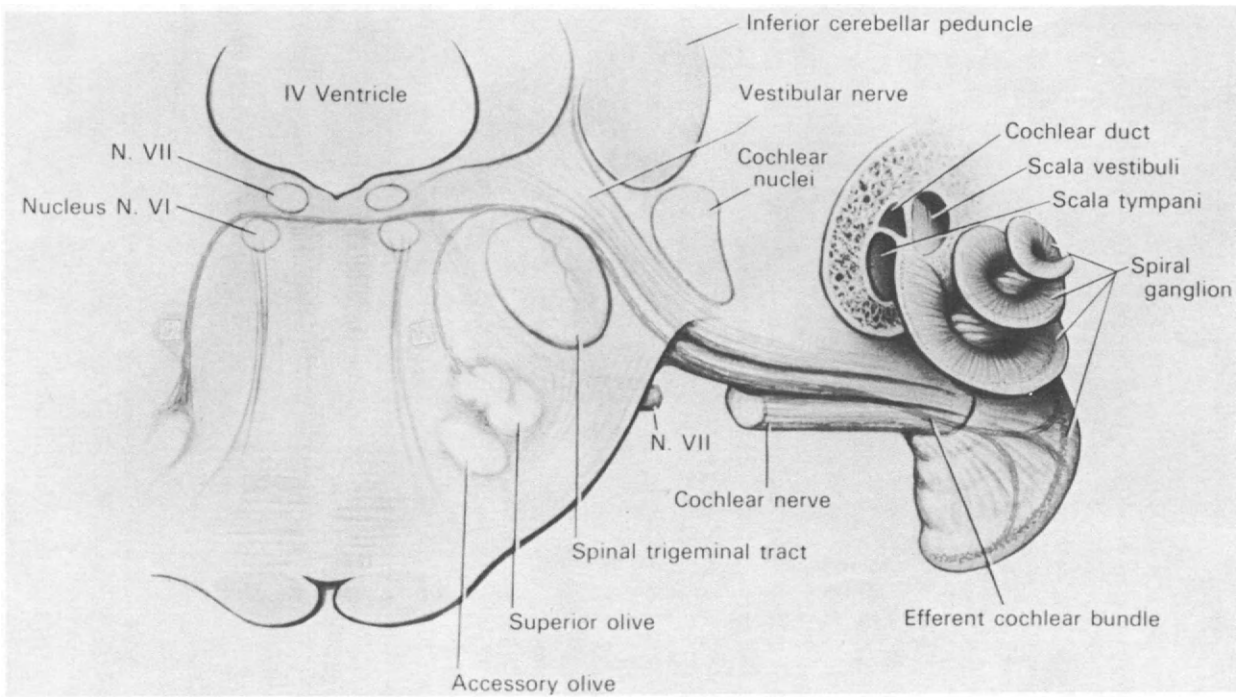

Figure 1.24 Crossed and uncrossed olivocochlear (Rasmussen's) bundle. (Based on Rasmussen [48]. From M. B. Carpenter [43], *Human Neuroanatomy,* 7th ed., © 1976, The Williams and Wilkins Co., Baltimore.)

REFERENCES

1. G. Bekesy, *Experiments in Hearing*, McGraw-Hill, New York, 1960.
2. B. J. Anson and J. A. Donaldson, *The Surgical Anatomy of the Temporal Bone and the Ear*, Saunders, Philadelphia, 1967.
3. E. G. Wever and M. Lawrence, *Physiological Acoustics*, Princeton Univ. Press, Princeton, N.J., 1954.
4. J. A. Donaldson and J. M. Miller, Anatomy of the ear, in *Otolaryngology* (M. M. Paparella and D. A. Shumrick, Eds.), Vol. 1: *Basic Sciences and Related Disciplines*, Saunders, Phiadelphia, 1973, pp. 75–110.
5. E. Garner, D. J. Gray, and R. O'Rahilly, *Anatomy*, 4th ed., Saunders, Philadelphia, 1975.
6. H. Helmholtz, Die Mechanik der Gehörknöchelchen und des Trommelfells, *Pflüg. Arch. ges. Physiol. I*, 1–60 (1868).
7. G. Bekesy, Zur Physik des Mittelohres und über das Hören bei fehlerhaftem Trommelfell, *Akust. Zeitschr. I*, 13–23 (1936).
8. O. Jepsen, Middle-ear muscle reflexes in man, in *Modern Developments in Audiology*, (J. Jerger, Ed.), Academic, New York, 1963, pp. 194–239.
9. E. Borg, On the neuronal organization of the acoustic middle ear reflex: A physiological and anatomical study, *Brain Res. 49*, 101–123 (1973).
10. M. Lawrence, Inner ear physiology, in *Otolaryngology* (M. M. Paparella and D. A. Shumrick, Eds.), Vol. 1: *Basic Sciences and Related Disciplines*, Saunders, Philadelphia, 1973, pp. 275–298.
11. C. A. Smith, O. H. Lowry, and M-L Wu, The electrolytes of the labyrinthine fluids, *Laryngoscope 64*, 141–153 (1954).
12. S. Rauch and A. Köstlin, Aspects chemiques de l'endolymphe et de la perilymphe, *Pract. Otol. 20*, 287–291 (1958).
13. S. Rauch, Biochemische Studien zum Hörvorgang, *Pract. Otol. 25*, 81–88 (1963).
14. I. Kirakae, Y. Nomura, M. Nagakura, Y. Matsuba, and S. Sugiura, A consideration on the circulation of perilymph, *Ann. Otol. 70*, 337–343 (1961).
15. R. Kimura, Experimental blockage of the endolymphatic duct and sac and its effect on the inner ear of the guinea pig, *Ann. Otol. 76*, 664–687 (1967).
16. H. Engström and J. Wersäll, Is there a special nutritive cellular system around the organ of Corti? *Ann. Otol. 62*, 507–512 (1953).
17. H. Engström, The cortilymph, the third lymph of the inner ear, *Acta Morphol. Neerl. Scand. 3*, 195–204 (1960).
18. W. R. Zemlin, *Speech and Hearing Science: Anatomy and Physiology*, Prentice-Hall, Englewood Cliffs, N.J., 1968.
19. H. Davis, Advances in the neurophysiology and neuroanatomy of the cochlea, *J. Acoust. Soc. Amer. 34*, 1377–1385 (1962).
20. J. R. Brobeck, Ed., *Best and Taylor's Physiological Bases of Medical Practice*, 9th ed., Williams and Wilkins, Baltimore, 1966.
21. G. Bourne and J. Danielli, Eds., *International Review of Cytology*, Vol. III, Academic, New York, 1958.
22. H. Spoendlin, *The Organization of the Cochlear Receptor*, Karger, Basel, 1966.

23. M. Lawrence, Effects of interference with terminal blood supply on the organ of Corti, *Laryngoscope 76*, 1318–1337 (1966).
24. D. J. Lim, Fine morphology of the tectorial membrane: Its relationship to the organ of Corti, *Arch. Otol. 96*, 199–215 (1972).
25. D. J. Lim, Scanning electron microscopic morphology of the ear, in *Otolaryngology* (M. M. Paparella and D. A. Shumrick, Eds.), Vol. 1: *Basic Sciences and Related Disciplines*, Saunders, Philadelphia, 1973, pp. 527–550.
26. G. Rasmussen, Studies of the VIIIth cranial nerve of man, *Laryngoscope 50*, 67–83 (1940).
27. C. Fernandez, The innervation of the cochlea (guinea pig), *Laryngoscope 61*, 1152–1172 (1951).
28. H. Spoendlin, Degeneration behavior in the cochlear nerve, *Arch Klin. Exp. Ohren. Nasen. Kehlkopfheilkd 200*, 275–291 (1971).
29. H. Engstrom, H. W. Ades, and A. Andersson, *Structural Pattern of the Organ of Corti*, Almqvist & Wiksell, Stockholm, 1966.
30. H. Spoendlin, Neuroanatomical basis of cochlear coding mechanisms, *Audiology 14*, 383–407 (1975).
31. H. Spoendlin, Neural connection of the outer hair-cell system, *Acta Otol. 87*, 381–387 (1979).
32. H. Engstrom, On the double innervation of the inner ear sensory epithelia, *Acta Otol. 49*, 109–118 (1958).
33. C. A. Smith and F. S. Sjöstrand, Structure of the nerve endings of the guinea pig cochlea by serial sections, *J. Ultrasound Res. 5*, 523–556 (1961).
34. C. A. Smith and G. L. Rasmussen, Recent observations on the olivo-cochlear bundle, *Ann. Otol. 72*, 489–506 (1963).
35. H. Spoendlin and R. R. Gacek, Electromicroscopic study of the efferent and afferent innervation of the organ of Corti, *Ann. Otol. 72*, 660–686 (1963).
36. K. Chow, Numerical estimates of the auditory central nervous system of the monkey, *J. Comp. Neurol. 95*, 159–175 (1951).
37. M. B. Carpenter, *Core Text of Neuroanatomy*, 2nd ed., Williams and Wilkins, Baltimore, 1978.
38. T. P. S. Powell and W. M. Cowan, An experimental study of the projection of the cochlea, *J. Anat. 96*, 269–284 (1962).
39. N. Moskowitz and J. C. Liu, Central projections of the spiral ganglion of the squirrel monkey, *J. Comp. Neurol. 144*, 335–344 (1972).
40. W. T. Barnes, H. W. Magoun, and S. W. Ranson, The ascending auditory projection in the brainstem of the monkey, *J. Comp. Neurol. 79*, 129–152 (1943).
41. J. A. Ferraro and J. Minckler, The human auditory pathways: A quantitative study: The human lateral lemniscus and its nuclei, *Brain Language 4*, 277–294 (1977).
42. J. A. Ferraro and J. Minckler, The human auditory pathways: A quantitative study: The brachium of the inferior colliculus, *Brain Language 4*, 156–164 (1977).
43. M. D. Carpenter, *Human Neuroanatomy*, 7th ed., Williams and Wilkins, Baltimore, 1976.

44. D. N. Pandya, M. Hallett, and S. Mukherjee, Intra- and inter-hemispheric connections of neocortical auditory system in the rhesus monkey, *Brain Res. 14*, 49–65 (1969).
45. E. A. Karol and D. N. Pandya, The distribution of the corpus callosum in the rhesus monkey, *Brain 94*, 471–486 (1971).
46. G. L. Rasmussen, The olivary peduncle and other fiber projections of the superior olivary complex, *J. Comp. Neurol. 84*, 141–219 (1946).
47. G. L. Rasmussen, Further observations of the efferent cochlear bundle, *J. Comp. Neurol. 99*, 61–74 (1953).
48. G. L. Rasmussen, Efferent fibers of the cochlear nerve and cochlear nucleus, in *Neural Mechanisms of the Auditory and Vestibular Systems* (G. L. Rasmussen and W. F. Windle, Eds.), Thomas, Springfield, Ill., 1960.
49. G. D. Luk, D. K. Morest, and N. M. McKenna, Origins of the crossed olivocochlear bundle shown by an acid phophatase method in the cat, *Ann. Otol. 83*, 382–392 (1974).
50. G. L. Rasmussen, Anatomical relationships of the ascending and descending auditory systems, in *Neurological Aspects of Auditory and Vestibular Disorders* (W. S. Field and B. R. Alford, Eds.), Thomas, Springfield, Ill., 1964.
51. G. L. Rasmussen, Efferent connections of the cochlear nucleus, in *Sensorineural Hearing Processes and Disorders* (A. B. Graham, Ed.), Little, Brown, Boston, 1967.

2
Hearing Theories and Sensory Action

The next four chapters are concerned with how information is transmitted and processed at various levels of the auditory system. Before proceeding, we shall briefly review the traditional theories of hearing as well as some principles of sensory receptor action. These two topics should provide a general framework within which the student can consider the material that follows.

THEORIES OF HEARING

The study of the auditory system is practically and historically intertwined with the theories of hearing. Broadly speaking, these theories fall into two general categories — place (resonance) theories and frequency (temporal, periodicity) theories — as well as the combined place–frequency theory. A detailed review is beyond the current scope or intent; the student is referred to Wever's [1] classic work, *Theory of Hearing*, for an excellent critical review of these theories.

Place Theories

Although place (or resonance) theories existed since the beginning of the 1600s, modern versions began with the resonance theory of Helmholtz [2] in the late 1800s. Helmholtz relied to a large extent upon Ohm's auditory "law" and Müller's doctrine of specific nerve energies. Ohm's auditory law states that the ear performs a Fourier analysis upon complex periodic sounds; i.e., that it breaks the complex wave down into its components regardless of their phase relationships. A major problem

with Ohm's auditory law is that it precludes temporal analysis. We shall see, however, that the auditory system is sensitive to temporal as well as frequency parameters. Müller's doctrine refers to the specificity of the different senses. It states that the neural signal coming from the ear is interpreted as sound whether the actual stimulus was a tone or a blow to the head; the eye elicits a visual image whether the stimulus is light or pressure on the eyeball, etc. This doctrine appears to hold on the periphery although there are dramatic commonalities among the various senses in terms of their fundamental principles of operation (see Action of Sensory Receptors, below) and central mechanisms.

The mechanism of the place theory proposed by Helmholtz is relatively simple and straightforward. It assumes that the basilar membrane is composed of a series of segments, each of which resonates in response to a particular frequency. Thus, an incoming stimulus results in the vibration of those parts of the basilar membrane whose natural frequencies correspond to the components of the stimulus. Since these resonators are arranged by place along the cochlear partition, the precise place of the vibrating segment would signal the existence of a component at the natural frequency of that location. Nonlinear distortions introduced by the ear (such as combination tones due to the interaction of two stimulus tones, or harmonics of the stimulus tone) are viewed as being generated by a nonlinear response of the middle ear mechanism. These distortion products are then transmitted to the cochlea, where they cause vibrations at the places whose resonant frequencies correspond to the frequency of the combination tone (or harmonic). The distortion product is thus perceived as though it were present in the original signal.

Such a strict place theory is faced with several serious problems. To begin with, in order to account for the sharp frequency tuning of the inner ear, the theory demands that the basilar membrane contain segments which are under differing amounts of tension in a manner analogous to the tension on variously tuned piano strings. However, Bekesy [3] demonstrated that the basilar membrane is under no tension at all.

A second problem is that strict place theory cannot account for the perception of the "missing fundamental" (Chap. 12), the phenomenon in which the presence of harmonics of a tone (e.g. 1100, 1200, and 1300 Hz) results in the perception of the fundamental frequency (100 Hz) even though the latter is not physically present.

Place theory is also plagued by the relationship between the sharpness of a system's tuning and the persistence of its response. In order for the ear to achieve the required fine frequency discriminations, the various segments of the basilar membrane must be sharply tuned. In other

words, they could each respond only to a very narrow range of frequencies. A segment could not respond to higher or lower frequencies, or else the necessary discriminations would be impossible. The problem is that such a narrowly tuned system must have very low damping — its response will take a relatively long time to die away after the stimulus stops. In other words, if there were sharp tuning of the resonators along the basilar membrane, then their responses would persist long after the stimulus has ceased. This situation would cause an interminable echo in our ears, precluding any functional hearing. On the other hand, if the resonators were less sharply tuned, they would not have the persistence problem, but they would be unable to support the necessary fine frequency discriminations.

Place theory ascribes distortions to the middle ear. However, considerable research has revealed that the middle ear's response is astoundingly linear, and that most nonlinear distortions are attributable to the cochlea (Chaps. 3 and 4). This fact is obviously inconsistent with strict place theory.

A variety of other place theories followed that of Helmholtz [2]. Of particular interest is the traveling wave theory of Nobel laureate Georg von Bekesy [3]. Bekesy's traveling wave theory has been confirmed by many investigations using a variety of approaches, and is discussed in detail in Chap. 4.

Frequency Theories

Frequency (or temporal) theory proposes that the peripheral hearing mechanism does not perform a frequency analysis, but rather that it transmits the signal to the central auditory nervous system for processing. Such theories have been referred to as "telephone theories" by analogy with the manner in which a telephone signal is transmitted en toto. Though there are several such theories, Rutherford's [4] telephone theory, proposed not long after Helmholtz's theory, has been the best known. This theory states that the cochlea is not frequency-sensitive along its length, but rather that all parts respond to all frequencies. The job of the hair cells is simply to transmit all parameters of the stimulus waveform to the auditory nerve, and analysis is performed at higher levels.

Since a neuron can only respond in an all-or-none manner, the only way in which it can of itself transmit frequency information is to discharge the same number of times per second as there are cycles in the stimulus (e.g., it must fire 720 times per second to transmit a 720 Hz tone). Frequency theory thus presumes that auditory nerve fibers can

fire fast enough to represent this information. There is no problem at low frequencies; however, an upper limit on the number of discharges per second is imposed by the absolute refractory period of the neuron. The absolute refractory period is the time required after discharging for the cell to reestablish the polarization it needs to fire again; it lasts about 1 msec. The fiber cannot fire during the absolute refractory period, no matter how intensely stimulated. This period is followed by a relative refractory period during which the neuron will respond provided the stimulus is strong enough. The 1 msec absolute refractory period corresponds to a maximum firing rate of 1000 times per second. Thus, simple frequency theory is hard pressed to explain how sounds higher in frequency than about 1000 Hz can be transmitted by the auditory nerve and perceived by the listener.

A second problem of the telephone theories is that damage to the apical parts of the cochlea result in high-frequency hearing loss. This is contradictory to frequency theory, which states that the different parts of the cochlea are not differentially sensitive to frequency. Furthermore, there is extensive data which demonstrates considerable frequency selectivity along the cochlear partition (Chap. 4).

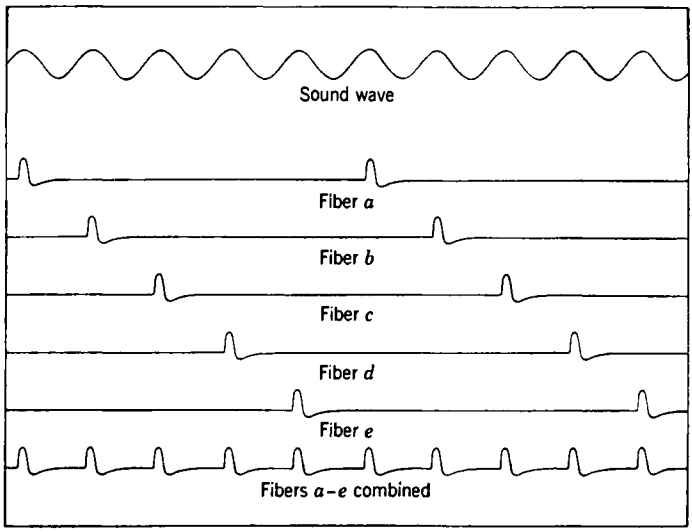

Figure 2.1 Diagrammatic representation of the volley principle (see text). [From E. G. Wever [1], p. 167, Fig. 33. *Theory of Hearing*, Dover, New York (1949).]

Rutherford's telephone theory was by no means the only frequency theory of hearing, but it serves to demonstrate the fundamental approach and associated problems. The most important modification of frequency theory was the volley principle advanced by Wever [1]. Instead of suggesting that any one neuron must carry all of the information burden, the volley principle states that groups of fibers cooperate to represent the stimulus frequency in the auditory nerve. This is shown in Fig. 2.1. The sinusoid (sound wave) at the top of the figure has a frequency too high to be represented by a series of spike discharges from a single auditory nerve fiber. Instead, fibers work in groups so that in the total response of the group there is a spike corresponding to each cycle of the stimulus. This cooperation is accomplished by having each individual neuron respond to cycles separated by some interval. In Fig. 2.1 this interval is every 5 cycles. Thus, fiber a discharges in response to cycles 1, 6, and 11; fiber b to cycles 2, 7, and 12; fiber c to 3, 8, and 13; etc. The result is that each cycle is represented by a spike in the combined response of the fiber group (bottom line in the figure). We shall see in Chap. 5 that the discharges of auditory nerve fibers do appear to follow the periodicity of the stimulus for frequencies as high as about 4000–5000 Hz. However, these responses are probabilistic rather than one-to-one.

Place-Volley Theory

Even at this early point it should be apparent that neither the place nor frequency theory alone can explain the selectivity of the ear. Both are operative. A frequency or periodicity mechanism is most important for low frequencies, while a place mechanism is paramount for high-frequency representation [1,5]. The question is not one of where the "cutoff points" are, since these do not exist. As we shall see in the following chapters, place coding below approximately 300–400 Hz is too broad to reasonably account for frequency discrimination, and periodicity coding is not supported for frequencies above roughly 4000–5000 Hz. Frequency coding in the wide range between these two extremes appears to occur by the interaction of both mechanisms.

ACTION OF SENSORY RECEPTORS

The auditory system is one of several specialized sensory systems. Although there is specificity of the senses at the periphery, there are nevertheless remarkable similarities among them. The following is a brief overview of sensory receptor action (6,7,8) with particular reference to the ear.

Davis [6] proposed a general plan of sensory receptor action, which is outlined schematically in Fig. 2.2. This model describes how external stimulus energy is transmitted and coded into a form which is usable by the central nervous system. The sensory neuron is common to all sensory systems, although specialized receptor cells (sense organ) and accessory structures may or may not be present, and are different for the various senses. In the ear, the conductive mechanisms and the parts of the cochlea other than the hair cells constitute the accessory structures. The hair cells are the specialized receptors for hearing.

The accessory structures assist in the action of the sense organ, but do not actually enter into the sensory process per se. In other words, the conductive and cochlear structures help the hair cell to do its job, but are not themselves receptors. In the ear, the accessory structures carry out a large variety of vital functions. They receive, amplify, and analyze the sound stimulus, and convert it to a form usable by the hair cells (Chaps. 3 and 4). They also perform a feedback or inhibitory function via the acoustic reflex (under the control of efferent neurons), and protect the sensory receptors from external damage.

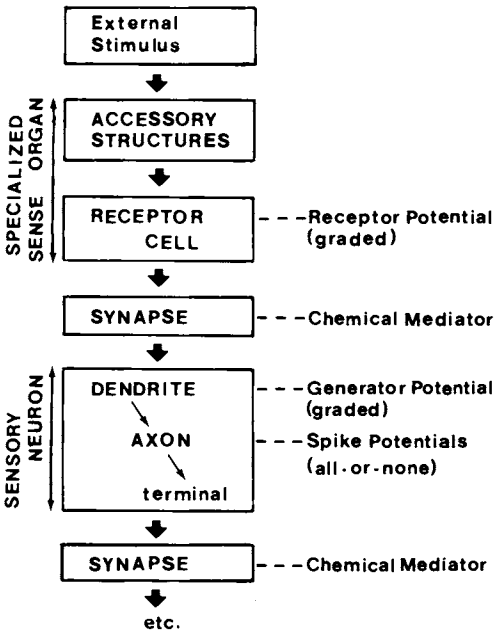

Figure 2.2 General plan of sensory receptor action (efferent feedback not shown). (Adapted from Davis [6].)

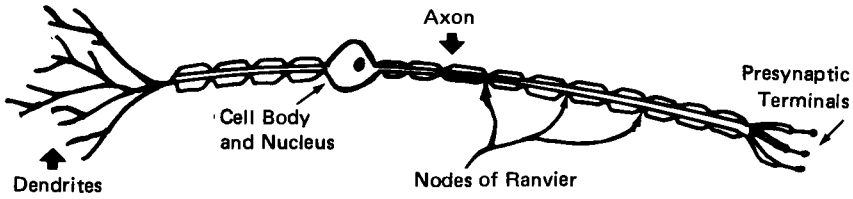

Figure 2.3 Typical bipolar sensory neuron.

The receptor cell transduces the stimulus and transmits it to the sensory neuron — an electrochemical event. An electrical potential associated with this process can be detected in the hair cells. This is their receptor potential. The receptor potential is graded, meaning that its magnitude depends upon the intensity of the stimulus. This potential reflects the activity of the sensory cell, but does not directly transmit information to the neuron. The receptor cell also emits a chemical mediator which is transmitted across the synapse between the hair cell and sensory neuron. It is this chemical mediator which excites the neuron.

A neuron has three major parts: the cell body, the dendrites, and the axon (Fig. 2.3). Most sensory neurons are bipolar — so called because the cell body is located part way along the axon. Auditory neurons are of this type. Communication across a synapse (between the hair cell and a neuron or between two neurons) is by the release and absorption of chemical mediators. Dendrites are excited chemically rather than electrically. On the other hand, axons are excited electrically, so that communication along the length of the neuron is electric in the form of all-or-none action (spike) potentials. Instead of being conducted along the length of the neuron, the spike discharges jump from one node of Ranvier to the next, which speeds the process considerably.

The dendrite receives an amount of chemical mediator from the hair cell, which elicits a graded postsynaptic potential. The postsynaptic potential is called a generator potential because it provides the electrical stimulus which triggers the all-or-none spike discharges from the axon. When the magnitude of the generator potential is great enough, it activates the initial segment of the axon, which in turn produces the spike potential (nerve impulse). The material in Chap. 5 on the activity of the auditory nerve is based upon these action potentials. The spikes travel down the axon to its terminus, where the presynaptic endings emit a chemical mediator. This chemical mediator crosses the synaptic junction to excite the dendrites of the next neuron, and the process is repeated.

REFERENCES

1. E. G. Wever, *Theory of Hearing*, Dover, New York, 1949.
2. H. Helmholtz, *Die Lehre von den Tonempfindugen* (English translation: A. J. Ellis, *On the sensation of tones,* 1885).
3. G. Bekesy, *Experiments in Hearing*, McGraw-Hill, New York, 1960.
4. W. Rutherford, A new theory of hearing, *J. Anat. Physiol. 21,* 166–168 (1886).
5. E. G. Wever and M. Lawrence, *Physiological Acoustics*, Princeton Univ. Press, Princeton, N.J., 1954.
6. H. Davis, Some principles of sensory receptor action, *Physiol. Rev. 4*, 391–416 (1961).
7. H. Grundfest, The general electrophysiology of input membrane in electrogenic excitable cells, in *Handbook of Sensory Physiology* (W. R. Lowenstein, Ed.), Vol. 1: *Principles of Receptor Physiology,* Springer-Verlag, New York, 1971, pp. 136–165.
8. J. Tonndorf, Davis-1961 revisited: Signal transmission in the cochlear hair cell-nerve junction, *Arch. Otol. 101*, 528–535 (1975).

3
Conductive Mechanism

This chapter deals with the routes over which sound is conducted to the inner ear. The first section is concerned with the usual air conduction path through the outer and middle ear. The second section briefly discusses the bone conduction route. In the last section, we shall go into some detail about the acoustic reflex, which is a topic of growing scientific and clinical interest.

AIR CONDUCTION ROUTE

Outer Ear

Pinna

The pinna has traditionally been credited with funneling sounds into the ear canal and enhancing localization. It has been demonstrated, however, that hearing sensitivity is not affected when the pinna is excluded from sound conduction by bypassing it with tubes to the canal and by filling its depressions [1]. Thus, the sound collecting/funneling function is not significant for the human pinna.

The pinna's greatest contribution to hearing is probably in sound source localization. The obvious example occurs when a pet turns its ears toward a ringing doorbell. This action is impossible for man, whose extrinsic ear muscles are vestigial. However, head turning in both humans and animals orients the ears relative to the stimulus, which, along with differences in the intensity and time of arrival of a sound at the two ears, permits accurate localization. The pinna itself contributes to localization in man, particularly when listening with only one ear (monaurally) and

when the sound source is in the median plane of the head rather than off to one side. Pinna effects are important in these two situations because monaural hearing precludes the use of the interaural differences available during binaural hearing, and because these interaural differences are minimized in the median plane, where the sound source is equidistant from both ears.

The depressions and ridges of the pinna contribute to median-plane localization [2–5], as is shown by increases in the number of localization errors made when the various depressions are filled. The contribution of the pinna to localization is due to variations in the stimulus spectrum caused by the structure of the pinna. Blauert [6] suggested that the pinna acts as a filter that attenuates or passes frequencies depending on their direction. Batteau [7] showed that sounds reflected from the structures of the pinna cause time delays between the reflected and direct sounds on the order of under 300 μsec. Since these very small delays are detectable [4], they may be used in median-plane and monaural localization. Frequencies over 4000 Hz are particularly important in these effects.

Ear Canal

The eardrum is located at the end of the ear canal rather than flush with the skull surface. Sounds reaching the drum are thus affected by the acoustic characteristics of the canal, as well as by those of the pinna. Sounds at the eardrum are also affected by the azimuth—the angle at which the ear is oriented toward the sound source (Fig. 3.1). This effect is due to the reflection and diffraction of sounds from the head and other body parts, and also to the acoustic shadow cast by the head.

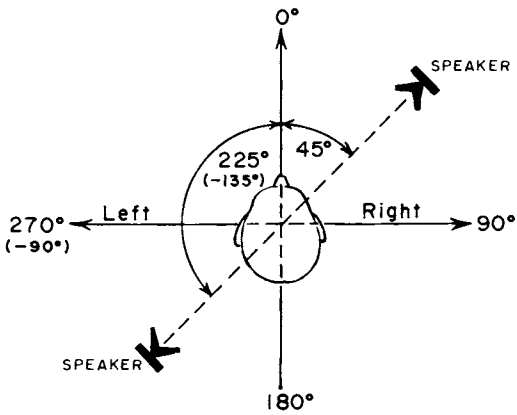

Figure 3.1 Azimuth angles around the head. The loudspeaker toward the right is at an azimuth of 45°; the one at the left has an azimuth of 225° (or −135°).

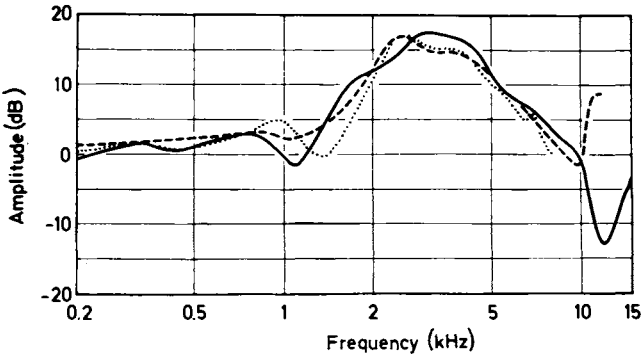

Figure 3.2 Ear-canal transfer functions at 0° azimuth (Winer and Ross [18] dotted; Shaw [9] dashed; Mehrgardt and Mellert [10] solid). (From Mehrgardt and Mellert [10]. Permission *J. Acoust. Soc. Amer.*)

A head shadow occurs when the head is between the sound source and the ear being investigated, and may be thought of as analogous to an eclipse. The signal coming from the left speaker (at 225° azimuth) in Fig. 3.1 will be attenuated at the right ear because the head is, so to speak, blocking the path. This head shadow is significant for frequencies over 1500–2000 Hz because their wavelengths are small compared to the size of the head.

The ear canal may be conceived of as a tube open at one end and closed at the other. Such a tube resonates at the frequency with a wavelength four times the length of the tube. Since the human ear canal is about 2.3 cm long, its resonance should occur at the frequency of wavelength 9.2 cm, i.e., at about 3800 Hz. One could test this hypothesis by directing a known sound into a sound field, and then monitoring the sound pressure at the eardrum of a subject sitting in that sound field. This test has been done in various ways in many studies. Figure 3.2 shows the results of three studies [8–10] in the form of transfer functions. (Transfer functions show the relationship between the input to a system and its output. See Dallos [11] for a complete discussion.) The figure shows how sounds presented from a speaker in front of the subject are affected by the ear canal. The common finding of these studies is a wide resonance peak in the vicinity of 4000 Hz. The peak is relatively smooth and extends from below 2000 Hz to above 5000 Hz. It does not resemble the sharp resonance of a rigid tube. However, the ear canal is an irregular rather than a simple tube, and the drum and canal walls are absorptive rather than rigid. These factors introduce damping. Also, group (averaged) data are shown in the figure. Since resonant frequency

depends on canal length, variations in ear canal length among subjects will widen and smooth the averaged function. The important point, however, is that the resonance characteristics of the canal serve to boost the level of sounds in the mid-high frequencies by as much as ~15 dB at the eardrum compared to the sound field.

The sound field to eardrum transfer function varies with azimuth. Figure 3.3 shows data by Shaw [9] at 45° azimuth increments. Positive azimuth angles denote sounds coming from the same side as the test ear, whereas negative azimuths indicate that the sound source is on the

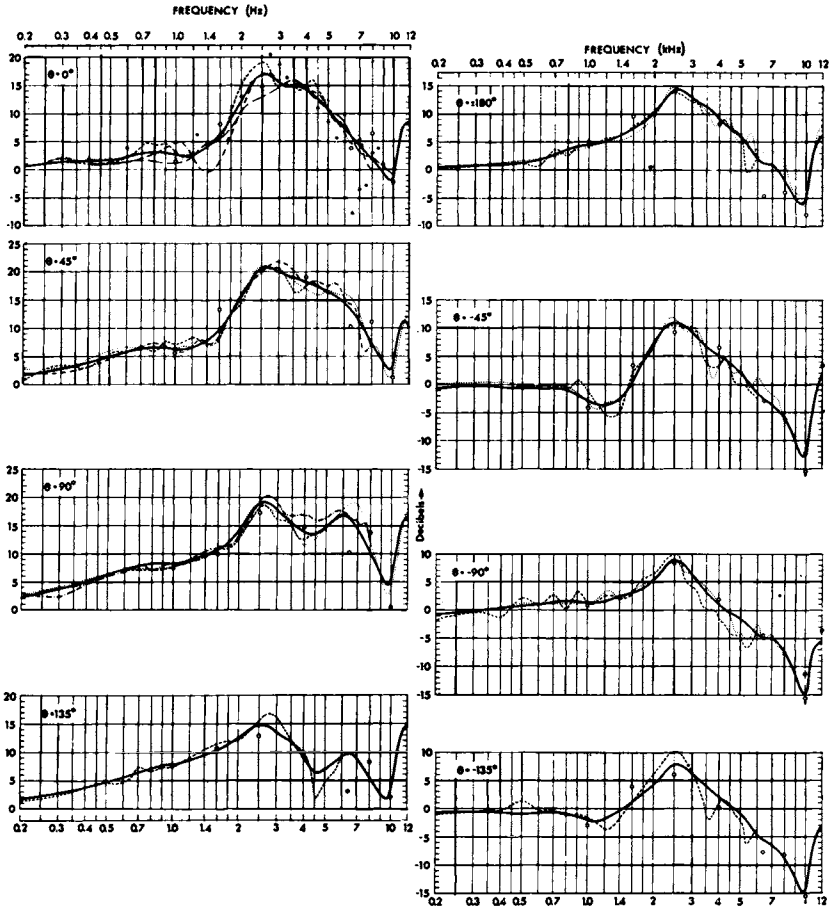

Figure 3.3 Ear canal transfer functions at various azimuths. (From Shaw [9]. Permission *J. Acoust. Soc. Amer.*)

opposite side of the head. In particular, one should note that there is a steady increase in sound pressure going from $-45°$ to $+45°$. Physical factors of this type provide the basis for sound localization in space.

Middle Ear

Sound reaches the ear by way of the air, a gas. On the other hand, the organ of Corti is contained within the cochlear fluids, which are physically comparable to seawater. This difference between these media is of considerable import to hearing, as the following example will show. Suppose you and a friend are standing in water at the beach. He is speaking, and in the middle of his sentence you dunk your head under the water. However loud and clear your friend's voice was a moment ago, it will be barely, if at all, audible while your head is submerged. Why?

The answer is really quite straightforward. Air offers less opposition to the flow of sound energy than does seawater. In other words, air has less impedance than water. We shall discuss impedance in a moment, but for now just think of it as the opposition to energy flow. Since the water's impedance is greater than that of the air, there is an impedance mismatch at the boundary between them. Airborne sound thus meets a substantial increase in opposition to its flow at the water's surface, and much of the energy is reflected back rather than being transmitted through the water. The impedance mismatch between air and cochlear fluids has the same effect. The middle ear serves as an impedance-matching transformer that permits sound energy to be effectively transmitted from the air to the cochlea. Let us briefly review some aspects of impedance before discussing how the middle ear performs its function.

Impedance

A mechanical representation of impedance and its components is shown in Fig. 3.4a. Impedance Z is the interaction of resistance R, positive reactance Xm, and negative reactance Xs. These components are respectively related to friction, mass, and stiffness. Friction is represented in the figure by the ridged surface across which the block (mass) is moved. Friction causes a portion of the energy applied to the system to be converted into heat; this dissipation of energy into heat is termed resistance. Resistance is not related to frequency, and occurs in phase with the applied force. It is shown on the x-axis in Fig. 3.4c.

Reactance is the storage, as opposed to the dissipation, of energy by the system. Mass (positive) reactance Xm is associated with the mass of the system (shown by the block in the figure). Since all mass has the property of inertia, the application of a force F to the mass M causes the

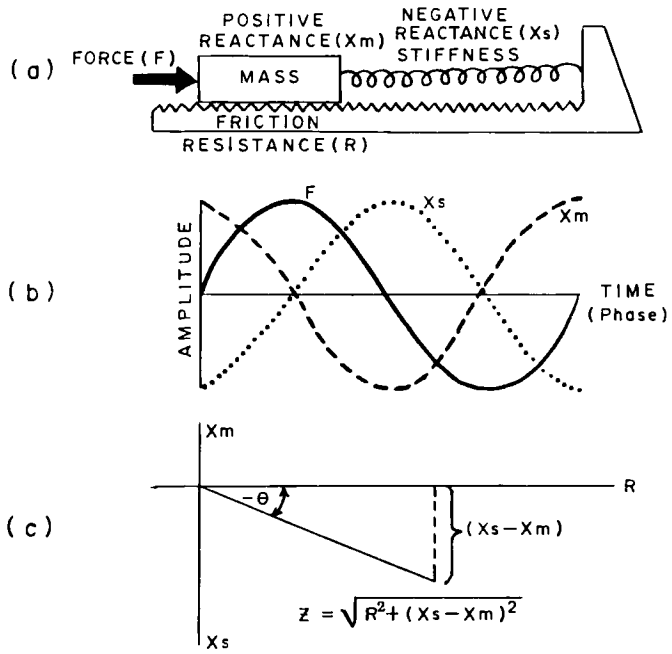

Figure 3.4 Impedance and its components (see text).

mass to accelerate according to the formula F = MA (where A is acceleration). If the force is applied sinusoidally, then the mass reactance is related to frequency as Xm = M2πf, where f is frequency. Thus, the magnitude of Xm is directly proportional to frequency (the higher the frequency the greater the mass reactance). Since acceleration precedes force by a quarter-cycle, Xm will lead the applied force in phase by 90° (Fig. 3.4b). This is why Xm is termed positive reactance, and its value is shown in the positive direction on the y-axis (Fig. 3.4c).

Stiffness (negative) reactance Xs is represented by the spring in Fig. 3.4a. We will represent the spring's stiffness as S. Applying a force compresses (displaces) the spring according to the formula F = SD, where D is the amount of displacement. When the stimulus is applied sinusoidally, Xs is related to frequency as Xs = S/2πf. In other words, the amount of stiffness reactance is inversely proportional to frequency (Xs goes down as frequency goes up). Since displacement follows force by a quarter-cycle, Xs lags behind the applied force in phase by 90° (Fig. 3.4b). It is thus called negative reactance and is plotted downward on the y-axis (Fig. 3.4c).

Since the stiffness and mass components are 180° out of phase (Fig. 3.4b), a system's net reactance is equal to the difference between them (Xs − Xm). This relation is shown in Fig. 3.4c for the condition where Xs exceeds Xm, which is the case for the normal ear. Notice that the overall impedance Z results from the interaction between the resistance R and the net reactance. The negative phase angle $-\Theta$ in the figure shows that the net reactance is negative. The relationship between impedance, resistance, and reactance is thus

$$Z = \sqrt{R^2 + (Xs - Xm)^2}$$

Looking at the effect of frequency, we find that

$$Z = \sqrt{R^2 + (S/w - Mw)^2}$$

where w represents $2\pi f$. The implication is that frequency counts. Since Xm is proportional to frequency while Xs is inversely proportional to frequency, they should be equal at some frequency. This is the system's resonant frequency, at which the reactance components cancel each other out, leaving only the resistance component.

As in any other system, the impedance of the middle ear is due to its stiffness, mass, and resistance. Figure 3.5 is a block diagram of the middle ear with respect to its impedance [12], revised from Zwislocki's earlier model [13]. One may think of the upper row of boxes as the line of energy flow from the eardrum to the cochlea, and of the boxes coming down from them as the ways energy is shunted from the system. The first box represents the middle ear cavities, which contribute significantly to the stiffness of the system. The next two boxes, "eardrum/malleus" and "eardrum (decoupled)," should be thought of together. The former represents the proportion of sound energy transmitted from the drum to the malleus. It includes the inertia of the malleus; the elasticity

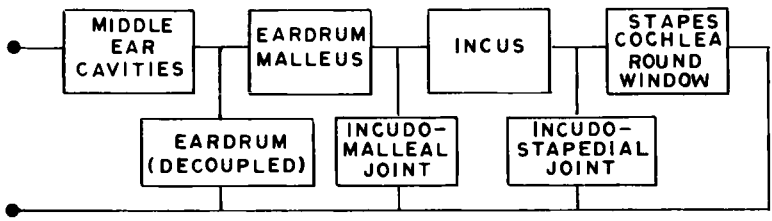

Figure 3.5 Block diagram of the middle ear. [From Zwislocki [12]. In *Acoustic Impedance and Admittance—The Measurement of Middle Ear Function* (A. S. Feldman and L. A. Wilber, Eds.), Williams and Wilkins, Baltimore (©1976).]

of the drum, tensor tympani muscle, and mallear ligaments; and the friction caused by the strain on these structures. "Eardrum (decoupled)" is the proportion of energy diverted from the system when the drum vibrates independently (decoupled from) the malleus, which occurs particularly at high frequencies. The box labeled "incus" is the effective mass of the incus and the stiffness of its supporting ligaments. The energy lost at the two ossicular joints is represented by "incudomalleal joint" and "incustapedial joint," which shunt energy off the main line of the diagram. The last box shows the effects of the stapes, cochlea, and round window in series. The attachments of the stapes as well as the round window membrane contribute to the stiffness component. Most of the ear's resistance is due to the cochlea. Zwislocki [14] has pointed out that a major effect of this resistance is to smooth out the response of the middle ear by damping the free oscillations of the ossicular chain.

Middle Ear Transformer Mechanism

The ratio between the impedances of the cochlear fluids and the air is approximately 4000:1. To find out how much energy would be transmitted from the air to the cochlea without the middle ear, we apply the simple formula $T = 4r/(r + 1)^2$, where T is transmission and r is the ratio of the impedances. The result is ~0.001. In other words, only 0.1% of the airborne energy would be transmitted to the cochlea, while 99.9% would be reflected back. This corresponds to a 30 dB drop going from air to cochlea.

The middle ear thus "steps up" the level of airborne sound to overcome the impedance mismatch between the air and cochlear fluids. Early place theory [15] held that the middle ear transformer was the source of various nonlinearities in hearing, such as combination tones. These distortion products of the middle ear's hypothetical nonlinear response were ostensibly transmitted to the cochlea, where the nonlinearities were analyzed according to the place principle as though they were present in the original signal. However, Wever and Lawrence [16] have demonstrated that the middle ear performs its function with elegant linearity, and we must accordingly regard it as a linear transformer, and look elsewhere (to the cochlea) for the sources of nonlinear distortions.

Several factors, discussed below, contribute to the transformer function of the middle ear. They include the areal ratio of the eardrum to the oval window, the curved-membrane mechanism of the drum, and the level action of the ossicular chain.

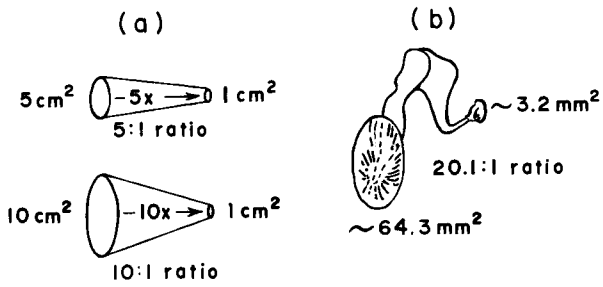

Figure 3.6 The areal ratio. General principle (a), and areal ratio of the middle ear ossicles (b).

Areal Ratio

We know that pressure P is force F per unit area A: $P = F/A$. If we therefore exert the same pressure over two areas, one of which is 5 times larger than the other, then the pressure on the smaller surface will be 5 times greater. Examples of this effect are shown in Fig. 3.6a.

The area of the human eardrum is roughly 64.3 mm^2, while that of the oval window is about 3.2 mm^2 [16]. Thus the drum:oval window ratio is ~20.1:1 (Fig. 3.6b). If we assume for the moment that the ossicles act as a simple rigid connection between these two membranes, then there would be a pressure amplification by a factor of 20.1 going from the eardrum to the oval window.

Eardrum Mechanism

Helmholtz [15] suggested that the eardrum contributes to the effectiveness of the middle ear transformer by lever action according to the curved membrane principle (Fig. 3.7). The drum's rim is firmly attached to the annulus, and curves down to the attachment of the malleus, which is mobile, as in the figure. A given force increment ΔF thus displaces the membrane with greater amplitude than it displaces the manubrium. Since the products of force and distance (amplitude of displacement) on both legs of the lever are equal ($F1D1 = F2D2$), the smaller distance traveled by the manubrium is accompanied by a much greater force. In this way, Helmholtz proposed that lever action of the eardrum would result in an amplification of force to the ossicles.

Subsequent experiments led to the abandonment of this principle, since studies of drum movement were inconsistent with it, and since Helmholtz's results were not replicated [16]. Bekesy [18] used a capacitive probe to measure human eardrum displacement at various fre-

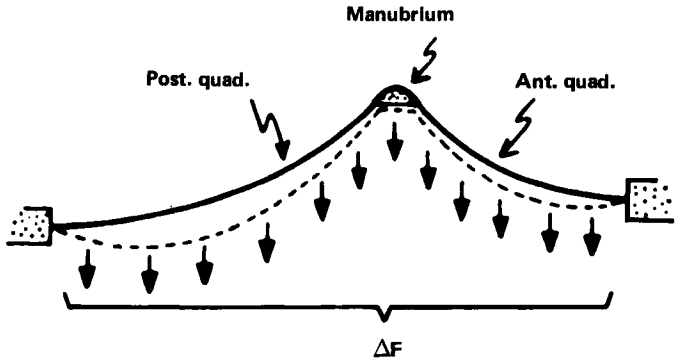

Figure 3.7 Curved membrane principle. (From Tonndorf and Khanna [17]. Permission *Ann. Otol.*)

quencies. The capacitive probe uses a very fine wire as one plate of a capacitor and the drum surface as the other plate. Sound causes the drum to vibrate, which in turn varies its distance from the wire. If a current is passed through this capacitor, the movement of the drum will affect current flow. Monitoring the current flow at different spots on the drum enabled Bekesy to determine its displacement with considerable accuracy.

Figure 3.8a shows Bekesy's results for frequencies below 2000 Hz. For lower frequencies the drum moved as a stiff plate or piston, hinged superiorly at the axis of the ossicles. The greatest displacement occurred inferiorly. Bekesy attributed the drum's ability to move in this manner, without significant deformation, to a highly elastic or loose-fitting fold at its inferior edge (Fig. 3.8b). Above about 2400 Hz the membrane's stiffness broke down, and movement of the manubrium lagged behind that of the membrane rather than being synchronized with it. The stiffly moving portion of the drum had an area of 55 mm^2 out of a total area of 85 mm^2. This area seemed to constitute an "effective area" for the drum of about 70% of its total area. Applying the drum's effective area to the overall areal ratio (20.1 × 0.7) would result in an effective areal ratio of 14.1:1.

The eardrum's role was reevaluated by Tonndorf and Khanna [17], who used time-averaged holography to study drum movement in the cat. Holography is an optical method that reveals equal-amplitude contours as alternating bright and dark lines on the vibrating membrane. Figure 3.9 shows the isoamplitude contours for a 600 Hz tone. These contours show that the drum actually does not have an effective

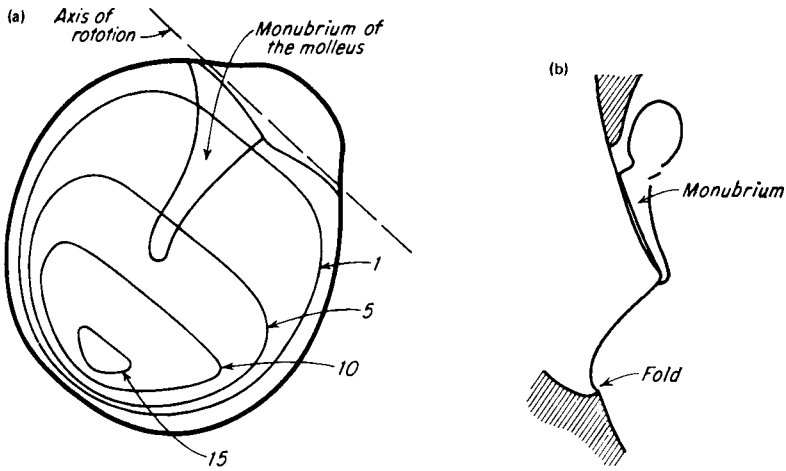

Figure 3.8 Equal relative eardrum displacement curves for frequencies up to 2000 Hz (a). Cross section of the drum showing a loose-fitting inferior edge (b). [From G. Bekesy, *Experiments in Hearing* (©1960 McGraw-Hill). Used with permission of McGraw-Hill Book Company.]

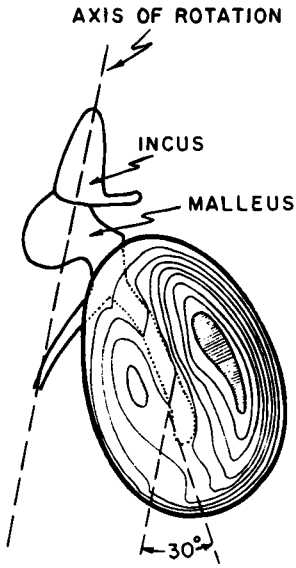

Figure 3.9 Vibration patterns of the cat's eardrum in response to a 600 Hz tone. (From Tonndorf and Khanna [17]. Permission *Anal. Otol.*)

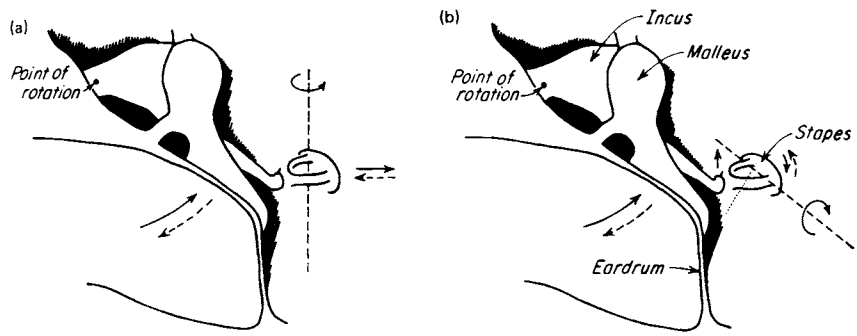

Figure 3.10 Stapes movement at (a) moderate and (b) intense levels in cadavers. [From G. Bekesy, *Experiments in Hearing* (©1960 McGraw-Hill). Used with permission of McGraw-Hill Book Company.]

area that moves like a stiff plate. Instead, there are two areas of peak displacement, consistent with Helmholtz's curved-membrane concept. This mode of vibration is seen up to about 1500 Hz. The pattern becomes more restricted at higher frequencies, and increasingly complex subpatterns occur in the vibrations as frequency rises above approximately 3000 Hz. Thus, the entire drum area contributes to the areal ratio, as opposed to just an effective portion of it. In addition, the curved membrane principle contributes to the transformer ratio (by a factor of 2.0 in the cat). If we accept an areal ratio of 39.8:1.4 = 34.6 for the cat [16], then the middle ear transfer ratio as of this point becomes $34.6 \times 2.0 = 69.2{:}1$. This value must be multiplied by the lever ratio of the ossicles to arrive at the final transformer ratio of the middle ear.

Ossicular Lever

Helmholtz [15] proposed that nonlinear distortions are introduced by the ossicular chain, and are largely due to what he conceived of as a cog-wheel articulation between the malleus and incus. This situation would allow for relative movement in one direction at the malleoincudal joint. The resulting distortions would stimulate the cochlea at places corresponding to the receptors for those frequencies, as though they were present in the original signal. Barany [19] demonstrated, however, that except during intense stimulation these two bones are rigidly fixed at the malleoincudal joint and move as a unit in response to sound stimulation.

Bekesy [18] has reported that the stapes moves differently in response to moderate and intense stimulation in human cadavers (Fig. 3.10). At moderate intensities, the stapes footplate rocks with a

piston-like motion in the oval window, with greater amplitude anteriorly (Fig. 3.10a). Intense stimulation results in rotation of the footplate around its longitudinal axis (Fig. 3.10b). Rocking of the stapes around the longitudinal axis substantially reduces the energy transmitted to the cochlea, which most likely serves as a protective device. However, the cat stapes has been observed to maintain essentially piston-like movements even at very high intensities, at least for low frequencies [20].

The ossicular chain rotates around an axis, illustrated in Fig. 3.11, corresponding to a line through the long process of the malleus and the short process of the incus [19]. The ossicular chain is delicately balanced around its center of gravity so that the inertia of the system is minimal [19]. The ossicles act as a lever about their axis. The lever ratio is on the order of 1.3 in humans and 2.2 in cats. However, the actual lever ratio is smaller, because of the interaction of the curviture of the drum and the length of the ossicular chain lever [17].

Recall that the drum curves more toward the umbo, and may be regarded as a curved string. The transformer ratio of a curved string decreases as the curviture becomes stronger (1/curviture). On the other hand, the transformer ratio of the ossicular lever increases with length. Note in Fig. 3.12 that the ossicular lever is long (with respect to the point of attachment of the malleus on the drum) where the curviture is strong, and that it is short where the curviture is small. This interaction results in an essentially constant lever ratio, with a value of about 1.4 for the cat ossicular chain.

We may now apply the ossicular lever ratio to the intermediate solution of 69.2 obtained above for the middle ear transfer ratio. The final value is $69.2 \times 1.4 = 96.9{:}1$. Therefore, the total transformer ratio, found using the areal ratio, the curved membrane mechanism, and the ossicular lever, is 96.9:1, which corresponds to 39.7 dB. This closely

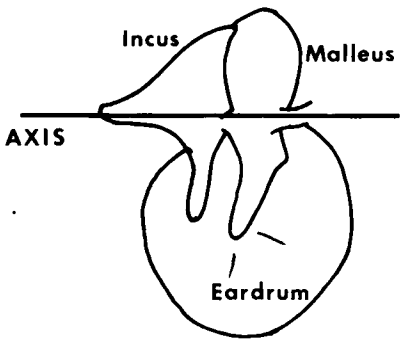

Figure 3.11 Axis of the ossicular chain. (After Barany [19].)

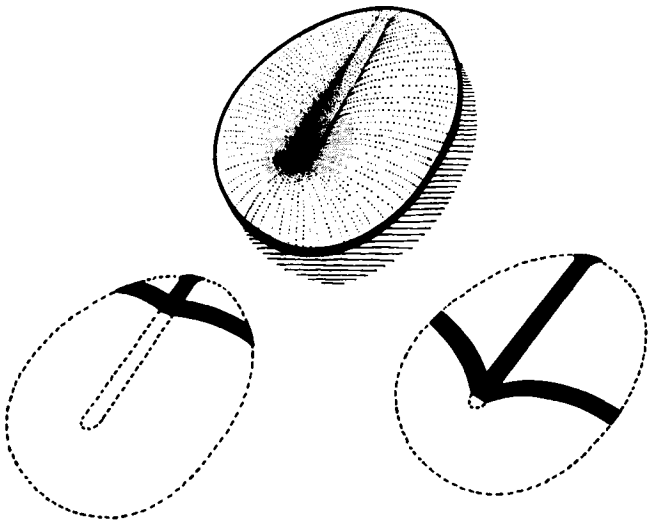

Figure 3.12 Interaction between the length of the ossicular chain and the inverse of drum curvature according to the relation (1/small curvature) (short lever) ~ const. ~ (1/strong curvature) (long lever). (From Tonndorf and Khanna [17]. Permission *Ann. Otol.*)

approximates the 40 dB loss that results when the cat's middle ear is entirely obliterated [21].

Middle Ear Response

The middle ear transfer function is to a large extent responsible for the shape of minimal audibility curves (see Chap. 9). These curves show the amount of sound energy needed to reach the threshold of hearing as a function of frequency.

Bekesy [18] reported that the resonant frequency of the middle ear is in the 800–1500 Hz region. Recall that resonance occurs when mass and stiffness reactance are equal, cancelling out. Impedance is then entirely composed of resistance, and accordingly the opposition to energy flow is minimal at the resonant frequencies. Møller [22] found the major resonance peak of the middle ear to be about 1200 Hz, with a smaller resonance peak around 800 Hz. Normal ear reactance and resistance values obtained by Zwislocki [14] are shown in Fig. 3.13. Note that the ear's impedance results primarily from negative reactance up to about 800 Hz. This effect is due to the middle ear mechanism itself, which is stiffness-controlled below the resonant frequency. There is

virtually no reactance between 800 and 6000 Hz, indicating that energy transmission from the drum to cochlea is maximal in this range. Positive reactance takes over at higher frequencies as a result of the effective mass of the drum and ossicles. We thus expect sound transmission through the middle ear to be frequency-dependent with emphasis on the mid-frequencies; and the minimal audibility curves of the ear should reflect this relation.

The open circles in Fig. 3.14 show the middle ear transfer function based on data from anesthesized cats [11,20]. The filled circles are the behavioral thresholds of waking cats in a sound field [23]. The binaural threshold and transfer function are generally similar, but the threshold curve is steeper at low frequencies and flatter at high. This may reflect several factors [11,24,25,26]: First, since the thresholds show mean group data, the curve is probably somewhat flattened by the averaging among subjects; the transfer function is from a single representative animal. In addition there is a peak at around 4000 Hz in the middle ear response of anesthesized animals, which is much smaller when they are awake due to damping of the system by the tonus of the stapedius muscle. A second factor has to do with the effects of head diffraction, the

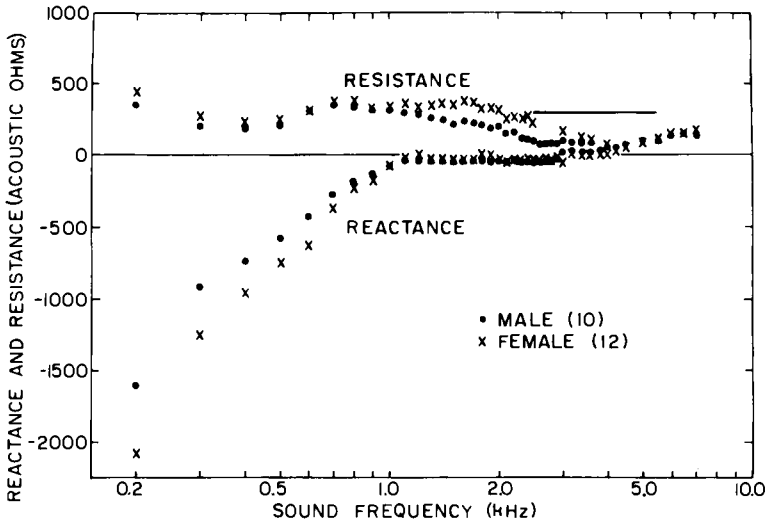

Figure 3.13 Acoustic resistance and reactance at the eardrum as a function of frequency. [From Zwislocki [14], in *The Nervous System*, Vol. 3: *Human Communication and Its Disorders* (D. B. Tower and E. L. Eagles, Eds.), ©1975, Raven Press, New York.]

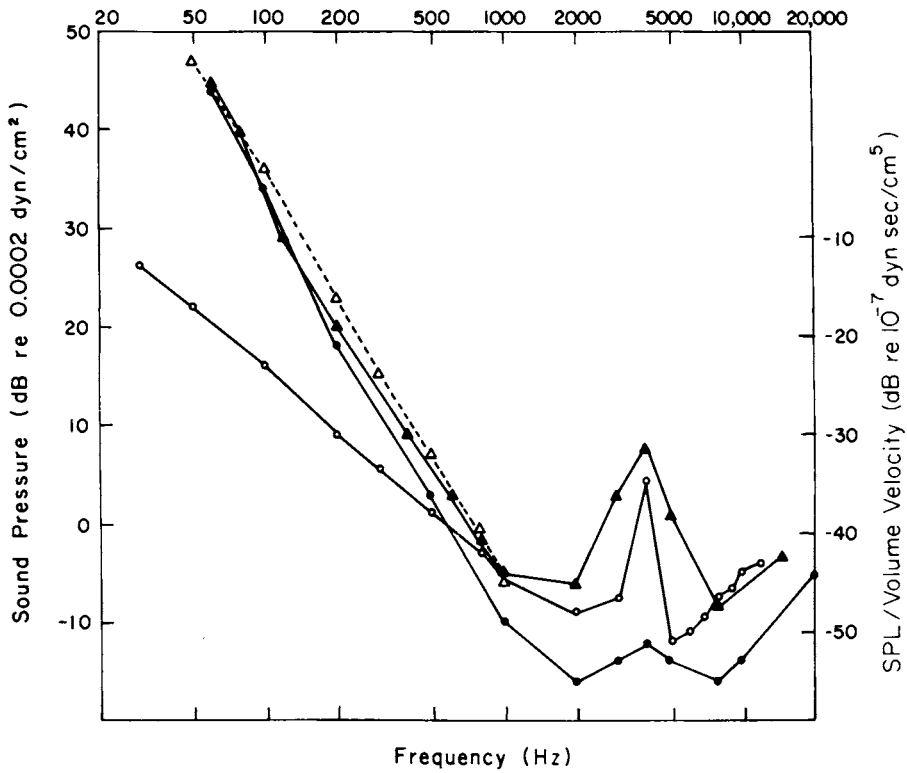

Figure 3.14 Middle ear transfer function (open circles) compared to behavioral thresholds (closed circles) and SPL at the drum (closed triangles). Open triangles show the transfer function corrected for cochlear input impedance. (From Dallos [11], *The Auditory Periphery*, ©1973, Academic Press, New York.)

pinna, and the ear canal, as discussed in the section on the Outer Ear in this chapter. These effects are accounted for by viewing the behavioral thresholds in terms of the SPL near the eardrum at threshold, as is shown by the filled circles in Fig. 3.14. The relationship of the threshold curve to the transfer function is closer when considered in these terms. The disparity between the transfer function and thresholds below about 1000 Hz is reconciled by correcting the transfer function for the input impedance of the cochlea (open triangles).

These factors show a remarkable degree of correspondence between the middle ear transfer function and the threshold curve, at least for the cat. Reasonable agreement between the transfer function based upon a model of the middle ear and the threshold curve has also been shown for

humans [14]. Thus, we find that the impedance matching job of the middle ear is accomplished quite well for the mid-frequencies, although the frequency-dependent nature of the middle ear reduces its efficiency at higher and lower frequencies.

BONE CONDUCTION ROUTE

Until now we have dealt with the usual route from the air to the cochlea. The discussion would be incomplete, however, without at least a brief consideration of bone conduction—the transmission of sound to the cochlea by the bones of the skull. For this to occur, a sound must be strong enough to cause the bones to vibrate, or else the stimulus must be delivered by way of a vibrator applied to the skull. The impedance mismatch between air and bone is even greater than between air and cochlear fluid: An airborne sound must exceed the air conduction threshold by at least 50–60 dB before the bone conduction threshold is reached [27]. Direct stimulation with a vibrator is routinely employed in audiological evaluations to separate hearing losses attributable to the outer and/or middle ear from those due to impairments of the sensorineural mechanisms.

Two experiments prove that both air conduction and bone conduction initiate the same traveling waves in the cochlea (see Chap. 4). Bekesy [28] showed that air- and bone-conduction signals cancel one another when their phases and amplitudes are appropriately adjusted. Lowy [29] demonstrated that this cancellation occurs in the cochlea, since repetition of the Bekesy experiment on guinea pigs resulted in cancellation of the cochlear microphonic. (The cochlear microphonic is an electrical potential that reflects the activity of the basilar membrane; see Chap 4.) The implications of these experiments are monumental, since they demonstrate that the final activity in the cochlea is the same regardless of the mode of entry of the sound. Furthermore, this result gives support to the use of bone conduction as an audiological tool the results of which can validly be compared with those of air conduction in determining the locus of a lesion (assuming appropriate calibration of both signals).

Bekesy [28] found that below 200 Hz the human skull vibrates as a unit (Fig. 3.15a). At about 800 Hz the mode of vibration changes (Fig. 3.15b), and the front and back of the head vibrate in opposite phase to one another, with a nodal line of compression between them. Above 1600 Hz the head begins to vibrate in four segments (Fig. 3.15c).

Tonndorf [30,31] demonstrated that the mechanism of bone conduction includes contributions from the outer, middle, and inner ear. For clarity, let us look at these components beginning with the inner ear.

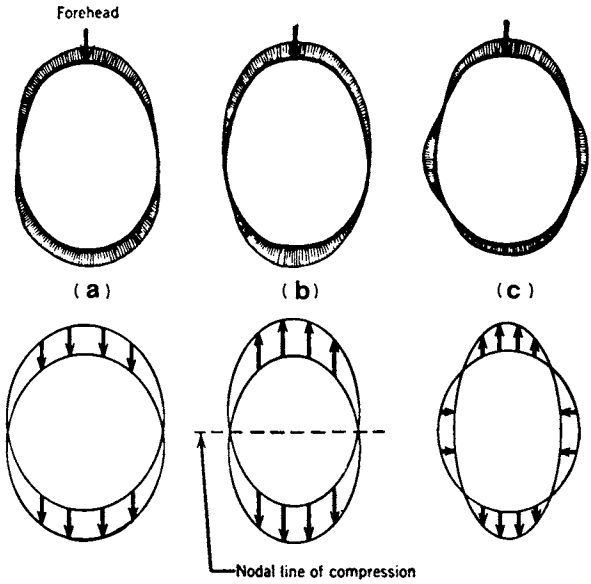

Figure 3.15 Patterns of skull vibration at (a) 200 Hz, (b) 800 Hz, and (c) 1600 Hz, with bone conduction stimulus applied for forehead. (After Bekesy [28].)

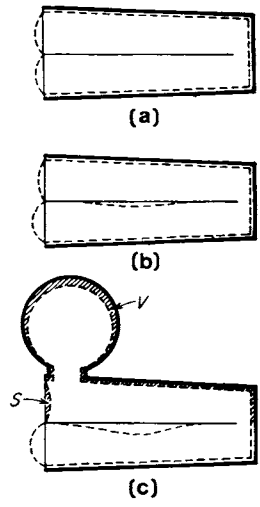

Figure 3.16 Compressional bone conduction (see text). (After Bekesy [28].)

Compressional bone conduction is illustrated in Fig. 3.16. Vibration of the temporal bone results in alternate compression and expansion of the cochlear capsule. Since the cochlear fluids are incompressible, there should be bulging at compliant points. Bulging would occur at the oval and round windows without displacement of the cochlear partition if both windows were equally compliant (Fig. 3.16a). However, since the round window is much more compliant than the oval window, compression of the cochlear capsule pushes the fluid in the scala vestibuli downward, displacing the basilar membrane (Fig. 3.16b). This effect is reinforced since the sum of the surface area V of the vestibule and the surface area of the scala vestibuli is greater than that of the scala tympani (Fig. 3.16c). The possibility of a "third window" for the release of this pressure is provided by the cochlear aqueduct.

Tonndorf [32], however, found that the primary mechanism of bone conduction in the inner ear involves distortional vibrations of the cochlear capsule which are synchronous with the signal (Fig. 3.17). Since the volume of the scala vestibuli is greater than that of the scala tympani, these distortions result in compensatory displacements of the cochlear partition even in the absence of compliant windows. The above-mentioned "window effects" modify the distortional component.

The contribution of the middle ear to bone conduction was demonstrated by Barany [19]. Recall that the ossicles move side-to-side rather than front-to-back. Barany found that for low frequencies bone conduction was maximal when a vibrator was applied to the side of the head and was minimal when it was applied to the forehead. This occurs because the lateral placement vibrates the skull in the direction of ossicular movement (Fig. 3.18a) whereas frontal placement vibrates the skull perpendicular to their movement (Fig. 3.18b). In other words, the signal was transmitted best when it was applied in the direction of rotation of the ossicular chain about its axis. The mechanism is as follows: Since the

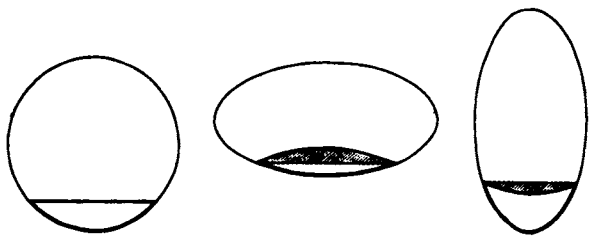

Figure 3.17 Effect of distortional vibrations on displacement of the cochlear partition. (From Tonndorf [32]. Permission *J. Acoust. Soc. Amer.*)

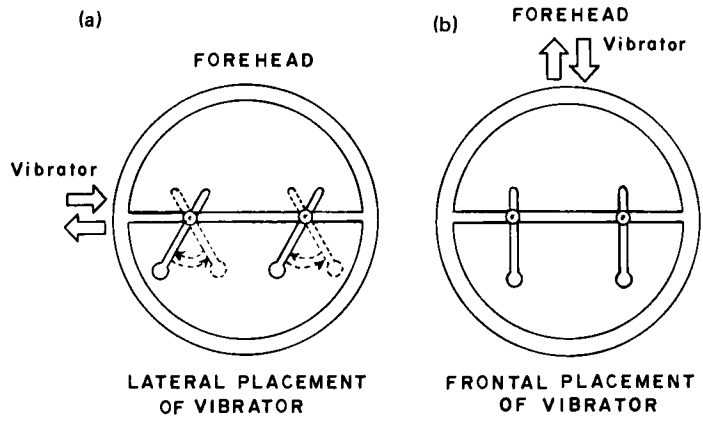

Figure 3.18 Inertial bone conduction: (a) lateral placement of the vibrator; (b) forehead placement. (Abstracted from Barany [19].)

ossicles are suspended analogously to pendulums, as shown in Fig. 3.18, their inertia cuases them to move relative to the skull when the latter is vibrated. *Inertial* bone conduction, then, stimulates the cochlea by the relative movement of the skull and ossicles, the effect of which is a rocking motion of the stapes at the oval window.

The middle ear component of bone conduction is of particular interest in otosclerosis, a disorder in which hearing loss results from fixation of the stapes in the oval window. A hearing loss results because the fixated stapedial footplate cannot effectively transmit energy to the cochlea. Although one might expect bone conduction to be impaired at low frequencies, impairment occurs instead at about 2000 Hz. This phenomenon is called "Carhart's notch" [33]. Bone conduction is impaired at 2000 Hz because this is the resonant frequency of the ossicular chain in man (the resonant frequency of the ossicular chain is species-specific [30]).

The contribution of the outer ear to bone conduction is most readily observed during occlusion of the cartilaginous part of the ear canal. Vibration of the skull under these conditions leads to the radiation of sound energy into the enclosed canal from its walls. These radiations then stimulate the eardrum and finally the cochlea along the familiar air conduction route [30]. This enhancement of the low frequencies for bone conduction is not seen when the ear is unoccluded, because the open ear canal acts as a high-pass (low-cut) filter. The occlusion effect occurs because the lows are not lost when the outer ear canal is closed off, so that sensitivity for these frequencies is enhanced.

In summary, the bone conduction mechanism appears to be primarily due to distortion of the inner ear capsule, to the relative movements of the skull and ossicles due to the inertial lag of the latter, and to the sound radiated into the ear canal from its walls. Tonndorf et al. [30] found in cats that the outer and middle ear components are dominant below about 1000 Hz, but that all three mechanisms are about equally important in the 1000–6000 Hz range.

ACOUSTIC REFLEX

Early experiments on dogs revealed bilateral tensor tympani contractions when either ear was stimulated by intense sound [34,35]. It was later demonstrated that the stapedius muscle also responds to intense sound stimulation in cats and rabbits [36]. However, whether the acoustic reflex *in man* is due to contractions of one or of both intratympanic muscles has been the subject of some controversy.

Direct observation through perforated eardrums revealed stapedius muscle contractions in man as a consequence of intense sound stimulation [37–39]. Terkildsen [40,41] indirectly examined middle ear muscle activity by monitoring changes in air pressure in the ear canal in response to sound stimulation. Stapedius contraction would result in an outward movement of the drum, while tensor concentration would pull the drum inward (see Chap. 1). The drum movement in turn, results in changes in ear canal pressure. Terkildsen could thus infer the nature of muscle activity by monitoring the air pressure in the ear canal during sound stimulation. Most of his subjects manifested an outward deflection of the drum, suggesting that the stapedius muscle was active. There were, however, some cases showing tensor activity as well (inward drum displacement). Similar findings were reported by Mendelson [42,43].

Perhaps the greatest contribution to what is known about the acoustic reflex (AR) comes from measurements of acoustic impedance at the plane of the eardrum using an impedance bridge. The mechanical acoustic impedance bridge was first applied to the study of the AR by Metz [44], and was improved and made clinically efficient by Zwislocki [45]. An electroacoustic impedance bridge was introduced by Terkildsen and Nielsen [46]. Almost all AR research since the introduction of the electroacoustic bridge has used this method or variations of it. The principle is straightforward: Since contractions of the intratympanic muscles stiffen the middle ear system (including the drum), the impedance is increased. (The reflex primarily affects the compliance component of impedance rather than resistance because the system is stiffened.) It is this change in acoustic impedance that is measured by the bridge.

Let us now consider how the ear's impedance can be used to infer information about the activities of one intratympanic muscle vs. the other. (Recall that the AR is bilateral, so we can stimulate one ear and monitor the impedance change in the other.) In the normal ear we really cannot tell whether the AR is due to contractions of the stapedius and/or of the tensor. However, if the reflex is present when the stapedius muscle is intact but not when the muscle (or its reflex arc or attachment) is impaired, then we may conclude that the stapedius muscle contributes to the AR in man. The same argument applies to the tensor tympani. (Remember at this point that the stapedius is innervated by the seventh cranial nerve and the tensor by the fifth; and that the former inserts via its tendon on the more medial stapes while the latter connects to the more lateral malleus.)

Several studies have demonstrated that the acoustic reflex is absent when there is pathology affecting the stapedius muscle [47–49]. However, the AR was still obtained in two cases of tensor tympani pathology [47]. It might be added at this point that a measurable tensor reflex does occur as part of a startle reaction to very intense sound [50], or in response to a jet of air directed to the eye [48]. Based on these observations, one is drawn to conclude that in man the AR to sound stimulation is at least primarily a stapedius reflex.

Reflex Parameters

Several parameters of the AR should be discussed before describing its effect upon energy transmission through the ear, or its suggested roles within the auditory system. The possible relationship between the AR and loudness will also be covered.

We have been referring to the AR as occurring in response to "intense" sound stimulation. Let us now examine just how intense a sound is needed. With a reasonable amount of variation among studies, the AR thresholds in response to pure tone signals from 250 to 4000 Hz range from about 85 to 100 dB SPL [51–57]. The reflex threshold is approximately 20 dB lower (better) when the eliciting stimulus is wide-band noise [55,57–59]. In general, the AR is obtained at a lower intensity when it is monitored in the ear being stimulated instead of in the opposite ear [60].

The lower reflex threshold for noise than for tones suggests that the AR is related to the bandwidth of the stimulus. Flottrop et al. [61] studied this relationship by measuring AR thresholds elicited by successively wider bands of noise and complex tones. They found that the increasing bandwidth did not cause the threshold to differ from its value

for a pure tone activator *until* a certain bandwidth was exceeded. At this point there was a clear-cut break, after which increasing the bandwidth resulted in successively lower reflex thresholds. This study was replicated at 1000 Hz with similar results [62]. These findings suggest that there is a *critical band* for the AR, above which widening the bandwidth results in lower thresholds. Confirmation was obtained by Djupesland and Zwislocki [63], who found that increasing the separation in frequency between the two tones in a two-tone complex caused a lowering of the reflex threshold once a particular bandwidth was exceeded.

Although the existence of a critical bandwidth was a constant finding in the noise and two-tone studies, the thresholds were lower for noise. Since this difference might be due to differences in the spectral density of the two types of stimuli, Popelka et al. [64] replicated this work using tonal complexes made up of many components equally spaced in frequency (on a log scale). Their data (Fig. 3.19) confirm the critical band

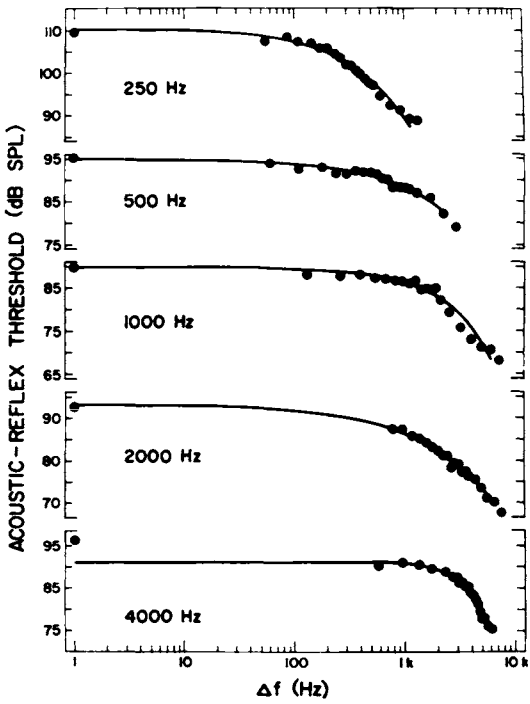

Figure 3.19 Critical bandwidths for acoustic reflex thresholds at center frequencies of 250–4000 Hz. Dots to the left of functions show corresponding pure tone reflex thresholds. (From Popelka et al. [64]. Permission *J. Acoust. Soc. Amer.*)

phenomenon. The width of the critical band (shown by the break from the horizontal on each function in Fig. 3.19) increases with center frequency. It is important to note that the critical bandwidth for the AR is substantially wider than the psychoacoustic critical bands discussed in the chapters to follow.

The AR does not occur instantaneously upon presentation of the activating signal. Instead, a measurable impedance change is observed only after a latency, the length of which depends on both the intensity and frequency of the stimulus. Metz [65] found that this latency decreased from about 150 msec at 80 dB above the threshold of hearing [80 dB sensation level (SL)] to 40 msec at 100 dB SL in response to a 1000 Hz activating signal. Møller [66] reported latencies of 25–130 msec for 500 and 1500 Hz pure tones. As a rule, latencies were shorter for 1500 Hz than for 500 Hz tones. Dallos [67] found a similar inverse relationship between activator intensity and reflex latency for white noise. These findings were confirmed and expanded upon by Hung and Dallos [68], who found that AR latencies were shorter for noise signals than for pure tones, with the longest latencies for tones below 300 Hz. The shortest latencies, on the order of 20 msec, were in response to noise activators.

These measurements were based upon changes in acoustic impedance. However, the latency of the impedance change is a measure of the latency of the *mechanical* response of the middle ear rather than of the neural transport time for the reflex arc alone [69]. Zakrisson et al. [70] found that the electromyographic (EMG) response of the stapedius muscle in man has a latency as short as 12 msec. They also reported that the EMG "threshold" is about 6 dB lower than that for the impedance change measured as the lower stimulus level needed to yield 10% of the maximum response. Since we are concerned with the effect of the AR on the transmission of energy through the middle ear, we are most interested in the mechanical-response latency. However, one should be aware that changes in muscle potentials occur in the stapedius prior to the measured impedance change.

Hung and Dallos [68] reported that some of their subjects demonstrated a "latency relaxation" at the onset of the AR response. That is, there was an actual *decrease* in impedance prior to the impedance increase. This decrease was routinely observed by Silman et al. [57]. It appears that it reflects a partial relaxation of the stapedius muscle before contraction.

We have already seen that AR latency shortens with increasing stimulus intensity. Similarly, increasing stimulus level also causes an increase in reflex magnitude (i.e., in the amount of impedance change associated

with the reflex) [57,65,66], and a faster rise time of the reflex response [67,68]. We shall call the relationship between stimulus intensity and reflex magnitude the growth function of the reflex.

The AR growth function has been studied in response to pure tone and wide-band and narrow-band noise activating signals [52,56,57,67,68★. Figure 3.20 shows the growth functions obtained from four subjects [68]. It illustrates that the growth of AR magnitude is essentially linear for pure tones as high as about 120 dB SPL. The functions for wide-band noise are essentially linear up to approximately 110 dB SPL. These data are substantially supported by the other studies cited. Thus, AR magnitude tends to increase linearly with a stimulus intensity of 85–120 dB SPL for tones and roughly 70–110 dB SPL for wide-band noise. Saturation occurs at higher levels.

Møller [52,60] reported steeper reflex growth functions with increasing frequency in the 300–1500 Hz range. Flottrop et al. [61] found greater impedance changes at 250 Hz than at 4000 Hz. They also reported that although signals at 1000 and 2000 Hz elicited the same maximum reflex magnitude as at 250 Hz, about 10 dB more (re: reflex threshold) was needed for the two higher frequencies. Furthermore, while some studies have suggested that a 2000 Hz tone elicits the greatest impedance change [71,72], others suggest that 1000 Hz and wide-band stimuli produce maximal responses [56,73]. On the other hand, Borg and Møller [74] found no significant differences in the slopes of the AR growth functions in the 500–3000 Hz range for laboratory animals. It thus appears that any clear-cut relationship between activator frequency and reflex magnitude is still undetermined.

Two temporal parameters of the AR are particularly interesting. One deals with stimulus durations below about one second and the other has to do with long stimulus durations. The former is temporal summation and the latter is adaptation.

Temporal summation deals with the relationship between stimulus duration and intensity when the time frame is less than about 1 sec (see Chap. 9). It is most easily understood by example. Suppose a subject's threshold for a tone that lasts 200 msec happens to be 18 dB. Will the threshold remain at 18 dB when the same tone is presented for only 20 msec? It is found that when the 20 msec tone is used the threshold changes to 28 dB. (A similar trade-off is needed to maintain the stimulus at a constant loudness.) This illustrates the general psychoacoustic observation that when a signal is shortened by a factor of 10 (e.g., from 200 to 20 msec), the signal level must be increased by as much as 10 dB to offset the decade decrease in duration. This is understandably called a time-intensity trade.

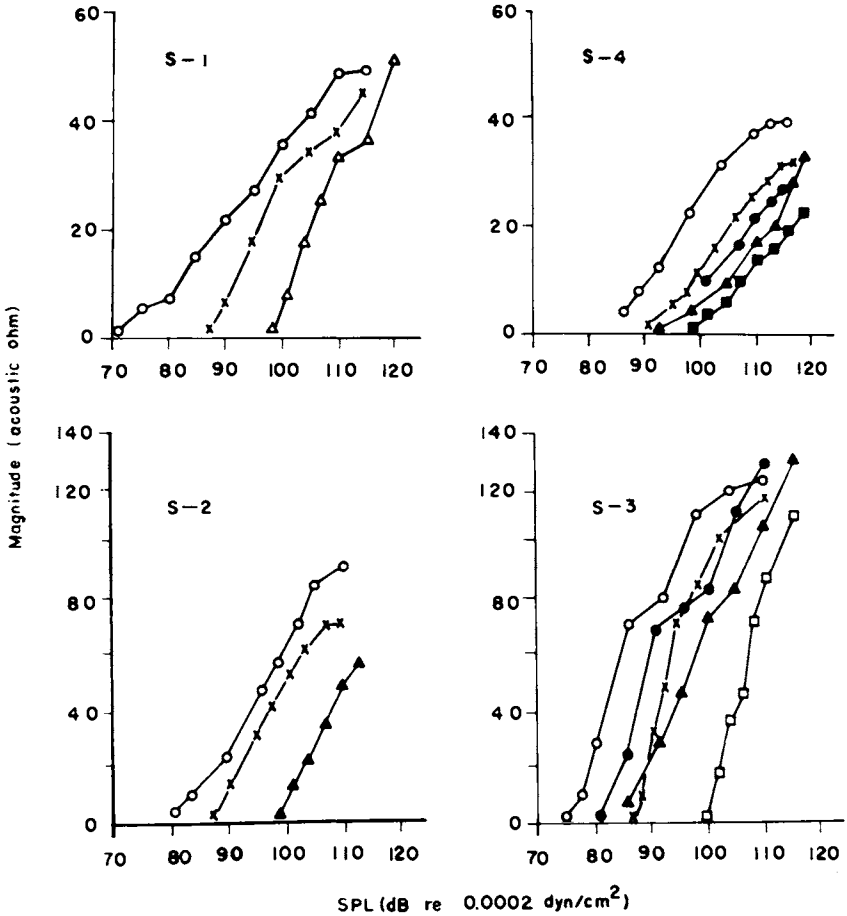

Figure 3.20 Reflex growth functions for wide-band noise (open circles) and pure tones (250 Hz, open squares; 300 Hz, filled squares; 500 Hz, filled triangles; 600 Hz, open triangles; 1000 Hz, crosses; 1500 Hz, filled circles). (From Hung and Dallos [67]. Permission *J. Acoust. Soc. Amer.*)

Temporal summation also occurs for the AR threshold and magnitude. However, it appears that the amount of intensity change needed to counteract a given decrease in stimulus duration is greater for the AR than for psychoacoustic phenomena. Unfortunately, there are rather large differences between the results of various studies reporting the intensity needed to offset a given duration change. For example, decreasing the duration of a 2000 Hz tone from 100 to 10 msec was offset by an

increase in stimulus level by about 25 dB in one study [75] as opposed to roughly 15 dB in another [76]. In the 500–4000 Hz range, Djupesland et al. [77] studied the time-intensity trade-off relation for the AR with one-octave wide-noise bands. They used as their comparison point the stimulus level/duration needed to maintain the reflex magnitude at half the maximum impedance change. Djupesland et al. found that a 10-fold decrease in duration was offset by a 20–25 dB increase in signal level. In contrast, Gnewikow (1974) [78] has found that a 12–23 dB intensity increase was needed to offset decade reductions in duration for 500 and 4000 Hz pure tones. Jerger et al. [78] found less temporal summation than the other studies, as shown in Fig. 3.21 for 500–4000 Hz. Notice that the amount of temporal integration increases with frequency, which is a common finding among the studies.

These discrepancies have yet to be resolved. However, Jerger et al. [78] suggested that at least some of the differences are due to problems associated with the visual detection threshold (in which the experimenter looks for the smallest noticeable impedance change on a meter or oscilloscope) and with constant percentage of maximum impedance change [77] methods of obtaining the data.

Let us now examine stimulus durations of several seconds and longer. Early studies on laboratory animals showed that the degree of

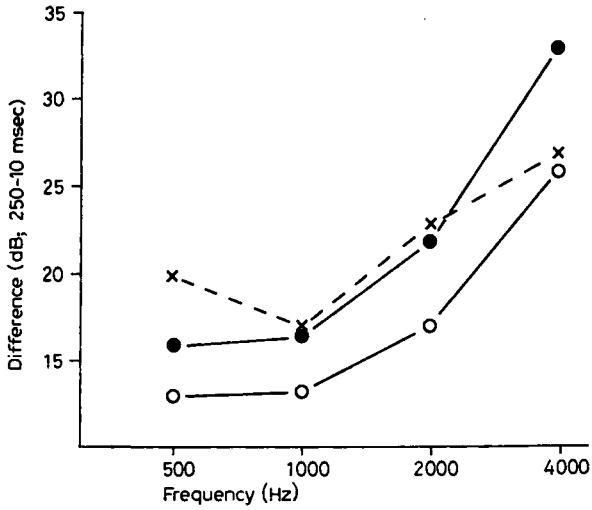

Figure 3.21 Temporal summation for the acoustic reflex at four frequencies obtained by Jerger et al. [78] (open circles); Gnewikow (1974) (filled circles); and Woodford et al. [76] (crosses). [From Jerger, Mauldin, and Lewis, Temporal summation of the acoustic reflex, *Audiology 16,* 177–200 (1977).]

muscle contraction due to the AR decreases as stimulation is prolonged [36,79]. This decrease is referred to as reflex decay or adaptation; it has been demonstrated as a decrease in reflex magnitude in man by numerous investigators [67,72,80–83]. In spite of wide differences between subjects, the common finding is that reflex adaptation increases as the frequency of the pure tone stimulus is raised.

Particular attention should be given to the findings of Kaplan et al. [72], who studied reflex adaptation to pure tones of 500–4000 Hz at 6, 12, and 18 dB above the reflex threshold. Figure 3.22 summarizes their median data at the three SLs re: reflex threshold, with frequency as the parameter. There is greater reflex adaptation as frequency increases. Also, adaptation tends to begin sooner after stimulus onset for higher frequencies. These data are normalized in Fig. 3.23, in which the point

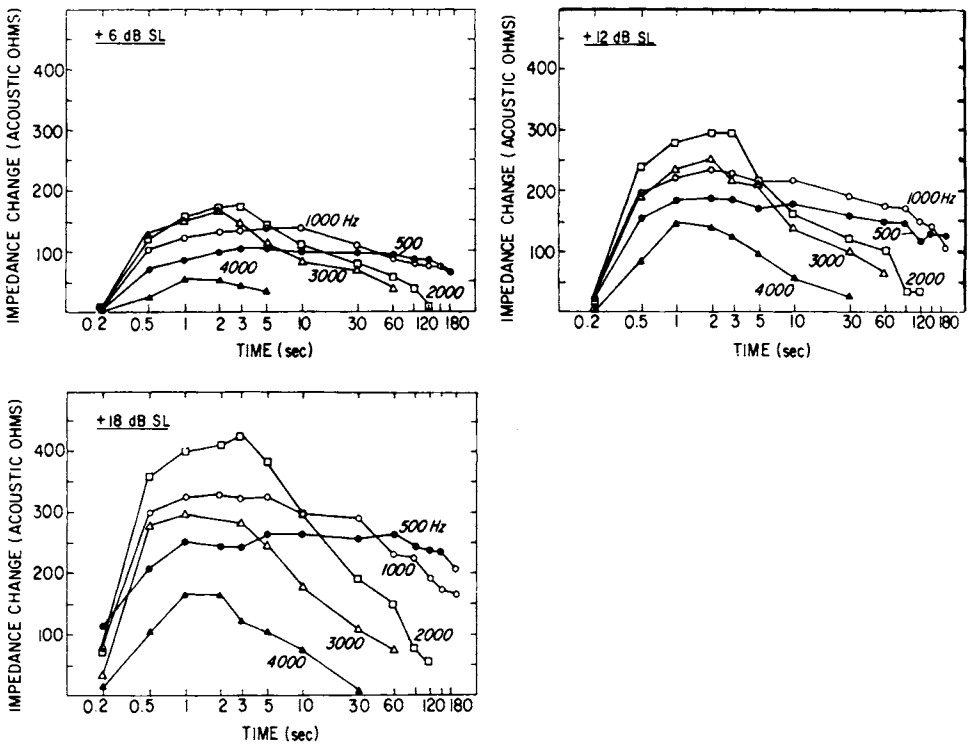

Figure 3.22 Median absolute impedance change (acoustic ohms) as a function of time for three suprareflex threshold levels. (From Kaplan et al. [72]. Permission *Ann. Otol.*)

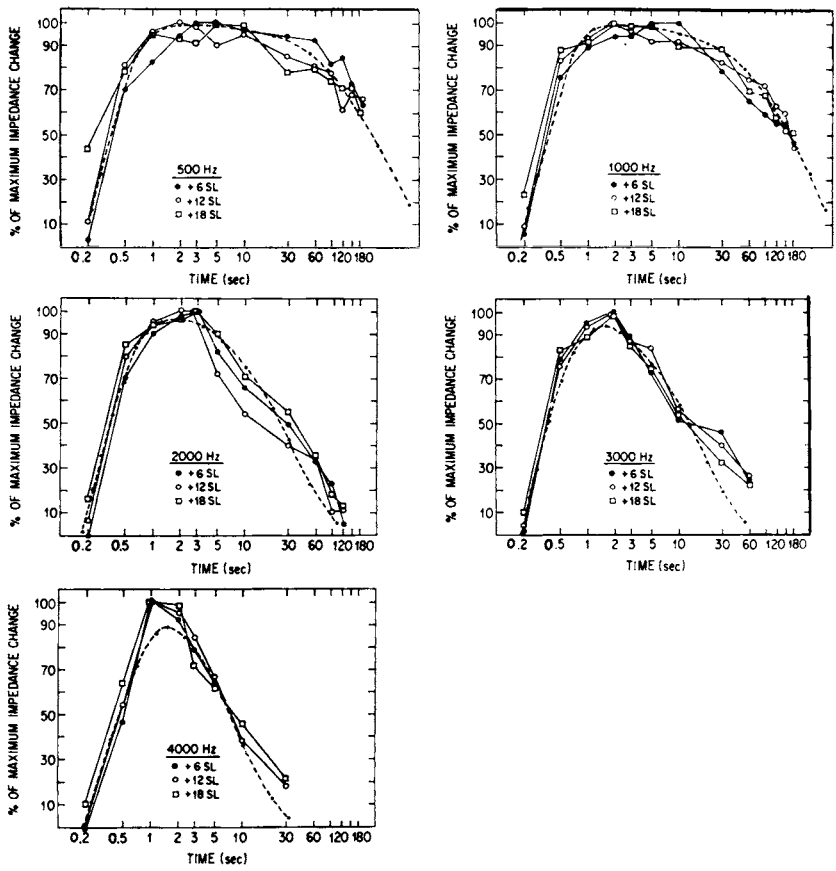

Figure 3.23 Median normalized impedance change (percent) as a function of time for each frequency. (From Kaplan et al. [72]. Permission *Ann. Otol.*)

of greatest impedance change is given a value of 100% and the other points are shown as percentages of the maximum impedance change. In this plot the data are shown separately at each frequency with the suprathreshold level as the parameter. In addition to clearly showing the frequency effect, Fig. 3.23 demonstrates that the course of the adaptation function is similar at various levels above reflex threshold, at least up to +18 dB.

Tietze [81,82] proposed that the course of AR adaptation could be described by the time constants of reflex rise time and adaptation. These time constants refer respectively to how long it takes for reflex magni-

tude to attain 63% of the maximum value (rise) and then to decrease to 63% of it (adaptation); both are measured from the moment of AR onset. The time constants are functions of frequency. Tietze's [82] model describes reflex adaptation by the formula

$$Zn = \frac{1}{1 - \tau_{an}/\tau_{ab}} \left[\exp(-t/\tau_{ab}) - \exp(-t/\tau_{an})\right]$$

where Zn is the normalized maximum impedance change; τ_{an} and τ_{ab} are the time constants of reflex rise and adaptation, respectively; and t is the time (sec) of reflex rise from onset. Kaplan et al. [72] applied this formula to their data, using the ratio τ_{an}/τ_{ab} to generate the dotted lines in Fig. 3.23. (Note that, except at high frequencies where rapid adaptation reduces the maximum impedance change, τ_{an} is generally quite small relative to τ_{ab}.) As Fig. 3.23 shows, Kaplan's data support the exponential adaptation predicted by Tietze's model.

Loudness is the perceptual correlate of the intensity of the acoustic stimulus (see Chap. 11); other things being equal, loudness increases as stimulus level is increased. An aberration of the intensity-loudness relationship is found in patients with cochlear disorders. Once the signal is increased above the impaired threshold of such an ear, the loudness grows at a faster rate than normal in response to increasing stimulus level. This is called loudness recruitment. What has the AR to do with loudness and recruitment? Since both loudness and the AR are related to stimulus intensity, at least some association between them is understandable. The question is whether the AR is a function of loudness.

Metz [51] obtained AR thresholds at lower SLs from patients with presumably cochlear sensorineural hearing loss than from normal subjects. In other words, the spread between the threshold of hearing and the AR threshold was smaller for those with the hearing losses. Since the impaired subjects also had loudness recruitment, Metz proposed that the lower SL of the reflex reflected recruitment. In other words, it was argued that the AR was elicited by the loudness of the signal. Jepsen [53] reported that, while the auditory thresholds of his subjects increased with age, the reflex thresholds tended to decrease, especially at 1000 Hz and above. He attributed this to presbycusis (hearing loss associated with aging), and assumed that it was due to recruitment.

The relationship between loudness and the AR, however, is not nearly as clear-cut as these early findings suggest. Ross [84] compared the equal-loudness and equal-reflex contours of four subjects. The contours were similar for two, while the others had systematic differences between the loudness and AR contours (the loudness curves were flatter). Ross suggested that the latter two equal-loudness contours

might have been aberrant, but in fact they were quite similar to those reported by Fletcher and Munson [85], among others.

Margolis and Popelka [55] compared the loudness of a variety of stimuli set to the AR threshold level. There was a range of about 17 dB in the loudness levels, suggesting that the AR is not a manifestation of a critical loudness level. Block and Wightman [86] suggested that the loudness-reflex relationship is supported by their finding of similarly shaped equal-loudness and equal-reflex contours. However, they often found that the same reflex magnitude was elicited by stimulus levels as much as 10 dB apart. Such a spread corresponds to a doubling of loudness [87]; in this light their findings appear to support rather than refute those of Margolis and Popelka. The substantially wider critical bands for the AR than for loudness discussed previously provide a further basis for questioning the concept that the AR is loudness-based.

Returning to patient data, Beedle and Harford [88] found steeper AR growth functions for normal ears than for ears with cochlear dysfunction. This result is, of course, inconsistent with a loudness basis for the reflex. The impact of their data, however, was impaired by the fact that their normal subjects averaged 24 years old, compared to 47 years old for the pathological group. The reflex growth functions of age-matched normal and hearing-impaired subjects were studied by Silman et al. [57]. Examples of their findings are shown in Fig. 3.24. If the

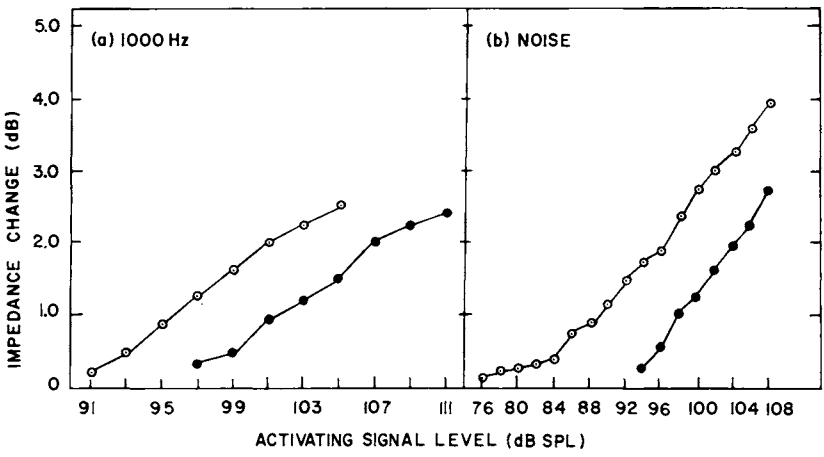

Figure 3.24 Reflex growth as a function of SPL for 1000 Hz (a) and a broadband noise (b). Open symbols, normal ears; closed symbols, impaired ears. (Adapted from Silman et al. [57].)

reflex were loudness determined, then the function for the impaired group, although displaced along the intensity axis, would approach the normal curves at higher stimulus levels. This would occur because loudness recruitment would result in equal loudness for both groups at equal suprathreshold levels. Equal loudness would in turn elicit equal reflex magnitudes at those levels. As the figure shows, this is not the case.

Acoustic Reflex and Sound Transmission

It is apparent that the AR is largely stimulus dependent. Furthermore, it should not be surprising that the parameters of the reflex response reflect the sensory (and neural) processing of relatively intense stimulation. It is equally apparent that the AR is a feedback or control mechanism — although its exact purpose(s) remains unclear. Given these points, what effect does the AR have on energy transmission through the conductive mechanism?

Recall that the AR stiffens the conductive mechanism so that sound is reflected at the eardrum [67]. Since the effect of stiffness is inversely related to frequency, we would expect the AR to affect middle ear transmission most strongly at low frequencies.

Smith [89] and Reger [90] compared the pure-tone thresholds of human subjects to the thresholds during voluntary contractions of the middle ear muscles. Thresholds were shifted by about 20–40 dB at 125–500 Hz, and by 15 dB at 1000 Hz, and there was little or no change at higher frequencies. While the expected frequency relation is maintained, voluntary contractions may not yield the same transmission loss as the AR [90]. Perhaps the most impressive data on how the AR affects middle ear transmission come from animal studies.

The cochlear microphonic (CM) is an electrical potential of the cochlea which is proportional to the intensity of the stimulus over a wide dynamic range. This is discussed in Chap. 4. Changes in the magnitude of the CM will thus reflect remarkably well the transmission loss resulting from the AR. Simmons [91] found that, while the AR in cats shifted the CM response in a manner that decreased with frequency up to 7000 Hz, the effect of the reflex was greatest below 2000 Hz and was only slight at higher frequencies. Figure 3.25 shows the change in the CM due to stapedius muscle contraction in the cat [92]. Møller [92] found that impedance change data and CM findings in response to the AR were within 5 dB over a substantial portion of the frequency range. The figure shows that the AR affects primarily the frequencies below about 2000 Hz. Similar low-frequency changes have been reported in response to electrically elicited contractions of the tensor tympani [93].

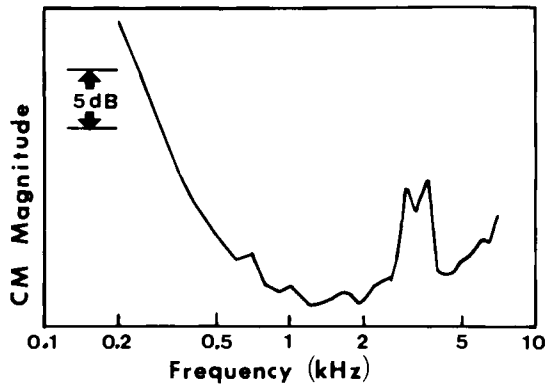

Figure 3.25 Effect of acoustic reflex on transmission in the cat's middle ear. (Adapted from Møller [92].)

Several recent studies [94–97] have dealt with the effects of middle ear muscle contractions on absolute thresholds and loudness in humans. Rabinowitz [94] has reported a 10 dB change in the low-frequency transmission characteristics of the middle ear due to the AR. However, Morgan and Dirks [96] found that reflex-eliciting stimulus levels up to 100 dB SPL did not result in a change in the loudness of a test tone presented to the opposite ear. Stimuli greater than 100 dB SPL caused a decrease in loudness for low-frequency tones but not for 1500 Hz. Furthermore, Morgan et al. [97] demonstrated that the absolute thresholds are not affected by the AR. These findings do not fully support earlier observations. It may be that there is no clear relationship between the amount of impedance change and the amount of threshold shift caused by the reflex [97]. This lack of correspondence between the animal and human data might also reflect species differences.

Reflex Theories

There are many theories and speculations about the purpose of the AR. Space and scope preclude more than a cursory review here, and the reader is urged to consult Jepsen [53] and Simmons [24].

Since the reflex is elicited by relatively high stimulus levels, and its magnitude grows with increasing stimulus level, one would expect that a primary purpose of the AR would be protection of the cochlea from damaging stimulation. This "protection theory" is weakened by the facts that the latency and adaptation of the reflex cause it to respond too slowly to sudden sounds, and make it inefficient against prolonged

sounds. Also, sounds intense enough to elicit the AR are virtually nonexistent in nature. Nevertheless, the protection offered by the AR is a beneficial side effect if not the main purpose.

The "accommodation theory" states that the action of the two middle ear muscles modifies the conductive mechanism so as to optimize the absorption of sound energy. According to the "fixation theory," the intratympanic muscles help keep the ossicles in proper position and appropriately rigid, particularly at high frequencies when acceleration is great. Other theories have asserted that these muscles contribute to changes in labyrinthine pressure and to the formation of aural overtones.

Simmons [24] found a sharp antiresonance peak around 4000 Hz in the middle ear transmission function of cats whose intratympanic muscles were removed, as well as in normal cats under anesthesia. Normal, waking cats whose intratympanic muscles have normal tonus showed greatly reduced dips in this region. Simmons reasoned that tonus of the middle ear muscles smooths the frequency response of the conductive system. He suggested that modulation of muscle tonus would have the effect of enhancing attention by changing the intensity and frequency characteristics of environmental sounds. This modulation would be analogous to the constant motion of the extraocular muscles in vision. Also, several skeletal movements as well as unexpected and novel environmental sounds elicit the AR. Since the reflex mainly attenuates low frequencies, and since most of an organism's own physiologic noises are low in frequency, such a reflex response would have the effect of reducing the animal's internal noise. A better signal-to-noise ratio would result, which is of obvious importance to the survival of any species, whether predator or prey. This idea agrees with Borg's [69] position that the qualitative purpose of the AR is to attenuate low-frequency sounds, thereby improving the auditory system's dynamic range.

REFERENCES

1. G. Bekesy and W. A. Rosenblith, The mechanical properties of the ear, in *Handbook of Experimental Psychology* (S. S., Stevens Ed.), Wiley, New York, 1958, pp. 1075–1115.
2. M. Gardner, Some monaural and binaural facets of median plane localization, *J. Acoust. Soc. Amer. 54,* 1489–1495 (1973).
3. M. Gardner and R. Gardner, Problem of localization in the median plane: Effect of pinnae cavity occlusion, *J. Acoust. Soc. Amer. 53,* 400–408 (1973).
4. D. Wright, J. Hebrank, and B. Wilson, Pinna reflections as cues for localization, *J. Acoust. Soc. Amer. 56,* 957–962 (1974).

5. J. Hebrank and D. Wright, Spectral cues used in the localization of sound sources in the median plane, *J. Acoust. Soc. Amer. 56*, 1829–1834 (1974).
6. J. Blauert, Sound localization in the median plane, *Acoustica 22*, 205–213 (1969).
7. D. W. Batteau, The role of the pinna in human localization, *Proc. Roy. Soc (London) B168*, 158–180 (1967).
8. F. M. Wiener and D. A. Ross, The pressure distribution in the auditory canal in a progressive sound field, *J. Acoust. Soc. Amer. 18*, 401–408 (1946).
9. E. A. G. Shaw, Transformation of sound pressure level from the free field to the eardrum in the horizontal plane, *J. Acoust. Soc. Amer. 56*, 1848–1861 (1974).
10. S. Mehrgardt and V. Mellert, Transformation characteristics of the external human ear, *J. Acoust. Soc. Amer. 61*, 1567–1576 (1977).
11. P. Dallos, *The Auditory Periphery*, Academic, New York, 1973.
12. J. Zwislocki, The acoustic middle ear function, in *Acoustic Impedance and Admittance — The Measurement of Middle Ear Function* (A. S. Feldman and L. A. Wilber, Eds.), Williams & Wilkins, Baltimore, 1976, pp. 66–77.
13. J. Zwislocki, Analysis of the middle-ear function. Part I: Input impedance, *J. Acoust. Soc. Amer. 34*, 1514–1523 (1962).
14. J. Zwislocki, The role of the external and middle ear in sound transmission, in *The Nervous System* (D. B. Tower and E. L. Eagles, Eds.), Vol. 3: *Human Communication and Its Disorders*, Raven, New York, 1975, pp. 45–55.
15. H. Helmholtz, Die Mechanik der Gehörknöchelchen und des Trommelfells, *Pflüg. Arch. Ges. Physiol. I*, 1–60 (1868).
16. E. G. Wever and M. Lawrence, *Physiological Acoustics*, Princeton Univ. Press, Princeton, N.J., 1954.
17. J. Tonndorf and S. M. Khanna, The role of the tympanic membrane in middle ear transmission, *Ann. Otol. 79*, 743–753 (1970).
18. G. Bekesy, *Experiments in Hearing*, McGraw-Hill, New York, 1960.
19. E. Barany, A contribution to the physiology of bone conduction, *Acta Otol.*, Suppl. 26 (1938).
20. J. Guinan and W. T. Peake, Middle ear characteristics of anesthesized cats, *J. Acoust. Soc. Amer. 41*, 1237–1261 (1967).
21. J. Tonndorf, S. M. Khanna, and B. J. Fingerhood, The input impedance of the inner ear in cats, *Ann. Otol. 75*, 752–763 (1966).
22. A. R. Møller, Improved technique for detailed measurements of the middle-ear impedance, *J. Acoust. Soc. Amer. 32*, 250–257 (1960).
23. J. D. Miller, C. S. Watson, and W. P. Covell, Deafening effects of noise on the cat, *Acta Otol.*, Suppl. 176 (1963).
24. F. B. Simmons, Perceptual theories of middle ear function, *Ann. Otol. 73*, 724–740 (1964).
25. F. N. Wiener, R. R. Pfeiffer, and A. S. N. Backus, On the pressure transformation by the head and auditory meatus of the cat, *Acta Otol. 61*, 255–269 (1966).
26. P. Dallos, Low-frequency auditory characteristics: Species dependencies, *J. Acoust. Soc. Amer. 48*, 489–499 (1970).

27. G. Bekesy, Vibrations of the head in a sound field and its role in hearing by bone conduction, *J. Acoust. Soc. Amer. 20*, 749–760 (1948).
28. G. Bekesy, Zur Theories des Hörens bei der Schallaufnahme durch Knochenleitung, *Poggendorff's Annln Phys. Chem. 13*, Ser. 5, 111–136 (1932).
29. K. Lowy, Cancellation of the electrical cochlear response with air- and bone-conduction, *J. Acoust. Soc. Amer. 14*, 156–158 (1942).
30. J. Tonndorf et al., Bone conduction: Studies in experimental animals; A collection of papers, *Acta Otol.*, Suppl. 213 (1966).
31. J. Tonndorf, A new concept of bone conduction, *Arch Otol. 87*, 49–54 (1968).
32. J. Tonndorf, Compressional bone conduction in cochlear models, *J. Acoust. Soc. Amer. 34*, 1127–1132 (1962).
33. R. Carhart, Clinical application of bone conduction audiometry, *Arch. Otol. 51*, 798–808 (1950).
34. V. Hensen, Beobachtungen über die Thätigkeit des Trommellspanners bei Hund und Katze, *Arch. Anat. Physiol. II*, 312–319 (1878).
35. J. Pollack, Über die Function des Musculus tensor tympani, *Med. Jahrbuch 82*, 555–582 (1886).
36. T. Kato, Zur Physiologie der Binnenmuskeln des Ohres, *Pflüg. Arch. ges. Physiol. 150*, 569–625 (1913).
37. E. Lüscher, Die Function des Musculus stapedius beim Menschen, *Zeitsch. Hals-Nasen-und-Ohrenheilkunde 23*, 105–132 (1929).
38. J. R. Lindsay, H. Kobrack, and H. B. Perlman, Relation of the stapedius reflex to hearing sensitivity in man, *Arch. Otol. 23*, 671–678 (1936).
39. A. B. Potter, Function of the stapedius muscle, *Ann. Otol. 45*, 639–643 (1936).
40. K. Terkildsen, Movements of the eardrum following inter-aural muscle reflexes, *Arch. Otol. 66*, 484–488 (1957).
41. K. Terkildsen, Acoustic reflexes of the human musculus tensor tympani, *Acta Otol.*, Suppl. 158 (1960).
42. E. S. Mendelson, A sensitive method for registration of human intratympanic muscle reflexes, *J. Appl. Physiol. 11*, 499–502 (1957).
43. E. S. Mendelson, Improved method for studying tympanic reflexes in man, *J. Acoust. Soc. Amer. 33*, 146–152 (1961).
44. O. Metz, The acoustic impedance measured on normal and pathological ears, *Acta Otol.*, Suppl. 63 (1946).
45. J. Zwislocki, An acoustic method for clinical examination of the ear, *J. Speech Hearing Res. 6*, 303–314 (1963).
46. K. Terkildsen and S. S. Nielsen, An electroacoustic measuring bridge for clinical use, *Arch. Otol. 72*, *339*–346 (1960).
47. O. Jepsen, Studies on the acoustic stapedius reflex in man: Measurements of the acoustic impedance of the tympanic membrane in normal individuals and in patients with peripheral facial palsy, Thesis, Universitetsforlaget, Aarhus, Denmark, 1955.
48. I. Klockhoff, Middle-ear muscle reflexes in man, *Acta Otol.*, Suppl. 164 (1961).
49. A. Feldman, A report of further impedance studies of the acoustic reflex, *J. Speech Hearing Res. 10*, 616–622 (1967).

50. G. Djupesland, Middle ear muscle reflexes elicited by acoustic and non-acoustic stimulation, *Acta Otol.*, Suppl. 188 (1964).
51. O. Metz, Threshold of reflex contractions of muscles of the middle ear and recruitment of loudness, *Arch. Otol. 55*, 536–593 (1952).
52. A. Møller, The sensitivity of contraction of tympanic muscle in man, *Ann. Otol. 71*, 86–95 (1962).
53. O. Jepsen, The middle ear muscle reflexes in man, in *Modern Developments in Audiology* (J. Jerger, Ed.), Academic, New York, 1963, pp. 194–239.
54. J. Jerger, Clinical experience with impedance audiometry, *Arch. Otol. 92*, 311–321 (1970).
55. R. Margolis and G. Popelka, Loudness and the acoustic reflex, *J. Acoust. Soc. Amer. 58*, 1330–1332 (1975).
56. R. H. Wilson and L. M. McBride, Threshold and growth of the acoustic reflex, *J. Acoust. Soc. Amer. 63*, 147–154 (1978).
57. S. Silman, G. Popelka, and S. A. Gelfand, Effect of sensorineural hearing loss on acoustic stapedius reflex growth functions, *J. Acoust. Soc. Amer. 64*, 1406–1411 (1978).
58. L. Dutsch, The threshold of the stapedius reflex for pure tone and noise stimuli, *Acta Otol. 74*, *248*–251 (1972).
59. J. L. Peterson and G. Liden, Some static characteristics of the stapedial muscle reflex, *Audiology 11*, 97–114 (1972).
60. A. Møller, Bilateral contraction of the tympanic muscles in man examined by measuring acoustic impedance-change, *Ann. Otol. 70*, 735–753 (1961).
61. G. Flottrop, G. Djupesland, and F. Winther, The acoustic stapedius reflex in relation to critical bandwidth, *J. Acoust. Soc. Amer. 49*, 457–461 (1971).
62. G. Popelka, R. Karlovich, and T. Wiley, Acoustic reflex and critical bandwidth, *J. Acoust. Soc. Amer. 55*, 883–885 (1974).
63. G. Djupesland and J. Zwislocki, On the critical band in the acoustic stapedius reflex, *J. Acoust. Soc. Amer. 54*, 1157–1159 (1973).
64. G. Popelka, R. Margolis, and T. Wiley, Effects of activating signal bandwidth on acoustic-reflex thresholds, *J. Acoust. Soc. Amer. 59*, 153–159 (1976).
65. O. Metz, Studies on the contraction of the tympanic muscles as indicated by changes in impedance of the ear, *Acta Otol. 39*, 397–405 (1951).
66. A. Møller, Intra-aural muscle contraction in man examined by measuring acoustic impedance of the ear, *Laryngoscope 68*, 48–62 (1958).
67. P. Dallos, Dynamics of the acoustic reflex: Phenomenological aspects, *J. Acoust. Soc. Amer. 36*, 2175–2183 (1964).
68. I. Hung and P. Dallos, Study of the acoustic reflex in human beings: I. Dynamic characteristics, *J. Acoust. Soc. Amer. 52*, 1168–1180 (1972).
69. E. Borg, Dynamic characteristics of the intra-aural muscle reflex, in *Acoustic Impedance and Admittance — The Measurement of Middle Ear Function* (A. S. Feldman and L. A. Wilber, Eds.), Williams & Wilkins, Baltimore, 1976, pp. 236–299.
70. J. E. Zakrisson, E. Borg, and S. Blom, The acoustic impedance change as a measure of stapedius muscle activity in man: A methodological study with electromyography, *Acta Otol. 78*, 357–364 (1974).

71. S. Silman, Growth function of the stapedius reflex in normal hearing subjects and in subjects with sensorineural hearing loss due to cochlear dysfunction, Ph.D. Dissertation, New York University, New York, 1976.
72. H. Kaplan, S. Gilman, and D. Dirks, Properties of acoustic reflex adaptation, *Ann. Otol. 86*, 348–356 (1977).
73. D. Cunningham, Admittance values associated with acoustic reflex decay, *J. Amer. Aud. Soc. 1*, 197–205 (1976).
74. E. Borg and A. Møller, The acoustic middle ear reflex in man and in anesthesized rabbits, *Acta Otol. 65*, 575–585 (1968).
75. G. Djupesland and J. Zwislocki, Effect of temporal summation on the human stapedius reflex, *Acta Otol. 71*, 262–265 (1971).
76. C. Woodford, D. Henderson, R. Hamernick, and A. Feldman, Threshold-duration function of the acoustic reflex in man, *Audiology 14*, 53–62 (1975).
77. G. Djupesland, A. Sundby, and G. Flottrop, Temporal summation in the acoustic stapedius reflex mechanism, *Acta Otol. 76*, 305–312 (1973).
78. J. Jerger, L. Mauldin, and N. Lewis, Temporal summation of the acoustic reflex, *Audiology 16*, 177–200 (1977).
79. R. Lorente de Nó, The function of the central acoustic nuclei examined by means of the acoustic reflexes, *Laryngoscope 45*, 573–595 (1935).
80. E. Borg, A quantitative study of the effects of the acoustic stapedius reflex on sound transmission through the middle ear, *Acta Otol. 66*, 461–472 (1968).
81. G. Tietze, Zum zeitverhalten des akustischen reflexes bei reizung mit dauertonen, *Arch. Klin. Exp. Ohren. Nasen. Kehlkpfheilkd 193*, 43–52 (1969).
82. G. Tietze, Einge eigenschaften des akustischen reflexes bei reizung mit tonimpulsen, *Arch. Klin. Exp. Ohren. Nasen. Kehlkopfheilkd 193*, 53–69 (1969).
83. T. L. Wiley and T. S. Karlovich, Acoustic reflex response to sustained signals, *J. Speech Hearing Res. 18*, 148–157 (1975).
84. S. Ross, On the relation between the acoustic reflex and loudness, *J. Acoust. Soc. Amer. 43*, 768–779 (1968).
85. H. Fletcher and W. A. Munson, Loudness: Its definition, measurement and calculation, *J. Acoust. Soc. Amer. 5*, 82–108 (1933).
86. M. G. Block and F. L. Wightman, A statistically based measure of the acoustic reflex and its relation to stimulus loudness, *J. Acoust. Soc. Amer. 61*, 120–125 (1977).
87. S. S. Stevens, The direct estimation of sensory magnitude — loudness, *Amer. J. Psychol. 69*, 1–25 (1956).
88. R. K. Beedle and E. R. Harford, A comparison of acoustic reflex growth in normal and pathological ears, *J. Speech Hearing Res. 16*, 271–280 (1973).
89. H. D. Smith, Audiometric effects of voluntary contraction of the tensor tympani muscle, *Arch. Otol. 38*, 369–372 (1943).
90. S. N. Reger, Effect of middle ear muscle action on certain psycho-physical measurements, *Ann. Otol. 69*, 1179–1198 (1960).
91. F. B. Simmons, Middle ear muscle activity at moderate sound levels, *Ann. Otol. 68*, 1126–1143 (1959).

92. A. R. Møller, An experimental study of the acoustic impedance and its transmission properties, *Acta. Otol. 60*, 129–149 (1965).
93. A. Starr, Regulatory mechanisms of the auditory pathway, in *Modern Neurology* (S. Locke, Ed.), Little, Brown & Co., Boston, 1969.
94. W. M. Rabinowitz, Acoustic-reflex effects on the input admittance and transfer characteristics of the human middle-ear, unpublished Ph.D. dissertation, MIT, Cambridge, Mass., 1976.
95. M. Loeb and A. J. Riopelle, Influence of loud contralateral stimulation on the threshold and perceived loudness of low-frequency tones, *J. Acoust. Soc. Amer. 32*, 602–610 (1960).
96. D. E. Morgan and D. D. Dirks, Influence of middle ear muscle contraction on pure tone suprathreshold loudness judgments, *J. Acoust. Soc. Amer. 57*, 411–420 (1975).
97. D. E. Morgan, D. D. Dirks, and C. Kamm, The influence of middle-ear contraction on auditory threshold for selected pure tones, *J. Acoust. Soc. Amer. 63*, 1896–1903 (1978).

4
Cochlear Mechanism

The last chapter described the route to the cochlea. We shall now concentrate directly on the sensory mechanism. The cochlea may be conceived of as a transducer which converts the vibratory stimulus into a form usable by the nervous system. However, this is far from the whole picture. The cochlea performs a considerable amount of analysis and is also the major source of aural distortion. In this section we shall first look into how the cochlear mechanism operates, considering first the traveling wave and then hair cell stimulation. We will then examine the electrical potentials of the cochlea. With these aspects as a basis, we will turn to the probable mechanisms of cochlear distortion and frequency selectivity.

TRAVELING WAVES

Resonance theory envisioned the basilar membrane to be under varying degrees of tension along its length in order to account for frequency tuning by place (Chap 2). However, Bekesy [1]* demonstrated that the basilar membrane is not under tension at all. Instead, its elasticity per unit area is essentially uniform, while the width of the membrane increases going from the base to apex (see Fig. 1.12). This widening of the membrane results in a gradation of the stiffness along the cochlear partition such that the membrane is about 100 times stiffer near the stapes

*Much of Bekesy's work is reproduced in his book *Experiments in Hearing* [1]. For simplicity, this reference will be cited whenever possible.

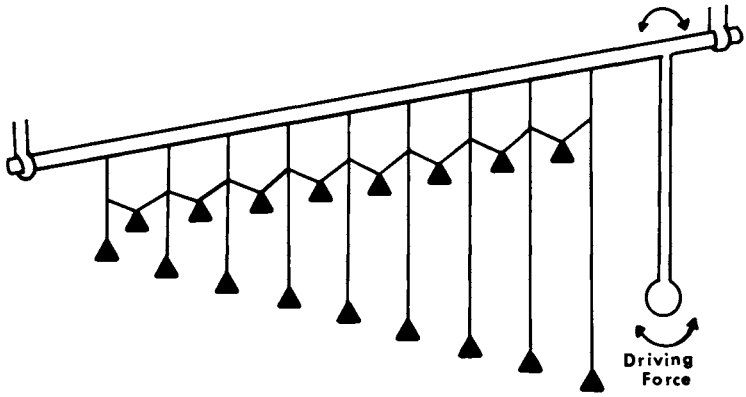

Figure 4.1 Pendulum analogy of traveling wave motion. (Adapted from various drawings by Bekesy [1].)

than at the helicotrema. Because of this stiffness gradient, stimulation of the cochlea results in the formation of a pressure wave that travels from the base toward the apex. In fact, the traveling wave proceeds toward the helicotrema regardless of where the stimulus is applied.

Before examining the details of the traveling wave, let us explore why it occurs. To begin with, the wavelengths of all audible sounds are much larger than the length of the outstretched cochlea. The result is that the pressure exerted on the cochlear partition is uniform over its length. The stiffness gradient of the basilar membrane causes it to act as a series of low-pass filters. Thus, no matter where applied, successively higher frequencies can only initiate vibrations of the cochlear duct closer and closer to the base (where they fall within the passband) [2]. Since the partition's impedance is composed of both stiffness and resistive components* toward the base, and virtually only of resistance toward the apex, the traveling wave is propagated up the partition from places of greater impedance to places of lesser impedance. The speed of the traveling wave decreases with distance from the stapes as it proceeds up the cochlear duct [1].

The pendulum analogy suggested by Bekesy [1] should make the nature of the traveling wave clear. Suppose there is a rod to which a series of progressively longer pendulums are attached (Fig. 4.1). Each pendulum has its own natural frequency; the shorter the string the higher the resonant frequency. We may think of each pendulum as representing a place along the basilar membrane, with the length of the pendulum string corresponding to the stiffness gradient. A driving force is supplied by swinging the heavy pendulum rigidly attached to the

*Refer to Chap. 3 for an overview of the components of impedance.

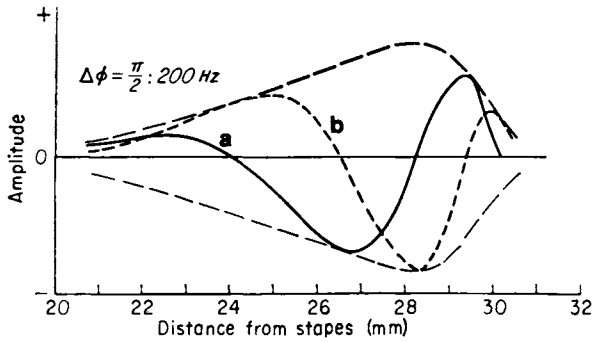

Figure 4.2 Traveling wave pattern for a 200 Hz tone. [From G. Bekesy [1], *Experiments in Hearing* (© 1960, McGraw-Hill). Used with permission of McGraw-Hill Book Company.]

rod. If the rod is rotated back and forth at a particular frequency, the resulting stimulus is applied over the entire rod just as the pressure from a sound stimulus is exerted over the entire cochlear duct. The motion of the rod will cause the pendulum to swing at the stimulus frequency. Of course, the closer the natural frequency of a particular pendulum is to the frequency being applied to the rod, the larger will be its amplitude of swing. There will thus be a particular pendulum that swings with maximum amplitude for each frequency applied to the rod; and changing the frequency at which the rod rocks will move the location of maximum swing to the pendulum whose resonant frequency corresponds to the new stimulus frequency. Note at this point that each pendulum is connected to its neighbor, so that the vibrations of the pendulums interact. The different string lengths cause phase differences between the pendulums, which produces waves. The result is that a sinusoidal motion applied to the rod causes a wave which travels from shorter (higher frequency) to longer (lower frequency) pendulums, with the maximum of the wave occurring at the pendulum which resonates at the frequency of the stimulus.

Let us now proceed to the vibration pattern of the basilar membrane in response to sinusoidal stimulation. The abscissa in Fig. 4.2 is distance (in mm) from the stapes along the basilar membrane, and the ordinate is amplitude of membrane displacement. Two types of information are shown in this figure. The outer dashed lines represent the envelope of the traveling wave as a whole. This envelope outlines the displacement of the cochlear partition during an entire cycle of the wave. Note that the displacement pattern builds gradually with distance from the stapes, reaches a maximum in the vicinity of 28–29 mm, and then decays

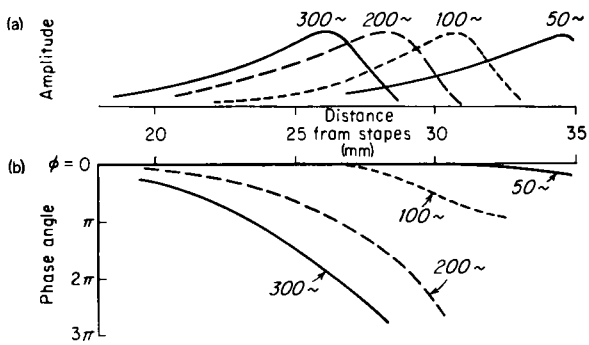

Figure 4.3 Traveling wave envelopes (a) and phase curves (b) for several low frequencies. [From G. Bekesy [1], *Experiments in Hearing* (© 1960 McGraw-Hill). Used with permission of McGraw-Hill Book Company.]

rapidly beyond the point of maximal displacement. The peak of the traveling wave envelope occurs at the place along the basilar membrane where vibration is greatest in response to the stimulus frequency (200 Hz in this case). The traveling wave envelopes for several low frequencies are shown in Fig. 4.3. Observe that these low frequencies result in a displacement pattern covering most of the basilar membrane, although the places of maximum vibration move toward the base with increasing frequency. Standing waves do not arise because there are virtually no reflections from the apical end of the cochlear duct. For very low frequencies (50 Hz) the entire membrane vibrates in phase, so that no traveling wave arises. For higher frequencies, however, notice that there is an increasing phase lag with distance from the stapes; this lag reflects the increasing propagation time and shortening wavelength as the wave proceeds toward the apex.

Figure 4.2 also shows the peak-to-peak amplitudes of membrane displacement at two discrete phases of the wave cycle. For simplicity, assume that the solid line *a* occurs at 0° (time zero) and that the dashed line *b* occurs at 90° (1/4 cycle later). The difference between the two instantaneous displacements depends on what phase of the cycle the wave is in. A full cycle would include a complete set of instantaneous displacement patterns back and forth within the traveling wave envelope, which would begin and end at the solid curve *a* we have designated as our reference for 0° phase. If one imagines this series of instantaneous displacement curves in rapid succession (as in a motion picture with a successive phase in each frame) then the resulting image would be a traveling wave with a maximum at the place shown by the peak of the envelope.

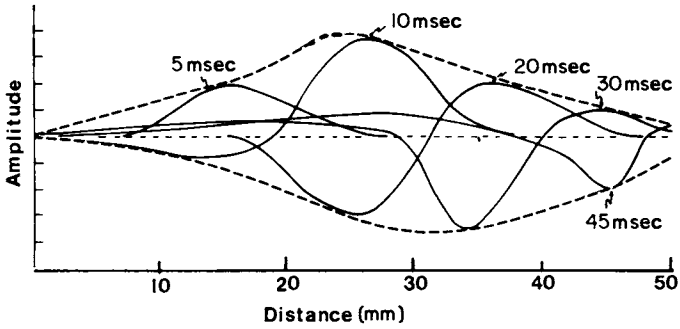

Figure 4.4 Traveling wave in a cochlear model (impulsive stimulus). (From Tonndorf [3]. Permission *J. Acoust. Soc. Amer.*)

How does the cochlear partition respond to an impulsive stimulus, such as a click? Think again of the pendulum analogy. Since an impulsive stimulus, at least ideally, contains energy at all frequencies, it will cause each pendulum to vibrate at its own resonant frequency. Recall that the stiffness of the basilar membrane decreases with distance from the stapes (which is analogous to the increasing lengths of the pendulum strings), and that wave movement is in the direction of decreasing stiffness. Therefore, the vibratory pattern over time is a traveling wave which migrates from the equivalents of shorter (higher frequency) to the longer (lower frequency) pendulums. As expected, the traveling wave envelope is not nearly as peaked as when there is a sinusoidal stimulus. Figure 4.4 shows the envelope and several instantaneous displacements of such a traveling wave in a cochlear model [3]. Such models of the cochlea as the one in which these data were generated maintain the known properties of the cochlear duct, and greatly facilitate experimental manipulations and measurements.

As we shall see in the next section, stimulation of the cochlear hair cells is accomplished by a shearing of the cilia. This is caused by the relative movement of the basilar and tectorial membranes about different axes. Bekesy [1] observed that the nature of these shearing forces changes at different points along the traveling wave envelope. As shown in Fig. 4.5, the shearing vibrations basal to the point of maximal displacement (toward the stapes) were found to be in the radial direction. In other words, the shearing forces were across the partition. The shearing vibrations apical to the peak of the traveling wave (toward the helicotrema) were in the longitudinal direction; i.e., in the direction followed by the cochlear duct.

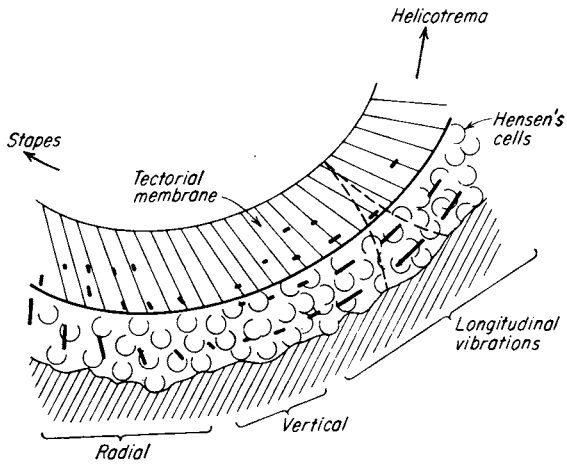

Figure 4.5 Radial and longitudinal shear near the traveling wave peak. [From G. Bekesy [1], *Experiments in Hearing* (©1960). Used with permission of McGraw-Hill Book Company.]

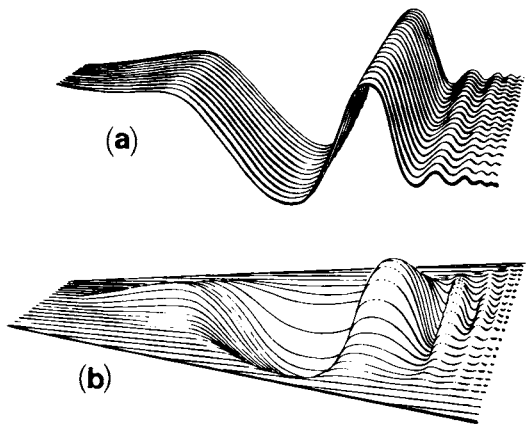

Figure 4.6 Vibration patterns of the basilar membrane as if it were like a vibrating ribbon (a), and when constrained by its lateral attachments (b). Apical end toward the left. (From Tonndorf [4]. Permission *J. Acoust. Soc. Amer.*)

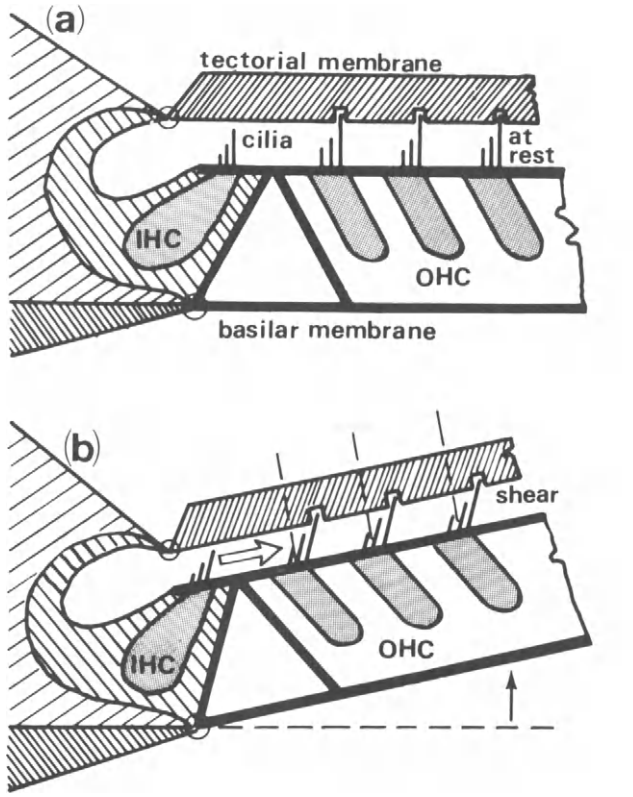

Figure 4.7 Relative position at rest (a) and during movement (b) of the basilar and tectorial membranes. [Adapted from Ryan and Dallos [5]. Courtesy of Little, Brown and Company (©1976).]

Tonndorf [4] explained how this change from radial to longitudinal shear comes about. Figure 4.6a shows how the basilar membrane might move if it were like a freely vibrating ribbon. However, this is not the case. Instead, the membrane is constrained on both sides by its attachments to the walls of the cochlea. The result is a vibration pattern more like the one in Fig. 4.6b, which is based upon the instantaneous pattern in a cochlear model. Vibration under these lateral constraints induces primarily radial forces on the stapedial (high frequency) side of the traveling wave peak and longitudinal forces on the helicotrema (low frequency) side.

HAIR CELL STIMULATION

Our primary task is to determine exactly what it is that stimulates the sensory cells of the organ of Corti. Recall (Fig. 1.14) that the hair cells (HCs) are seated on the basilar membrane, with the reticular lamina on top, and that their cilia are associated in different ways with the tectorial membrane. The stereocilia of the inner hair cells (IHCs) are not embedded in the tectorial membrane at all; whereas the longer (outermost) stereocilia of the outer hair cells (OHCs) are in intimate contact with the overlying membrane. (The shorter OHC cilia do not make contact with the tectorial membrane.) Moreover, whereas the basilar membrane extends from the bottom of the osseous spiral lamina, the tectorial membrane is hinged at the upper lip of the limbus. As a result, the axis of rotation is different for these two membranes (Fig. 4.7). Thus, displacement of the cochlear partition results in the reticular and tectorial membranes being moved relative to each other; so that there is a shearing force upon the cilia between them. This shearing force results in stimulation of the HCs [1]. The OHCs are stimulated by shearing forces in the radial direction, while the IHCs are stimulated by longitudinal shearing. The import of the difference between IHCs and OHCs will be made apparent below.

Outer Hair Cells

The direction of cochlear duct displacement, and thus the direction in which the cilia are sheared, is determined by the motion of the stapes footplate at the oval window. Inward motion displaces the duct downward toward the scala tympani, causing the hairs to bend in one direction; whereas outward motion deflects the partition up toward the scala vestibuli (Fig. 4.7b), causing the cilia to be sheared in the other direction.

The importance of the direction of shear is demonstrated in Fig. 4.8, which shows that sensory HCs are directionally sensitive to the bending of their stereocilia. Bending of the stereocilia toward the longer kinocilium results in excitation (measured by the activity of the associated neuron), while bending in the opposite direction is inhibitory. Although cochlear OHCs do not have a kinocilium, they seem to have a centriole representing a rudimentary kinocilium at the base of the "W" arrangement of the stereocilia. Since displacement of auditory HC cilia toward the centriole has been shown to result in excitation [7], one would expect bending toward the base of the "W" to cause excitation of the associated auditory neuron(s). The base of the "W" faces toward the outside of the

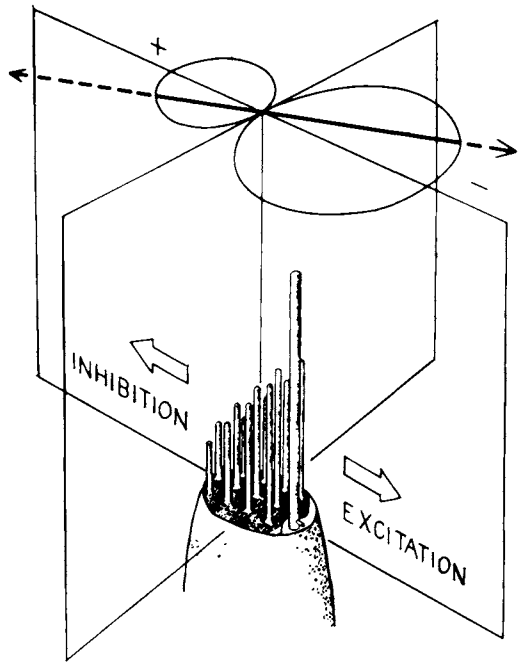

Figure 4.8 Directional sensitivity of sensory hair cells. (From Flock [6]. Permission Springer-Verlag, Inc.)

cochlear duct (away from the modiolus). Therefore, bending of the cilia in this direction occurs when the cochlear partition is deflected upward (Fig. 4.7b). We shall see in the next chapter that auditory nerve fibers do, in fact, discharge in a manner coordinated with elevation of the cochlear partition.

The exact mechanism of HC stimulation is still controversial. One explanation suggests that the cilia are actually bent. Another theory regards the cilia as being rigidly affixed to the cuticular plate, so that shearing forces displace the latter in a hingelike manner. Regardless of this, however, it is clear that the OHCs are stimulated by radial shearing forces which are proportional to the displacement of the basilar membrane [8,9].

Inner Hair Cells

In the light of what we have learned about the OHCs, the arrangement of the IHCs would lead us to expect that excitation would result when their cilia are displaced radially. However, as already mentioned, Bekesy

[1] found that the IHCs are responsive to shearing forces in the longitudinal rather than the radial direction. Recall also that the IHC cilia do not make actual contact with the tectorial membrane, as do the longer cilia of the OHCs. These factors suggest that the IHCs may be stimulated by a different mechanism than the OHCs. This idea is supported by the observations that the thresholds of the IHCs are about 30–40 dB poorer than for the OHCs, and that they are less sensitive to drug-induced damage (ototoxicity) [9,10].

What, then, is the mechanical stimulus for the IHCs? To study this, Dallos and associates compared the cochlear microphonics (see the section Cochlear Electrical Potentials, below) generated by the IHCs and OHCs of the guinea pig [9]. It is possible to differentiate the responses of the two cell groups, since the normal cochlear microphonic is derived chiefly from the OHCs, and ototoxic drugs tend to obliterate these same cells. Thus, the cochlear, microphonic responses of the two cell groups were separated by measuring the responses before and after the animals were injected with an ototoxic drug (kanamycin). The output of the OHCs was found to be proportional to basilar membrane *displacement*, whereas the response of the IHCs was proportional to the *velocity* of basilar membrane displacement. In other words, the OHCs respond to the amount of displacement and the IHCs respond to the rate of change. The reason for this difference appears to lie in the different relationships of the inner and outer HC cilia to the tectorial membrane. Since the OHC cilia attach to the tectorial membrane, an effective stimulus is provided by the relative movements of the reticular and tectorial membranes, which depend upon basilar membrane displacement. The IHC cilia, on the other hand, stand free of the tectorial membrane. Their stimulus is provided by the drag imposed by the surrounding viscous fluid as the basilar membrane is displaced; the greater the velocity of basilar membrane displacement, the greater the drag exerted upon the cilia.

COCHLEAR ELECTRICAL POTENTIALS

Several electrical potentials may be recorded from various parts of the cochlea and its environs. References have already been made to the cochlear microphonic, which is one of these. We might think of four general kinds of potentials, including (1) resting, (2) receptor, (3) generator, and (4) action (spike) potentials. Actually, (1) and (2) are cochlear in origin, while (3) and (4) are neural. Resting potentials are the positive and negative direct current (dc) polarizations of various tissues and their surrounding fluids. Receptor and generator potentials were dichotomized by Davis [11]. A receptor potential is an electrical response from a

receptor cell (e.g., a cochlear HC) which results when the cell is stimulated. The cochlear microphonic is a receptor potential. Note that the presence of a receptor potential does not mean that the nervous system is aware of the stimulus; it just reflects the fact that the HC itself has responded. The generator potential, on the other hand, is the electrical signal that actually triggers the transmission of a spike potential along the axon of a neuron. The latter is the all-or-none spike discharge of the nerve fiber and is called the action potential (see Chap. 2).

We may also think of these potentials in terms of whether they are derived from a single cell or from many cells. This is an important dichotomy because compound potentials, which include the contributions of many cells at various distances from the electrode, may include responses to different phases of the same stimulus (or to different stimuli) at different times and in different ways. Thus, the electrode "sees" a much different picture than it would if it were recording from a single cell.

The method of measurement chosen determines whether single-cell or compound (gross) potentials are recorded. The differences among methods boil down to a dichotomy between microelectrodes and gross electrodes. Basically, microelectrodes are electrodes small enough to impale an individual cell, so that its activity can be motioned in relative isolation from other cells. Microelectrodes often have diameters much smaller than 1 μm. Gross electrodes on the other hand, are not small enough to enter a cell. They include electrodes ranging in size from those which can be inserted into a nerve bundle (which are sometimes themselves called microelectrodes) to the large surface electrodes used in EEG and other clinical applications. Gross electrodes are used in two general ways. The first method uses a single active electrode to measure the electrical activity in an area (relative to a ground electrode). The other method, that of differential electrodes, uses two active electrodes the signals of which are added, subtracted, or averaged, depending upon the specific need.

Endocochlear and Intracellular Potentials

The unusual chemical situation in the cochlear duct was discussed in Chap. 1. Recall that endolymph is among the very few extracellular fluids to be high in potassium; and that the proper ionic balance of the cochlear fluids has a profound effect upon the functioning of the organ of Corti.

Bekesy [1] measured the cochlear resting potentials in the guinea pig by inserting an electrode into the perilymph of the scala vestibuli (which

he set as a 0 mV reference), and then advancing it down through the scala media and the organ of Corti, and into the scala tympani. He found a positive 50–80 mV resting potential within the scala media. As the electrode was advanced through the organ of Corti, the voltage dropped from about +50 mV to about −50 mV, and then returned to near zero as the electrode passed through the basilar membrane into the perilymph of the scala tympani. Subsequent experiments [12] revealed the scala media potential to be about +100 mV. This positive resting voltage is called the endocochlear potential (EP). The EP is greatest near the stria vascularis [13]; and this is the only area that keeps a positive potential when the endolymph is drained from the cochlear duct [14]. Furthermore, the EP is still present after damage to or removal of the HCs, and is found even in the waltzing guinea pig, whose HCs are congenitally absent [14,15]. The stria vascularis is thus the source of the EP.

The negative resting potential in the organ of Corti was confirmed by Tasaki et al. [13], who found its magnitude to be about −60 to −70 mV. The origin of this potential has been the subject of some controversy, some saying that it is an intracellular potential and others positing that it is extracellular (from the cortilymph). Two strong arguments lead one to conclude that it is an intracellular potential [16]. First, this potential cannot be recorded for more than a few minutes, which suggests that it is due to an individual cell which stops functioning after a while because of the damage caused by the invading electrode. The potential would be recordable for longer durations if it were extracellular. Second, if cortilymph had a negative potential, there could be no polarization difference between the insides of the HCs and the auditory nerve fibers and their surrounding fluid, because the functioning of these cells depends upon the presence of a polarization difference across their membranes.

The net result of the positive EP and negative intracellular potential is an electrical polarity difference of 160 mV or more across the reticular membrane. This gradient may be thought of as an amplifier, because the total potential gradient is far greater than what would result from the intracellular potentials of the HCs alone.

Cochlear Microphonic

In 1930, Wever and Bray [17] reported that if the electrical activity picked up from the cat's auditory nerve is amplified and directed to a loudspeaker, then one can talk into the animal's ear and simultaneously hear himself over the speaker. This result demonstrated that the electrical potentials being monitored were a faithful representation of the stimulus waveform.

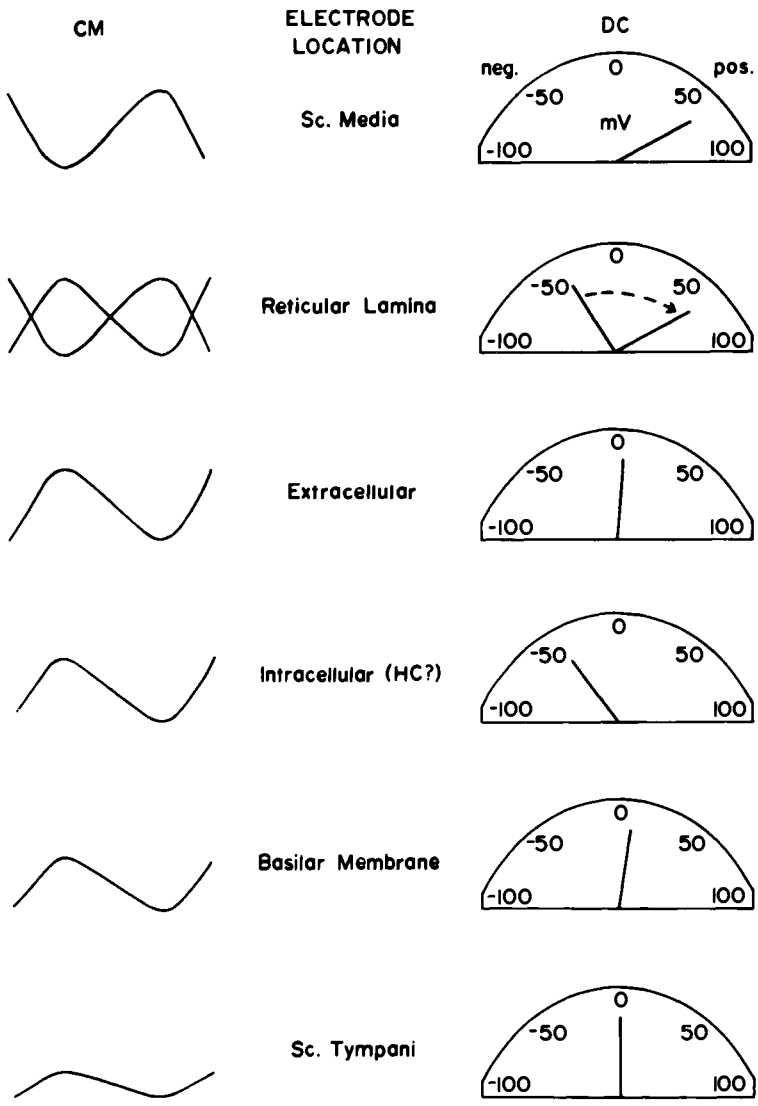

Figure 4.9 Polarity reversal of the cochlear microphonic and DC resting potentials recorded by an electrode being advanced through the cochlea. (Modified after Tasaki et al. [13].)

Wever and Bray originally thought that they were monitoring the auditory nerve alone. A major problem with this early interpretation is that no other nerve mimics the stimulus waveform, and why should the auditory nerve be the only exception? It was soon shown that the eighth-nerve action potential was not the only signal being recorded. Instead, the Wever–Bray effect was actually due to an ac electrical potential being picked up by the electrodes placed near the nerve [18,19]. It was found, for example, that the response was stronger at the round window than at the nerve, and that the Wever–Bray effect could still be demonstrated even if the nerve was destroyed or anesthesized. Such findings demonstrated that the ac potential which reflects the stimulus with such remarkable fidelity is generated by the cochlea; and Adrian [18] coined the term cochlear microphonic (CM) to describe it.

The relationship between the HCs and the CM is firmly established [16,20]. Bekesy [1] demonstrated that CMs are elicited by basilar membrane deflections. An interesting experiment revealed that the CM is generated at the cilia-bearing ends of the HCs [13,21]. This experiment was based on the fact that the polarity of an electrical potential is out of phase when measured from opposite sides of its generator. To locate the generator of the CM, the polarity of the potential was monitored in an electrode which was advanced through the cochlea from the scala tympani toward the scala vestibuli (Fig. 4.9). The polarity of the CM reversed when the electrode crossed the reticular lamina, suggesting that this is the location of the CM generator. The site of CM polarity reversal was localized to the cilia-bearing ends of the HCs (at the reticular lamina) because the dc resting potential changed dramatically to the positive EP at this point, indicating that the reticular membrane had been impaled.

Cochlear microphonics are generated by both inner and outer HCs. However, there is convincing evidence [10,16,22,23] that the potentials from the OHCs are greater and have lower thresholds. For example, Dallos [23] found that the CMs of normal guinea pigs were elicited by stimuli 30–40 dB lower than those in animals whose OHCs were destroyed with kanamycin.

How is the CM generated? There are several theories, but Davis' variable-resistance model has enjoyed the widest acceptance. Think of the sources of the cochlear resting potentials as biological batteries, generating a current flowing through scala media, the basilar membrane, and the scala tympani. (One pole of the biological battery goes to the cochlear blood supply, completing the circuit.) A sound stimulus would then be represented electrically (the CM) if it caused the resistance to current flow to change in accordance with the stimulus waveform [1,24,25]. Davis [25] proposed that this variable resistance is provided by

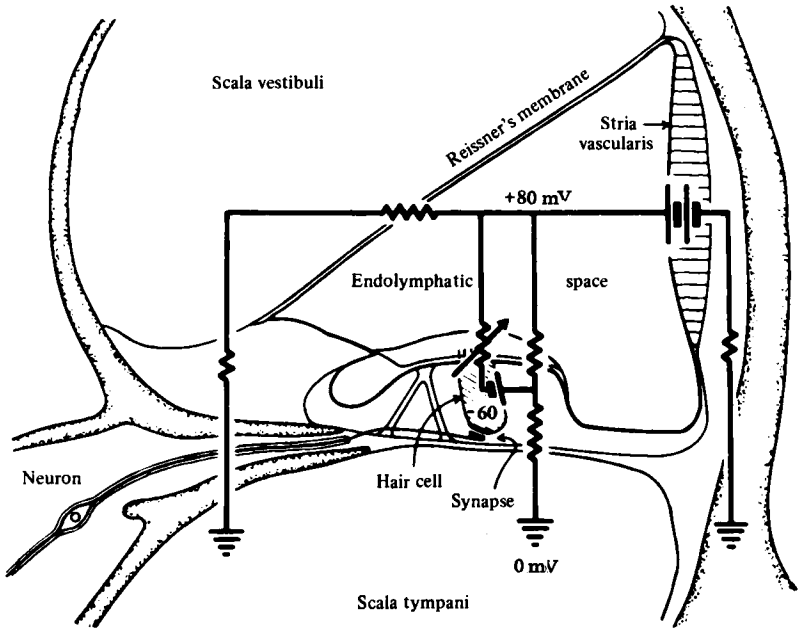

Figure 4.10 Davis' variable resistance model (schematic). [From *Hearing and Deafness, Third Edition,* by Hallowel Davis and S. Richard Silverman, Editors. Copyright 1947, © 1960, 1970.by Holt, Rinehart & Winston, Inc. Reprinted by permission of Holt, Rinehart & Winston.]

the movements of the HC cilia, which occur in response to basilar membrane displacement. In other words, the movements of the cilia modulate the resistance, which in turn sets up an alternating current. This ac potential is monitored as the CM. The amount of current (CM magnitude) depends on the forces exerted upon the hairs, which is ultimately determined by the intensity of the sound stimulus. Figure 4.10 shows a recent version of the model.

Davis' resistance model, while not directly tested, remains the most widely accepted explanation of CM generation. It is strongly supported by the observation that changes in CM polarity and magnitude result from experimentally induced anoxia and electrical currents [27]. This evidence is supportive because, for example, one would expect an artificially induced current to enhance the one produced by the physiological battery, resulting in a larger CM.

Two alternative models will be mentioned by way of examples. Wever [20] suggested that the CM is generated by the HC itself, rather than by the interaction of the HCs with other structures. Another mech-

anism was suggested by Katsuki [28]. He cited measurements by Tanaka et al. [29] which revealed that the CM within the guinea pig HC was small, whereas the CM was largest in the scala media. This result suggests that the origin of the CM might lie outside of the HC, and casts doubt on the generally accepted concept of CM phase reversal at the reticular lamina. Katsuki proposed that the CM may be generated by ion exchanges on the surface of the HCs or within the tectorial membrane.

The CM is a graded potential — its magnitude changes as the stimulus intensity is raised or lowered. This is shown by the CM input–output (I–O) function (Fig. 4.11). The magnitude of the CM increases linearly with stimulus level over a range of roughly 60 dB [30], as is shown by the straight line segment of the I–O function. This classic figure shows the linear response extending down to about 0.4 mV, but subsequent studies have recorded CM magnitudes as low as a few thousandths of a microvolt [20]. With adequate stimulation, the CM does not appear to have a threshold, per se; instead the smallest response seems to grow out of the noise floor, being limited by the recording method, equipment, and physiologic noise. Saturation occurs as the stimulus level is raised beyond the linear segment of the I–O function, as shown by the flattening of the curve. Increasing amounts of harmonic distortion occur in this region. Raising the intensity of the stimulus even further causes overloading, in which case the overall magnitude of the CM actually decreases and damage ensues.

Distribution of the CM

Recall that large CM responses are recorded from the round window [17]. This placement was used extensively by Wever and Lawrence [30]. Similar responses are recorded at the cochlear promontory [31]. It is thus important to know what contributes to the round-window response. To study this, Misrahy et al. [32] successively destroyed sections of the guinea pig's cochlea, beginning at the apex and working downward, while monitoring the CM in the basal turn. They found that the upper turns did not make significant contributions to the CM in the basal turn. It may thus be concluded that the CM recorded at the round window is for the most part derived from the activity of the basal turn.

Another approach to studying the distribution of the CM is to place electrodes along the cochlea to obtain CMs from more or less restricted locations. The space and time distribution of the CM along the cochlear spiral was first reported with this method by Tasaki et al. [33]. They inserted pairs of differential electrodes into the scalae tympani and vestibuli of the guinea pig. One electrode pair was placed in each of the four turns. This method allowed them to separate the CM, which is of oppo-

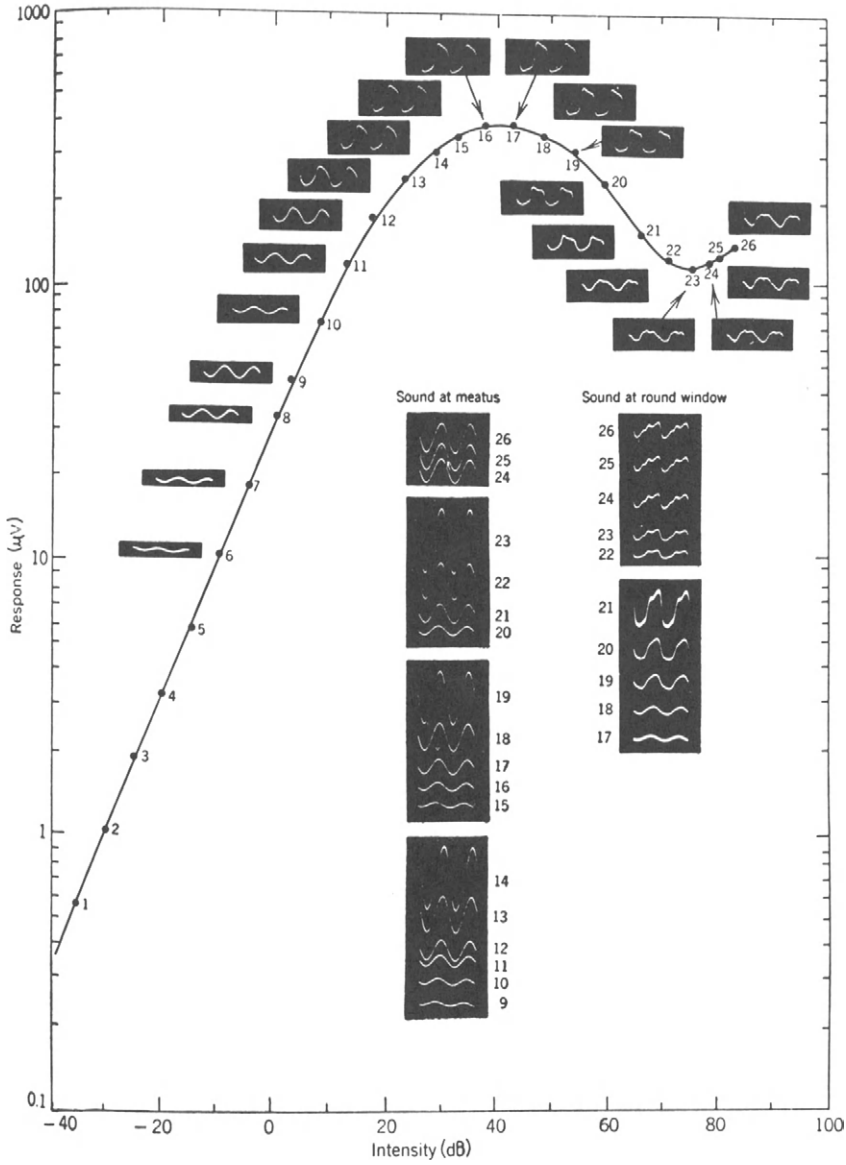

Figure 4.11 Cochlear microphonic input–output function (cat) for a pure tone stimulus. Inserts along the curve are CM responses at points on the function (the apparent decrease in size between points 9 and 10 is due to a change in oscilloscope sensitivity, not a drop in CM magnitude). Inserts below the function are sound levels at the ear canal and round window. [From Ernest Glen Wever and Merle Lawrence, *Physiological Acoustics* (copyright 1954 by Princeton University Press): Fig. 62, p. 167. Reprinted by permission of Princeton University Press.]

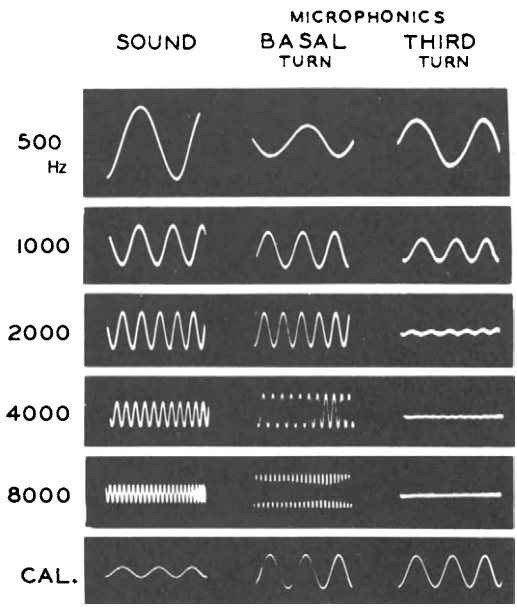

Figure 4.12 Cochlear microphonics recorded with differential electrodes in the first and third turns of the guinea pig cochlea. [From Davis [34], in *Neural Mechanisms of the Auditory and Vestibular Systems* (G. L. Rasmussen and W. F. Windle, Eds.), 1960. Courtesy of Charles C Thomas, Publisher, Springfield, Illinois.]

site polarity in the two scalae, from the auditory nerve action potential (AP), which is always negative. (Addition of the two out-of-phase signals will cancel out the CM and enhance the AP, whereas subtraction will remove the AP and enhance the CM.) They found that the space–time distribution of the CM was consistent with the propagation pattern of the traveling wave. Low-frequency signals had large CMs at the apical turn and minimal responses at the base; while high frequencies had maximal CMs at the base and no response at the apex (Fig. 4.12). They also found that the velocity of the signal was very high in the first turn and smaller toward the apex.

Of particular interest are the findings of Honrubia and Ward [35]. They used microelectrodes in the scala media to measure the distribution of CM responses along the cochlear duct. The electrodes were precisely placed at intervals determined from place–frequency maps of the cochlear partition [1,36]. Tones were then presented at fixed intensities, and CM magnitude was measured at various distances along the

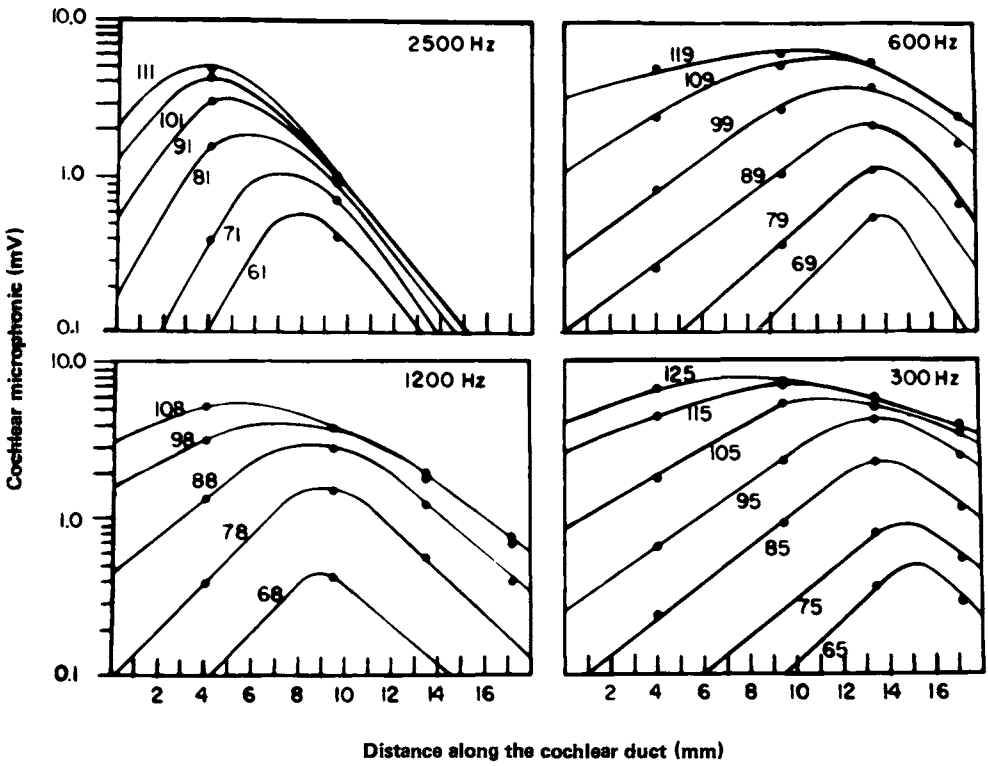

Figure 4.13 Cochlear microphonic magnitude as a function of distance along the basilar membrane for four frequencies presented at various intensities. (From Honrubia and Ward [35]. Permission *J. Acoust. Soc. Amer.*)

cochlear duct. For example, the CM was measured at each electrode site in response to a 1200 Hz tone at 78 dB, a 2500 Hz tone at 101 dB, etc. Figure 4.13 shows typical results, with stimulus level as the parameter. Consider first the distribution of CMs at the lowest stimulus levels (most sensitive curves). Consistently with the traveling wave envelope, the peaks of these curves occur closer to the apex as frequency decreases. However, the CM curves do not line up exactly with the basilar membrane displacement curve. This discrepancy is shown clearly in Fig. 4.14, in which the CM curve is wider and less peaked than the basilar membrane tuning curve. The difference is largely due to the fact that the electrode "sees" CMs generated by thousands of HCs in its general vicinity rather than by just those at its precise point of insertion [37]. In other words, the electrode really monitors a weighted average of many CMs, which has the effect of flattening the curve somewhat. We will return to this point.

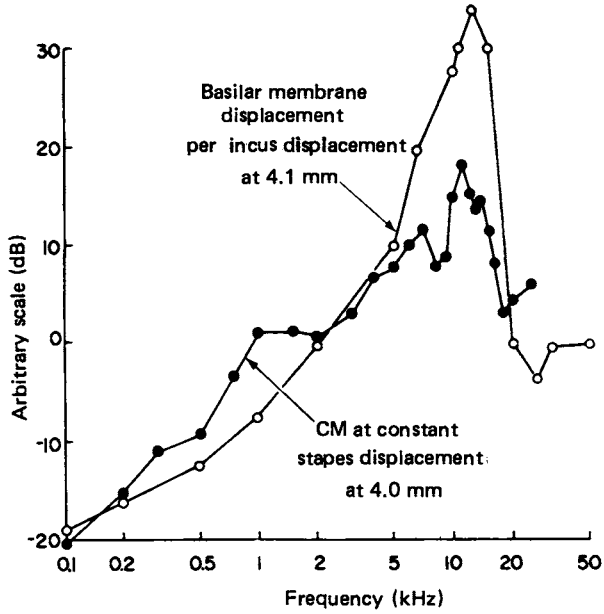

Figure 4.14 Comparison of basilar membrane tuning curves [36] and the CM curve at similar locations in the guinea pig cochlea. (From Dallos et al. [37]. Permission *J. Acoust. Soc. Amer.*)

Returning to Fig. 4.13, we see that CM magnitude increases with stimulus level for each frequency. Also, the place of maximum CM magnitude shifts downward toward the base as intensity increases. This may at first seem inconsistent with the place principle. However, the basal shift of maximum CM response is probably due to the wider range over which the more basal generators respond linearly [16]. In other words, as stimulus intensity increases, the CMs from the most sensitive place along the basilar membrane become saturated sooner than do the responses from more basal regions. Thus, CMs generated toward the base continue to increase in magnitude when those from the most sensitive place have already become saturated. The place of maximal CM response therefore shifts downward along the cochlear partition (upward in frequency).

Since the electrode picks up CMs from generators along the duct in a manner which is spatially weighted (closer HCs are better represented than distant ones), the potentials from lower-frequency (apical) generators tend to be better represented than higher-frequency (basal) generators. This stiuation also occurs for the following reasons [38,39]: Recall that phase changes more for high frequencies than for low with listance along the cochlear partition. Many CMs from higher-frequency

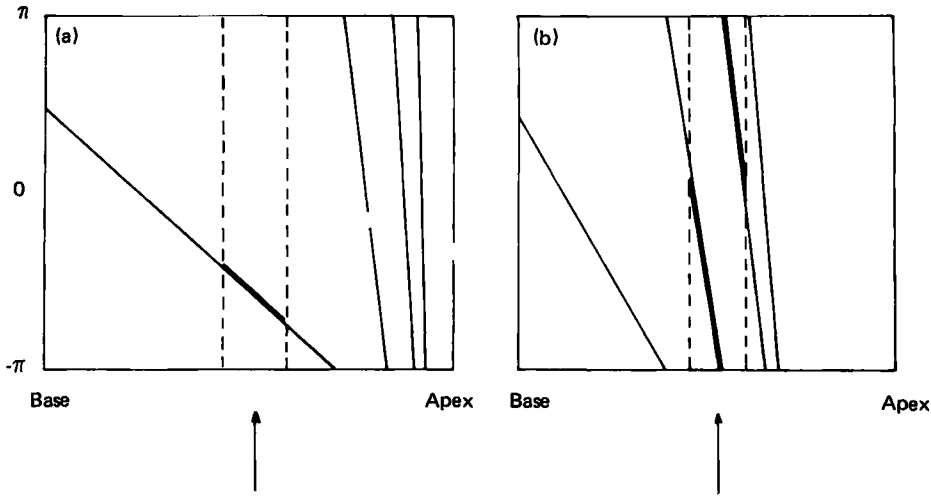

Figure 4.15 Cochlear microphonic phase along the cochlear duct for a low frequency (a) and a high frequency (b). Arrows show electrode sites. Dotted lines show the areas "seen" by the electrode. The CM sources are in phase in (a) and out of phase in (b). [From Dallos [16], *The Auditory Periphery*, ©1973, Academic Press, New York.]

generators will thus arrive at the electrode out of phase due to the rapid phase change, which will cause some of these potentials to cancel one another (Fig. 4.15). On the other hand, the essentially inphase lower frequency CMs do not cancel. The result is deemphasis of the higher frequency regions, so that the electrode in effect gives a low-pass filtered picture of the response.

Summating Potential

The summating potential (SP) was first described by Davis et al. [40] and Bekesy [1] in 1950. Unlike the CM, which is an ac potential, the SP is a shift in the dc baseline in response to sound stimulation. Bekesy called this shift "dc fall." Subsequent research revealed that it may be a positive (SP^+) or negative (SP^-) baseline shift [41], and we shall see that SP polarity is associated with the traveling wave envelope. This association is in contrast to the CM, which reflects the instantaneous displacement of the cochlear partition. Like the CM, the SP is a graded potential — it increases in amplitude as stimulus level is raised [42].

The precise origin of the SP is still debatable. Since the SP is a dc response to sound stimuli, which are ac events, it is by definition nonlinear [43]. It may, in fact, be the result of nonlinear process in the

cochlea [38,44,45]. This does not imply, however, that the SP is just an artifact. It is definitely a physiologic response [46], and there is good reason to believe that it is produced by (or is at least related to) HC activity. Since SP polarity reverses from one side of the cilia-bearing ends of the HCs to the other [21], in a manner reminiscent of the effect previously described for the CM [13], it is possible that this is the site of SP generation. On the other hand, there is at least preliminary evidence [47] that the average component of the SP (see below) has a neural origin, and may actually be the generator potential of the auditory nerve.

Davis et al. [10] produced relatively selective damage to the OHCs of guinea pigs with streptomycin and surgery. Although there was also damage of the IHCs and a few exceptional cases, they suggested that the SP^- is produced by the IHCs while the SP^+ is related to the OHCs. However, Dallos and Bredberg [16] found that both positive and negative SPs (measured as the potential difference between the scalae vestibuli and tympani, see below) are greatly affected by OHC destruction. Furthermore, Stopp and Whitfield [48] observed both SP^+ and SP^- in pigeon cochleas, which do not have differentiated inner and outer HCs.

Honrubia and Ward [42] measured the SP and CM simultaneously in each turn of the guinea pig cochlea, using electrodes located in the scala media. We have already seen that the envelope (distribution of amplitude along the length of the cochlear duct) of the CM is a reasonable representation of the traveling wave envelope. Honrubia and Ward found that the SP was positive on the basal side of the CM envelope and negative on the apical side. This result suggests that the SP is positive on the basal side of the traveling wave and becomes negative apical to the traveling wave peak. Such a correspondance between the SP^- and the traveling wave peak supports the association, which has been suggested by several researchers [49,50], of the SP^- with auditory nerve excitation. Honrubia and Ward also proposed that the SP^+ may inhibit neural excitation along the basal part of the traveling wave. This would restrict the part of the traveling wave envelope capable of causing neural excitation to the peak region, which would "sharpen" frequency selectivity.

Dallos and colleagues [16,23,43,51,52] used a somewhat different recording approach, which makes the important distinction between the average potential of both the scala vestibuli and the scala tympani on the one hand and the potential gradient (difference) across the cochlear partition on the other. This at first complicated distinction is clarified in Fig. 4.16. One electrode is in the scala vestibuli and the other is in the scala tympani. Each one registers the SP at the same cross-sectional plane along the cochlea, but from opposite sides of the scala media. Subtracting the SP in the scala tympani (ST) from the SP in the scala vestibuli

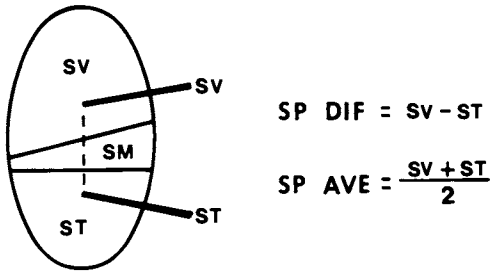

Figure 4.16 Recording DIF and AVE components of the summating potential from scala vestibuli (SV) and scala tympani (ST). (Modified from various drawings by Dallos.)

(SV) (SV − ST) gives the potential difference of the SP across the coclear partition. This difference is called the DIF component. The average (AVE) component is obtained by simply averaging the SPs from both scalae [(SV + ST)/2]. The AVE component thus expresses the common properties (common mode) of the SP on both sides of scala media.

The spatial arrangements of the DIF SP and AVE SP are shown superimposed upon the traveling wave envelope in Fig. 4.17. Note that the DIF component becomes negative in the vicinity of the peak of the traveling wave envelope, a situation which resembles the spatial distribution of SP^+ and SP^- discussed above. The polarity of the AVE component is essentially the reverse, being positive around the traveling wave peak and negative elsewhere. Figure 4.18 shows the relationships among the DIF and AVE components of the SP; the CM; and the traveling wave. These results were obtained from the third turn of one guinea pig

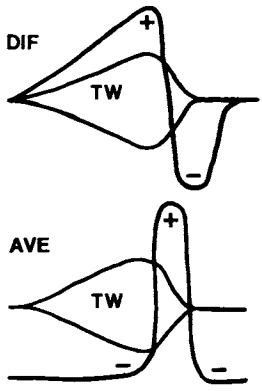

Figure 4.17 Spatial relationships of AVE SP and DIF SP to the traveling wave (TW) envelope. (Modified from Dallos [43], with permission.)

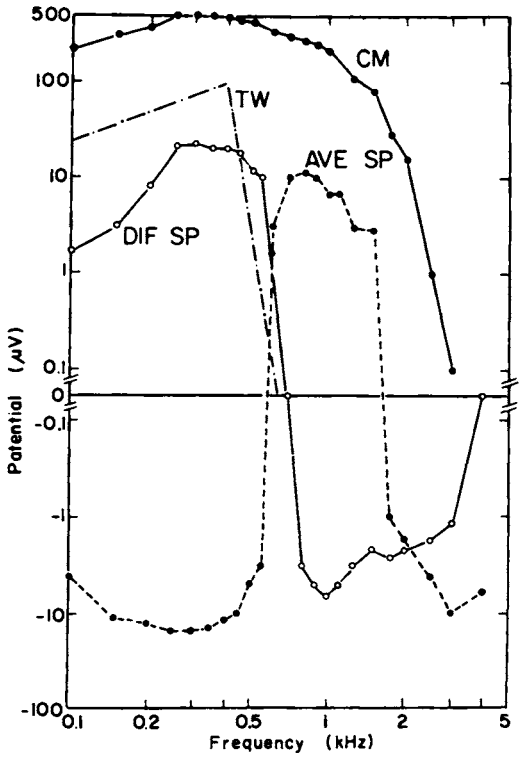

Figure 4.18 Amplitude as a function of frequency for the DIF and AVE summating potentials, cochlear microphonic (CM), and traveling wave (TW) for a 400 Hz stimulus (50 dB SPL). (From Dallos et al. [52]. Permission *Acta Otol.*)

cochlea, in which the electrodes were 14.5 mm from the base (the approximate location for 400 Hz). The traveling wave peaks at 400 Hz, with a rising slope of 6 dB per octave below the peak and a falling slope of −100 dB per octave above it. As in Fig. 4.17, the DIP SP is negative and the AVE SP is positive in the vicinity of the traveling wave peak. Note that the SP curves are more localized (at least for relatively low intensities) than the CM. Also, Dallos [23] has shown that the SP tends to be tuned (responds with greatest amplitude) about two-thirds of an octave higher in frequency than the microphonic.

DISTORTIONS IN THE COCHLEA

Pure place theory attributed aural distortions to the conductive mechanism. However, the middle ear system has been shown to be remark-

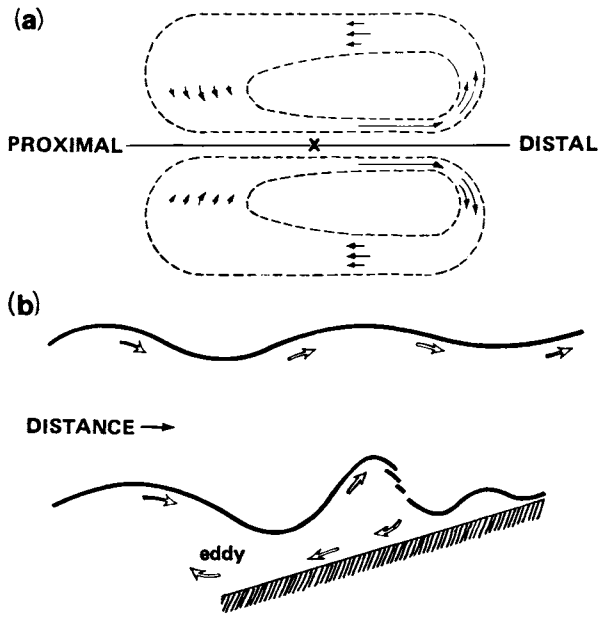

Figure 4.19 (a) Eddy currents in a cochlear model. The traveling wave peak is represented by the X. (From Tonndorf [55]. Permission *J. Acoust. Soc. Amer.*) (b) Surface waves in open water and in the production of eddies (undertow) when striking a surface. [From Tonndorf [54], Cochlear nonlinearities, in *Basic Mechanisms in Hearing* (A. R. Møller, Ed.), ©1973, Academic Press, New York.]

ably linear over a wide dynamic range [30,53]. Actually, no part of the ear is completely linear [16]; but the greatest proportion of nonlinear distortions are derived from the cochlea [2,16,30,54].

Bekesy [1] reported that eddy currents developed in opposite directions in the two scalae of a cochlear model (Fig. 4.19a). The locations of the eddies were related to the traveling wave maximum, being toward the base for high frequencies and toward the apex for low. As is shown by the lengths of the arrows in Fig. 4.19a, Tonndorf [55] found that the eddy currents accelerated in the narrower distal end and slowed in the wider proximal end. This effect occurs because, if the same amount of fluid per unit time is to flow through a narrow channel as through a wide one, then it must move faster in the narrow channel. The development of these eddy currents is analogous to the formation of undertow when seawater hits a beach (Fig. 4.19b). These eddies are inherently nonlinear; however, they are a reflection rather than the cause of the distortion process [56]. The distortion occurs because the inertia of the fluid and the restoring force of the basilar membrane are not exactly out of phase at high stimulus levels.

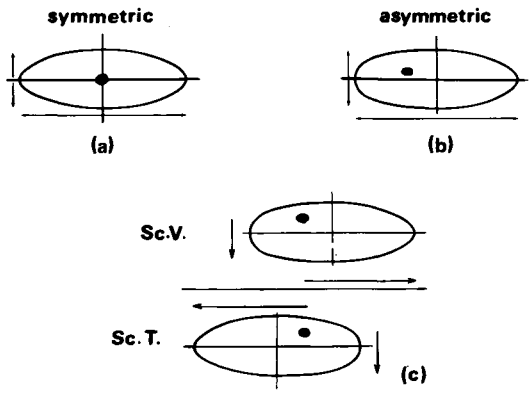

Figure 4.20 Fluid particle orbits in a cochlear model. (From Tonndorf [57]. Permission *J. Acoust. Soc. Amer.*)

Tonndorf [2,54,57] suggested two mechanisms of distortion based on cochlear hydromechanics (mechanical properties of the cochlear fluids). Amplitude-dependent distortions occur at high stimulus levels, for which eddy currents appear. As stimulus level increases the elliptical eddies begin to flatten, which is a form of peak clipping. In addition, other eddies develop at places along the partition corresponding to the harmonics of the original (primary) tone. This second mechanism is amplitude independent; i.e., it does not depend on stimulus level. Amplitude-independent distortion is schematically represented in Fig. 4.20. If two primary tones (500 and 510 Hz) are presented, a difference tone (or beat frequency) will be heard (510 − 500 = 10 Hz beat frequency). The beat frequency is caused by a distortion due to movements of individual fluid particles in elliptical orbits that are asymmetric (Fig. 4.20b) rather than symmetric (Fig. 4.20a). This asymmetry causes the particle orbits in the scalae vestibuli and tympani to be displaced relative to one another in a "push–pull" manner (Fig. 4.20c). The resultant energy at the difference (beat) frequency, which is lower in frequency than the primary tones, then proceeds up the cochlear partition (forming a new traveling wave if the beat frequency is high enough).

It is undisputed that distortions at high intensities are related to hydrodynamic nonlinearities which are amplitude dependent. However, CM data are inconsistent with the amplitude-independent mechanism at low and moderate levels. An alternate source of distortion is within the HC, and is most likely associated with the electrical generation of the CM [16,30]. This idea is supported by CM studies [16,53,58–61], and is likely because the variable-resistance model of CM production is inherently nonlinear [16].

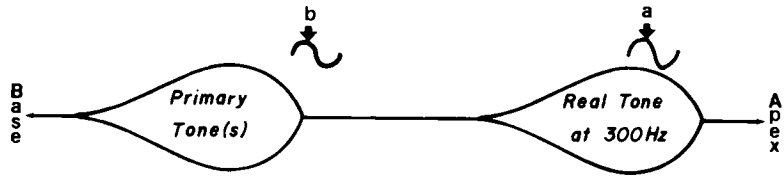

Figure 4.21 Highly schematic locations of maximum CM responses for a 300 Hz distortion product mediated by a traveling wave (point a), and not mediated by a traveling wave (point b).

Suppose that a distortion component were due to cochlear hydromechanics (mediated by a traveling wave). Such a traveling wave would result in basilar membrane displacement with a peak at the same place as would be produced by a real tone of the same frequency as the distortion product. Thus, a 300 Hz distortion product would produce the same traveling wave as a real 300 Hz tone; and the associated CM pattern would reach a maximum at the 300 Hz point (point a in Fig. 4.21). On the other hand, if the CMs associated with the distortion are not localized at point a, then it is unlikely that the distortion is the result of a traveling wave caused by a hydrodynamic nonlinearily. Several studies [16,53,59–61] demonstrated that the CMs corresponding to various kinds of distortion at low and moderate levels of stimulation were localized near the primary tones (point b) rather than at the place of a real tone of the same frequency (point a). This result suggests that distortions at low and moderate levels may be due to nonlinearities in the mechanoelectrical transducer; i.e., may be associated with the production of CMs at the HC level.

Two findings support this theory. Dallos [16] found that the magnitudes of CMs due to distortion products were greater than the magnitudes of CMs associated with actual tones of the same frequency. Dallos, et al. [58] found direct evidence that lower-level distortions are related to the electrical rather than the hydromechanical process. They polarized the cochlear partition with a dc current, which affects the electrical process that generates the CM. If the distortion products were due to traveling waves (hydrodynamic), then they would be affected by the dc current in the same way as the primary tones, because both were produced prior to the effects of the dc bias. However, if the distortion is due to the electrical process at the HC, then it should be affected differently than the primaries because one is produced prior to the electrical process and the other is the result of it. Figure 4.22 shows an example of the results of this experiment. The CMs associated with the primary and the distortion product were affected (a) differently at

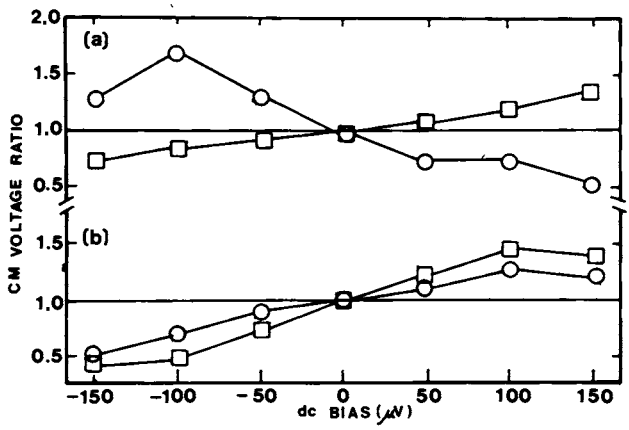

Figure 4.22 Effects of electrical bias current on CMs produced by primary tones (squares) and distortion components (circles) at low (a) and high (b) levels. (Redrawn from the data of Dallos et al. [58].)

low stimulus levels (67 dB SPL), but (b) similarly at high levels. The low-level distortion thus seems to result from electrical processes, whereas the high-level distortion is hydromechanical.

These studies suggest that distortions in the cochlea are due to a two-stage process, one mechanism operating at low, and the other at high, stimulus levels. The higher-level distortion products, which result from a chain of events similar to that for a directly presented sound (cochlear hydrodynamics), are the most certainly audible. Since the lower-level distortion products do not appear to occur by the same route, there is still some question as to their audibility [61].

COCHLEAR FREQUENCY SELECTIVITY

Basilar membrane displacement reaches a peak near the apex of the cochlea in response to low frequencies and near the base for higher frequencies. In other words, the traveling wave causes a displacement pattern which is tuned as a function of distance along the basilar membrane. In the following chapters we shall see that fine frequency discriminations are made in psychoacoustic experiments, and that auditory nerve fibers are very sharply tuned. Can the mechanical displacement pattern in the cochlea account for this remarkable degree of frequency selectivity? If the mechanical displacement pattern of the basilar membrane is too wide to account for the narrowly tuned neural responses, then we must seek other sharpening mechanisms.

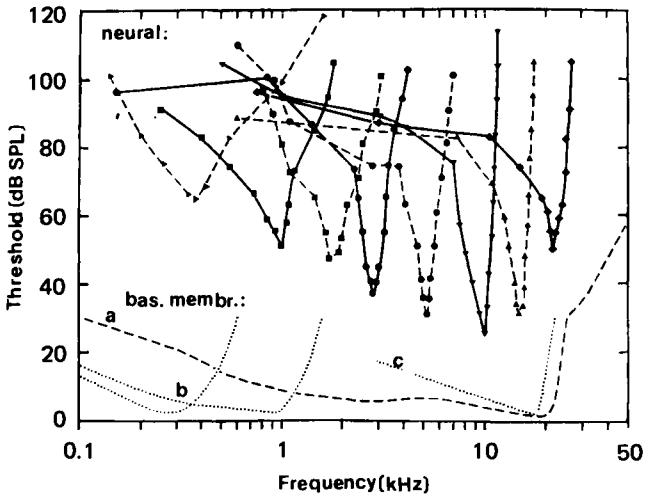

Figure 4.23 Upper curves: neural tuning curves (guinea pig). Lower curves: mechanical tuning (guinea pig) from (a) Wilson and Johnstone [36] at 20 kHz, (b) Bekesy [1] at 270 and 900 Hz, and (c) Johnstone et al. [62]. [From Evans [63], The sharpening of cochlear frequency selectivity, *Audiology 14*, 419–442 (1975).]

The key question is whether the sharpness of cochlear tuning approaches that of the auditory nerve. The upper curves in Fig. 4.23 show a set of response areas (tuning curves) for auditory nerve fibers in the guinea pig. Neural tuning curves will be discussed in Chap. 5. For now, note their sharp peaks, which indicate that auditory nerve fibers respond best to a limited range of frequencies around a characteristic frequency. The low-frequency slopes of the neural tuning curves range from about 100 to 300 dB per octave, and the high-frequency slopes are approximately −100 to −600 dB per octave in the cat [64–66], guinea pig [67] and monkey [68].

The sharpness of tuning can also be described by a value Q, which is the ratio of the center (characteristic) frequency to the bandwidth. For a particular center frequency, the narrower the bandwidth is, the larger is the Q value (Fig. 4.24). Since it is difficult to determine the half-power points (which usually define the bandwidth) of physiological tuning curves, it has become standard practice to use the points on the curve 10 dB down from the peak ($Q_{10_{dB}}$). Auditory nerve fibers have $Q_{10_{dB}}$ values of about 2–10 for the mid-frequencies (Fig. 4.25). With this fact in mind, let us proceed to examine the nature of frequency selectivity in the cochlea.

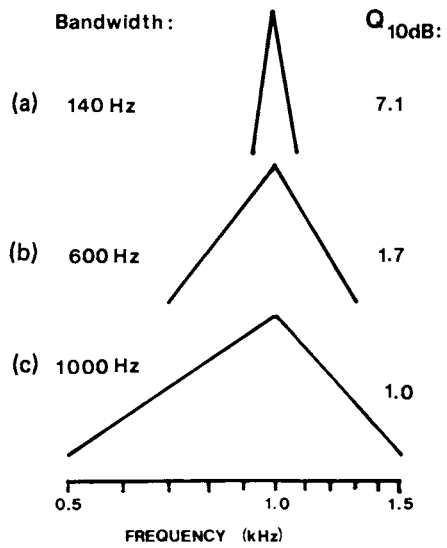

Figure 4.24 Q_{10dB} for three "tuning curves" centered at 1000 Hz. These curves were obtained as follows: (a) 1000/140 = 7.1; (b) 1000/600 = 1.7; (c) 1000/1000 = 1.0.

Bekesy [1] was the first to describe mechanical tuning in the cochlea. He used a binocular microscope to observe basilar membrane displacement patterns in response to high-intensity stimuli. Figure 4.26 shows some of the basilar membrane tuning curves obtained by Bekesy. These curves are in terms of relative amplitude, so that the peak is assigned a value of 1.0 and the displacements at other points along the curves are in proportions of the peak amplitude. The tuning curves tend to become sharper with increasing frequency, but the low-frequency slopes of about 6 dB per octave are far from the sharpness of neural tuning. This fact is illustrated in Fig. 4.23, in which Bekesy's data curve (b) are compared with neural tuning curves having similar characteristic frequencies.

Although Bekesy's data were of monumental significance, several experimental limitations make it difficult to compare these early findings with neural tuning data: (1) Access to the cochlear partition was limited to the apical areas, so that Bekesy's observations were restricted to relatively low frequencies. (2) The observations were obtained visually. (3) Absolute displacements were not specified. (Absolute values were reported in other contexts, however.) (4) Very high stimulus levels (roughly 120 dB SPL) were used, while neural data are available at much lower, including threshold, levels.

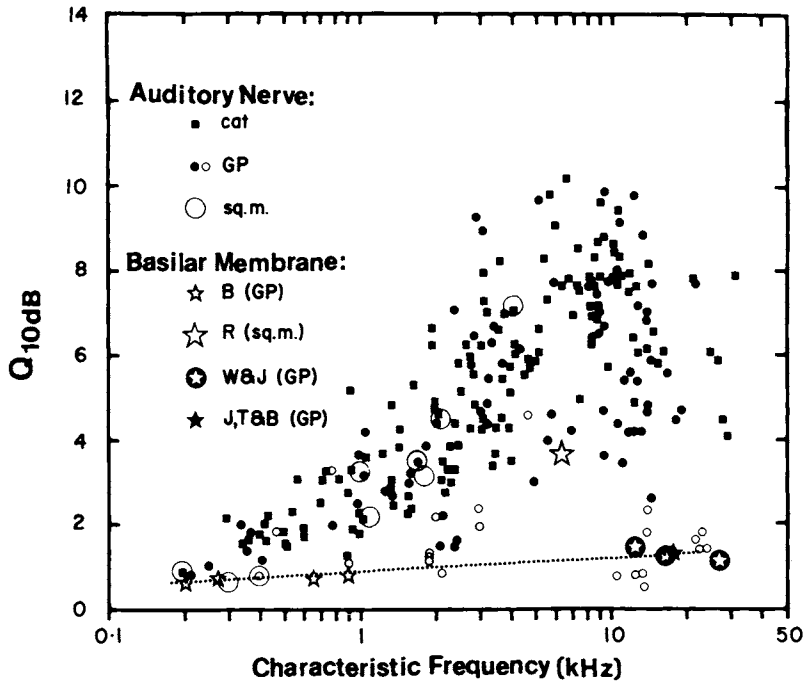

Figure 4.25 $Q_{10_{dB}}$ for auditory nerve and basilar membrane tuning curves. GP, guinea pig; sq. m., squirrel monkey. [From Evans and Wilson [66], Frequency selectivity in the cochlea, in *Basic Mechanisms in Hearing* (A. R. Møller, Ed.), ©1973, Academic Press, New York.]

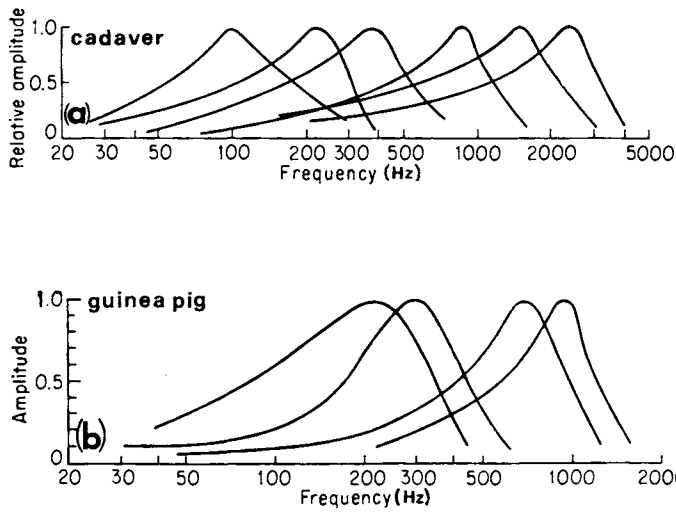

Figure 4.26 Basilar membrane tuning curves in (a) cadaver and (b) guinea pig. [From G. Bekesy [1], *Experiments in Hearing* (©1960, McGraw-Hill). Used with permission of McGraw-Hill Book Company.]

In 1967, Johnstone and Boyle [69] reported on the mechanical tuning of the guinea pig cochlea at a high frequency (about 18 kHz). Instead of visual observation, they used the Mössbauer technique. In this method, a radioactive source is placed on the basilar membrane and an absorber of the radiation is situated nearby. Vibration of the basilar membrane (on which the source is located) causes the amount of gamma rays absorbed to be modulated in a manner related to the vibration. Basilar membrane displacement is calculated based on these data. The Mössbauer method permits the use of stimulus levels as low as 60–70 dB SPL. Johnstone and Boyle found that the mechanical tuning of the basilar membrane is considerably sharper at 18 kHz than Bekesy's tuning curves at lower frequencies. Tuning in this area (about 1.2 mm from the stapes) had a 13 dB per octave slope on the low-frequency side of the peak and a slope of −70 dB per octave on the high-frequency side.

These results might at first seem to contradict Bekesy's findings at lower frequencies. Recall, however, that Bekesy's tuning curves became sharper as frequency increased (Fig. 4.26). Tonndorf and Khanna [70] calculated Q for both sets of data and found that the values fell along the same straight line. This result suggests that Bekesy's data are consistent with those of Johnstone and Boyle, which simply reflect sharper tuning along the basilar membrane as the base is approached.

There have been several studies of basilar membrane tuning using the Mössbauer [62,71–75] and other [36,76,77] techniques. Although data on specific aspects of the mechanical tuning curves vary, the important point is the considerable agreement among studies on the degree of mechanical tuning. This agreement is shown by the dotted line in Fig.4.25, which joins the values of $Q_{10_{dB}}$ for several studies. Examples of mechanical tuning curves are shown in the lower part of Fig. 4.23 and in Fig. 4.27.

Wilson and Johnstone [36,77] measured basilar membrane tuning with a capacitance probe. This was a miniature version of the one used by Bekesy [1] (Chap. 3). Their findings are of particular interest because they confirmed the Mössbauer data with another technique, and also because the capacitance probe enabled them to measure basilar membrane vibrations for stimuli as low as 40 dB SPL. This level is in the range used in obtaining neural tuning curves, so that the problem of comparing mechanical tuning at high levels to neural tuning at low levels is overcome.

The more recent measurements of mechanical tuning have revealed high-frequency slopes as great as about −350 dB per octave, but the low-frequency slopes do not become much steeper than roughly 24 dB per

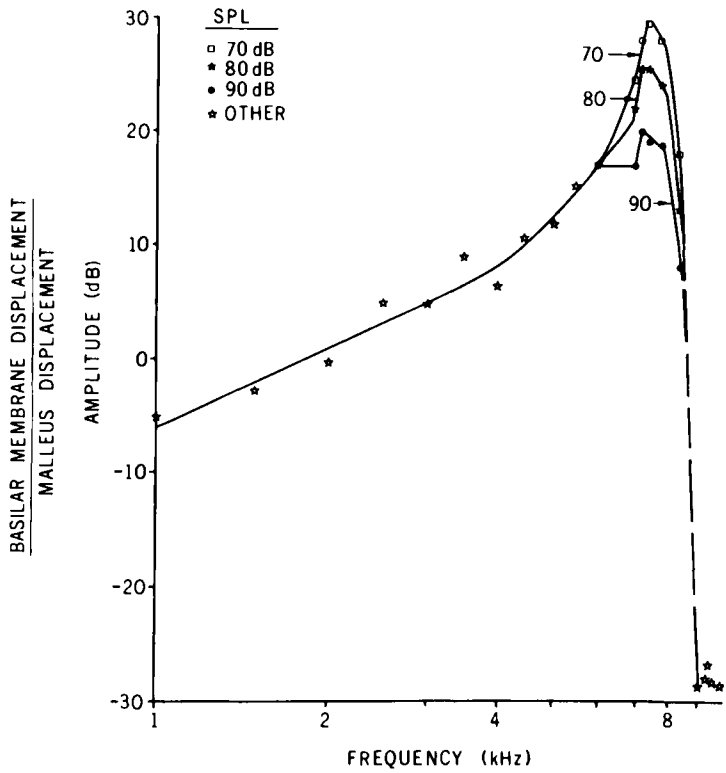

Figure 4.27 Basilar membrane tuning curves at 70, 80, and 90 dB SPL for the squirrel monkey. Lack of overlap around the peak indicates nonlinearity. (From Rhode [73]. Permission *J. Acoust. Soc. Amer.*)

octave. Clearly, these results cannot account for the sharpness of the neural tuning data. The difference in terms of $Q_{10_{dB}}$ is dramatically evident in Fig. 4.25.

A persistent inconsistency in the basilar membrane data should be mentioned before examining other possible sources of cochlear sharpening. Rhode and colleagues [73–75] repeatedly demonstrated that basilar membrane vibration is nonlinear in the squirrel monkey. The nonlinearity appears as a lack of overlapping of the tuning curve peaks at different stimulus levels (Fig. 4.27). The three curves in the figure would have overlapped exactly if basilar membrane vibration were linear, since the amplitude (in dB) is derived from the ratio of basilar-membrane-to-malleus displacement. In a linear system this ratio would stay the same for different stimulus levels. This nonlinearity has

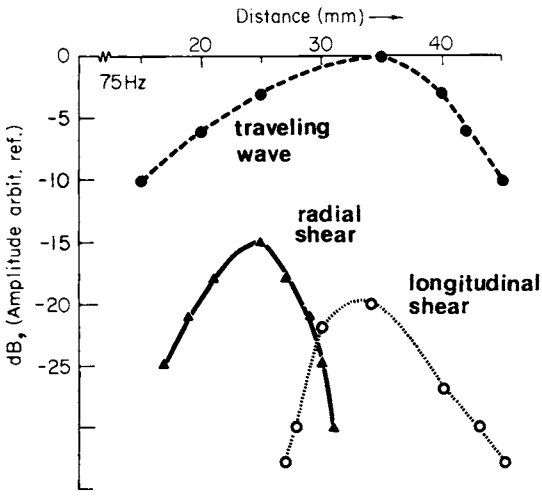

Figure 4.28 Traveling wave, radial shear, and longitudinal shear envelopes. (From Khanna et al. [78]. Permission *J. Acoust. Soc. Amer.*)

been the subject of considerable debate because it is not observed in the guinea pig. There is reason to believe, however, that it is species-specific to the squirrel monkey.

A mechanism that might enhance mechanical frequency sharpening in the cochlea was suggested by Khanna et al. [78]. Recall that the traveling wave causes longitudinal and radial shearing forces upon the organ of Corti. Using a computer simulation, Khanna et al. demonstrated that the shearing motion is more localized than the traveling wave pattern, so that the tuning curves associated with the shearing forces would reveal sharpened frequency selectivity (Fig. 4.28). Although these findings suggest that shearing forces sharpen the mechanical response, they still do not appear to account for the selectivity of the neural tuning curves.

Since the perilymph was partially drained in some of the above-mentioned experiments, Robertson [79] questioned the validity of comparing basilar membrane and neural tuning data. He found that perilymph drainage changed neural tuning curves, and suggested that the absence of perilymph from the cochlea might have resulted in widening of the basilar membrane tuning curves. However, wide mechanical (relative to neural) tuning is a finding common to studies of cochlear frequency selectivity regardless of whether perilymph was drained. Evans and Wilson [80] demonstrated that wide mechanical and sharp neural tuning curves were present in the same animals during

various amounts of perilymph drainage. It has thus been shown that sharp neural tuning occurs in the *same* animals in which wide basilar membrane tuning is also observed. It appears that another source of cochlear sharpening must be sought.

Evans and Wilson [65,66] hypothesized that a "second filter,"* existing between the basilar membrane and neural responses, accounts for the sharpness of neural tuning. They found that the sharpness of neural tuning curves is reduced when the metabolism of the cochlea is disturbed by interference with its oxygen supply or through chemical influences [63,66,67,81,82]. The resulting neural tuning curves, in fact, resemble basilar membrane tuning curves. It is unlikely that this result is due to mechanical changes in the cochlea. Further evidence for a second filter comes from other observations: Auditory nerve fibers with the same characteristic frequency vary in their degrees of tuning. Neural frequency selectivity in guinea pigs appears to increase as thresholds decrease. In some guinea pigs, a number of broadly tuned neurons are interspersed among sharply tuned units. These observations are inconsistent with sharp mechanical tuning because such a system would have *uniform* tuning of the neural fibers. It appears, then, that a second-filter mechanism probably does exist, and that it has two distinctive characteristics: First, the second filter is vulnerable to metabolic disturbances. Second, it is likely that the second-filter units are "private" to single auditory neurons or to groups of them, as opposed to being shared by a large number of neural fibers.

What (and where) is the second filter? A seemingly good candidate for this mechanism is lateral inhibition [83]. If one thinks of a sensory field, such as the retina of the eye, one can imagine a strongly stimulated area surrounded by a weakly stimulated area. Lateral inhibition is the suppression of activity in the weakly stimulated neural units by the presence of more intense stimulation in the adjoining area. The effect is to sharpen the contrast between the two areas. A frequently encountered visual example of lateral inhibition is the sharply defined border between adjoining black and white bars (Mach bands). Lateral inhibition has been proposed as a sharpening mechanism in hearing [1,84,86,87] but there is good evidence that it does not contribute to auditory frequency selectivity [65,66,88–91].

*One might note that there has been some recent evidence suggesting that the wide tuning curves like those reported by Bekesy are due to studying dead animals whereas in vivo mechanical tuning curves may be sharp enough to account for neural tuning. If this turns out to be so, then the second filter concept would be weakened.

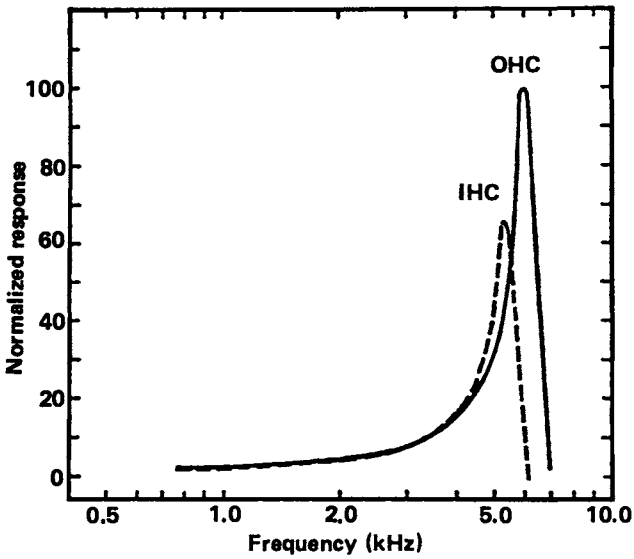

Figure 4.29 Hypothetical responses from inner and outer HCs as a function of frequency. (From Zwislocki [92]. Courtesy of *Acoustica.*)

It appears reasonable to locate the second filter to the cochlear HCs, or to some process(es) associated with them. The mechanism, however, remains elusive. For example, if the OHCs are in some way responsible for the sharply tuned segments of neural tuning curves, then the widened tuning seen after anoxia (oxygen starvation) could be attributed to the loss of more susceptible OHCs. Two interesting sharpening models based upon IHC–OHC interactions were proposed by Zwislocki [92,93] and Manley [94].

Zwislocki [92,93] suggested that frequency sharpening might result from an out-of-phase relationship between the inner and outer HCs. Auditory nerve recordings indicated that the IHCs were excitatory when the basilar membrane was displaced upward and that the OHCs were excitatory when the partition moved downward [95,96]. These data would mean that the two sets of HCs operate in phase opposition. The data also suggested that the IHCs and OHCs receive essentially the same frequency distribution except for a slight frequency shift between them (Fig. 4.29). Since the outer and inner HCs are out of phase, the resulting cancellation would cause the signal delivered to the auditory nerve to be equal to the difference between the two frequency responses shown in

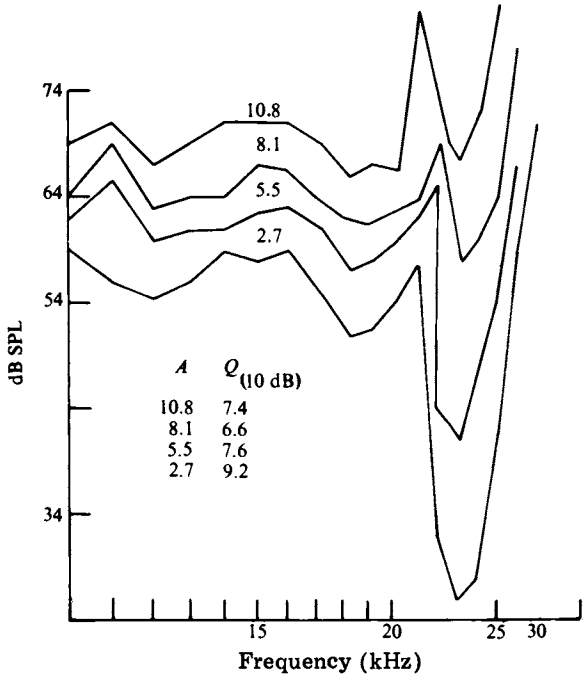

Figure 4.30 Intracellular SP tuning curves for an IHC with CF of 23 kHz. SP amplitude (A) in mV is shown along with values of $Q_{10_{dB}}$. (From Russell and Sellick [97]. Courtesy of *Nature*.)

the figure, which would have the effect of sharpening frequency selectivity. This is an extremely interesting model for sharpening; however, it relies to a large extent upon a neural interaction between the two sets of HCs, which probably does not occur [97].

Manley's [94] frequency sharpening model uses a different mechanism. Recall that most of the ear's afferent innervation is to the IHCs, while most of the inhibitory innervation from Rasmussen's bundle terminates at the OHCs. Manley proposed that all afferent signals are from the IHCs alone. The IHCs are also affected by the production of SP^- from the OHCs, which has an inhibitory effect. At low stimulus levels near the characteristic frequency (CF) of a particular auditory nerve fiber, there is effective stimulation of the neuron but little production of SP^-. At frequencies other than the CF more energy is needed to elicit a neural response from the nerve fiber, and this increased stimulus level also results in a much larger inhibitory SP^-. (Recall that SP magnitude is related to stimulus level.) There is thus

antagonism between an increase in the mechanical signal (basilar membrane displacement) on the one hand and the increasing SP^- that it causes on the other, for frequencies other than the neuron's CF. The neuron will not fire until the mechanical signal overcomes the suppressive effect of the SP^-. Sharper tuning results because this mechanism is frequency dependent (frequencies other than the CF require higher stimulus levels, which in turn produce a larger SP^-, etc.).

Of special interest is a recent study by Russell and Sellick [98] that has provided a provocative mechanism which may underlie the second filter. They measured the intracellular SPs of guinea pig IHCs as a function of frequency and intensity. Figure 4.30 shows a typical SP tuning curve for an IHC with a CF of about 23 kHz. At high stimulus levels, SP tuning is more or less like the mechanical data. However, as stimulus level decreases, the reduction in SP amplitude A is accompanied by a dramatic increase in frequency selectivity. Thus, when SP amplitude is 2.7 mV (within 10 dB of threshold), the value of $Q_{10_{dB}}$ is 9.2. This is as sharp as neural tuning. Whether this is the actual mechanism of the second filter is, of course, subject to further investigation. It does, however, represent the most sharply tuned preneural data obtained as of this writing.

REFERENCES

1. G. Bekesy, *Experiments in Hearing*, McGraw-Hill, New York, 1960.
2. J. Tonndorf, Cochlear mechanics and hydro-mechanics, in *Foundations of Modern Auditory Theory* (J. Tobias, Ed.), Vol. 1, Academic, New York, 1970, pp. 203–254.
3. J. Tonndorf, Response of cochlear models to aperiodic signals and to random noises, *J. Acoust. Soc. Amer. 32*, 1344–1355 (1960).
4. J. Tonndorf, Shearing motion in scala media of cochlear models, *J. Acoust. Soc. Amer. 32*, 238–244 (1960).
5. A. Ryan and P. Dallos, Physiology of the inner ear, in *Hearing Disorders* (J. L. Northern, Ed.), Little, Brown, Boston, 1976, pp. 89–101.
6. A. Flock, Sensory transduction in hair cells, in *Handbook of Sensory Physiology* (W. R. Lowenstein, Ed.), Vol. 1: *Principles of Receptor Physiology*, Springer-Verlag, New York, 1971, pp. 396–441.
7. O. Lowenstein and J. Wersäll, A fundamental interpretation of the electron-microscopic structure of the sensory hairs in the cristae of the elasmobranch raja clavata in terms of directional sensitivity, *Nature 184*, 1807–1808 (1959).
8. W. S. Rhode and C. D. Geisler, Model of displacement between opposing points on tectorial membrane and reticular lamina, *J. Acoust. Soc. Amer. 42*, 185–190 (1967).

9. P. Dallos, M. C. Billone, J. D. Durrant, C-Y. Wang, and S. Raynor, Cochlear inner and outer hair cells: Functional differences, *Science 177*, 356–358 (1972).
10. H. Davis, B. Deatherage, B. Rosenblut, C. Fernandez, R. Kimura, and C. A. Smith, Modificatiaon of cochlear potentials by streptomycin poisoning and by extensive venous obstruction, *Laryngoscope 68*, 596–627 (1958).
11. H. Davis, Some principles of sensory physiology, *Physiol. Rev. 4*, 391–416 (1961).
12. W. T. Peak, H. S. Sohmer, and T. F. Weiss, Microelectrode recordings of intracochlear potentials, *M.I.T. Res. Lab. Electronics Quart. Prog. Rep. 94*, 293–304 (1969).
13. I. Tasaki, H. Davis, and D. H. Eldredge, Exploration of cochlear potentials in guinea pig with a microelectrode, *J. Acoust. Soc. Amer. 26*, 765–773 (1954).
14. I. Tasaki and C. S. Spiropoulos, Stria vascularis as a source of endocochlear potential, *J. Neurophysiol. 22*, 149–155 (1959).
15. T. Konishi, R. A. Butler, and C. Fernandez, Effect of anoxia on cochlear potentials, *J. Acoust. Soc. Amer. 33*, 349–356 (1961).
16. P. Dallos, *The Auditory Periphery*, Academic, New York, 1973.
17. E. G. Wever and C. W. Bray, Action currents in the auditory nerve in response to acoustic stimulation, *Proc. Natl. Acad. Sci. 16*, 344–350 (1930).
18. E. D. Adrian, The microphonic action of the cochlea: An interpretation of Wever and Bray's experiments, *J. Physiol. 71*, 28–29 (1931).
19. H. Davis, A. Derbyshire, M. Lurie, and L. Saul, The electric response of the cochlea, *Ann. Otol. 68*, 665–674 (1934).
20. E. G. Wever, Electrical potentials of the cochlea, *Physiol. Rev. 46*, 102–126 (1966).
21. T. Konishi and T. Yasuna, Summating potential of the cochlea in the guinea pig, *J. Acoust. Soc. Amer. 35*, 1448–1452 (1963).
22. J. Hawkins, The ototoxicity of kanamycin, *Ann. Otol. 68*, 698–715 (1959).
23. P. Dallos, Cochlear potentials and cochlear mechanics, in *Basic Mechanisms in Hearing* (A. R. Møller, Ed.), Academic, New York, 1973, pp. 335–376.
24. C. S. Hallpike and A. F. Rawdon-Smith, The Wever-Bray phenomenon: Origin of the cochlear effect, *Ann. Otol. 46*, 976–990 (1937).
25. H. Davis, A model for transducer action in the cochlea, *Cold Spring Harbor Symp. Quant. Biol. 30*, 181–190 (1965).
26. H. Davis, Anatomy and physiology of the auditory system, in *Hearing and Deafness* (H. Davis and S. R. Silverman, Eds.), 3rd ed., Holt, Rinehart, and Winston, New York, 1970, pp. 47–82.
27. V. Honrubia and P. H. Ward, Dependence of the cochlear microphonic and summating potential on the endocochlear potential, *J. Acoust. Soc. Amer. 46*, 388–392 (1969).
28. Y. Katsuki, The origin of the microphonic potential, in *Hearing and Davis: Essays Honoring Hallowell Davis* (S. K. Hirsh, D. H. Eldridge, I. J. Hirsh, and S. R. Silverman, Eds.), Washington Univ. Press, St. Louis, 1976, pp. 25–35.

29. Y. Tanaka, A. Asanuma, K. Yanagisawa, and Y. Katsuki, DC potential and distribution of the cochlear microphonic potential [in Japanese], *Audiol. Jpn. 18*, 241–245 (1975).
30. E. G. Wever and M. Lawrence, *Physiological Acoustics*, Princeton Univ. Press, Princeton, N. J., 1954.
31. J. J. Eggermont, D. W. Odenthal, P. N. Schmidt, and A. Spoor, Electrocochleography: Basic principles and clinical application, *Acta Otol.*, Suppl. 316 (1974).
32. G. S. Misrahy, K. M. Hildreth, E. W. Shinabarger, and W. J. Gannon, Electrical properties of the wall of endolymphatic space of the cochlea (guinea pig), *Amer. J. Physiol. 194*, 396–402 (1958).
33. I. Tasaki, H. Davis, and J. P. Legouix, The space-time pattern of the cochlear microphonic (guinea pig) as recorded by differential electrodes, *J. Acoust. Soc. Amer. 24*, 502–518 (1952).
34. H. Davis, Mechanisms of excitation of auditory nerve impulses, in *Neural Mechanisms of the Auditory and Vestibular Systems* (G. L. Rasmussen and W. F. Windle, Eds.), Thomas, Springfield, Ill., 1960, pp. 21–39.
35. V. Honrubia and P. H. Ward, Longitudinal distribution of the cochlear microphonics inside the cochlear duct (guinea pig), *J. Acoust. Soc. Amer. 44*, 951–958 (1968).
36. J. P. Wilson and J. R. Johnstone, Capacitive probe measures of basilar membrane vibration, paper at Symposium on Hearing Theory, Institute for Perception Research, Eindhoven, The Netherlands, 1972, pp. 172–181.
37. P. Dallos, M. S. Cheathan, and J. Ferraro, Cochlear mechanics, nonlinearities, and cochlear potentials, *J. Acoust. Soc. Amer. 55*, 597–605 (1974).
38. I. C. Whitfield and H. F. Ross, Cochlear microphonic and summating potentials and the outputs of individual hair cell generators, *J. Acoust. Soc. Amer. 38*, 126–131 (1965).
39. I. C. Whitfield, *The Auditory Pathway*, Arnold, London, 1967.
40. H. Davis, C. Fernandez, and D. R. McAuliffe, The excitatory process in the cochlea, *Proc. Natl. Acad. Sci. 36*, 580–587 (1950).
41. H. Davis, B. H. Deatherage, D. H. Eldredge, and C. A. Smith, Summating potentials of the cochlea, *Amer. J. Physiol. 195*, 251–261 (1958).
42. V. Honrubia and P. H. Ward, Properties of the summating potential of the guinea pig's cochlea, *J. Acoust. Soc. Amer. 45*, 1443–1449 (1969).
43. P. Dallos, Summating potentials of the cochlea, in *Physiology of the Auditory System* (M. B. Sachs, Ed.), Natl. Educational Consultants, Baltimore, 1971, pp. 57–67.
44. J. R. Johnstone and B. M. Johnstone, Origin of the summating potential, *J. Acoust. Soc. Amer. 40*, 1405–1413 (1966).
45. A. M. Engebretson and D. H. Eldridge, Models for the nonlinear characteristics of cochlear potentials, *J. Acoust. Soc. Amer. 44*, 548–554 (1968).
46. G. A. Misrahy, E. W. Shinabarger, and K. M. Hildreth, Studies on factors affecting the summating potential, WADC Rep. 57–467 (1957).
47. P. Dallos and M. A. Cheatham, Generator potentials: Are they identifiable from gross cochlear potentials? *J. Acoust. Soc. Amer. 56*, Suppl., S10 (1974).

48. P. E. Stopp and I. C. Whitfield, Summating potentials of the avian cochlea, *J. Physiol. 175*, 45p–46p (1964).
49. N. Y.-S. Kiang and W. T. Peake, Components of electrical responses, *Ann. Otol. 69*, 448–458 (1960).
50. J. Fex, Efferent inhibition in the cochlea related to hair-cell dc activity: Study of postsynaptic activity of the crossed olivo-cochlear fibers in the cat, *J. Acoust. Soc. Amer. 41*, 666–675 (1967).
51. P. Dallos, Z. G. Shoeny, M. A. Cheatham, Cochlear summating potentials: Composition, *Science 170*, 641–644 (1970).
52. P. Dallos, Z. G. Shoeny, and M. A. Cheatham, Cochlear summating potentials: Descriptive aspects, *Acta Otol.*, Suppl. 302 (1972).
53. R. H. Sweetman and P. Dallos, Distribution pattern of cochlear combination tones, *J. Acoust. Soc. Amer. 45*, 58–71 (1969).
54. J. Tonndorf, Cochlear nonlinearities, in *Basic Mechanisms in Hearing* (A. R. Møller, Ed.), Academic, New York, 1973, pp. 11–47.
55. J. Tonndorf, The hydrodynamic origin of aural harmonics in the cochlea, *Ann. Otol. 67*, 754–774 (1958).
56. J. Zwislocki, Discussion of cochlear nonlinearities (by Tonndorf), in *Basic Mechanisms in Hearing* (A. R. Møller, Ed.), Academic, New York, 1973.
57. J. Tonndorf, Nonlinearities in cochlear hydrodynamics, *J. Acoust. Soc. Amer. 47*, 579–591 (1970).
58. P. Dallos, Z. G. Shoeny, D. W. Worthington, and M. A. Cheatham, Cochlear distortion: Effects of direct current polarization, *Science 164*, 449–451 (1969).
59. P. Dallos and R. H. Sweetman, Distribution pattern of cochlear harmonics, *J. Acoust. Soc. Amer. 45*, 37–46 (1969).
60. P. Dallos, Combination tone $2f_1 - f_h$ in microphonic potentials, *J. Acoust. Soc. Amer. 46*, 1437–1444 (1969).
61. D. W. Worthington and P. Dallos, Spacial patterns of cochlear difference tones, *J. Acoust. Soc. Amer. 49*, 1818–1830 (1971).
62. B. M. Johnstone, K. J. Taylor, and A. J. Boyle, Mechanics of the guinea pig cochlea, *J. Acoust. Soc. Amer. 47*, 504–509 (1970).
63. E. F. Evans, The sharpening of cochlear frequency selectivity in the normal and abnormal cochlea, *Audiology 14*, 419–442 (1975).
64. N. Y.-S. Kiang, *Discharge Patterns of Single Nerve Fibers in the Cat's Auditory Nerve*, MIT Press, Cambridge, Mass., 1965.
65. E. F. Evans and J. P. Wilson, Frequency resolving power of the cochlea: The effective bandwidth of cochlear nerve fibers, in *Proc. 7th Intl. Congr. on Acoustics*, Vol. 3, Akademiai Kiado, Budapest, 1971, pp. 453–456.
66. E. F. Evans and J. P. Wilson, Frequency selectivity of the cochlea, in *Basic Mechanisms in Hearing* (A. R. Møller, Ed.), Academic, New York, 1973, pp. 519–554.
67. E. F. Evans, The frequency response and other properties of single nerve fibers in the guinea pig cochlea, *J. Physiol. 226*, 263–287 (1972).
68. C. D. Geisler, W. S. Rhode, and D. T. Kenedy, 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 (1974).

69. B. M. Johnstone and J. F. Boyle, Basilar membrane vibration examined with the Mössbauer technique, *Science 158*, 389–390 (1967).
70. J. Tonndorf and S. M. Khanna, Displacement pattern of the basilar membrane: A comparison of experimental data, *Science 160*, 1139–1140 (1968).
71. B. M. Johnstone and K. Taylor, Mechanical aspects of cochlear function, in *Frequency Analysis and Periodicity Detection in Hearing* (R. Plomp and G. F. Smoorenburg, Eds.), Sijthoff, Leiden, 1970, pp. 81–93.
72. B. M. Johnstone and G. K. Yates, Basilar membrane tuning curves in the guinea pig, *J. Acoust. Soc. Amer. 55*, 584–587 (1974).
73. W. S. Rhode, Observations of the vibration of the basilar membrane in squirrel monkey using the Mössbauer technique, *J. Acoust. Soc. Amer. 49*, 1218–1231 (1971).
74. W. S. Rhode and L. Robles, Evidence from Mössbauer experiments for nonlinear vibration in the cochlea, *J. Acoust. Soc. Amer. 55*, 588–596 (1974).
75. L. Robles and W. S. Rhode, Nonlinear effects in the transient response of the basilar membrane, in *Facts and Models in Hearing* (E. Zwicker and E. Terhardt, Eds.), Springer-Verlag, New York, 1974.
76. L. U. E. Kohllöfel, A study of basilar membrane vibrations, *Acoustica 27*, 49–89 (1972).
77. J. P. Wilson and J. R. Johnstone, Basilar membrane and middle ear vibration in guinea pig measured by capacitive probe, *J. Acoust. Soc. Amer. 57*, 705–723 (1975).
78. S. M. Khanna, R. E. Sears, and J. Tonndorf, Some properties of longitudinal shear waves: A study by computer simulation, *J. Acoust. Soc. Amer. 43*, 1077–1084 (1968).
79. D. Robertson, Cochlear neurons: Frequency selectivity altered by perilymph removal, *Science 186*, 153–155 (1974).
80. E. F. Evans and J. P. Wilson, Cochlear tuning properties: Concurrent basilar membrane and single nerve fiber measurements, *Science 190*, 1218–1221 (1975).
81. E. F. Evans, Does frequency sharpening occur in the cochlea? paper at Symposium on Hearing Theory, Institute for Perception Research, Eindhoven, The Netherlands, 1972.
82. E. F. Evans, The effects of hypoxia on tuning of single fibers in the cochlear nerve, *J. Physiol. 238*, 65–67 (1974).
83. H. K. Hartline and F. Ratliff, Inhibitory interaction of receptor units in the eye of limulus, *J. Gen. Physiol. 40*, 357–376 (1957).
84. G. G. Furman and L. S. Frishoff, Model of neural inhibition in the mammalian cochlea, *J. Acoust. Soc. Amer. 36*, 2194–2207 (1964).
85. G. Bekesy, *Sensory Inhibition*, Princeton Univ. Press, Princeton, N. J., 1967.
86. E. C. Carterette, M. P. Friedman, and S. D. Lovell, Mach bands in hearing, *J. Acoust. Soc. Amer. 45*, 986–998 (1969).
87. T. Houtgast, Psychophysical evidence for lateral inhibition in hearing, *J. Acoust. Soc. Amer. 51*, 1885–1894 (1974).
88. J. Zwislocki, E. Buining, and J. Glantz, Frequency distribution of central masking, *J. Acoust. Soc. Amer. 43*, 1267–1271 (1968).

89. A. R. Møller, Studies of the damped oscillatory response of the auditory frequency analyzer, *Acta Physiol. Scand. 78*, 299–314 (1970).
90. J. P. Wilson and E. F. Evans, Grating acuity of the ear: Psychophysical and neurophysiological measures of frequency resolving power, in *Proc. 7th Intl. Congr. on Acoustics*, Vol. 3, Akademiai Kiado, Budapest, 1971.
91. H. Rainbolt and A. M. Small, Mach bands in auditory masking: An attempted replication, *J. Acoust. Soc. Amer. 51*, 567–574 (1974).
92. J. Zwislocki, A possible neuro-mechanical sound analysis in the cochlea, *Acoustica 31*, 354–359 (1974).
93. J. Zwislocki, Phase opposition between inner and outer hair cells and auditory sound analysis, *Audiology 14*, 443–455 (1975).
94. G. A. Manley, Cochlear frequency sharpening — A new synthesis, *Acta Otol. 85*, 167–176 (1978).
95. J. Zwislocki and W. G. Sokolich, Velocity and displacement responses in auditory-nerve fibers, *Science 182*, 64–66 (1973).
96. J. Zwislocki, Cochlear waves: Interaction between theory and experiments, *J. Acoust. Soc. Amer. 55*, 578–583 (1974).
97. J. Zwislocki, personal communication, 1979.
98. I. J. Russell and P. M. Sellick, Tuning properties of cochlear hair cells, *Nature 267*, 858–860 (1977).

5
Auditory Nerve

FREQUENCY CODING

Tuning Curves

The responses of single auditory nerve fibers to acoustic stimulation at various frequencies were reported by Galambos and Davis [1], and their results were confirmed and expanded upon by many others. The response areas of single neurons as a function of frequency are shown by their tuning curves (Fig. 5.1). A narrowly tuned cell responds to a very limited range of frequencies, whereas a broadly tuned cell responds to a much wider frequency range. Since an unstimulated neuron maintains an ongoing (spontaneous) discharge rate even in the absence of any apparent stimulation, its threshold may be determined by varying the stimulus level until the lowest intensity is reached at which the neuron responds above its spontaneous rate. An alternative approach is to present the stimulus at a fixed intensity, and to measure the number of spike potentials with which the unit responds at different stimulus frequencies. The former method measures the neuron's sensitivity, and the latter its firing rate, as functions of frequency. The frequency with the lowest threshold (or the greatest firing rate) is the best or characteristic frequency (CF) of the neuron.

The tuning curves of various single fibers in the auditory nerve of the cat are shown in Fig. 5.1. Frequency is along the x-axis and the level needed to reach the neuron's threshold is on the y-axis. Notice that each fiber will respond to a range of frequencies if the stimulus level is high

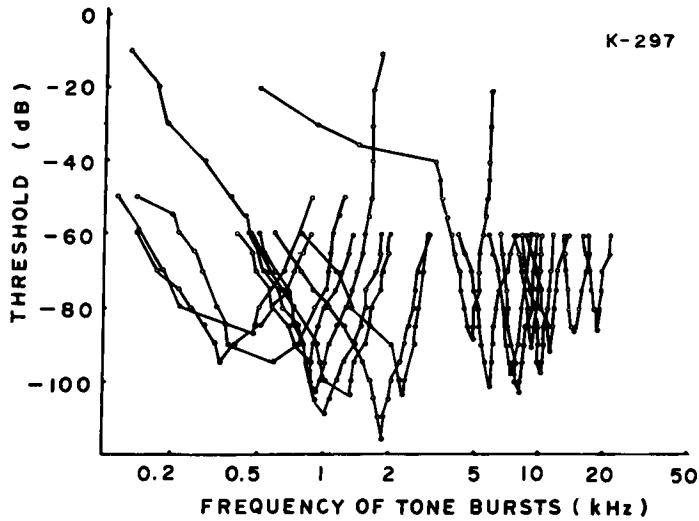

Figure 5.1 Tuning curves showing the response areas of single auditory nerve fibers in the cat. [Reprinted from *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* by N. Y.-S. Kiang by permission of The M.I.T. Press, Cambridge, Massachusetts (©1965, p. 87).]

enough. This frequency range extends considerably below the CF, but is quite restricted above it. In other words, a fiber responds readily to intense stimulation below its CF, but is only minimally responsive to stimulation above it. At lower intensity levels, the fibers are quite narrowly tuned to a particular frequency, as is shown by the V-shaped troughs around each CF.

Although stimulus amplitude may be expressed in SPL, the tuning curves become sharper and more regular when stimulus level is expressed as the degree of stapes footplate displacement in the oval window. This sharpening occurs because using stapes displacement as the reference omits middle ear effects (see Chap. 3), which are included when the stimulus is expressed as the SPL at the ear canal or the voltage applied to the earphone [3,4].

The sensitivity of a particular neuron generally falls at a rate of more than 25 dB per octave below the CF and well over 100 dB per octave above it. The low-frequency "tails" of higher-CF fibers actually extend very far below the CF; and phase-locked* responses for low-frequency

*"Phase locking" refers to a clear and fixed relationship between some aspect of the response and the phase (or time) of some aspect of the stimulus. The importance of this phenomenon will become clear in the following sections dealing with firing patterns.

stimuli at high levels have been demonstrated for these fibers [5]. (These low-frequency tails should not be surprising, since we know that much of the cochlear partition is affected at high intensities.)

The tuning of auditory nerve fibers reflects to a large extent the frequency analysis of the cochlea. However, since neural tuning curves are far sharper than those for the basilar membrane, there is probably a "second filter" between the mechanical and neural mechanisms which accounts for this difference [6]. This second filter is most likely located at the hair cell or at its interface with the neuron (see Chap. 4).

Firing Patterns

Single nerve fibers discharge spike potentials ("fire") in an all-or-none manner. The firing pattern is the manner in which a fiber discharges in response to a stimulus (or spontaneously in the absence of stimulation). The firing patterns of fibers can suggest how information is coded and transmitted in the auditory nervous system. The firing patterns of auditory nerve fibers also provide corroborative information about cochlear mechanics.

Various stimuli are used to provide different kinds of information about neural coding. Clicks (at least ideally) are discrete and instantaneous in time, with energy distributed equally throughout the frequency range. On the other hand, pure tones (also ideally) extend indefinitely in time, but are discrete in frequency. Thus, click stimuli lend themselves to the study of the temporal characteristics of the discharge pattern, while sinusoids can be used to study frequency-related aspects. Tone bursts can be used to investigate both frequency and temporal characteristics, since they are similar to clicks in their duration and to tones in their frequency specificity [7]. One should remember, however, that a tone burst is a compromise between the two extremes, so that it is really less discrete in time than a click and less discrete in frequency than a pure tone.

Responses to Clicks

One way to study the response of an auditory nerve fiber is to determine the probability that it will discharge under given circumstances. This is not as complicated as it sounds. Assume that we have placed a microelectrode into an auditory nerve fiber. We then present a click to the animal's ear and record the time between the click onset and the discharge of any spike potential which it may elicit. In other words, we record the time delay or latency of each spike from the start of the click. This procedure is repeated many times for the same fiber, and the number of spikes that occurrred after each latency is counted. These data are plotted on a graph called a poststimulus time (PST) histogram,

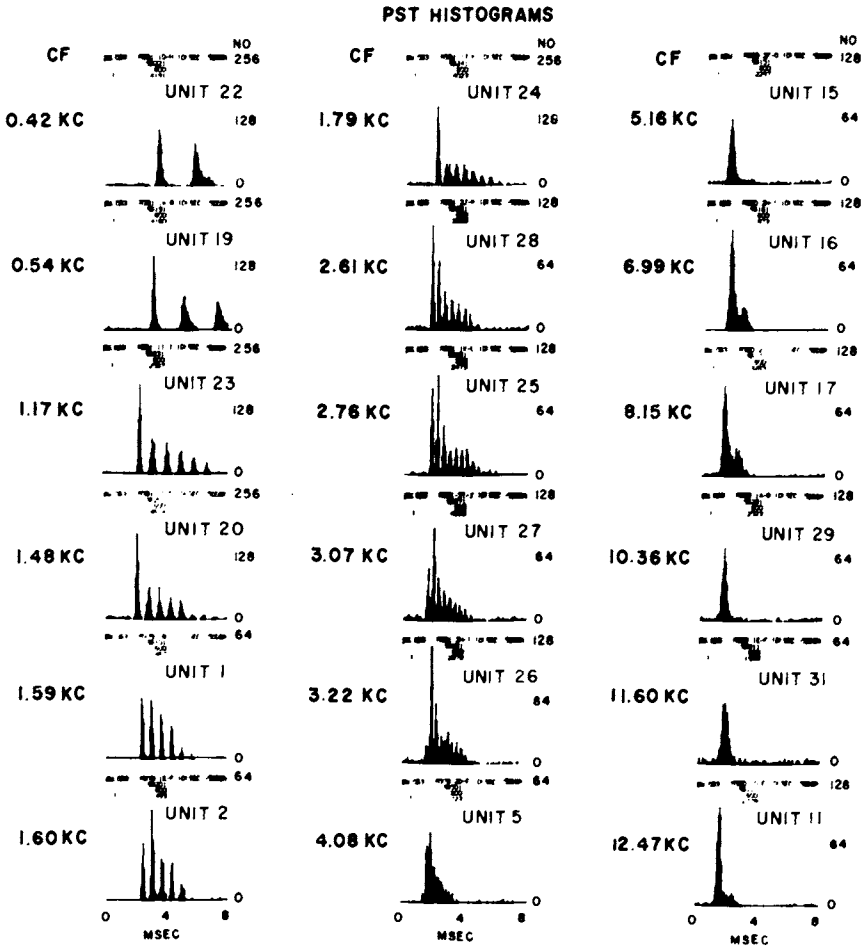

Figure 5.2 PST histograms for 18 auditory nerve fibers of a single cat in response to clicks. [Reprinted from *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* by N. Y.-S. Kiang by permission of The M.I.T. Press, Cambridge, Massachusetts (©1965, p. 28).]

in which the latencies are shown on the abscissa and the number of spikes on the ordinate (Fig. 5.2). If many spikes occurred at a latency of 2 msec, for example, then we can say that there is a high probability that the fiber will respond at 2 msec. If concentrations of spikes (modes) also occur at other latencies, we can say that there is a good probability that the fiber will respond at these other latencies as well as at the first one.

Figure 5.2 shows the PST histograms obtained by Kiang [2] for 18 auditory nerve fibers in one cat. The CF of each fiber is given to the left of its histogram. Note that fibers with lower CFs have multiple peaks, while those with higher CFs have single peaks. Three other observations are important. First, the latency to the first peak decreases as the CF increases. Second, the interpeak interval (the time between successive modes) gets smaller as the CF increases. Closer crutiny reveals that the interpeak interval corresponds to the period of the characteristic frequency of a fiber (i.e., to 1/CF) in a relationship which constitutes a linear function. Lastly, baseline activity is actually reduced between the peaks. Kiang found that these time-locked multiple peaks and the reduced baseline activity were maintained even at very low levels of stimulation. Decreased baseline activity was, in fact, the initial response noted at the lowest stimulus levels.

If the PST histograms are compared for clicks of opposite polarity, the peaks resulting from rarefaction clicks are found to occur *between* those in response to condensation clicks [2]. That is, when the click's polarity is reversed (which means that drum, and eventually basilar membrane, deflection is reversed, see Chap. 4), then so are the times at which the peaks and dips occur in the PST histogram. Furthermore, the rarefaction phase is related to increased neural activity, in accordance with the earlier findings of Davis et al. [8] for stimuli up to 2000 Hz.

The traveling wave response to an impulsive stimulus such as a click travels up the cochlea at a speed which is quite fast at the basal end and slows down toward the apex [9]. We would therefore expect the PST histogram to reflect rapid and synchronous discharges of higher-frequency fibers originating from the basal turn. This effect is shown by the short latencies of the first peak for higher CFs. The longer latencies for successively lower CFs reflect the propagation time of the traveling wave up the cochlear partition to the places from which the fiber arise. Thus, the latency to the first peak represents a neural coding of the mechanical activity in the cochlea.

The interpeak interval also reflects the activity of the cochlea, since it is a function of the period of the frequency (1/f) in the click stimulus to which the fiber responds. Since the latencies of the peaks do not change with click level, Kiang suggested that deflections of the cochlear parti-

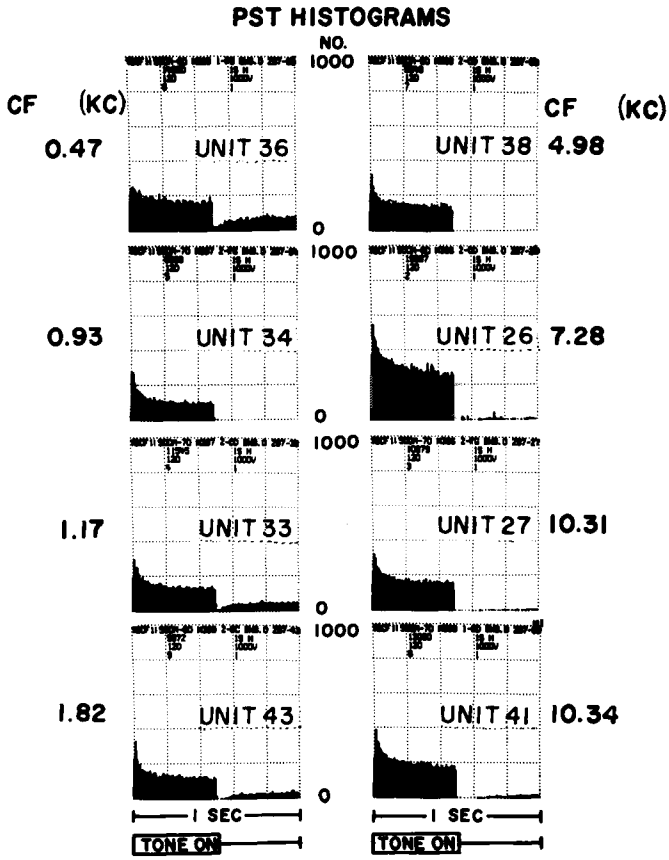

Figure 5.3 PST histograms in response to tone burst for fibers of different CFs. [Reprinted from *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* by N. Y.-S. Kiang by permission of The M.I.T. Press, Cambridge, Massachusetts (©1965, p. 69).]

tion in one direction result in increased neural activity, while neural activity is reduced relative to the baseline rate when the basilar membrane is deflected in the opposite direction. This interpretation is further supported by the reversal of peaks and dips for clicks of opposite polarity.

Responses to Tone Bursts

The responses of eight auditory nerve fibers to tone bursts are shown in Fig. 5.3. Each PST histogram shows the response pattern beginning 5 msec after tone burst onset. Note that there is an initial peak in

response to stimulus onset in each response pattern. The prominence of this peak is a function of the tone burst intensity. The peak is followed by a gradual decrease in the discharge rate until a stable level is attained. The stable rate continues until the tone burst is turned off, at which time activity drops sharply to a level below the spontaneous discharge rate. The spontaneous rate is then gradually reattained. The neural discharges are time-locked to individual cycles of the tone burst for fibers up to about 5000 Hz. (This effect is not seen in the figure because of the restricted time scale.)

Responses to Tones

Kiang [2] found that the discharges of auditory nerve fibers are time-locked to tonal stimuli up to about 4000–5000 Hz. This relation was demonstrated by the presence on the PST histogram of single peaks corresponding to individual cycles of the stimulus; and it is consistent with other evidence that auditory nerve fibers respond to the particular phase of the stimulus within this frequency range [10,11]. Furthermore, there is impressive evidence that auditory nerve fibers respond only to deflections of the basilar membrane in one direction (which is consistent with the click data), and that the timing of the firings corresponds to unilateral elevations of the partition [12,13]. (See Chap. 4.)

The relationship of the responses of single auditory nerve fibers to stimulus phase is illustrated in Fig. 5.4. These graphs are not PST histograms; instead they show the number of spikes discharged at various time intervals in response to 1 sec pure tones from 412 to 1600 Hz. The tones were presented at 80 dB SPL. A single fiber was monitored. It responded to frequencies between 412 and 1800 Hz, with its best responses between 1000 and 1200 Hz.

The dots under each histogram correspond to integral multiples of the period of the stimulus tone. Thus, in the upper left-hand graph of Fig. 5.4, with a frequency of 412 Hz and a period of 2427 μsec, the dots indicate 2427 μsec time intervals (2427 μsec, 4854 μsec, 7281 μsec, etc.). The spikes in each histogram cluster at a number of relatively discrete latencies, with fewer spikes at successively higher multiples of the period. Of primary significance is that the locations of the peaks closely correspond to integral multiples of the period for each stimulus frequency up to and including 1100 Hz. At higher frequencies, the period of the peaks become as low as 625 μsec (for 1600 Hz), although the latencies of the first peaks stay in the 800–900 μsec range. This minimum period reflects the fiber's refractory period.

These findings suggest that, at least for pure tones, a period–time code is used to transmit frequency information in auditory nerve fibers: The discharge pattern of the neuron is in cadence with the period of the

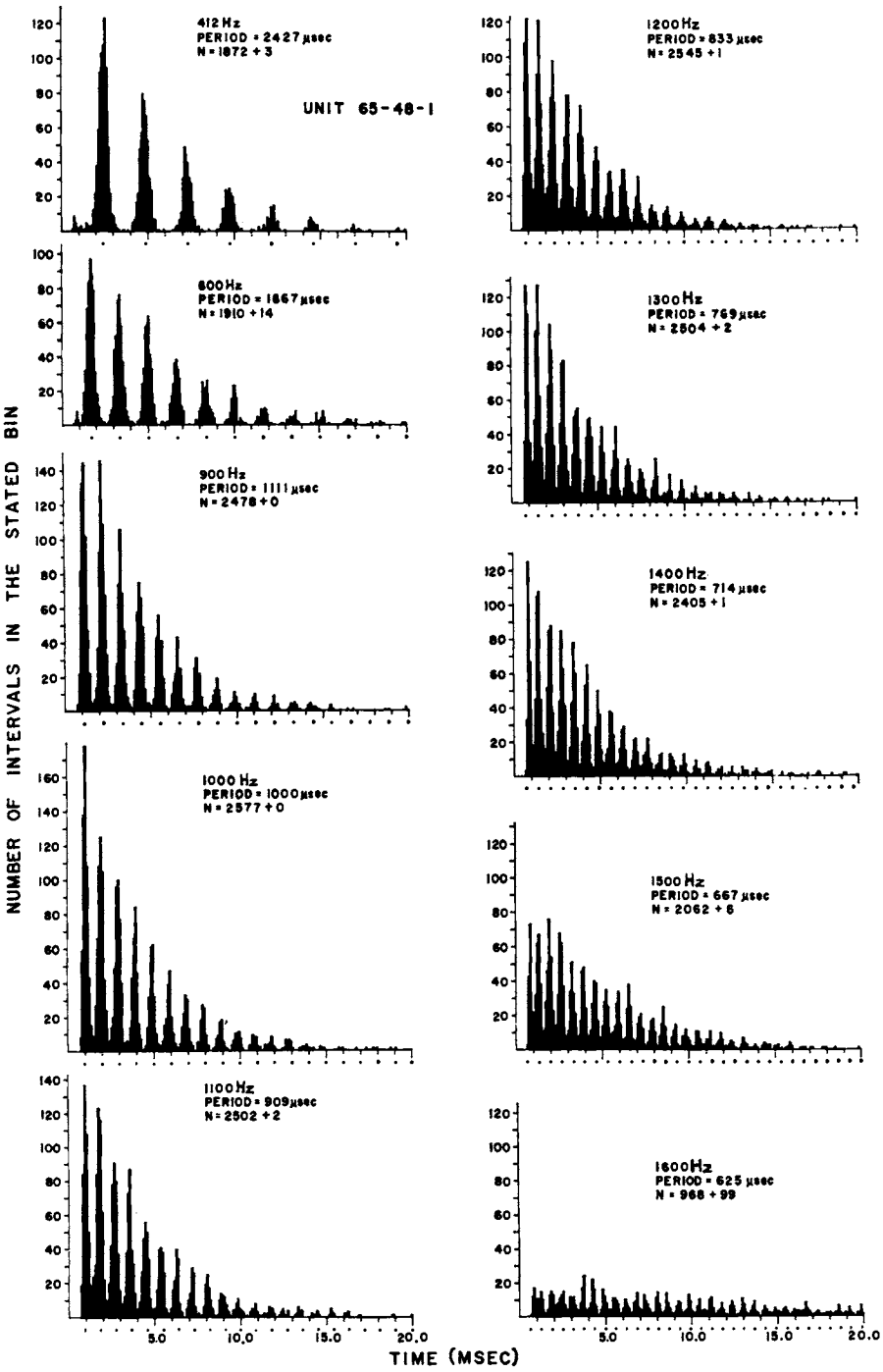

Figure 5.4 Interspike intervals for a single auditory neuron of a squirrel monkey in response to 1 sec tones at 80 dB SPL. Stimulus frequency is shown above each graph. The dots below the abscissa are integral multiples of the period of the stimulus. (N is the number of intervals plotted plus the number of intervals with values greater than shown on the abscissa.) (From Rose, Brugge, Anderson, and Hind [10]. Permission *J. Neurophysiol.)*

stimulus. The locking of the response pattern to the period of the sinusoid is maintained even if the stimulus frequency is not the CF of the fiber, and regardless of stimulus intensity. When the spike discharges were studied relative to a fixed point on the stimulus cycle, phase locking of the discharge pattern was found for frequencies as high as 5000 Hz. This result is not to suggest that place coding is unfounded. On the contrary, the important of place coding is demonstrated by the tuning curves in Fig. 5.1. Both mechanisms contribute to frequency coding.

Responses to Combinations of Tones

When an auditory nerve fiber is stimulated simultaneously by two relatively low-frequency tones, the resulting firing pattern will be phase locked to either: (1) the cycles of the first sinusoid, (2) the cycles of the second sinusoid, or (3) the cycles of both stimuli [11]. When the response pattern is to only one of the two original tones, the phase-locked response is the same as it would have been if that tone were presented alone. Which of the three response modes occurs is determined by the intensities of the two tones and by whether their frequencies lie within the fiber's response area. (See the section Two-Tone Inhibition.)

Brugge et al. [12] have reported the discharge patterns in response to complex periodic sounds. Their stimuli were made up of two relatively low-frequency primary tones combined in various ways, as shown in Fig. 5.5. The firing patterns are shown in period histograms, in which discharges are represented as a function of the period of the stimulus (i.e., as though they all occurred during one period of the complex wave). Again, we see clear-cut phase locking of the discharge pattern to the stimulus, reflecting neural coding of the mechanical activity along the cochlear partition in a straightforward manner. Note that the neural activity shown by the period histogram follows the shape of the stimulus waveform above the origin of the y-axis. Clearly, the major implications are that, for waveforms arising on regions of the cochlear partition able to excite a nerve fiber, only deflections in one direction result in neural excitation, and neural discharges occur at the times when the basilar membrane is deflected upward.

These findings were corroborated by a related study by Rose et al. [13], who found that when two tones resulted in a complex wave whose peaks are not equidistant, the spikes cluster about integral multiples of the period of the complex wave. The complex wave's period corresponds to its fundamental frequency, which, in turn, is the greatest common denominator of the two original tones. Note in this regard that when the ratio of the low to the high primary has a numerator of one less than the denominator, as for 1000 and 1100 Hz (1000/1100 = 10/11),

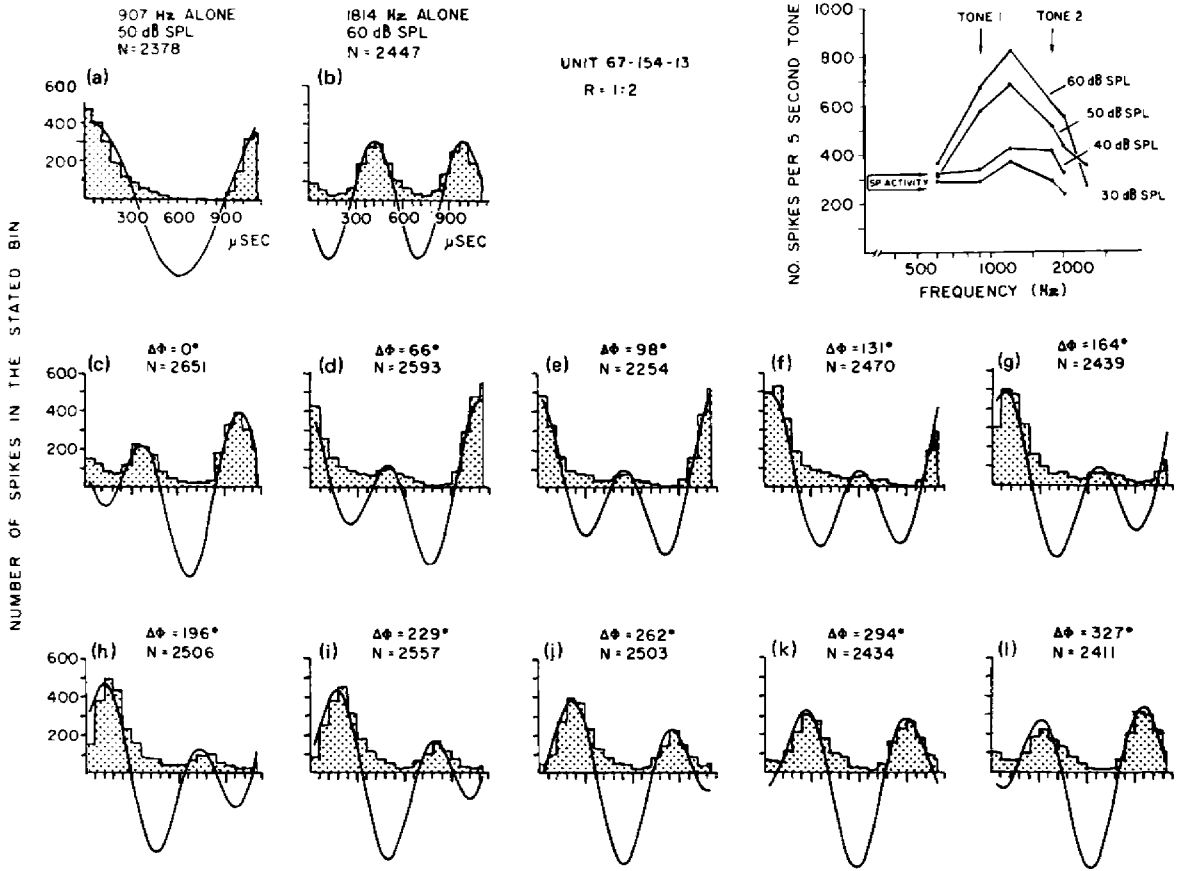

Figure 5.5 Complex tonal stimuli and period histograms of resulting discharge patterns. (a) and (b), primary tones; (c)–(1), complex tones; ϕ is phase shift between the primaries; upper right-hand graph shows the response area at various SPLs.) (From Brugge, Anderson, Hind, and Rose [12]. Permission *J. Neurophysiol.*)

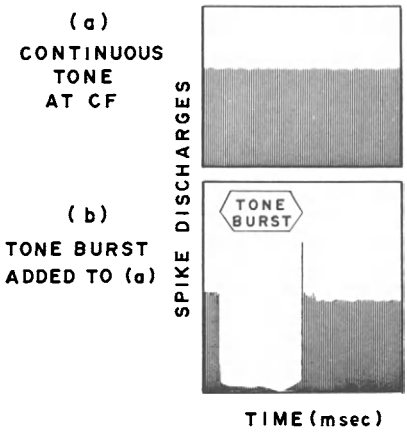

Figure 5.6 Two-tone inhibition (see text).

the fundamental is the difference tone (1100 − 1000 = 100 Hz). In this case, the spike discharges correspond to a period of 100 Hz, and the listener would perceive a tone corresponding to a pitch of 100 Hz. This provides one basis for the "missing fundamental" phenomenon discussed in Chap. 12.

Two-Tone Inhibition

Assume that a single auditory nerve fiber is discharging in response to a continuous tone at its CF (Fig. 5.6a). A second tone at a slightly different frequency is then added to the first. The presence of the second tone will actually cause a *decrease* in the firing rate (Fig. 5.6b). This decrease in the unit's firing rate due to the addition of the second tone is called two-tone inhibition, and has been repeatedly demonstrated in auditory nerve fibers [2,11,14,15]. Most studies of two-tone inhibition use tone bursts to inhibit a unit's response to a continuous tone at its CF, as shown schematically in Fig. 5.6b. As might be imagined, it would be cumbersome to "map out" the inhibitory area(s) of a neuron: An excessive number of discrete tone bursts would be needed at all of the frequencies and intensities being tested.

Instead of tone bursts, Sachs and Kiang [15] used a sweep frequency (SF) tone to inhibit the fiber's response to a continuous tone at its characteristic frequency (CTCF). (A sweep frequency tone is one which changes frequency continuously over time.) Figure 5.7 shows a series of PST histograms for a fiber with a CF of 22.2 kHz. The left column shows the firing patterns (as a function of log frequency) for the SF (inhibiting) tone alone. The frequency of the SF tone was smoothly changed from 6 kHz to 60 kHz and then back down to 6 kHz. There were thus two

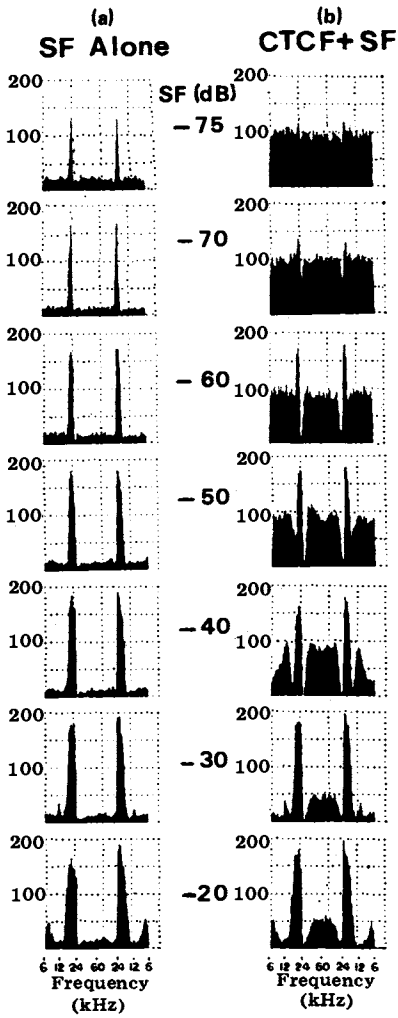

Figure 5.7 Discharge rates for (a) sweep frequency (SF) alone, and (b) SF added to a continuous tone at the fiber's characteristic frequency (CTCF) at various levels of the SF tone. The CTCF is 22.2 kHz at −75 dB. Duration from 6 kHz to 60 kHz and back to 6 kHz is 72 sec. Abscissa: log frequency for SF tone; ordinate: spike rate. (From Sachs and Kiang [15]. Permission *J. Acoust. Soc. Amer.*)

"approaches" by the SF tone to the CF: one from below and one from above. The histograms in the right-hand column of the figure are for the ongoing CTCF combined with the SF tone. Observe that the SF tone causes a reduction in the firing rate at frequencies near the CF. The series of histograms in the right-hand column reveals that a wider range of frequencies will inhibit the discharge rate as the intensity of the SF tone is increased (from −75 to −20 dB). The data compiled at various SF tone frequencies and intensities can be plotted in a manner similar to a tuning curve. An idealized representation is given in Fig. 5.8, where

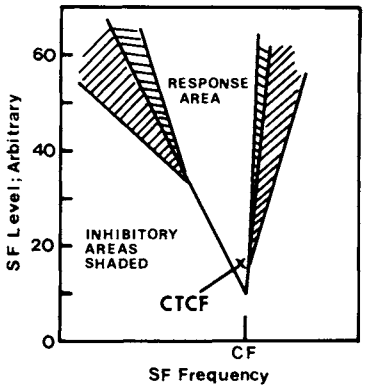

Figure 5.8 Idealized response and inhibitory areas in a two-tone inhibition experiment. (Adapted from Sachs and Kiang [15]. Permission *J. Acoust. Soc. Amer.*)

the cross-hatched areas show the combinations of SF frequencies and intensities that inhibit the firing of a fiber for a CTCF. This figure illustrates several aspects of two-tone inhibition.

Two-tone inhibition can be elicited by a suppressing (SF) tone either higher or lower than the CF, as long as its intensity and frequency are appropriate. Figure 5.8 shows that the high-frequency inhibitory area extends to within 10 dB of the level of the CTCF, although the low-frequency inhibitory area does not. In other words, inhibitory tones higher in frequency than the CF are effective at lower intensities than tones whose frequencies are below the CF. The intensity and frequency dependencies in two-tone inhibition have been described [16,17], but are beyond the scope of this book. It should be added that Sachs and Kiang found inhibitory areas for all 310 fibers they studied, and were able to demonstrate this effect at intensities near the threshold of the fiber.

Although two-tone inhibition certainly represents a nonlinearity in the ear's response, its origin is still not firmly established. One might suppose that it represents a neural sharpening of the tuning curves. However, two-tone inhibition more likely results from nonlinear cochlear processes. This origin is supported by the observation that two-tone inhibition is demonstrable for the cochlear microphonic [18–20], which is a preneural event. A model based upon the directional sensitivity of the hair cells [21] can account for many aspects of two-tone inhibition. Another model [22] also can explain much of the data, and results in the generation of the cubic combination tone ($2f_1 - f_2$) as a cross-product. This combination tone has been observed in the cochlear microphonic [23]. Furthermore, the latency of two-tone inhibition is too short to result from efferent feedback, which suggests some interaction at the cochlear level [24].

INTENSITY CODING

Determining how intensity is coded by the auditory neural system is a formidable problem. The most difficult aspect is to explain how the auditory nerve is able to subserve an intenstiy continuum covering 120+ dB. Although there are several ways in which intensity information might be coded, the precise mechanism is still unresolved.

The differences between the inner and outer hair cells provide one possible basis for neural intensity coding. Recall that the IHCs are close to the hinge-point of the basilar membrane, while the OHCs are seated more distally. It also appears that some of the cilia of the OHCs are embedded in the undersurface of the tectorial membrane, whereas the IHC cilia are not. The result is that greater mechanical forces are exerted upon the outer than upon the inner HCs in response to movement of the basilar membrane. This fact would lead one to suspect that there are differences in the mode of stimulation for the two cell groups; and there are data indicating that the IHCs respond to the velocity of basilar membrane displacement, whereas the OHCs respond to displacement magnitude [25]. It has also been show by several cochlear microphonic measurements that the OHCs are about 35–50 dB more sensitive than the IHCs [26]. That such a mechanism could subserve at least some degree of neural intensity coding is supported by the innervation patterns reported by Spoendlin [27]. As described in Chap. 1, Spoendlin found that 95% of the neural supply goes to the IHCs while only 5% innervates the OHCs. At least equally important is his observation that the neuron:HC ratio is *convergent* (1:10) upon the auditory nerve from the OHCs, whereas this ratio is *divergent* (20:1) upon the nerve from the IHCs. The convergent neuron:HC ratio would allow for spatial summation of stimulation from the OHCs, which, in turn, could theoretically support a graded intensity coding in the auditory nerve (at least for the 5% of the fibers from these hair cells).

This supposition is supported by Pfeiffer and Kim's [28] findings for the click-elicited firing patterns of single auditory nerve fibers in cats. They found that the 1258 response patterns that they studied could be put into two categories (populations I and II) upon several bases. The fibers in population I, which made up 93% of the sample, were essentially insensitive to changes in signal level. However, the 7% constituting population II showed considerable increases in the number of peaks in their PST histograms as a function of signal level. Although one must remember that no data directly support the correlation between these findings and Spoendlin's, the similarities are impressive.

Even if we assume for the moment that Spoendlin's and Pfeiffer and Kim's data both describe the same fibers, and that the spatial summation

hypothesis is accurate, we are still faced with the problem that a complete absence of OHCs results in a hearing loss of only about 50 dB [26]. Thus, the OHC–population II concept could conceivably explain part of the ear's dynamic range, but certainly not all 120+ dB of it. We must thus look elsewhere.

The first neural responses at barely threshold intensities appear to be a decrease in spontaneous activity [2], and phase locking of spike discharges to the stimulus cycle [10]. This is not to say that the fiber will fire in response to every cycle of a near-threshold stimulus. Rather, even though the overall discharge rate may not be significantly greater than the spontaneous level, those spikes that do occur will tend to be locked in phase with the stimulus cycle [29]. This effect might provide some degree of stimulus intensity coding, since there is evidence that fibers with higher spontaneous rates have lower thresholds [2,30]. However, Kiang [4] has reported that, for any particular animal, the range of thresholds among fibers with similar CFs is quite narrow. We shall return to this point below.

It is, of course, reasonable to consider the intensity range over which a single auditory nerve fiber is responsive to intensity change. In their classic experiment, Galambos and Davis [1] found that the dynamic range of auditory nerve fibers is only about 20–40 dB. In other words, the discharge rate increases with stimulus intensity from threshold up to a level 20–40 dB above it. At higher intensities the spike rate either levels off or decreases. Obviously, a single fiber cannot accommodate the 120+ dB range from minimal audibility to the upper usable limits of hearing. However, if there were a set of units with graded thresholds, they could cooperate to produce the dynamic range of the ear [31]. For example, if there were four fibers with similar CFs having dynamic ranges of 0–40, 30–70, 60–100, and 90–130 dB, respectively, then they could easily accommodate the ear's dynamic range.

Findings by Sachs and Abbas [32] suggest a wider dynamic range in fibers with similar CFs than previously thought (Fig. 5.9), and that the saturation rates are flatter for low-threshold fibers and steeper for units with higher thresholds.

More recently, Liberman [33] reported three relatively distinct groups of fibers with respect to spontaneous discharge rate and threshold (Fig. 5.10). These groups covered a threshold range as great as 70 dB for fibers with similar CFs in the same cat. The fibers of the first group (group c in Fig. 5.10) had high spontaneous rates (over 18/sec), and tended to have the lowest relative thresholds. Within this group, however, the thresholds were essentially the same regardless of the actual spontaneous rate. That is, the thresholds were within ±5 dB whether the spontaneous rate was 20/sec or 100/sec. High-spontaneous-

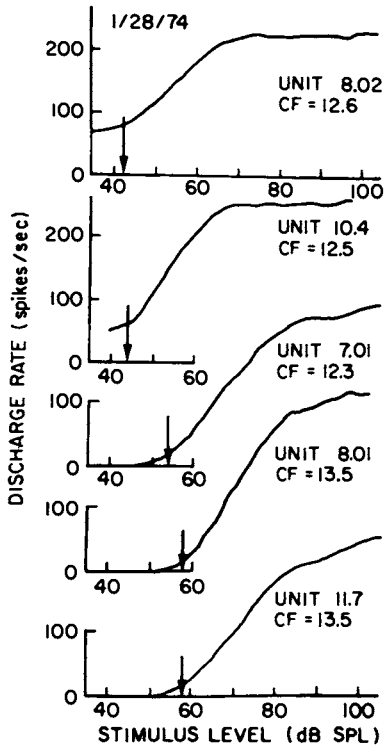

Figure 5.9 Discharge rate as a function of stimulus level for five fibers with similar CFs in the same cat. Arrows are fiber thresholds. (From Sachs and Abbas [32]. Permission *J. Acoust. Soc. Amer.*)

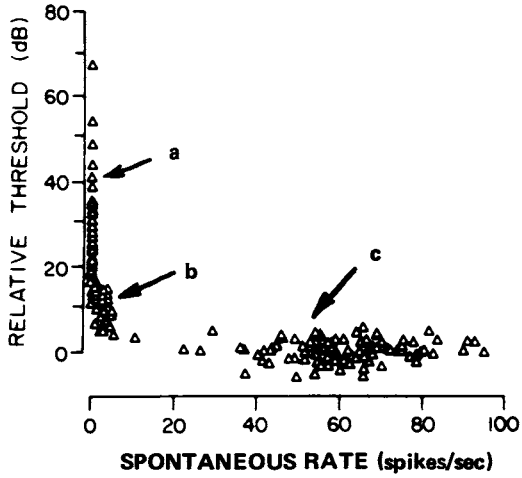

Figure 5.10 Relationship between spontaneous rate and relative thresholds for fibers with low (a), medium (b), and high (c) spontaneous firing rates. (From Liberman [33]. Permission *J. Acoust. Soc. Amer.*)

rate fibers made up about 61% of those sampled. Medium-spontaneous-rate fibers (over 0.5/sec) comprised about 23% of the fibers. These fibers covered a threshold range of roughly 15 dB (group b in Fig. 5.10). The remaining 16% had low spontaneous rates (under 0.5/sec). These units (group a in Fig. 5.10) had the highest thresholds, covering about 50 dB. However, there was no clear correlation between their spontaneous rates and their thresholds.

One might speculate that the low-spontaneous-rate fibers correspond to Spoendlin's longitudinal fibers, or to Pfeiffer and Kim's population II units. However, one must be careful not to overgeneralize prematurely. Recall that Spoendlin identified only 5% of the auditory nerve fibers as longitudinal, compared to the 16% comprising the low-spontaneous-rate group. Pfeiffer and Kim found that population II fibers range in spontaneous rate from less than 1/sec to over 80/sec, whereas the low-spontaneous-rate fibers have rates less than 0.5/sec. Furthermore, Liberman referred to preliminary observations suggesting that the low-spontaneous-rate and population II fibers are not the same.

While the data of Sachs and Abbas and of Liberman suggest that a wider dynamic range than predicted by Kiang's threshold ranges could be subserved, the reason for the discrepancies are not fully apparent; and it would still seem unlikely that a 120+ dB dynamic range could be supported on the basis of threshold differences among fibers alone.

It is apparent that none of the explanations just discussed fully accounts for the width of the range over which the ear can discriminate intensity changes. However, it is reasonable to consider the possibility that these factors do make at least some contribution. An alternative explanation might be based upon several things we already know, and is not inconsistent with the factors already discussed. For example, we know that auditory nerve fiber activity reflects the pattern of cochlear excitation in a rather straightforward manner. Figure 5.11 shows the effect of stimulus intensity on the response area of a single auditory neuron whose CF is 2100 Hz. Note that as intensity is increased, the number of spikes per second also increases, as does the frequency range to which the fiber responds. Also, the frequency range over which the fiber responds tends to increase more below than above the CF.

Keeping this in mind, recall also that the basilar membrane may be conceived of as an array of elements which are selectively responsive to successively lower frequencies going from base to apex. Now refer to Fig. 5.12, by Whitfield [34]. Frequency is represented along the horizontal axis from left to right; the ordinate respresents spike rate. The hypothesis for neural coding is as follows. Figure 5.12a represents the

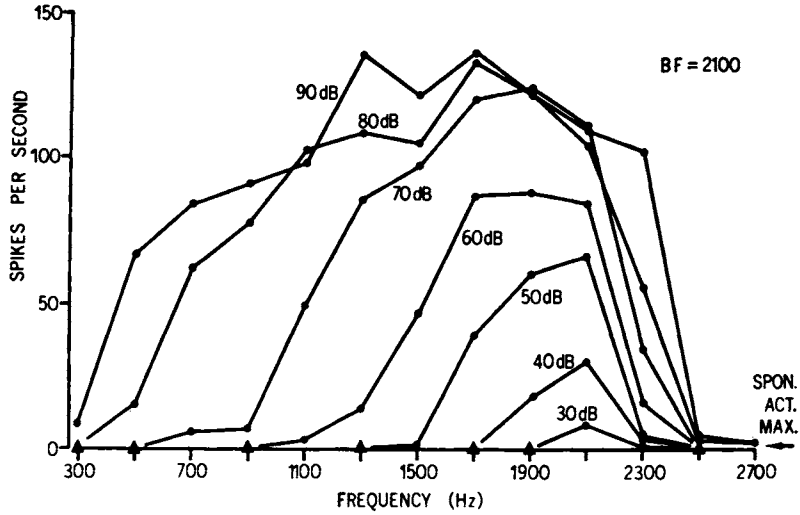

Figure 5.11 Effect of stimulus intensity on the response area of a single auditory neuron (CF = 2100 Hz). (From Rose, Hind, Anderson, and Brugge [30]. Permission *J. Neurophysiol.*)

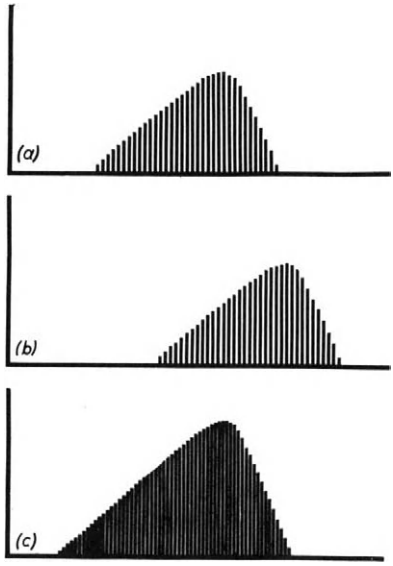

Figure 5.12 Hypothetical changes in the array of auditory nerve responses (see text). (Courtesy of I. C. Whitfield [34].)

response area resulting from stimulation at a given frequency. Figure 5.12b shows the discharge pattern for equivalent stimulation at a higher frequency. The frequency change is represented as a simple displacement of the hypothetical response area along the frequency axis, and is analogous to the movement of the traveling wave along the basilar membrane. Figure 5.12c represents the effect of increasing the stimulus level at the same frequency as in Fig. 5.12a. The position of the response area stays the same, but the number of responding fibers increases. Here is what hypothetically occurs: A particular number of fibers are responding in Fig. 5.12a. As the stimulus level increases, the fibers increase their spike rates (until saturation is reached). Although some fibers saturate, other fibers with similar CFs but different thresholds continue to increase their discharge rates as the level increases. As intensity continues to increase, the stimulus enters the excitatory areas of other fibers which respond to that frequency only at higher intensities. The overall effect, then, is for the intensity increment to be coded by increased overall firing rates, among more fibers, over a wider frequency range. (We shall see in the next section that increasing the stimulus level also results in greater synchrony of the individual neural discharges, so that the whole-nerve action potential has a shorter latency and greater magnitude.)

Such a model could employ all of the preceding factors in arriving at the overall change in the excitation pattern between Fig. 5.12a and Fig. 5.12c, and could account for the wide dynamic range of the ear. It does not, however, account for the observation in Fig. 5.11 that the peak of a fiber's response curve shifts in frequency as the stimulus level is increased [30]. The implication is that phase-locking to the stimulus cycle would be particularly important in maintaining frequency coding when the intensity increment is encoded by increases in density of discharges per unit time and by widening of the array of active fibers.

WHOLE-NERVE ACTION POTENTIALS

So far we have examined individual auditory nerve fibers. In this section we shall review some aspects of the whole-nerve, or compound, action potential (AP) of the auditory nerve. In addition to its contribution to an understanding of the auditory system, the whole-nerve AP is also clinically useful in the procedure called electrocochleography [35,36]. The whole-nerve AP, as its name suggests, is a composite of many individual fiber discharges. These more or less synchronous discharges are recorded by an extracellular electrode as a negative deflection. Recall from Chap 4 that the AP must be separated from the cochlear microphonic (CM), which is generally accomplished by an averaging procedure. (Recall that the CM is in opposite phase in the scalae vestibuli and tym-

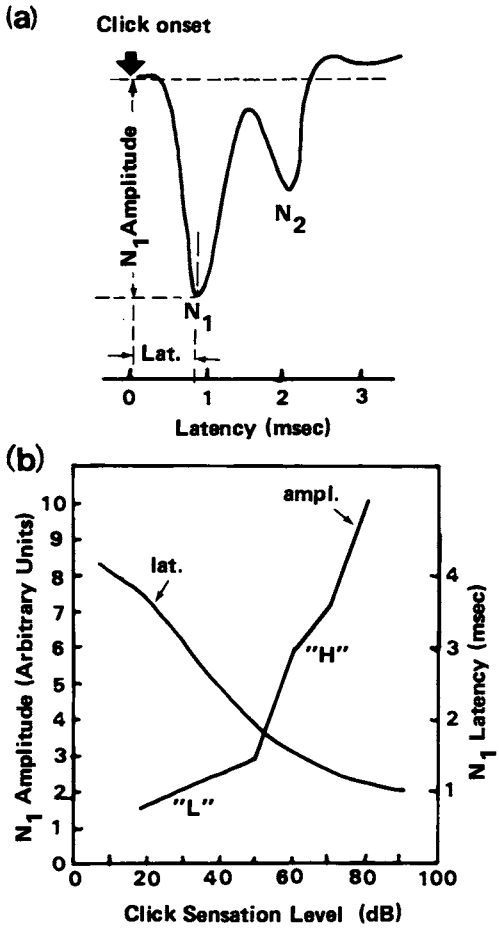

Figure 5.13 (a) Hypothetical whole-nerve AP response. (b) Amplitude and latency functions of the AP response based on Yoshie's [37] data. Amplitude is in arbitrary units where 10 is the amplitude of the response to a click at 80 dB SPL.

pani, whereas the AP is always negative. Thus, averaging the responses from the two scalae cancels the CM and enhances the AP.)

The AP is also frequently recorded from other locations than the inside of the cochlea, for example at the round window in animals, and at the promontory and even from the ear canal in humans. In these cases, averaging the responses has the important effect of enhancing the AP and reducing the random background noise (see Chap. 6 for a discussion of signal averaging).

As shown in Fig. 5.13a, the AP appears as a negative amplitude peak (N_1) at some latency following stimulus onset, and then as one or more

smaller peaks (N_2, N_3, etc.). Whole-nerve APs are elicited by transient stimuli with fast rise times, such as clicks or tone bursts. It is generally accepted that the transient stimulus leads to the synchronous firing of many fibers [38] exclusively from the basal turn of the cochlea [39]. The AP is attributed to the basal turn because the high speed of the traveling wave in this part of the cochlea causes a large number of receptors to be stimulated almost simultaneously, leading to the synchrony of the neural discharges. More apical parts of the cochlea are not thought to contribute because the longer travel time up the partition would cause firings from these regions to be nonsynchronous. However, there is some evidence that other parts of the cochlea may also contribute to the AP response [40].

The origin of the N_2 peak has been the subject of some controversy. Teas et al. [41] pointed out that the diphasic waveform (a negative deflection followed by a positive one) would result in the N_1 and N_2 peaks. Other suggested sources of these peaks have included more apical areas of the cochlea [42], repeated firings of the auditory nerve fibers [39], and discharges from the second-order neurons of the cochlear nucleus [43]. Eggermont [44] found that the distinct N_2 peaks found in the guinea pig response were absent in the human response. He attributed this to differences in the electroanatomy of the human and guinea pig, and suggested that the source of the peak might be cochlear nucleus discharges, which are more readily recorded in the guinea pig than in man. Experimental data, however, have shown that it is the positive deflection of the diphasic response which separates the AP into N_1 and N_2 peaks; and that this separation occurs within the internal auditory meatus [40,45]. These studies showed that when the central end of the auditory nerve (in the internal auditory canal) was deactivated chemically or mechanically, the result was a change in the AP wave so that the N_2 peak was absent. This loss is apparently due to the removal of the positive deflection of the diphasic response, which normally separates the N_1 and N_2 peaks.

Figure 5.13b shows that both the magnitude and latency of the AP depend on stimulus intensity [37,38]: Increasing the level of the stimulus increases the amplitude of the response and decreases its latency. This relationship is actually not a simple one-to-one correspondence between stimulus level and AP amplitude and latency. Notice that the amplitude curve in Fig. 5.13b is made up of two rather distinct line segments rather than of a single straight line. There is a portion with a shallow or low ("L") slope at lower stimulus levels, and a segment which becomes abruptly steep or high ("H") in slope as stimulus intensity continues to increase. Furthermore, the function is not necessarily monotonic; i.e., the relationship between stimulus level and AP amplitude is not always in the same direction for all parts of the curve. A small dip in AP ampli-

tude has been reported to occur at about the place where the L and H segments meet, before the curve resumes increasing with stimulus level [46].

The source of the two-part amplitude function has been thought to be the two populations of receptor cells in the cochlea [31,36]. That is, the L segment is interpreted as reflecting the output of the OHCs and the H portion is seen as derived from the IHCs. This interpretation is largely based on the findings that the IHCs are richly innervated compared to the OHCs, and that the OHCs have lower thresholds. A major problem with this viewpoint is that two distinct populations of auditory neurons, differing in threshold, would also be required. Unfortunately, the data supporting two such neural populations is sketchy. There are also experimental data which contradict the theory that different sets of receptors cause the two segments of the AP amplitude function. In particular, the amplitude function has been found to have a single slope over its entire length when the central end of the nerve within the internal auditory meatus is deactivated [45]. Deactivation of this part of the auditory nerve changes the shape of the AP waveform, also, by removing the positive deflection that separates the N_1 and N_2 peaks.

Alternate models to explain the two segments of the AP amplitude function have been proposed by Evans [47] and by Özdamar and Dallos [48]. These similar models explain the shape of the amplitude function on the basis of single-fiber tuning curves. Recall that the typical tuning curve has two parts: a narrowly tuned area around the CF at low levels, and a more widely tuned area extending downward in frequency as stimulus intensity is raised. This shape is shown in Fig. 5.14a. The important point is that the low-frequency tails make a neuron responsive to a wide frequency range provided the intensity is great enough. Consider the responses of various fibers to a 1000 Hz tone, represented by the vertical line at 1000 Hz in the figure. Point 1 represents the lowest level at which a 1000 Hz CF fiber responds. The remaining points (2 through 8) show the levels at which the 1000 Hz tone activates fibers with other CFs.

The 1000 Hz tone primarily crosses the low-frequency tails of the higher-CF tuning curves, but it also crosses the higher-frequency portions of some lower-CF tuning curves (e.g. point 6). Thus, a 1000 Hz tone activates more and more fibers with different CFs as its level increases. This situation is plotted in Fig. 5.14b, which shows *across*-fiber tuning curves. The numbers correspond to those in Fig. 5.14a. Across-fiber tuning curves show the levels at which fibers with various CFs respond to a particular frequency (1000 Hz here). The tip of the across-fiber tuning curve represents the lowest level at which *any* fiber responds

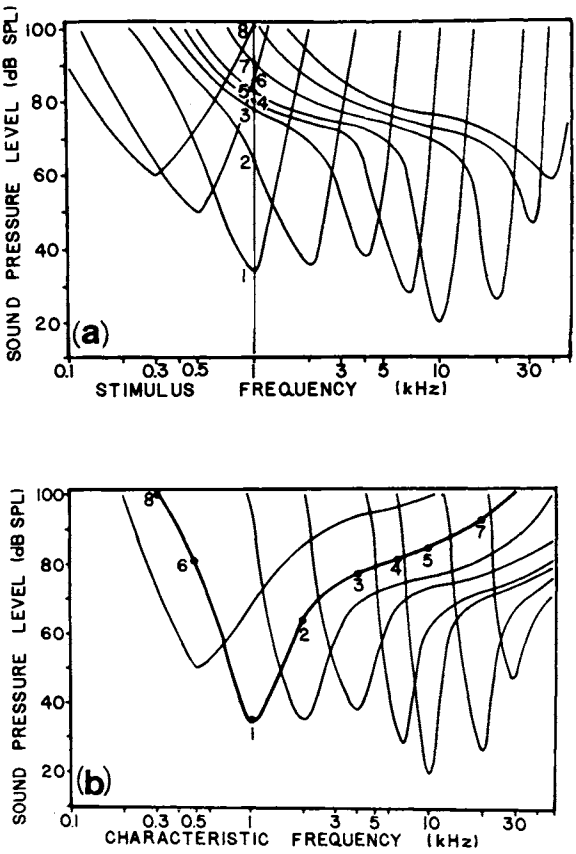

Figure 5.14 (a) Auditory nerve fiber tuning curves. (b) Across-fiber tuning curves. (From Özdamar and Dallos [48]. Permission *J. Acoust. Soc. Amer.*)

to that frequency. The width of the across-fiber tuning curve shows the *range* of neurons with different CFs that respond at a given stimulus level. For example, only the 1000 Hz CF fiber responds to 1000 Hz at 35 dB SPL; but fibers with CFs between about 500 and 10,000 Hz may respond when the same tone is presented at 80 dB SPL (i.e., between points 6 and 5).

Across-fiber tuning curves have high-frequency tails instead of the low-frequency tails of individual fiber tuning curves. This situation occurs because the low-frequency tails of an increasing number of high-frequency fibers respond as the intensity is raised (Fig. 5.14a).

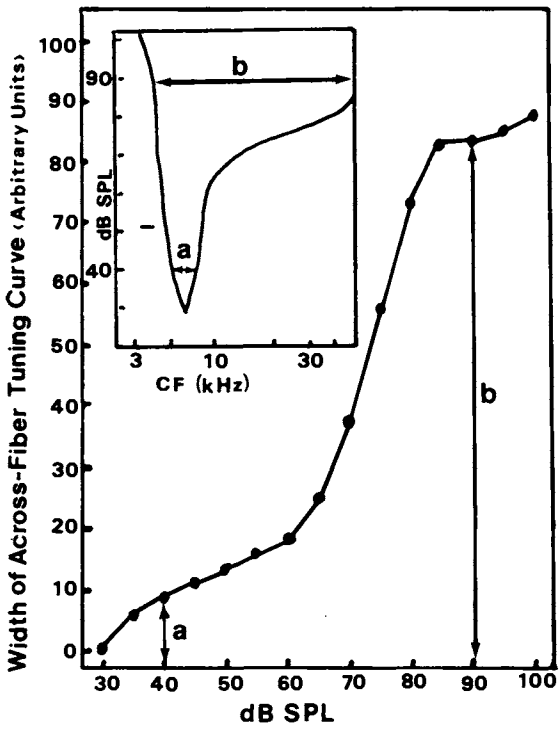

Figure 5.15 Width of an across-fiber tuning curve as a function of SPL. Insert: across-fiber tuning curve upon which the figure is based. Lines a and b correspond on both curves. (Adapted from Özdamar and Dallos [48]. Permission *J. Acoust. Soc. Amer.*)

How might this effect account for the two segments of the AP amplitude function? The insert in Fig. 5.15 shows the width of a hypothetical across-fiber tuning curve. Line a shows the width of the across-fiber tuning curve within the narrowly tuned segment. The width of the tuning curve at 90 dB SPL is shown by line b, which is in the widely tuned part of the curve. The main figure shows the width of the across-fiber tuning curve on the y-axis as a function of SPL. Lines a and b correspond to those in the insert. Only a small population of fibers with CFs close to the stimulus frequency respond at low intensity levels. At higher levels there is a dramatic increase in the number of responding fibers, as the stimulus level reaches the low-frequency tails of the other neurons. There is thus a slow increase in the width of the across-fiber tuning curve followed by a sharp increase as intensity rises. Comparison of Fig. 5.15 with Fig. 5.13b shows that this curve is remarkably similar to the AP ampli-

tude function. The portion of the curve in the vicinity of line a corresponds to the L portion of the AP amplitude function, while the part around line b corresponds to the steeper H segment. This correspondence provides a mechanism which might underlie the two segments of the compound AP amplitude function. (One might also consider this mechanism with respect to excitation models like the one in Fig. 5.12.) Recent physiological findings by Özdamar and Dallos [49] support this mechanism.

REFERENCES

1. R. Galambos and H. Davis, The response of single auditory nerve fibers to acoustic stimulation, *J. Neurophysiol. 6*, 39–57 (1943).
2. N. Y.-S. Kiang, Discharge Patterns of Single Fibers in the Cat's Auditory Nerve, M.I.T. Press, Cambridge, Mass., 1965.
3. N. Y.-S. Kiang, M. B. Sachs, and W. T. Peake, Shapes of tuning curves for single auditory-nerve fibers, *J. Acoust. Soc. Amer. 42*, 1341–1342 (1967).
4. N. Y.-S. Kiang, A survey of recent developments in the study of auditory physiology, *Ann. Otol. 77*, 656–675 (1968).
5. N. Y.-S. Kiang and E. C. Moxon, Tails of tuning curves of auditory-nerve fibers, *J. Acoust. Soc. Amer. 55*, 620–630 (1974).
6. E. F. Evans and J. P. Wilson, The frequency selectivity of the cochlea, in *Basic Mechanisms in Hearing* (A. R. Møller, Ed.), Academic, New York, 1973, pp. 519–554.
7. N. Y.-S. Kiang, Stimulus representation in the discharge patterns of auditory neurons, in *The Nervous System* (D. B. Tower and E. L. Eagles, Eds.), Vol. 3: *Human Communication and Its Disorders*, Raven, New York, 1975, pp. 81–96.
8. H. Davis, C. Fernandez, and D. R. McAuliffe, The excitatory process in the cochlea, *Proc. Natl. Acad. Sci.* 580–587 (1950).
9. G. Bekesy, *Experiments in Hearing*, McGraw-Hill, New York, 1960.
10. J. E. Rose, J. F. Brugge, D. J. Anderson, and J. E. Hind, Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey, *J. Neurophysiol. 30*, 769–793 (1967).
11. J. E. Hind, D. J. Anderson, J. F. Brugge, and J. E. Rose, Coding of information pertaining to paired low-frequency tones in single auditory nerve fibers of the squirrel monkey, *J. Neurophysiol. 30*, 794–816 (1967).
12. J. F. Brugge, D. J. Anderson, J. E. Hind, and J. E. Rose, Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds, *J. Neurophysiol. 32*, 386–401 (1969).
13. J. E. Rose, J. F. Brugge, D. J. Anderson, and J. E. Hind, Some possible neural correlates of combination tones, *J. Neurophysiol. 32*, 402–423 (1969).
14. M. Nomoto, N. Suga, and Y. Katsuki, Discharge pattern and inhibition of primary auditory nerve fibers in the monkey, *J. Neurophysiol. 27*, 768–787 (1964).

15. M. B. Sachs and N. Y. S. Kiang, Two-tone inhibition in auditory nerve fibers, *J. Acoust. Soc. Amer. 43*, 1120–1128 (1968).
16. P. J. Abbas and M. B. Sachs, Two-tone suppression in auditory-nerve fibers: Extension of a stimulus-response relationship, *J. Acoust. Soc. Amer. 59*, 112–122 (1976).
17. M. B. Sachs and P. J. Abbas, Phenomenological model for two-tone inhibition, *J. Acoust. Soc. Amer. 60*, 1157–1163 (1976).
18. R. R. Pfeiffer and C. E. Molnar, Cochlear nerve fiber discharge patterns: Relationship to cochlear microphonic, *Science 167*, 1614–1616 (1970).
19. J. P. Legouix, M. C. Remond, and H. B. Greenbaum, Inteference and two-tone inhibition, *J. Acoust. Soc. Amer, 53*, 409–419 (1973).
20. P. Dallos and M. A. Cheatham, Cochlear microphonic correlates of cubic difference tones, in *Facts and Models in Hearing* (E. Zwicker and E. Terhardt, Eds.), Springer-Verlag, New York, 1974.
21. H. Duifhuis, Cochlear nonlinearity and second filter: Possible mechanism and implications, *J. Acoust. Soc. Amer. 59*, 408–423 (1976).
22. R. R. Pfeiffer, A model for two-tone inhibition of single cochlear-nerve fibers, *J. Acoust. Soc. Amer. 48*, 1373–1378 (1970).
23. P. Dallos, Combination tone $2f_1 - f_h$ in microphonic potentials, *J. Acoust. Soc. Amer. 46*, 1437–1444 (1969).
24. G. G. Furman and R. R. Frishkopf, Model for neural inhibition in the mammalian cochlea, *J. Acoust. Soc. Amer. 36*, 2194–2201 (1964).
25. P. Dallos, M. C. Billone, J. D. Durrant, C. Y. Wang, and S. Raynor, Cochlear inner and outer hair cells: Functional differences, *Science 177*, 356–358 (1972).
26. P. Dallos, *The Auditory Periphery: Biophysics and Physiology*, Academic, New York, 1973.
27. H. Spoendlin, Neurophysiological basis of cochlear coding mechanisms, *Audiology 14*, 383–407 (1975).
28. R. R. Pfeiffer and D. O. Kim, Response patterns of single cochlear nerve fibers to click stimuli: Descriptions for cat, *J. Acoust. Soc. Amer. 52*, 1669–1677 (1972).
29. J. E. Hind, Physiological correlates of auditory stimulus periodicity, *Audiology 11*, 42–57 (1972).
30. J. E. Rose, J. E. Hind, D. J. Anderson, and J. F. Brugge, Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey, *J. Neurophysiol. 34*, 685–699 (1971).
31. H. Davis, Peripheral coding of auditory information, in *Sensory Communication* (W. A. Rosenblith, Ed.), M.I.T. Press, Cambridge, Mass., 1961, pp. 119–141.
32. M. B. Sachs and P. J. Abbas, Rate versus level functions for auditory-nerve fibers in cats: Tone-burst stimuli, *J. Acoust. Soc. Amer. 56*, 1835–1847 (1974).
33. M. C. Liberman, Auditory-nerve response from cats raised in a low-noise chamber, *J. Acoust. Soc. Amer. 63*, 442–455 (1978).
34. I. C. Whitfield, The Auditory Pathway, Arnold, London, 1967.
35. J. J. Eggermont, D. W. Odenthal, P. H. Schmidt, and A. Spoor, Electrocochleography: Basic principles and clinical applications, *Acta Otol.*, Suppl. 316 (1974).

36. J. J. Eggermont, Electrocochleography, in *Handbook of Sensory Physiology* (W. D. Keidel and W. D. Neff, Eds.), Vol. 3, Springer-Verlag, Berlin, 1976, pp. 625–705.
37. N. Yoshie, Auditory nerve action potential responses to clicks in man, *Laryngoscope 78,* 198–213 (1968).
38. M. H. Goldstein and N. Y.-S. Kiang, Synchrony of neural activity in electrical responses evoked by transient acoustic clicks, *J. Acoust. Soc. Amer. 30,* 107–114 (1958).
39. I. Tasaki, Nerve impulses in individual auditory nerve fibers, *J. Neurophysiol. 17,* 97–122 (1954).
40. J. P. Legouix and A. Pierson, Investigations on the source of whole-nerve action potentials from various places in the guinea pig cochlea, *J. Acoust. Soc. Amer. 56,* 1222–1225 (1974).
41. D. C. Teas, D. H. Eldredge, and H. Davis, Cochlear responses to acoustic transients: An interpretation of whole nerve action potentials, *J. Acoust. Soc. Amer. 34,* 1438–1459 (1962).
42. J. E. Pugh, Jr., D. J. Anderson, and P. A. Burgio, The origin of N_2 of the cochlear whole-nerve action potential, *J. Acoust. Soc. Amer. 53,* 325(A) (1973).
43. I. Tasaki, H. Davis, and R. Goldstein, The peripheral organization of activity, with reference to the ear, *Cold Spring Harbor Symp. Quant. Biol. 17,* 143–154 (1952).
44. J. J. Eggermont, Analysis of compound action potential responses to tone bursts in the human and guinea pig cochlea, *J. Acoust. Soc. Amer. 60,* 1132–1139 (1976).
45. J. P. Legouix, D. C. Teas, H. A. Beagley, and M. C. Remond, Relation between the waveform of the cochlear whole nerve action potential and its intensity function, *Acta Otol. 85,* 177–183 (1978).
46. N. Y.-S. Kiang, M. H. Goldstein, and W. T. Peake, Temporal coding of neural responses to acoustic stimuli, *Inst. Radio Eng. Trans. Inform. Theory IT-8,* 113–119 (1962).
47. E. F. Evans, The sharpening of cochlear frequency selectivity in the normal and abnormal cochlea, *Audiology 14,* 419–442 (1975).
48. Ö. Özdamar and P. Dallos, Input-output function of cochlear whole-nerve action potentials: Interpretation in terms of one population of neurons, *J. Acoust. Soc. Amer. 59,* 143–147 (1976).
49. Ö. Özdamar and P. Dallos, Synchronous responses of the primary auditory fibers to the onset of tone burst and their relation to compound action potentials, *Brain Res. 155,* 169–175 (1978).

6
Auditory Pathways

STIMULUS CODING

Tuning to best or characteristic frequencies (CFs) has been observed at all levels of the auditory system. This tuning is implicit in the tonotopic organization of these pathways; i.e., in the orderly arrangement of frequency by place. Such tuning curves have been reported for the cochlear nuclei [1], the superior olivary complex and trapezoid body [2–5], the inferior colliculus and medial geniculate [4,6], and the auditory cortex [7,8]. Particularly sharp tuning curves resembling those of the cochlear nerve have been reported at levels up to the superior olivary complex, while some nuclei even at these relatively low levels show broad tuning [5]. Both narrowly and broadly tuned fibers have been observed at higher levls. Although there is debate as to whether there is sharper tuning in individual auditory nerve fibers than in the cochlea (see Chaps. 4 and 5), it does not appear likely that further sharpening occurs at higher levels in the auditory system.

Auditory nerve firing patterns were discussed in some detail in the previous chapter. These primary units are exclusively excitatory. In other words, they always respond to stimulation. As we shall see, this and other characteristics of the response of auditory nerve fibers are not seen in all neurons at higher levels. In fact, discharge patterns drastically divergent from those of auditory nerve units are seen as low as the cochlear nuclei.

Stimulus coding in the cat's cochlear nuclei was studied extensively by Kiang et al. [1] and by Rose et al. [9]. Five groups of neurons relatively distinct in their discharge patterns were described in the cochlear nuclei

[1]. "Primary-like" units—units with firing patterns like those of auditory nerve fibers—were found in the interstitial nucleus (IN) and in the posteroventral cochlear nucleus (PVCN).

A second group, found primarily in the IN and PVCN but also observed in the dorsal cochlear nucleus (DCN), are called "choppers." They were labeled as such due to the chopped appearance of their PST histograms, which look as though segments have been chopped out. The chopping is due to the presence of preferred discharge times which are regularly spaced over time; the rate of chopping is related to the tone burst's level, duration, and rate. The response is greatest at the onset of the tone burst, and decreases over the time that the stimulus is left on.

The third group of neurons, the "on" units, are found almost exclusively in the IN and PVCN, although several are observed in the anteroventral cochlear nucleus (AVCN) and DCN. These fibers are particularly interesting because they respond to the onset of a tone burst with a momentary discharge of spikes, but do not respond to the remainder of the tone burst or to continuous tones.

A fourth type of unit is found only in the DCN. These units are called "pausers" because they respond after longer latencies than the other cells. At higher stimulus levels pausers have an initial discharge peak, then a silent interval followed by discharges that last for the rest of the tone burst.

Finally, the AVCN is composed almost completely of cells whose discharge patterns differ from other units in that their responses decay at an exponential rate.

The point being made is that responses at higher levels are more variable than the exclusively excitatory firing patterns of the auditory nerve fibers. For example, some superior olivary complex fibers respond to clicks with a single discharge, others with two, and still others with a train of firings [10,11]. "On" units, pausers, and fibers responding to the duration of the tone burst have been identified in the inferior colliculus [9]. Many fibers in the medial geniculate are "on" units; others are pausers, cells which fire upon stimulus offset, and cells that respond as long as the stimulus is present [12,13,14].

The responses of single neurons in the auditory cortex demonstrate the importance of stimulus novelty. That is, cortical neurons are more interested in change than in ongoing stimulation. Auditory cortical neurons tend to respond to stimulus onset and/or offset [7,8]. Hind [7] reported that the typical response in anaesthetized animals is composed of only one to four spikes rather than a sustained discharge. However, Brugge and Merzenich [15] found that the typical response of an unanesthesized animal is a high spike rate which adapts to a lower rate in under a second. This apparent inconsistency actually serves to highlight

the effects of anesthesia upon the responses of central auditory neurons. Of particular interest is that many auditory neurons will not respond to steady-state tones, but will respond to tones that are modulated in frequency [16]. Furthermore, many of these cells will respond only to a rise or to a fall in frequency. Other cortical neurons respond to movements of a sound source around the head [17].

BINAURAL CODING

A variety of auditory functions depend upon information from both ears, or binaural hearing. For example, the ability to localize a sound source is dependent upon differences in the intensity and time of arrival of the sound at the two ears. This topic will be discussed further in Chap. 13. The current question is how the auditory system encodes and uses information from the two ears.

Superior Olivary Complex

Inputs from both ears must obviously be available to a neural center if binaural information is to be processed. Referring back to Fig. 1.22, recall that the superior olivary complex (SOC) is the lowest level at which binaural information is available in a single way station. This binaural convergence is particularly observed at the medial superior olive (MSO), whose connections with the cochlear nuclei (CN) of both sides make it well suited for this function. Consider a representative cell in the MSO. It synapses on its medial surface with a fiber from the opposite CN and on its lateral surface with a fiber from the CN on the same side [18]. These connections are represented in Fig. 6.1. Thus, neurons in the right MSO receive primarily projections from the left CN on their left (medial) sides and inputs from the right CN on their right (lateral) surfaces.

Hall [19] found that most SOC neurons responded to click stimulation of the opposite ear only. Stimulation of the ipsilateral ear did not yield a response unless the opposite ear also was stimulated. Such cells might be called excitatory–inhibitory (EI), since a signal from one side causes an increase in the firing rate (excitation) while an input from the other side decreases the discharge rate (inhibition). If the ipsilateral (inhibitory) click preceded the click from the other (inhibitory) side, then the firing rate decreased. The firing rate also decreased if the click intensity was greater at the ipsilateral ear. Thus, excitation and/or inhibition of the cell was controlled by the relative intensities and times of stimulus arrival at the two ears.

Figure 6.1 reflects Hall's finding that contralateral stimulation is excitatory and ipsilateral stimulation is inhibitory. However, this situa-

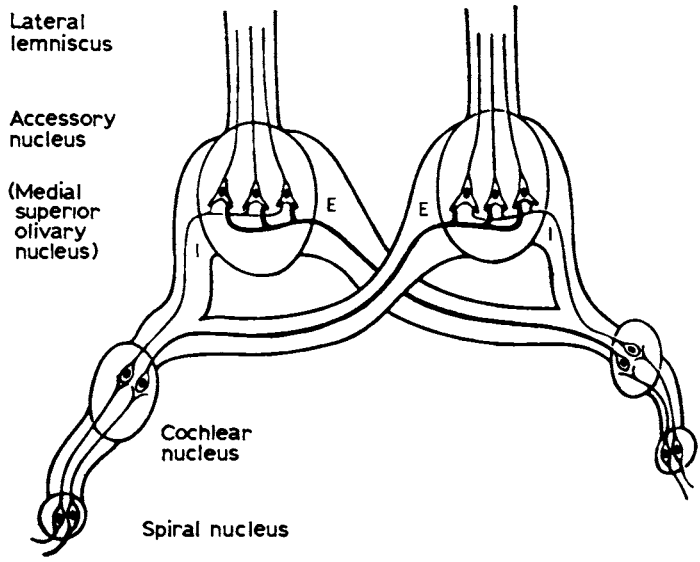

Figure 6.1 Schematic representation of the bilateral projections to MSO cells. The existence of only excitatory (E) contralateral inputs and inhibitory (I) ipsilateral ones is questioned by subsequent research (see text). (From Van Bergeijk [20]. Permission *J. Acoust. Soc. Amer.*)

tion was not observed for every fiber; and Moushegian et al [21] found that the opposite situation occurred in some click-stimulated fibers. They also found ipsilateral as well as contralateral excitation in response to tones [22]. More recently, they reported that ipsilateral and contralateral inputs each excite roughly half of the MSO neurons, while binaural clicks are inhibitory [23]. Using tone bursts, Goldberg and Brown [24] found two groups of SOC neurons responsive to high frequencies. One group was excitatory–excitatory (EE),—it responded to inputs from either ear. The second group was excitatory–inhibitory (EI), responding to one ear and being suppressed by the other. The EE neurons were sensitive to the average intensity of the tones at both ears, but not to a difference between the intensities at the two ears. The EI cells were sensitive to interaural intensity differences (IIDs), but not to the average binaural level.

Interestingly, Goldberg and Brown observed a different effect at low frequencies. Here, cells responded in a manner which was synchronized with the phase of the stimulus tone when there was a certain characteristic delay, or interaural time difference (ITD), between the ears. In other words, the discharge rate was related to the relative timing of the low-frequency tones at each ear. The cell's responses were maxi-

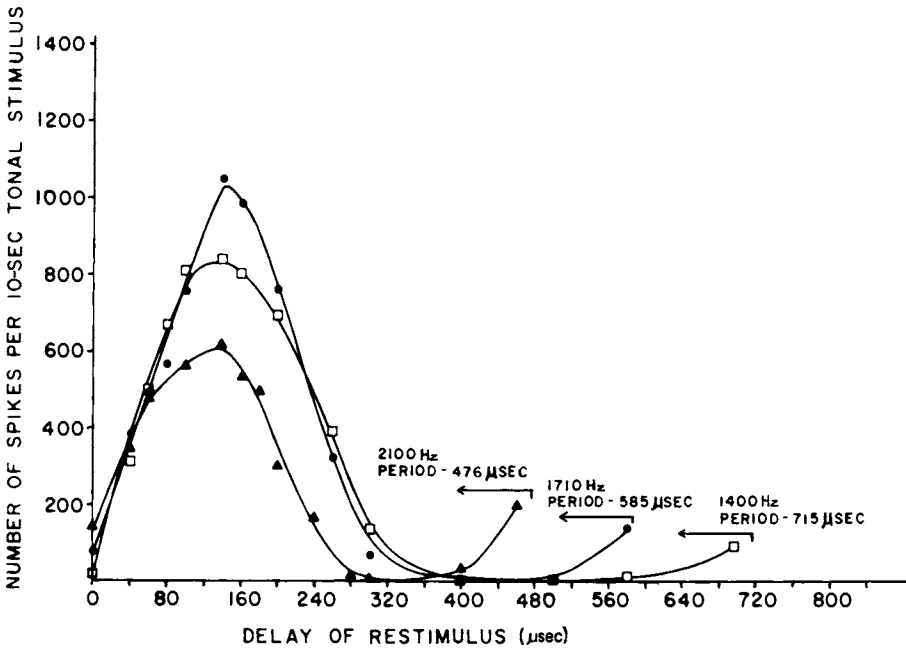

Figure 6.2 Responses of a single IC neuron with a characteristic delay of approximately 140 μsec to binaural stimulation at various low frequencies. The function shows a peak at the characteristic delay regardless of frequency. (From Rose et al [26]. Permission *J. Neurophysiol.*)

mized when the inputs arrived in phase and were minimized when they were out of phase. This result is consistent with other studies which found that SOC cells fire in a probabilistic manner that is in cadence with multiples of the period of the stimulus tone [23,25]. Thus, the phase locking of neural responses to stimulus cycle (period) seen at the cochlear nerve is maintained as high as the SOC; and it appears that these phase-locked volleys contribute to binaural analysis as well as to frequency coding.

In summary, SOC neurons code binaural information through an interaction of excitatory and inhibitory inputs, which are the result of intensity and time (phase) differences in the stimuli at the two ears.

Inferior Colliculus

Binaural information is also available to the inferior colliculus (IC). Rose et al. [26] found cells in the IC that are sensitive to ITDs at characteristics delays. In other words, the neuron responded maximally at a particular ITD independent of the intensity or frequency of the tone. Figure 6.2

shows the spike counts of a single IC neuron in response to various stimulus frequencies as a function of the ITD. Note that the number of discharges is greatest at an interaural time delay of about 140 μsec regardless of the frequency. If the time scale were extended, one would see additional peaks repeated at intervals equal to the period of the frequency of the stimulus tones. Thus, the peaks for the 2100 Hz tone would recur at 476 μsec intervals, while the period of the function for 1400 Hz would be 715 μsec. This period/spike function for ITDs is observed throughout the auditory system.

Other cells were found that were sensitive to IIDs. Rose et al. [26] suggested the ITD-sensitive cells play a role in low frequency localization, and that IID-sensitive cells are involved in high frequency localization (see Chap. 13).

The findings of Rose et al. have been essentially confirmed for tonal, click, and noise stimuli [27–29], with the general finding that contralateral inputs are excitatory and ipsilateral signals are inhibitory. In addition to IID- and ITD-sensitive cells, Benevento and Coleman [28] found IC neurons sensitive to both IID and ITD, as well as units not sensitive to either one. This result is also evidence that certain IC cells are particularly responsive to a sound source which is moved around the head [30].

In summary, neurons in the inferior colliculus respond to interaural intensity and time differences. These differences between the ears result in excitatory and inhibitory inputs to IC neurons, whose firing depends upon the interaction of these inputs. The IC and SOC thus respond to binaural stimulation in similar ways. However, there is an interesting distinction that should be added here. Although phase locking to low frequencies is the rule in the SOC, it occurs for less than a third of the cells in the IC [27].

Other Subcortical Nuclei

Most of the work on binaural coding has been concentrated upon the SOC and IC. However, responses to interaural time-of-arrival and intensity differences have also been found in the dorsal nucleus of the lateral lemniscus [31] and in the medial geniculate body [32,33]. Findings in the dorsal nucleus of the lateral lemniscus have been similar to those for the SOC. Medial geniculate neurons tend to respond with short-latency initial discharges (up to about 30 msec and/or a late discharge (over 500 msec). These units may be EE or EI, and some fibers respond to the direction of sound source movement. The longer-latency fibers may play a role in maintaining a short-term memory for interaural differences [33].

Auditory Cortex

The responses of single units to monaural and binaural stimulation in the auditory cortex [in the primary (AI) and secondary (AII) auditory areas] have been studied in laboratory animals [15,34,35]. Auditory cortex cells tend to exhibit onset or offset responses to clicks and tone or noise bursts, although some units yield a sustained discharge during tone bursts. Most units are excitatory–inhibitory. The cells tend to respond to contralateral but not to ipsilateral stimulation; and the response to a stimulus at the opposite ear is generally inhibited by ipsilateral stimulation. Cells in the auditory cortex are responsive to interaural intensity differences (but not to changes in the binaural level), as well as to interaural time differences.

In addition to single-cell responses, evoked surface (gross) potentials also reflect the resonses of the auditory cortex to ITDs [36]. For lower frequencies (below roughly 3 kHz), cortical auditory neurons respond maximally to ITDs at characteristic delays, and with a periodic function that is repeated at the period of the stimulus frequency. This situation is similar to the one seen at lower levels. There are also cells in AI that respond to the direction of sound source movement [17]. Thus, interaction of excitatory and inhibitory inputs contributes to the processing of binaural information (particularly with regard to localization) at levels up to and including the auditory cortex. This model is consistent with the localization difficulties demonstrated in animals whose auditory cortices have been ablated (as discussed later in this chapter).

TONOTOPIC ORGANIZATION

One of the most interesting aspects of the auditory pathways is the relatively systematic representation of frequency at each level. That is, there is a virtual "mapping" of the audible frequency range within each nuclear mass; neurons most sensitive to high frequencies are in one area, those sensitive to lows are in another part, and those sensitive to intermediate frequencies are located successively between them. This orderly representation of frequency according to place is called tonotopic organization.

It is generally accepted that high frequencies are represented basally in the cochlea, tapering down to low frequencies at the apex. This tonotopic arrangement is continued in the eighth nerve, where apical fibers are found toward the core of the nerve trunk and basal fibers on the outside and on the inferior margin [37], as shown in Fig. 6.3.

Lewy and Kobrack [38] found that basal axons enter the dorsal part of the dorsal cochlear nucleus, and apical fibers enter the ventral cochlear nucleus (VCN) as well as part of the DCN. The ventrolateral termi-

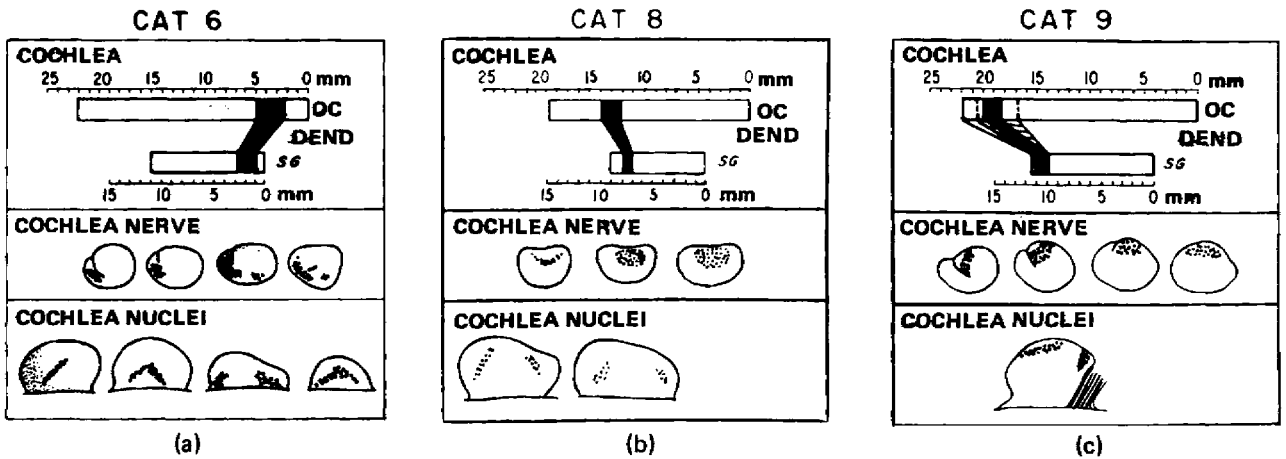

Figure 6.3 Tonotopic relations between the organ of Corti (OC), spiral ganglia (SG), cochlear nerve, and cochlear nuclei based on degeneration observed after selective lesions of the OC and SG. Solid lines indicate the induced lesions; crosshatched and dotted areas are resulting degenerations: (a) basal, (b) middle, and (c) apical turns of different cats. (From Sando [37]. Permission *Acta Otol.*)

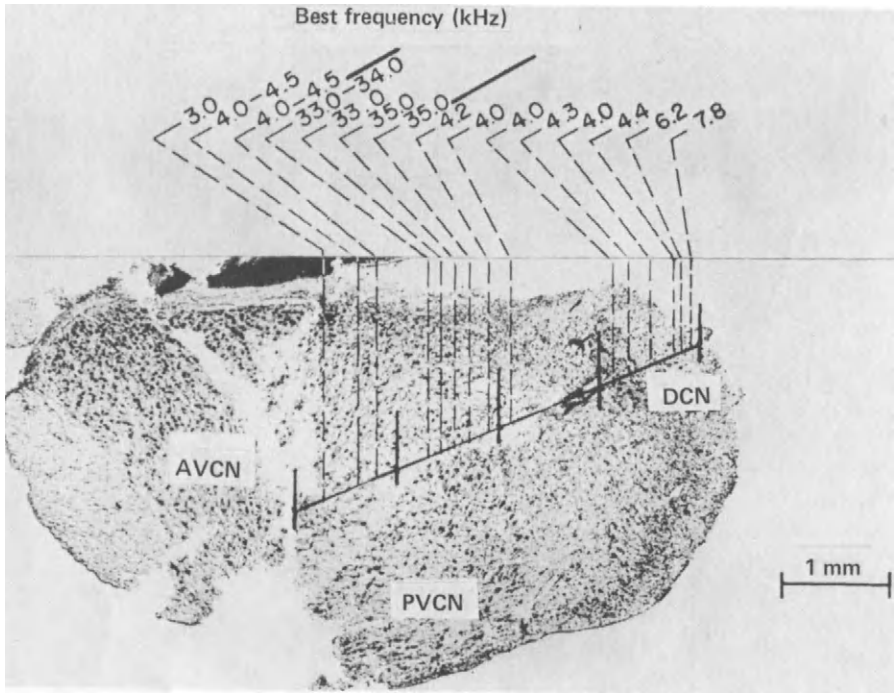

Figure 6.4 Tonotopic arrangement of anteroventral (AV), posteroventral (PV), and dorsal (DC) cochlear nuclei of a cat (saggital section on left side). Each nuclear group has its own sequence from low- to high-frequency representation. (From Rose, Galambos, and Hughes [9], Microelectrode studies of the cochlear nuclear nuclei of the cat, *Bull. Johns Hopkins Hosp. 104*, 211–251, ©1959, Johns Hopkins University Press.)

nations of the fibers from the apical turn and the dorsomedial insertions of the basal fibers were confirmed by Sando's [37] degeneration studies.

Experimenting on cats, Rose and colleagues [9,39] used microelectrodes to study the frequency sensitivity of the cochlear nuclei. They advanced the microelectrode through the cochlear nuclei of an anesthetized cat, and measured the number of neural discharges in response to tone pips of various frequencies. At a given intensity, a neuron responds with the greatest number of spikes at its characteristic frequency (CF). Rose et al. found that each of the three nuclear groups of the cochlear nuclear complex has a complete frequency mapping in a dorsoventral direction; i.e., low frequencies ventrally and high frequencies dorsally in each division of the cat's CN (Fig. 6.4).

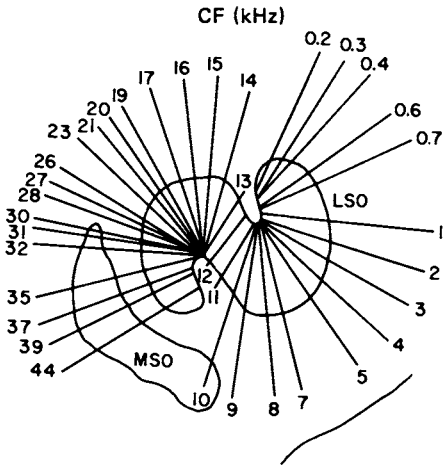

Figure 6.5 Tonotopic arrangement of the cells in the lateral superior olivary complex of the cat. Characteristic frequencies (kHz) taper from high ventromedially to low dorsomedially. (From Tsuchitani and Boudreau [40]. Permission *J. Neurophysiol.*)

Rose [39] hypothesized that the three complete representations of frequency in each part of the CN could be subserved by an arrangement in which each axon distributes to each of the nuclei. Sando's [37] findings are particularly interesting in this regard (Fig. 6.3). He found that eighth nerve axons bifurcate into an anterior branch to the VCN and a posterior branch to the DCN. Fibers from the basal turn distribute dorsomedially in the CN, while fibers with more apical origins in the cochlea distribute ventrolaterally. Furthermore, the apical fibers of the posterior branch were found more inferiorly in the DCN than the corresponding fibers of the anterior branch in the VCN. On the other hand, the basally derived fibers in the posterior branch are distributed more superiorly in the DCN than the corresponding anterior branch fibers in the VCN. Middle turn fibers in both branches terminate at similar levels in each nucleus. Thus, Rose's hypothesis appears to be borne out.

Tsuchitani and Boudreau [40] studied the responses of single neurons in the S-shaped lateral nucleus of the cat's superior olivary complex (SOC). This structure appears as a backward "S" (Fig. 6.5). The higher CFs were found in the curve lying downward toward the dorsomedial curve of the S. Goldberg and Brown [41] studied the distribution of CFs in the dog's medial SOC. By advancing a microelectrode

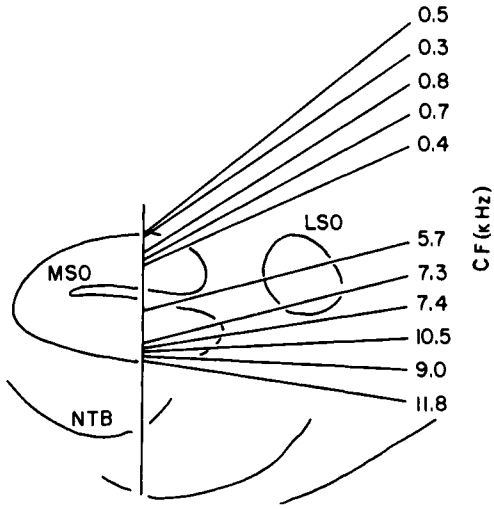

Figure 6.6 Distribution of neurons with high (ventral) and low (dorsal) characteristic frequencies in the medial superior olivary nucleus of the dog. (From Goldberg and Brown [41]. Permission *J. Neurophsiol.*)

through the U-shaped nucleus, they found that the neurons with higher CFs are in the ventral leg of the structure and those with lower CFs are in the dorsal leg (Fig. 6.6).

Aitkin et al. [42] found a tonotopic organization in both the ventral and dorsal nuclei of the cat's lateral lemniscus (LL). Neurons which have high CFs were found ventrally and those with low CFs were found dorsally in both nuclei. It is interesting to note that Ferraro and Minckler [43] found that the nuclei of the human LL are somewhat dispersed into scattered cell clusters among the lemniscal fibers, and that a clear-cut demarcation between them could not be found. However, the effect (if any) of this dispersed nuclear arrangement upon the tonotonicity of the human LL is not yet determined.

The tonotopic arrangement of the cat's inferior colliculus (IC) was studied by Rose et al. [6]. They inserted a microelectrode into the dorsomedial surface of the IC, and advanced it in a ventromedial direction (Fig. 6.7). The electrode thus passed through the external nucleus and into the central nucleus. Within the central nucleus, cells with increasing CFs were encountered as the electrode proceeded ventromedially into the nucleus. The opposite arrangement was observed in the external nucleus, i.e., the CFs went from high to low as the electrode was advanced ventromedially. However, the high-to-low frequency organi-

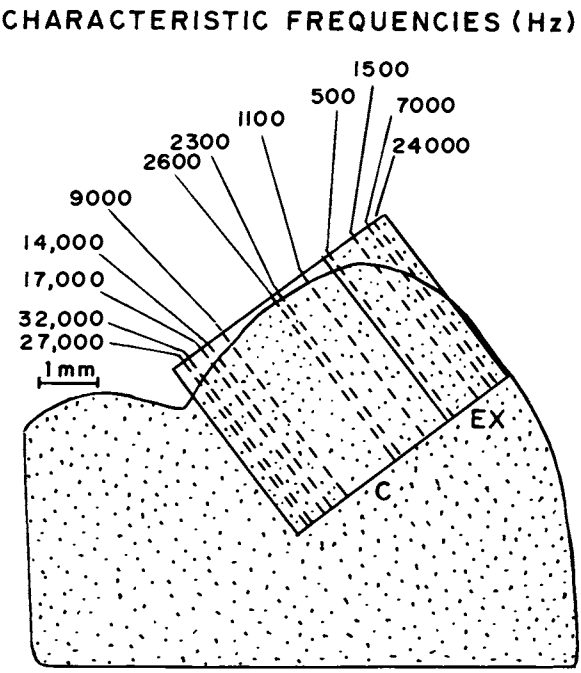

Figure 6.7 Tonotopic representation of characteristic frequencies in the central nucleus (C) and external nucleus (EX) of the inferior colliculus of the cat. (Based on Rose et al. [6].)

zation was found in only 9 out of 17 external nuclei. Thus, while there is most likely a tonotopic arrangement in the external nucleus, it is not as consistent as in the central nucleus of the IC.

The existence of a tonotopic organization in the medial geniculate body (MGB) was confirmed by Aitkin and Webster [44]. They found that lateral units in the pars principalis of the cat responded to low frequencies, while medial neurons were responsive to high. This was a significant finding because there was very little to support tonotonicity in the MGB prior to their report [45].

Tonotopic Organization of the Cortex

Woolsey and Walzl [46] measured the electrical potentials of the cat's cortex in response to stimulation of the auditory nerve. They found two auditory areas in each hemisphere, each with its own projection of the auditory nerve. A primary auditory area (AI) was identified in the mid-

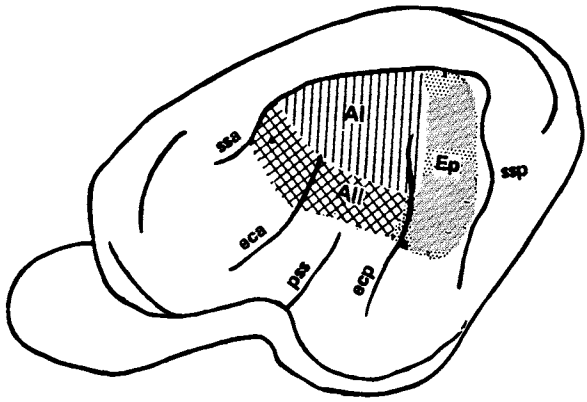

Figure 6.8 Cytoarchitecture (arrangement by cell types) of the auditory cortex of the cat (see text). (From Rose [48]. Permission The Wistar Press.)

dle ectosylvian area, and a secondary auditory area (AII) was found just ventral to AI. Stimulating the basal fibers of the auditory nerve resulted in responses in the rostral part of AI, whereas stimulation of fibers from the apical turn yielded responses from the caudal part. The opposite arrangement was observed in AII, i.e., the basal fibers were projected caudally and the apical fibers rostrally. This arrangement was essentially confirmed by Tunturi [47], who studied the cortical responses to tone burst stimuli in the dog. He later reported a third auditory area (AIII) under the anterior part of the suprasylvian gyrus.

Rose [48] described the cytoarchitecture (arrangement by cell types) of the cat's auditory cortex (Fig. 6.8). Rose identified AI as essentially triangular, with AII placed somewhat ventrally, and a posterior ectosylvian area (Ep) posteriorly. This result apparently contradicted Woolsey and Walzl's findings, since their electrophysiologic data had shown high-frequency responses from AII extending posteriorly to the place where this area ends anatomically [49]. This apparent discrepancy was resolved by the finding that both of these areas (AII and Ep) actually respond separately to apical stimulation, rather than just as one larger area (the original AII), as was previously supposed [50,51]. Furthermore, findings by Sindberg and Thompsen [52] suggested that the Ep also responds to low frequencies ventral to the part that responds to basal (high-frequency) stimulation. Several studies make it appear that the responses of AII and the Ep are mediated by the primary area, AI [51,53–55].

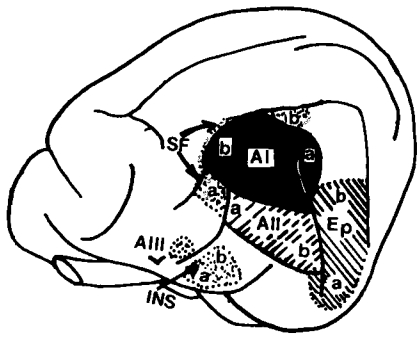

Figure 6.9 Summary of the tonotopic organization of the cortex. a indicates the projection of the apical part of the cochlea (low frequencies) on the cortex, and b indicates the basal projections (high frequencies). See text for details. [Adapted from Woolsey [51], in *Neural Mechanisms of the Auditory and Vestibular Systems* (G. L. Rasmussen and W. F. Windle, Eds.), 1960. Courtesy Charles C Thomas, Publisher, Springfield, Illinois.]

Woolsey [51] proposed a general summary of the tonotopic organization of the cortex based upon these and other experiments. This summary is shown in Fig. 6.9, and continues to be the accepted representation [56]. The figure shows complete representations of the cochlea in four cortical areas; as well as several other areas also responsive to auditory stimulation.

In AI, the tonotopic organization is from high frequencies rostrally in this area to low frequencies caudally. Area AII lies ventral to AI. Its frequency representation is less distinct than that of the primary area, and is also in the opposite direction. The ectosylvian area Ep has a frequency representation in which the basal part of the cochlea is projected above and the apical cochlea is projected below. Organized fiber connections between AI and the MGB have been observed in the cat [49,57] and monkey [58]. It appears that, while the MGB sends essential projections to AI, only sustaining projections are received by AII and the Ep [49,57,59].

The fourth cortical area with a complete projection of the cochlea is the sylvian fringe area [SF], which has a posterior representation of high frequencies and an anterior representation of low. Note in Fig. 6.9 that the SF lies on the surface on the anterior ectosylvian gyrus, sinks deep into the suprasylvian sulcus, and then reappears on the surface above the posterior ectosylvian sulcus. Woolsey [51] noted that the thalamic

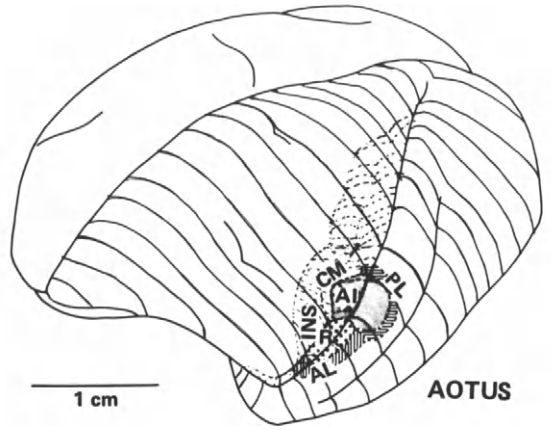

Figure 6.10 Primary (AI) and secondary auditory fields in the owl monkey *Aotus*, obtained by Imig et al. [63]. (CM, caudomedial field; PL, posterolateral field; AL, anterolateral field; R, rostrolateral field.) [From Brugge [56], in *The Nervous System, Vol. 3: Human Communication and Its Disorders* (D. B. Tower and E. L. Eagles, Eds.), ©1975, Raven Press, New York.]

connection of the SF is possibly with the pars principalis of the MGB, but that the specific relationship is not known.

Tunturi's AIII is situated in the head subdivision of somatic sensory area II, and is apparently connected with the posterior medial geniculate of the thalamus. Some degree of frequency representation has been identified in the insular area (INS) [60,61]. Woolsey [51] considers the thalamic connections of the INS to be undetermined.

Cortical tonotopic organization has also been demonstrated in primates. After removing the portions of parietal cortex covering this area, Merzenich and Brugge [62] exposed AI on the supratemporal gyrus of the monkey. Their electrophysiologic results showed that in this area high frequencies are represented caudomedially and lows rostrally. The secondary auditory cortex is made up of a belt of cortex around AI. As shown in Fig. 6.10 for the owl monkey [63], this belt includes a number of fields shown electrophysiologically to be responsive to auditory stimulation. Mesulam and Pandya [58] have demonstrated that the MGB sends a tonotopically arranged projection to the primary auditory cortex of the rhesus monkey.

Based upon findings that the somesthetic and visual cortices are organized in vertical columns, Abeles and Goldstein [64] undertook a study to determine whether the cat's primary auditory cortex is organized in columnar fashion, and whether it is organized by depth. They

reported that neurons narrowly tuned to the same or similar frequencies are found in clusters. They also found that the auditory cortex is not organized by depth; i.e., the cells in a given vertical column of cortex show the same CFs as a microelectrode is advanced deeper into the cortex. Thus, the tonotopic organization of AI has cells with essentially similar CFs arranged in vertical columns, and in bands horizontally. These isofrequency bands widen as stimulus intensity is increased [56]; i.e., as a stimulus gets more intense a wider band of cortex responds to it.

Most of the auditory areas of each cerebral hemisphere are connected by commissural pathways with their counterparts on the other side. This arrangement has been shown in both cats [65] and monkeys [66]. Also, most of the auditory areas on the same side are interconnected in both animals [50,66,67].

AVERAGED EVOKED AUDITORY POTENTIALS

Until now we have been primarily concerned with the responses of single cells, measured by inserting microelectrodes directly into the neural material. It is also possible to measure the neural responses evoked by sound stimulation by using electrodes placed on the surface of the brain or, more often, on the scalp. While this approach does not allow one to focus upon the activity of single neurons, it does permit the study of aggregate responses of various nuclei. The major advantage is that surface electrodes allow noninvasive study, and hence may readily be applied to diagnostic audiology. This section provides a brief description of these evoked responses.

Two problems are readily apparent when measuring evoked auditory responses. One is the very low intensity of any one response, and the other is the excessive noise in the form of ongoing neural activity. Both problems are overcome by obtaining an averaged response from many stimulus repetitions.

Suppose we measure the ongoing electrical response of the brain, or the electroencephalographic response (EEG). In the absence of stimulation, the EEG will be random at any particular moment. That is, if we randomly select and line up a large number of segments from an ongoing EEG, there will be about as many positive as negative values at the points being compared. The algebraic sum of these random values will be zero. Alternatively, suppose a group of cells fire at a given time (latency) after the onset of a click. If we always measure the activity at the same latency, then we would find that the responses at that point are always positive- (or negative-) going. Instead of averaging these responses out, algebraic summation will exaggerate them, since positives

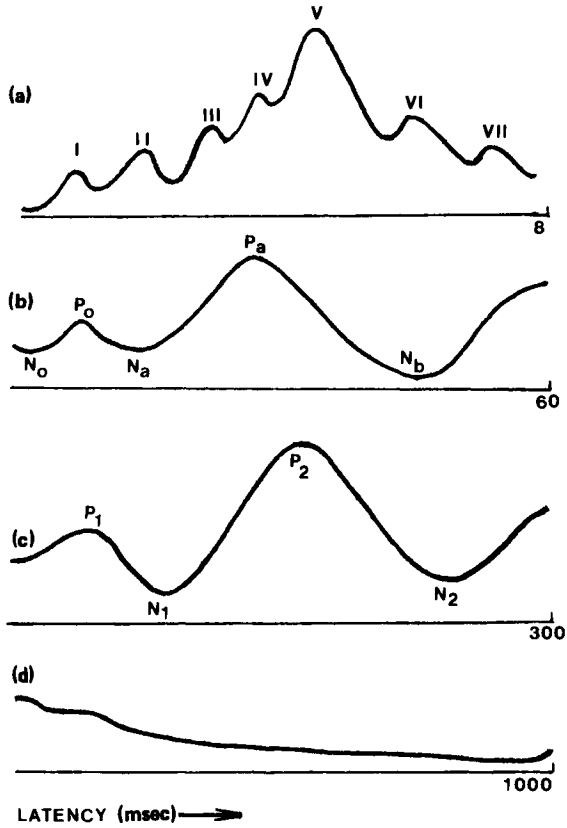

Figure 6.11 Idealized averaged auditory evoked responses: (a) brainstem response, (b) middle response, (c) late response, (d) contingent negative variation.

are always added to positives and negatives to negatives. Thus, the averaging (or summation) process improves the signal-to-noise ratio by averaging out the background activity (noise) and summating the real responses which are locked in time to the stimulus.

Auditory evoked responses are commonly described with regard to their response latencies. The earliest are the evoked brainstem responses (or "Jewett bumps") occurring within the first 8 msec of stimulus onset in both cats [68,69] and humans [70,71]. These appear as about seven successive peaks, the largest and most stable of which is peak V (Fig. 6.11a). The various peaks have been attributed on the basis of their latencies to successive nuclei in the auditory pathways [68,69]. Peak I is the negative compound action potential of the auditory nerve. The re-

maining peaks are positive. The second peak (II) is most likely due to activity of the cochlear nuclei, and possibly also to some double firings of the eighth nerve. Wave III has been attributed to the superior olive. The lateral lemniscus and/or inferior colliculus appear to elicit peak IV; and the inferior colliculus (and possibly higher centers) contribute to peak V. Peak VI is probably related to the medial geniculate. One should not arbitrarily attribute each peak to a single neural center, since Jewett and Williston have pointed out that the peaks beyond the first two almost definitely result from multiple generators [71].

The "middle response" (Fig. 6.11b) occurs at latencies between about 8 and 60 msec [72,73]. It appears that this response is neurogenic [73,74], but the locus of its origin is not apparent. R. Goldstein has been cited [74] as suggesting that the response may arise from the subcortical alerting (reticular) system.

The "late response" (Fig. 6.11c) is observed between roughly 70 and 300 msec [75–77], and has been the subject of extensive study. Although it is a diffuse response, there is reason to believe that it is derived from the primary auditory cortex [78,79].

The "contingent negative variation" (Fig. 6.11d) appears as a slow shift in the DC baseline after a latency of about 300 msec [80,81]. Unlike the other potentials, the contingent negative variation requires a mental or motor task. The response begins prefrontally and sweeps back toward the motor cortex.

The major value of these averaged auditory evoked potentials is in the audiological evaluation of clinical patients, an area beyond the current scope. The interested reader is referred to reviews by Hood [74], Reneau and Hnatiow [82], and Skinner and Glattke [83].

EFFECTS OF CORTICAL ABLATION

Which auditory functions are subserved at a cortical level? Equivalent to this question is the more testable one of which functions can still be accomplished (or relearned) after the auditory cortex has been removed, and which ones can be carried out only when the cortex is intact. This question has been investigated extensively in laboratory animals, particularly cats. Although the specific experimental paradigms have differed somewhat, the fundamental approach has been fairly consistent. This approach has been to bilaterally ablate the auditory cortex, and then to test the animal's ability to perform (or relearn) various sound discrimination tasks. The assumption is that discriminations unaffected by the ablations are subserved by subcortical and/or nonauditory cortical centers, whereas discriminations impaired after ablation must require

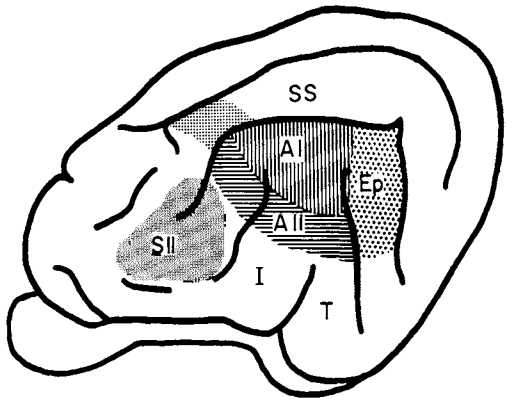

Figure 6.12 Cortical projection areas (in the cat) which were subjected to bilateral ablation in various sound discrimination studies: AI, auditory area I; AII, auditory area II; Ep, posterior ectosylvian area; I, insular area; T, temporal area; SSII, somatosensory area II; SS, suprasylvian area. [From Neff [84], in *Neural Mechanisms of the Auditory and Vestibular Systems* (G. L. Rasmussen and and W. F. Windle, Eds.), 1960. Courtesy of Charles C Thomas, Publisher, Springfield, Illinois.]

processing in the auditory cortices. Figure 6.12, by Neff [84], shows those auditory projection areas of the cat's cortex that have been ablated in various sound discrimination studies. Comparison with Fig. 6.9 reveals that these are essentially the same as the ones revealed by electrophysiologic and cytoarchitectural studies of the auditory cortex.

It is well established that certain sound discriminations can be accomplished after bilateral ablation of the auditory cortices, while others cannot. The bilaterally ablated animal *can* discriminate (1) the onset of a sound [85,86], (2) changes in tonal intensity [84,87,88], and (3) changes in the frequency of a tonal stimulus [86,89,90]. However, the type of frequency discrimination is critical, especially when the discrimination is between sequences of tones. Diamond et al. [91] presented bilaterally ablated animals with discrimination tasks between two sequences of three tones each. One sequence contained three low-frequency tones (LO-LO-LO), and the other group had two low tones separated by a higher frequency (LO-HI-LO). The LO-HI-LO sequence could be discriminated from the LO-LO-LO group, but the discrimination could no longer be performed when the task was reversed (i.e., when all low tones was the positive stimulus to be discriminated from the LO-HI-LO sequence).

Bilaterally ablated animals *cannot* discriminate (1) changes in tonal duration [92], (2) changes in the temporal pattern of a tonal sequence [93], or (3) sound localizations in space [94]. Masterton and Diamond [95] presented bilaterally ablated cats with a series of pairs of clicks separated by 0.5 msec, in which the first click of each pair went to one ear and the second click went to the other ear. If the first click is directed to the right ear (R-L) then the clicks are lateralized to the right; if the first click goes to the left ear (L-R) they are lateralized left (see Chap. 13). These workers found that bilaterally ablated cats could not discriminate the change from a series of L-R click pairs to R-L click pairs. In this regard, Neff [96] suggested that the auditory cortex plays an important role in the accurate localization of sounds in space.

Cats with bilateral ablations of the auditory cortex evince a startle response to sound, but do not exhibit the orientation response of reflexive head turning in the direction of the sound source found in normal cats [96]. Also, bilaterally ablated animals do appear to be able to push a response lever indicating whether a sound came from the right or left, but cannot approach the sound source (to get a reward) [97,98]. Neff and Casseday [99] suggested that the right–left distinction may be based upon different perceptual cues than spatial location. However, they explained that the auditory cortex is essential if the animal is to perceive spatial orientations and the relationships of its own body to the environment. In light of this fact, Strominger [100] found that localization is particularly mediated by area AI; and Sovijärvi and Hyvärinen [17] found that there are cells in the cat's primary auditory cortex that respond to the movement of a sound source in space.

There is also evidence that localization ability is impaired by unilateral ablation of the auditory cortex [101]. A recent experiment by Neff and Casseday [99] is especially interesting in this regard. They surgically destroyed one cochlea each in a group of cats, and then trained these unilaterally hearing animals to make sound localizations. Following this procedure, half of the animals were subjected to unilateral ablations of the auditory cortex opposite the good ear, and half were ablated ipsilateral to the good ear (opposite the deaf ear). Those with cortical ablations on the same side as the hearing ear were essentially unaffected in localization ability. The animals whose cortical ablations were contralateral to the good ear could not relearn the localization task. These findings demonstrate that localization behavior is affected by destruction of the auditory cortex opposite to a functioning ear. This result also suggests that excitation of the auditory cortex on one side serves as a cue that the opposite ear is stimulated. These results are consistent with single-cell data suggesting that contralateral stimulation is excitatory to cells in the auditory cortex.

Most studies on humans with central auditory lesions have employed speech tests. However, tests using nonverbal materials are particularly valuable since they are directly comparable with the results of animal studies [102]. Bilateral temporal lobe damage in humans has been shown to result in impaired ability to make temporal pattern discriminations [103–105]. Temporal lobe damage in humans has also been reported to result in impaired sound localization in space [98,103]. These results essentially confirm the animal studies. Clinical data also tend to support Neff and Casseday's [99] position that neural activity at the contralateral cortex provides a "sign" of the stimulated ear. For example, patients who underwent unilateral temporal lobectomies have reported localization difficulties on the side opposite to the removed cortex [106]; and electrical stimulation of the auditory cortex has resulted in hallucinations of sounds or hearing loss at the ear opposite to the stimulated cortex [107,108].

Neff [84,96] has proposed that in order for sound discriminations to be accomplished after bilateral ablation of the auditory cortex, it is necessary for the positive stimulus to excite neural units that were not excited by the previous (neutral) stimulus. Thus, stimulus onset, as well as intensity and frequency changes, are discriminated because new neural units are excited by the changes in each case. The same is true when the sequence LO-HI-LO is discriminated from LO-LO-LO. However, LO-LO-LO is not discriminated from LO-HI-LO because there is nothing new in the LO-LO-LO sequence that was not already present in LO-HI-LO; so that no new neural units are excited by the change. These functions may be processed below the level of the auditory cortex, or by other cortical areas. On the other hand, discriminations that involve the processing of serial orders or time sequences (i.e., discriminations in the temporal doma n) are subserved by the auditory cortex. Thus, they are obliterated when the auditory cortex is bilaterally removed.

REFERENCES

1. N. Y.-S. Kiang, R. R. Pfeiffer, W. B. Warr, and A. S. Backus, Stimulus coding in the cochlear nucleus, *Ann. Otol. 74*, 463–485 (1965).
2. J. D. Boudreau and C. C. Tsuchitani, Cat superior olivary S segment cell discharge to tonal stimulation, in *Contributions to Sensory Physiology* (W. D. Neff, Ed.), Academic, New York, 1970, pp. 143–213.
3. J. J. Guinan, B. E. Norris, and S. S. Guinan, Single auditory units in the superior olivary complex II: Locations of unit categories and tonotopic organization, *Intl. J. Neurosci. 4*, 147–166 (1972).
4. Y. Katsuki, Neural mechanisms of auditory sensation in cats, in *Sensory Communication* (W. A. Rosenblith, Ed.), MIT Press, Cambridge, Mass., 1961, pp. 561–583.

5. N. Y.-S. Kiang, D. K. Morest, D. A. Godfrey, J. J. Guinan, and E. C. Kane, Stimulus coding at caudal levels of the cat's auditory nervous system: I. Response characteristics of single units, in *Basic Mechanisms in Hearing* (A. R. Møller, Ed.), Academic, New York, 1973, pp. 455–478.
6. J. E. Rose, D. D. Greenwood, J. M. Goldberg, and J. E. Hind, Some discharge characteristics of single neurons in the inferior colliculus of the cat. I. Tonotopical organization, relation of spike-counts to tone intensity, and firing patterns of single elements, *J. Neurophysiol. 26*, 294–320 (1963).
7. J. E. Hind, Unit activity in the auditory cortex, in *Neural Mechanisms of the Auditory and Vestibular Systems* (G. L. Rasmussen and W. F. Windle, Eds.), Thomas, Springfield, Ill., 1960, pp. 201–210.
8. Y. Katsuki, N. Suga, and Y. Kanno, Neural mechanisms of the peripheral and central auditory systems in monkeys, *J. Acoust. Soc. Amer. 32*, 1396–1410 (1962).
9. J. E. Rose, R. Galambos, and J. R. Hughes, Microelectrode studies of the cochlear nuclei of the cat, *Bull. Johns Jopkins Hosp. 104*, 211–251 (1959).
10. R. Galambos, J. Schwartzkopff, and A. L. Rupert, Microelectrode study of superior olivary nuclei, *Amer. J. Physiol. 197*, 527–536 (1959).
11. A. L. Rupert, G. M. Moushegian, and M. A. Whitcomb, Superior olivary complex response patterns to monaural and binaural clicks, *J. Acoust. Soc. Amer. 39*, 1069–1076 (1966).
12. R. Galambos, Microelectrode studies on medial geniculate body of cat. III. Response to pure tones, *J. Neurophysiol. 15*, 381–400 (1952).
13. Y. Katsuki, T. Watanabe, and N. Maruyama, Activity of auditory neurons in upper levels of brain of cat, *J. Neurophysiol. 22*, 343–359 (1959).
14. L. M. Aitkin, C. W. Dunlop, and W. R. Webster, Click evoked response patterns in single units in the medial geniculate body of the cat, *J. Neurophysiol. 29*, 109–123 (1966).
15. J. F. Brugge and M. M. Merzenich, Patterns of activity of single neurons of the auditory cortex in monkey, in *Basic Mechanisms in Hearing* (A. R. Møller, Ed.), Academic, New York, 1973.
16. I. C. Whitfield and E. F. Evans, Responses of auditory cortical neurons to stimuli of changing frequency, *J. Neurophysiol. 28*, 655–672 (1965).
17. A. R. A. Sovijärvi and J. Hyvärinen, Auditory cortical neurons in the cat sensitive to the direction of sound source movement, *Brain Res. 73*, 455–471 (1974).
18. W. A. Stotler, An experimental study of the cells and connections of the superior olivary complex of the cat, *J. Comp. Neurol. 98*, 401–431 (1953).
19. J. L. Hall, Binaural interaction in the accessory superior olivary nucleus of the cat: An electrophysiological study of single neurons. *MIT Res. Lab. Electron*, Tech. Rept. 416, Cambridge, Mass, 1964.
20. W. A. van Bergeijk, Variation on a theory of Bekesy: A model of binaural interaction, *J. Acoust. Soc. Amer. 34*, 1431–1437 (1962).
21. G. M. Moushegian, A. L. Rupert, and M. A. Whitcomb, Medial-superior-olivary unit response patterns to monaural and binaural clicks, *J. Acoust. Soc. Amer. 36*, 196–202 (1964).

22. G. M. Moushegian, A. L. Rupert, and M. A. Whitcomb, Brainstem neuronal response patterns to monaural and binaural tones, *J. Neurophysiol. 27*, 1174–1191 (1964).
23. G. M. Moushegian, A. L. Rupert, and M. A. Whitcomb, Processing of auditory information by medial superior-olivary neurons, in *Foundations of Modern Auditory Theory* (J. V. Tobias, Ed.), Vol. 2, Academic, New York, 1972, pp. 263–299.
24. J. M. Goldberg and P. B. Brown, Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization, *J. Neurophysiol. 32, 613*–636 (1969).
25. G. M. Mousheigian, A. L. Rupert, and T. L. Langford, Stimulus coding by medial superior olivary neurons, *J. Neurophysiol. 30, 1239*–1261 (1967).
26. J. E. Rose, N. B. Gross, C. D. Geisler, and J. E. Hind, Some neural mechanisms in the inferior colliculus of the cat which may be relevant to localization of a sound source, *J. Neurophysiol. 29*, 288–314 (1966).
27. C. D. Geisler, W. S. Rhode, and D. W. Hazelton, Responses of inferior colliculus neurons in the cat to binaural acoustic stimuli having wide-band spectra, *J. Neurophysiol. 32*, 960–974 (1969).
28. L. A. Benevento and P. D. Coleman, Responses of single cells in cat inferior colliculus to binaural click stimuli: Combinations of intensity levels, time differences and intensity differences, *Brain Res. 17*, 387–405 (1970).
29. F. Flammino and B. M. Clopton, Neural response in the inferior colliculus of albino rat to binaural stimuli, *J. Acoust. Soc. Amer. 57*, 692–695 (1975).
30. J. A. Altman, Are there neurons detecting direction of sound source motion? *Exp. Neurol. 22*, 13–25 (1968).
31. J. F. Brugge, D. J. Anderson, and L. M. Aitkin, Responses of neurons in the dorsal nucleus of the lateral lemniscus of cat to binaural tonal stimulation, *J. Neurophysiol. 33*, 441–458 (1970).
32. L. M. Aitkin and C. W. Dunlop, Interplay of excitation and inhibition in the cat medial geniculate body, *J. Neurophysiol. 31*, 44–61 (1968).
33. J. A. Altman, J. Syka, and G. N. Shmigidina, Neuronal activity in the medial geniculate body of the cat during monaural and binaural stimulation, *Exp. Brain Res. 10*, 81–93 (1970).
34. J. F. Brugge, N. A. Dubrovsky, L. M. Aitkin, and D. J. Anderson, Sensitivity of single neurons in auditory cortex of cat to binaural tonal stimulation: Effects of varying interaural time and intensity, *J. Neurophysiol. 32*, 1005–1024 (1969).
35. J. L. Hall and M. H. Goldstein, Representation of binaural stimuli by single units in primary auditory cortex of unanesthesized cats, *J. Acoust. Soc. Amer. 43*, 456–461 (1968).
36. J. E. Hirsch, Effect of interaural time delay on amplitude of cortical responses evoked by tones, *J. Neurophysiol. 31*, 916–927 (1968).
37. I. Sando, The anatomical interrelationships of the cochlear nerve fibers, *Acta Otol. 59*, 417–436 (1965).
38. F. H. Lewy and H. Kobrak, The neural projection of the cochlear spirals of primary acoustic centers, *Arch. Neurol. Psychiat. 35*, 839–852 (1936).

39. J. E. Rose, Organization of frequency sensitive neurons in the cochlear nuclear complex of the cat, in *Neural Mechanisms of the Auditory and Vestibular Systems* (G. L. Rasmussen and W. F. Windle, Eds.), Thomas, Springfield, Ill., 1960, pp. 116–136.
40. C. Tsuchitani and J. D. Boudreau, Single unit analysis of cat superior olive S segment with tonal stimuli, *J. Neurophysiol. 29*, 684–697 (1966).
41. J. M. Goldberg and P. B. Brown, Functional organization of the dog superior olivary complex: An anatomical and physiological study, *J. Neurophysiol. 31*, 639–656 (1968).
42. L. M. Aitkin, D. J. Anderson, and J. F. Brugge, Tonotopic organization and discharge characteristics of single neurons in nuclei of the lateral lemniscus of the cat, *J. Neurophysiol. 33*, 421–440 (1970).
43. J. A. Ferraro and J. Minckler, The human auditory pathways: A quantitative study: The human lateral lemniscus and its nuclei, *Brain Language 4*, 277–294 (1977).
44. L. M. Aitkin and W. R. Webster, Tonotopic organization in the medial geniculate body of the cat, *Brain Res. 26*, 402–405 (1971).
45. I. C. Whitfield, *The Auditory Pathway*, Arnold, London, 1967.
46. C. N. Woolsey and E. M. Walzl, Topical projection of nerve fibers from local regions of the cochlea to the cerebral cortex of the cat, *Bull. Johns Hopkins Hosp. 71*, 315–344 (1942).
47. A. R. Tunturi, Audio frequency localization in the acoustic cortex of the dog, *Amer. J. Physiol. 141*, 397–403 (1944).
48. J. E. Rose, The cellular structure of the auditory region of the cat, *J. Comp. Neurol. 91*, 409–439 (1949).
49. J. E. Rose and C. N. Woolsey, The relations of thalamic connections, cellular structure and evokable electrical activity in the auditory region of the cat, *J. Comp. Neurol. 91*, 441–466 (1949).
50. C. B. Downman, C. N. Woolsey, and R. A. Lende, Auditory areas I, II and Ep: Cochlear representation, afferent paths and interconnections, *Bull. Johns Hopkins Hosp. 106*, 127–146 (1960).
51. C. N. Woolsey, Organization of cortical auditory system: A review and a synthesis, in *Neural Mechanisms of the Auditory and Vestibular Systems* (G. L. Rasmussen and W. F. Windle, Eds.), Thomas, Springfield, Ill., 1960, pp. 165–180.
52. R. M. Sindberg and R. F. Thompsen, Auditory responses in the ventral temporal cortex of cat, *Physiologist 2*, 108–109 (1959).
53. A. H. Ades, A secondary acoustic area in the cerebral cortex of the cat, *J. Neurophysiol. 6*, 59–63 (1943).
54. A. H. Ades, Functional relationships between the middle and posterior ectosylvian areas in the cat, *Amer. J. Physiol. 159*, 561 (1949).
55. J. L. Arteta, V. Bonnet, and F. Bremer, Répercussions corticales de la résponse d'aire acoustique primaire: L'aire acoustique secondaire, *Arch. Intl. Physiol. 57*, 425–428 (1950).
56. J. F. Brugge, Progress in neuroanatomy and neurophysiology of auditory cortex, in *The Nervous System* (D. B. Tower, Ed.), Vol. 3: *Human Communication and Its Disorders*, Raven, New York, 1975, pp. 97–111.

57. J. E. Rose and C. N. Woolsey, Cortical connections and functional organization of the thalamic auditory system of the cat, in *Biological and Biochemical Bases of Behavior* (H. F. Harlow and C. N. Woolsey, Eds.), University of Wisconsin Press, Madison, Wisc., 1958, pp. 127–150.
58. M. M. Mesulam and D. N. Pandya, The projections of the medial geniculate complex within the Sylvian fissure of the rhesus monkey, *Brain Res. 60,* 315–333 (1973).
59. I. T. Diamond, K. L. Chow, and W. D. Neff, Degeneration of caudal medial geniculate body following cortical lesion ventral to auditory area II in the cat, *J. Comp. Neurol. 109,* 349–362 (1958).
60. J. D. Loeffler, An investigation of auditory responses in insular cortex of cat and dog, Thesis, University of Wisconsin, Madison, Wisc., (1958).
61. J. E. Desmedt and K. Michelse, Corticofugal projections from temporal lobe in cat and their possible role in acoustic discrimination, *J. Physiol. (London) 147,* 17–18 (1959).
62. M. M. Merzenich and J. F. Brugge, Representation of the cochlear partition on the superior temporal plane of the macaque monkey, *Brain Res. 50,* 275–296 (1973).
63. T. J. Imig, M. A. Ruggero, L. M. Kitzes, and J. F. Brugge, Organization of auditory cortex in the owl monkey (*Aotus trivirgatus*), *J. Acoust. Soc. Amer. 56,* S523 (1974).
64. M. Abeles and M. H. Goldstein, Functional architecture in cat primary auditory cortex: Columnar organization and organization according to depth, *J. Neurophysiol. 33,* 172–187 (1970).
65. I. T. Diamond, E. G. Jones, and T. P. S. Powell, Interhemispheric fiber connections of the auditory cortex of the cat, *Brain Res. 11,* 177–193 (1968).
66. D. N. Pandya, M. Hallett, and S. K. Mukherjee, Intra- and interhemispheric connections of the neocortical auditory system in the rhesus monkey, *Brain Res. 14,* 49–65 (1969).
67. I. T. Diamond, E. G. Jones, and T. P. S. Powell, The association connections of the auditory cortex of the cat, *Brain Res. 11,* 560–579 (1968).
68. D. L. Jewett, Volume-conducted potentials in response to auditory stimuli as detected by averaging in the cat, *EEG Clin. Neurophysiol. 28,* 609–618 (1970).
69. J. S. Buchwald and C.-M. Huang, Far-field acoustic response: Origins in the cat, *Science 189,* 382–384 (1975).
70. D. L. Jewett, M. N. Romano, and J. S. Williston, Human auditory evoked potentials: Possible brainstem components detected on the scalp, *Science 167,* 1517–1518 (1970).
71. D. L. Jewett and J. S. Williston, Auditory-evoked far fields averaged from the scalp of humans, *Brain 94,* 681–696 (1971).
72. C. Geisler, L. Frishkopf, and W. Rosenblith, Extracranial responses to acoustic clicks in man, *Science 128,* 1210–1211 (1958).
73. H. Ruhm, E. Walker, and H. Flanigin, Acoustically evoked potentials in man: Mediation of early components, *Laryngoscope 77,* 806–822 (1967).
74. D. C. Hood, Evoked cortical response audiometry, in *Physiological Measures of the Audio-Vestibular System* (L. J. Bradford, Ed.), Academic, New York, 1975, pp. 349–370.

75. H. Davis, Some properties of the slow cortical response in humans, *Science 146*, 434 (1964).
76. G. McCandless and L. Best, Evoked responses to auditory stimulation in man using a summating computer, *J. Speech Hearing Res. 7*, 193–202 (1964).
77. G. McCandless and L. Best, Summed evoked responses using pure tone stimuli, *J. Speech Hearing Res. 9*, 266–272 (1966).
78. H. G. Vaughan, The relationship of brain activity to scalp recordings of event-related potentials, in *Average Evoked Potentials: Methods, Results and Evaluation* (E. Donchin and D. Lindsay, Eds.), NASA, SP191, Washington, D. C., 1969.
79. H. G. Vaughan and W. Ritler, The sources of auditory evoked responses recorded from the human scalp, *EEG Clin. Neurophysiol. 28*, 360–367 (1970).
80. W. G. Walter, The convergence and interaction of visual, auditory and tactile responses in human nonspecific cortex, *Ann. N.Y. Acad. Sci. 112*, 320–361 (1964).
81. W. G. Walter, Effects on anterior brain responses of an expected association between stimuli, *J. Psychosomat. Res. 9*, 45–49 (1965).
82. J. P. Reneau and G. Z. Hnatiow, *Evoked Response Audiometry: A Topical and Historical Review*, Univ. Park Press, Baltimore, Md., 1975.
83. P. Skinner and T. J. Glattke, Electrophysiologic response audiometry: State of the art, *J. Speech Hearing Dis. 42*, 179–198 (1977).
84. W. D. Neff, Role of the auditory cortex in sound discrimination, in *Neural Mechanisms of the Auditory and Vestibular Systems* (G. L. Rasmussen and W. F. Windle, Eds.), Thomas, Springfield, Ill. 1960, pp. 211–216.
85. K. D. Kryter and H. W. Ades, Studies on the function of the higher acoustic nerve centers, *Amer. J. Psych. 56*, 501–536 (1943).
86. D. R. Meyer and C. N. Woolsey, Effects of localized cortical destruction on auditory discriminative conditioning in cat, *J. Neurophysiol. 15*, 149–162 (1952).
87. D. H. Raab and H. W. Ades, Cortical and midbrain mediation of a conditioned discrimination of acoustic intensities, *Amer. J. Psych. 59*, 59–83 (1946).
88. M. R. Rosenzweig, Discrimination of auditory intensities in the cat, *Amer. J. Psych. 59*, 127–136 (1946).
89. R. A. Butler, I. T. Diamond, and W. D. Neff, Role of auditory cortex in discrimination of changes in frequency, *J. Neurophysiol. 20*, 108–120 (1957).
90. J. M. Goldberg and W. D. Neff, Frequency discrimination after bilateral ablation of auditory cortical areas, *J. Neurophysiol. 24*, 119–128 (1961).
91. I. T. Diamond, J. M. Goldberg, and W. D. Neff, Tonal discrimination after ablation of auditory cortex, *J. Neurophysiol. 25*, 223–235 (1962).
92. D. P. Scharlock, W. D. Neff, and N. L. Strominger, Discrimination of tonal duration after ablation of cortical auditory areas, *J. Neurophysiol. 28*, 673–681 (1965).
93. I. T. Diamond and W. D. Neff, Ablation of temporal cortex and discrimination of auditory patterns, *J. Neurophysiol. 20*, 300–315 (1957).

94. W. D. Neff, J. F. Fisher, I. T. Diamond, and M. Yela, Role of auditory cortex in discrimination requiring localization of sound in space, *J. Neurophysiol. 19*, 500–512 (1956).
95. R. B. Masterton and I. T. Diamond, Effects of auditory cortex ablation on discrimination of small binaural time differences, *J. Neurophysiol. 27*, 15–36 (1964).
96. W. D. Neff, Auditory discriminations affected by cortical ablations, in *Sensorineural Hearing Processes and Disorders* (A. B. Graham, Ed.), Little, Brown, Boston, Mass., 1967, pp. 201–210.
97. H. E. Heffner and R. B. Masterton, The contributions of auditory cortex to sound localization in the monkey, *J. Neurophysiol. 38*, 1340–1358 (1975).
98. W. D. Neff, I. T. Diamond, and J. H. Casseday, Behavioral studies of auditory discrimination: Central nervous system, in *Handbook of Sensory Physiology* (W. D. Keiday and W. D. Neff, Eds.), Vol. 5: *Auditory System*, Springer-Verlag, New York, 1975, pp. 307–400.
99. W. D. Neff and J. H. Casseday, Effects of unilateral ablation of auditory cortex on monaural cat's ability to localize sounds, *J. Neurophysiol 40*, 44–52 (1977).
100. N. L. Strominger, Subdivisions of auditory cortex and their role in localization of sound in space, *Exp. Neurol. 24*, 348–362 (1969).
101. I. C. Whitfield, J. Cranford, R. Ravizza, and I. T. Diamond, Effects of unilateral ablation of auditory cortex in cat on complex sound localization. *J. Neurophysiol. 35*, 718–731 (1972).
102. W. D. Neff, The brain and hearing: Auditory discriminations affected by brain lesions, *Ann. Otol. 86*, 500–506 (1977).
103. J. Jerger et al., Bilateral lesions of the temporal lobe, *Acta Otol.*, Suppl. 258 (1969).
104. F. Lhermitte et al., Etude de troubles perceptifs auditifs dans les lesions temporales bilaterales, *Rev. Neurol. 124*, 329–351 (1971).
105. T. A. Karaseva, The role of the temporal lobe in human auditory perception. *Neuropsychology 10*, 227–231 (1972).
106. W. Penfield and J. Evans, Functional deficits produced by cerebral lobectomies, *A. Res. Nerv. Ment. Dis. Proc. 13*, 352–377 (1934).
107. W. Penfield and T. Rasmussen, *The Cerebral Cortex: A Clinical Study of Localization of Function*, MacMillan, New York, 1950.
108. W. Penfield and H. Jasper, Epilepsy and the Functional Anatomy of the Human Brain, Little, Brown, Boston, 1954.

7
Psychoacoustic Methods

Psychophysics is concerned with how we perceive the physical stimuli impinging upon our senses. The branch of psychophysics that deals with the perception of sound is psychoacoustics. In defining this term we make a sharp distinction between the physical stimulus and the psychological response to it. We may think of the sound presented to our ears as the stimulus and of what we hear as the response. For example, what we hear as loudness is the perceptual correlate of intensity. Other things being equal, a rise in intensity is perceived as an increase in loudness. Pitch corresponds to sound frequency: Other things being equal, pitch gets higher as frequency increases.

If there were a simple one-to-one correspondence between the physical parameters of sound and how they are perceived, then we could quantify what we hear directly in terms of the attributes of the sound. That would mean that all physically existing sounds could be heard, that all changes in them would be discriminable, and that any change in stimulus magnitude would result in a perceptual change of the same magnitude. This is not the case. It is thus necessary to describe the manner in which sound is perceived, and to attempt to explain the underlying mechanisms of the auditory process. This is the province of psychoacoustics.

MEASUREMENT METHODS

Establishing relationships between the sound presented and how the subject perceives it is a primary goal. To accomplish this goal, the experi-

menter contrives a special situation designed to home in on the relation of interest. An experimental situation is used to avoid the ambiguities of presenting a stimulus and, in effect, asking the open-ended question "what did you hear?" Instead, the stimulus and response are clearly specified, and then some aspect of the stimulus (intensity, frequency, etc.) is manipulated. The subject's task is to respond in a predetermined manner so that the experimenter can get an unambiguous idea of what was heard. For example, one may vary the intensity of a tone and ask the subject whether it was heard during each presentation. The lowest level at which the sound is heard (the transition between audibility and inaudibility) might be considered an estimate of absolute sensitivity. Alternatively, two tones might be presented, one of which is varied in frequency. The subject is asked whether the varied tone is higher (or lower) in pitch, and the smallest perceivable frequency difference — the just noticeable difference (jnd) — might be considered an estimate of differential sensitivity.

We must also distinguish between what the subject actually hears and the manner in which he responds. The former is *sensory capability* (or *sensitivity*) and the latter is *response proclivity*. We are interested in sensory capability. Response proclivity reflects not only the subject's sensitivity, but also the biases and criteria that affect how he responds. We therefore try to select measurement methods and techniques which minimize the effects of response bias. An excellent discussion of the many details to be considered in psychacoustic experiments is given in Robinson and Watson [1]. In this chapter, we shall be concerned with classical psychophysical methods, adaptive techniques, and some aspects of scaling. Chapter 8 covers signal detection.

Classical Methods

There are three classical psychophysical methods: limits, adjustment, and constant stimuli. Each has its individual advantages and disadvantages, as well as sharing some pros and cons with the other methods.

Method of Limits

In the method of limits, the stimulus is under the experimenter's control and the subject simply responds after each presentation. Suppose we are interested in the absolute sensitivity (threshold) for a particular sound. The sound is presented at a level expected to be well above threshold. Since it is clearly audible, the subject responds by saying that he heard the sound (+ in Fig. 7.1). The level of the sound is then decreased by a discrete amount (2 db in Fig. 7.1) and presented again. This process is repeated until the subject no longer perceives the

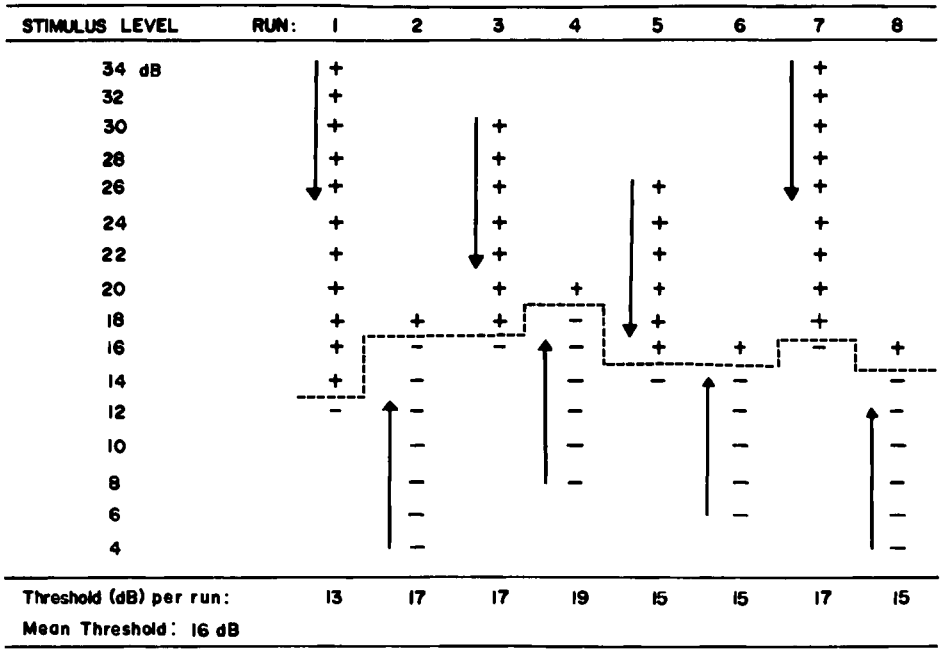

Figure 7.1 Method of limits.

sound (−), at which point the series (or run) is terminated. This example involves a descending run. In an ascending series, the sound is first presented at a level known to be below the threshold and is increased in magnitude until a positive (+) response is obtained. The odd-numbered runs in Fig. 7.1 are descending series and the even-numbered runs are ascending. Since the crossover between "hearing" and "not hearing" lies somewhere between the lowest audible level and the highest inaudible one, the "threshold" for each series may be taken as the halfway point between them. The subject's threshold is obtained by averaging the threshold levels across runs. This average is 16 dB for the data in Fig. 7.1.

Several forms of response bias are associated with the method of limits. Since a series either ascends or descends, and is terminated by a change in response, the subject may anticipate the level at which his response should change from "no" to "yes" in an ascending run and from "yes" to "no" in a descending series. Anticipation thus results in a lower (better) ascending threshold because the subject anticipates hearing the stimulus, and a higher (poorer) descending threshold since he

anticipates not hearing it. An opposite affect is caused by habituation. Here, the subject does not change his response from "no" to "yes" during an ascending run until the actual threshold is exceeded by a few trials (raising the measured threshold level); and he continues to respond "yes" for one or more descending trials after the sound has actually become inaudible (lowering the measured threshold level). These biases may be minimized by using an equal number of ascending and descending test runs in each threshold determination. These runs may be presented alternatively (as in the figure) or randomly. A second way to minimize these biases is to vary the starting levels for the runs. Both tactics are illustrated in Fig. 7.1.

The method of limits is also limited in terms of step size and inefficiently placed trials. Too large a step size reduces accuracy because the actual threshold may lie anywhere between two discrete stimulus levels. For example, a 10 dB step is far less precise than a 2 dB increment; and the larger the increment between the steps, the more approximate the result. Too large a step size may place the highest inaudible presentation at a level with a 0% probability of response, and the lowest audible presentation at a level with a 100% probability of response. The 50% point (threshold) may be *anywhere* between them! To make this point clear, consider the psychometric functions in Fig. 7.2. A psychometric function shows the probability (percentage) of responses for different stimulus levels. Figure 7.2a shows the psychometric function for a particular sound. It is inaudible (0% responses) at 13 dB and is always heard (100% responses) at 21 dB. It is customary to define the threshold as the level at which the sound is heard 50% of the time (0.5 probability). The threshold in Fig. 7.2a is thus 17 dB. Suppose we try to find this threshold using a 10 dB step size, with increments corresponding to 14 dB, 24 dB, etc. Notice that this step size essentially includes the whole psychometric function, so that we do not know where the responses change from 0% to 100%; nor do we know whether they do so in a rapid jump (a step function) or along a function where gradual changes in the proportion of "yes" responses correspond to gradual changes in stimulus level. The result is low precision in estimating the location of the 50% point. However, a large step size is convenient in that it involves fewer presentations (and thus shorter test time), since responses go from "yes" to "no" in very few trials, each of which is either well above or well below threshold.

A smaller step size permits a more precise estimate of threshold because the reversals from "yes" to "no" (and vice versa) are better placed (closer) in relation to the 50% point. The relationship of a 2 dB step to the psychometric function is shown in Fig. 7.2c, which gives the probability of a response in 2 dB intervals. Notice that these points are

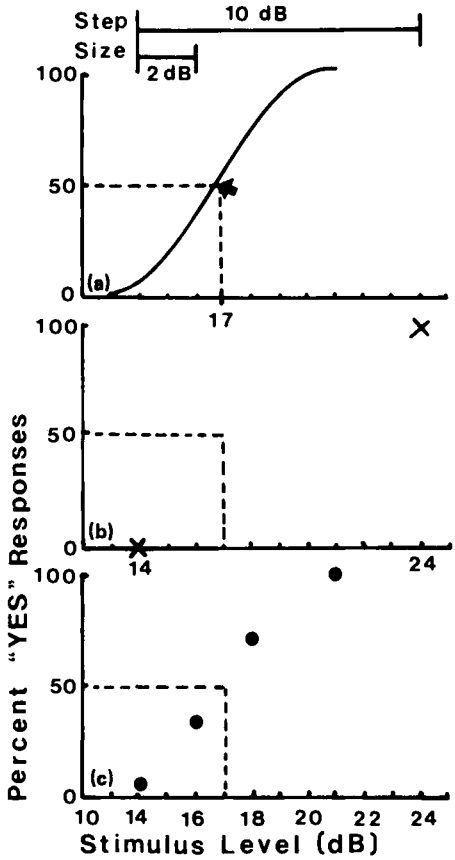

Figure 7.2 (a) Psychometric function showing 50% threshold at 17 dB (arrow). (b) Responses (x's) of 0% at 14 dB and 100% at 24 dB with a 10 dB step size. (c) Percent responses at various levels with a 2 dB step. The 50% threshold is shown on each graph. The 2 dB and 10 dB step sizes are illustrated at the top of the figure.

better placed than those for the 10 dB step size in Fig. 7.2b. For this reason, even though there may be "wasted" presentations due to test levels well above or below the threshold, the method of limits with an appropriate step size is still popular. This is particularly true in pilot experiments and in clinical evaluations, both of which take advantage of the speed with which thresholds are estimated by the method of limits. The clinical method of limits [2], however, is actually a hybrid technique with characteristics of the staircase method, discussed below.

Table 7.1 Results of Method of Limits in a Hypothetical Discrimination Experiment

	Run			
Stimulus level (dB)	1	2	3	4
60	Louder			
59	Louder		Louder	
58	Louder		Louder	Louder
57	Equal	Louder	Louder	Equal
56	Equal	Equal	Equal	Equal
55	Equal	Equal	Equal	Equal
54	Equal	Equal	Equal	Softer
53	Softer	Equal	Softer	Softer
52		Softer		Softer
51		Softer		Softer
50		Softer		

The method of limits may also be used to determine differential thresholds. In this case, two stimuli are presented in each trial, and the subject is asked whether the second is greater than, less than, or equal to the first with respect to some parameter. The first stimulus is held constant, and the second is varied by the experimenter in discrete steps. The procedure is otherwise the same as for determining thresholds, although the findings are different. Suppose the subject is to make an equal loudness judgment. The method of limits would result in a range of intensities in which the second stimulus is louder than the first, a range in which the second is softer, and a range in which the two sounds appear equal. In Table 7.1 the averge upper limen (halfway between "higher" and "equal") is 57 dB, and the average lower limen (between "equal" and "lower") is 53.5 dB. The range between these values is an interval of uncertainty that is 3.5 dB wide. Although there is a range of "equal" judgments, we may estimate the "equal" level to lie halfway between the upper and lower limens, at 55.25 dB. This level is commonly referred to as the point of subjective equality (PSE). The jnd or difference limen (DL) is generally estimated as one-half of the uncertainty interval, or 1.75 dB for the data in Table 7.1

Method of Adjustment

The adjustment method differs from the method of limits in two ways. First, the stimulus is controlled by the subject instead of by the experimenter. In addition, the level of the stimulus is varied con-

tinuously rather than in discrete steps. As in the method of limits, the level is adjusted downward from above threshold until it is just inaudible, or increased from below threshold until it is just audible. Threshold is taken as the average of the just audible and just inaudible levels. To obtain an estimate of differential sensitivity, the subject adjusts the level of one sound until it is as loud as a standard sound, or adjusts the frequency of one sound until it has the same pitch as the other.

The stimulus control (generally a continuous dial) must be unlabeled and should have no detents that might provide tactile cues that could bias the results. Furthermore, it is common practice to insert a second control between the subject's dial and the instrumentation, allowing the experimenter to vary the starting point of a test series by an amount unknown to the subject. This procedure avoids biases due to positioning of the response dial and to the use of dial settings as "anchors" from one series to the next. Even with these precautions, however, it is difficult for the experimenter to exercise the same degree of control over the procedure as in the method of limits. Furthermore, the subject may change his criterion of audibility during test runs, introducing another hard-to-control bias into the method of adjustment.

Just as anticipation and habituation affect the results obtained with the method of limits, stimulus persistence (preseveration) biases the results from the method of adjustment. Persistence of the stimulus means that a lower threshold is obtained on a descending run because the subject continues to turn the level down below threshold as though the sound were still audible. Thus, we may think of this phenomenon as persistence of the stimulus, or as perseveration of the response. In an ascending trial, the absence of audibility persists, so that the subject keeps turning the level up until the true threshold is passed by some amount, which has the opposite effect of raising the measured threshold level. These biases may be minimized by using both ascending and descending series in each measurement. Another variation is to have the subject "bracket" his threshold by varying the level up and down until a just audible sound is perceived. After the ending point is recorded, the experimenter may use the second stimulus control discussed above to change the starting level by an amount unknown to the subject, in preparation for the next trial.

Method of Constant Stimuli

The method of constant stimuli (or constants) involves the presentation of various stimulus levels to the subject in random order. Unlike the methods of limits and adjustments, the method of constants is a nonsequential procedure. In other words, the stimuli are not presented in an ascending or descending manner. A range of intensities is selected

Table 7.2 Threshold of a Tone Using the Method of Constant Stimuli

Stimulus level (dB)	Number of responses	Percent of responses
11	50	100
10	50	100
9	47	94
8	35	70
7	17	34
6	3	6
5	0	0
4	0	0

which, based upon previous experience or a pilot experiment, encompasses the threshold level. A step size is selected, and the stimuli are then presented to the subject in random order. During the experiment, an equal number of stimuli are presented at each level. The subject states whether there was a stimulus presentation during each test trial. (In a differential sensitivity experiment, the task would be to say whether the two items are same or different.)

In an experiment to determine the threshold for a tone by using the method of constant stimuli, one might randomly present tones in 1 dB increments between 4 dB and 11 dB, for a total of 50 trials at each level. Sample results are tabulated in Table 7.2. When these data are graphed

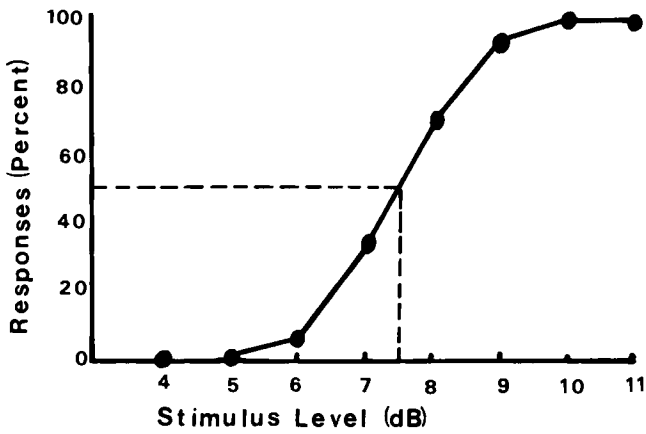

Figure 7.3 Psychometric function for method of constant stimuli. Threshold corresponds to 7.5 dB. (Data from Table 7.2.)

Table 7.3 Data from an Experiment on Differential Sensitivity for Intensity Using the Method of Constant Stimuli

Level of second tone (dB)	Percentage of louder judgments
70	100
68	95
66	85
64	70
62	55
60	35
58	20
56	10
54	8
52	5
50	0

in the form of a psychometric function (Fig. 7.3), the 50% point corresponds to 7.5 dB, which is taken as the threshold.

Table 7.3 shows the results of an experiment using the method of constants to find differential sensitivity for intensity. Two tones are presented and the subject is asked whether the second tone is louder or softer than the first. The intensity of the second tone is changed so that the various stimulus levels are presented randomly. Table 7.3 shows the percentage of presentations in which the subject judged the second tone to be louder than the first tone at each of the levels used. (The percentage of "softer" judgments is simply obtained by subtracting the percentage of "louder" judgments from 100%. Thus, the 60 dB presentations of the second tone were "softer" 100% − 35% = 65% of the time.) Figure 7.4 shows the psychometric function for these data. Since the intensity at which the second tone is judged louder 50% of the time is also the tone for which it was judged softer half of the time, the 50% point is where the two tones were perceived as equal in loudness. This is the PSE. In experiments of this kind, the 75% point is generally accepted as the threshold for "louder" judgments. (If we had also plotted "softer" judgments, then the 75% point on that psychometric function would constitute the "softer" threshold.) The DL is taken as the difference in stimulus values between the PSE and the "louder" threshold. For the data in Fig. 7.4 this difference is 64.8 dB − 61.5 dB = 3.3 dB.

The method of constant stimuli enables the experimenter to include "catch" trials over the course of the experiment. These are intervals during which the subjet is asked whether a tone was heard, when no tone was really presented. Performance on catch trials provides a estimate of

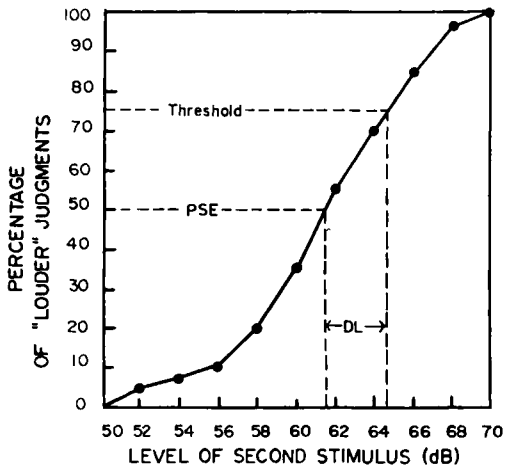

Figure 7.4 Psychometric function for a differential-sensitivity experiment showing the point of subjective equality (PSE), "higher" threshold, and difference limen (DL). (Data from Table 7.3.)

guessing, and performance on real trials is often corrected to account for this effect (see Chap. 8). This correction reduces, but does not completely remove, response biases from the results.

The method of constants has the advantage over the methods of limits and adjustments of greater precision of measurement; and as just mentioned has the advantage of allowing direct estimation of guessing behavior. However, it has the disadvantage of inefficiency, because a very large number of trials are needed to obtain the data. Most of these trial points are poorly placed relative to the points of interest (generally the 50% and 75% points), so that the method of constants costs heavily in time and effort for its accuracy. The prolonged test time increases the effects of subject fatigue and the difficulty of maintaining motivation to respond.

Adaptive Procedures

In an adaptive procedure, the level at which a particular stimulus is presented to the subject depends upon how the subject responded to the previous stimuli [3–5]. Broadly defined, even the classical method of limits may be considered an adaptive method because of its sequential character and the rule that stimuli are presented until there is a reversal in the subject's responses from "yes" to "no" or vice versa. However, the term adaptive procedures has come to be associated with methods that

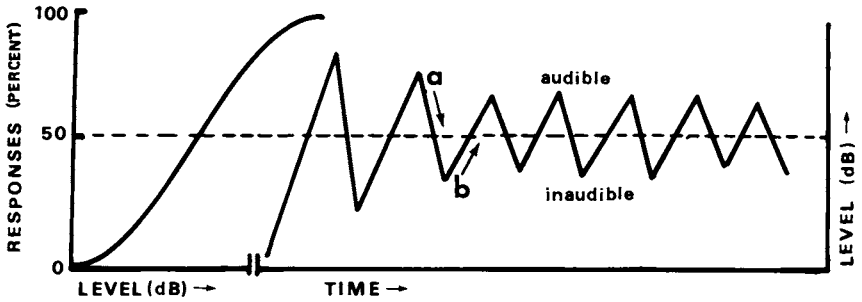

Figure 7.5 Bekesy's tracking method: (a) Intensity decreases as subject depresses the button when he hears the sound. (b) Intensity increases as subject releases the button when he cannot hear the sound. Excursion midpoints (50% level) correspond to the psychometric function at the left.

tend to *converge* upon the threshold level (or some other target point), and then place most of the observations around it. This procedure, of course, maximizes the efficiency of the method, because most of the test trials are close to the threshold rather than being "wasted" at some distance from it. It also has the advantage of not requiring prior knowledge of where the threshold level is located, since adaptive methods tend to home in on the threshold regardless of the starting point, and often include step sizes which are large at first and then become smaller as the threshold level is approached. As a result, both efficiency and precision are maximized.

Bekesy's Tracking Method

Bekesy [6] devised a tracking method which shares features both with the classical methods of adjustment and limits and with adaptive procedures. The level of the stimulus changes at a fixed rate (e.g. 2.5 dB/sec) under the control of a motor-driven attenuator, and the *direction* of level change is controlled by the subject via a pushbutton switch. The motor is also connected to a recorder which shows the sound level as a function of time (Fig. 7.5) or frequency. The pushbutton causes the motor to decrease the sound level when it is depressed and to increase the level when it is up. The subject is asked to press the button whenever he hears the tone and to release it whenever the tone is inaudible. Thus, the sound level is increased toward threshold from below when the tone is inaudible, and decreased toward threshold from above when the sound is heard. The threshold is thus tracked by the subject, and its value is the average of the midpoints of the excursions on the recording (once they are stabilized).

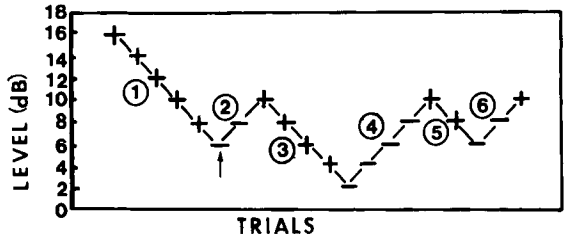

Figure 7.6 Simple up–down (staircase) method. Six runs are shown. Odd numbers are descending runs and even numbers are ascending runs. The first reversal (arrow) is generally omitted from the threshold calculation.

Tracking has the advantages of speed and reasonable precision. It is, of course, subject to several sources of response bias. At fast attenuation rates (intensity change speeds) the subject's reaction time can substantially affect the width of the tracking excursions and the precision of measurment. For example, if the tone increases and decreases in level at 5 dB/sec and a subject has a 2 sec reaction time, then the motor will have advanced the stimulus level (and pen position on the recorder) 10 dB above threshold before the button is finally depressed. Precision is improved and reaction time becomes less critical at reasonably slower attenuation rates, although the tracking may meander somewhat on the recording as the subject's criterion for threshold varies.

The tracking method has gained wide acceptance in clinical audiology. Bekesy audiometry has been used extensively in hearing screening programs, in the diagnosis of lesions in the auditory system, and in the identification of functional (nonorganic) hearing loss.

Up–Down (Staircase) Method

The simple up–down (staircase) method [7] involves increasing the stimulus level when the subject does not respond to a presentation and decreasing the intensity when there is a response. It differs from the method of limits in that testing does not stop when the responses change from "yes" to "no" or vice versa. As in the method of limits, the stimuli are varied in discrete steps.

Figure 7.6 shows the first six runs of a staircase procedure using a 2 dB step size. A "run" is a group of stimulus presentations between two response reversals. In other words, a descending run starts with a positive response and continues downward until there is a negative response; while an ascending run begins with a negative response and ends with a positive one. Since stimulus intensity is always increased after a negative (−) response and decreased after a positive (+) response, the

staircase method converges upon the 50% point on the psychometric function. The procedure is continued through at least six to eight reversals (excluding the first one), and the threshold value is then calculated as the average of the midpoints of the runs, or as the average of their peaks and troughs [3,8]. The latter method appears to give a somewhat better estimate. The precision of the method can be increased by first estimating the threshold with a larger step size, and then using a smaller step size (generally half that of the previous one) to locate the threshold in the vicinity of the first estimate [8]. For example, if the average of six runs using a 4 dB step is 10 dB, a second group of runs using a 2 dB step might begin at 10 dB in order to obtain a more precise estimate of the threshold.

The simple up–down method has several advantages and limitations [4]. It quickly converges upon the 50% point, so that most trials are efficiently placed close to the point of interest. It also has the advantage of being able to follow changes (drifts) in the subject's responses. On the other hand, the subject may bias his responses if he realized that the stimuli are being presented according to a sequential rule which depends on the way he responds. As with the method of limits, if the step size is too small a large number of trials are wasted, and if the step is too large they are badly placed for estimating the 50% point. Another limitation is that only the 50% point can be converged upon with the simple up–down rule.

PEST Procedure

Parameter estimation by sequential testing (PEST) is an adaptive procedure which uses changes in both the direction and step size of the stimulus to home in on the threshold [9]. As in the simple up–down method, positive responses are followed by decreases in stimulus level because the threshold is probably lower, and negative responses are followed by increases in intensity since the threshold is probably higher. The difference is that PEST includes a series of rules for doubling and halving the stimulus level depending upon the previous sequence of responses.

At each stimulus level, PEST in effect asks whether the threshold has been exceeded. The level is then changed so that the maximum amount of information is obtained from the next trial. To do this, the step size is varied in the manner specified in Fig. 7.7. Although it is most efficient to know the approximate location of the threshold range in advance, it is not essential. Suppose we begin testing at some value below threshold corresponding to point A in Fig. 7.7. Since the subject gives a negative response, the stimulus is presented at a higher level (B). This level also

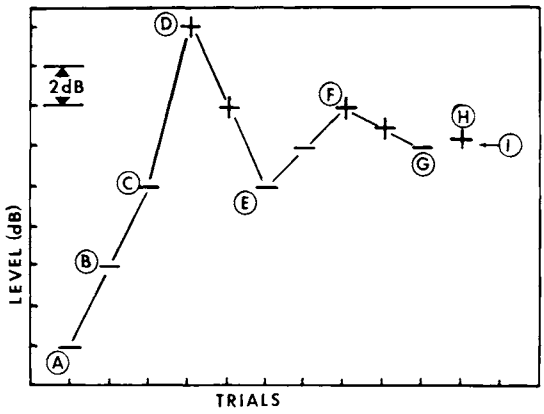

Figure 7.7 Obtaining threshold with PEST. (+) Positive response, (−) negative response. Letters are points discussed in the text. Point I is the estimate of threshold.

produces no response and the stimulus is raised by the same amount as previously, and is presented again (C). Since there is still no response, the stimulus level is again increased. However, PEST has a rule which states that if there is a negative response on two successive presentations in the same direction, then the step size is doubled for the next presentation. Thus, the next stimulus is presented at level D. The doubling rule insures that a minimal number of trials are wasted in finding the range of interest.

A positive response at level D indicates that the threshold has been exceeded. As in the staircase method, the direction of the trials is changed after a response reversal. However, the PEST procedure also halves the step size at this point. The halving rule causes the stimuli to be presented closer to the threshold value. Thus, precision is improved as the threshold is converged upon. Since D is followed by another positive response, the stimulus is then presented at a lower level (E). A negative response at E causes the direction of stimulus change to be changed again, and the step size is halved compared to the previous one. The stimulus is heard again at the next higher level (F), so the direction is changed again and the step size is again halved. Stimuli are now presented in a descending run until there is a negative response (G). Halving the step size and changing direction results in a positive response at H, indicating that the threshold lies somewhere between points G and H. Since this interval represents an acceptable degree of precision, the procedure is terminated. The level at which the *next* stimulus would have been presented is taken as the threshold. This level is point I, which lies

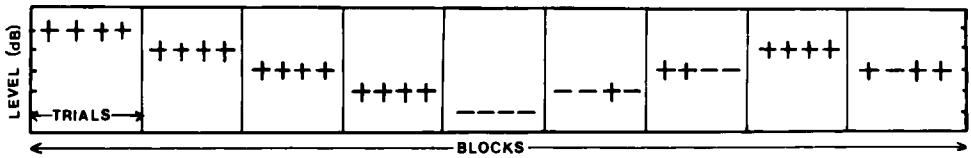

Figure 7.8 Convergence of the psychometric function upon the 75% point using BUDTIF.

halfway between levels G and H. Note on the scale for Fig. 7.7 that the step size between E and F is 2 dB, between F and G is 1 dB, and between G and H is 0.5 dB. This observation highlights the rapidity with which PEST results in a precise threshold estimate.

BUDTIF Procedure

Suppose we are interested in the 75% point of psychometric function. One way to converge upon the point is to modify the simple up–down procedure by replacing the single trial per stimulus level with a block of several trials per level. Then, by adopting three out of four positive responses (75%) as the criterion per level, the strategy will home in on the 75% point. If blocks of five were used with a four out of five criterion, then the 80% point would be converged upon. The procedure may be further modified by changing the response from yes–no to a two-alternative forced choice. In other words, the subject is presented with two stimulus intervals during each trial, and must indicate which of the intervals contains the stimulus. This is the block up–down temporal interval forced-choice (BUDTIF) procedure [10]. (Using the two-interval forced choice allows the experimenter to determine the proportion of responses to the no-stimulus interval — the "false alarm" rate. We shall see when the theory of signal detection is discussed in the next chapter that this distinction is important in separating sensitivity from bias effects.)

The BUDTIF procedure is illustrated in Fig. 7.8. Note that each block is treated as though it were one trial in a staircase procedure. Since the target point has been preselected as 75%, stimulus intensity is raised whenever there are less than three out of four correct responses in a block, and is decreased when all four are correct. Testing is terminated when three out of four correct responses are obtained. The last level is the target level (75% in this case). Notice that since blocks of trials are presented at each level, poor placement of the initial test level will cause many wasted trials in converging upon the target range.

A modification of BUDTIF replaces the two-alternative forced-choice paradigm with the familiar yes–no response. This adaptation is

called the block up–down yes–no (BUDYEN) method [11]. However, it is less satisfactory than its forced-choice predecessor because an estimate of false alarms is not obtained [12].

Transformed Up–Down Procedures

The simple up–down method converges on the 50% point of the psychometric function because each positive response leads to a decrease in stimulus level and each negative response leads to an intensity increase. If the up–down rule is modified so that stimulus level is changed only after *certain sequences* have occurred, then the procedure will home in on other points of the psychometric function [3,4]. These other target points depend upon the particular set of sequences chosen by the experimenter.

When the target is 50% on the psychometric function, as in the simple up–down method, the chances of a positive response to stimuli below the 50% point are very small. Similarly, it is likely that stimuli presented at levels well above the 50% point will frequently be heard. However, as the intensity corresponding to 50% is approached, the chances of positive and negative responses become closer and closer. *At* the 50% point, the probability of a positive response is the same as of a negative one. This is, of course, exactly what we mean by 50%. Now, suppose that the total probability of all responses is 1.00. If we call the probability of a positive response P, then the probability of a negative response would be (1 − P). At the 50% point

$$P = (1 - P) = 0.5$$

In other words, the probability of a positive response at the 50% point is 0.5, which is also the probability of a negative response. In effect, the simple up–down rule forces the intensity to the point on the psychometric function where the probabilities of positive and negative responses are equal (0.5 each).

Other target levels can be converged upon by changing the up–down rule so that the probabilities of increasing and decreasing stimulus intensity are unequal. This is done by setting the criteria for increasing stimulus level (the "up rule") to be a certain response sequence, and those for decreasing stimulus level (the "down rule") to be other response sequences. An example will demonstrate how the transformed up–down method works.

Suppose we are interested in estimating a point above 50% on the psychometric function, say 70%. To accomplish this, we would increase the stimulus level after a negative response (−) or a positive response followed by a negative one (+,−); and lower stimulus level after two

successive positives (+,+). In other words, we have established the following rules for changing stimulus level:

Up rule:	(−)	or	(+,−)
Down rule:	(+,+)		

As with the simple staircase rule, levels well above the target will often yield (+,+) responses, and those well below will frequently have (−) or (+,−) responses. However, at the target level, the probability of increasing the stimulus level will be

$$\underset{\text{Probability of } (-)}{(1 - P)} + \underset{\text{Probability of } (+,-)}{P(1 - P)}$$

and the probability of two successive positive responses (+,+) will be

$$P \times P \quad \text{or} \quad P^2$$

The up–down strategy will converge on the point where the up and down rules have the same probabilities (0.5). In other words, the probability of the transformed positive response (+,+) at the target is

$$P^2 = 0.5$$

Since we are interested in the probability P of a single positive response, which is the square root of P^2, we simply find the square root of $P^2 = 0.5$, and we obtain

$$P = 0.707$$

Converting to percent, the transformed procedure just outlined homes in on the 70.7% point of the psychometric function, which is a quite acceptable estimate of the 70% point.

To converge on the 29.3% of the psychometric function (which is a reasonable estimate of the 30% point), we might choose to increase the stimulus after a sequence of two successive negative responses (−,−), and to decrease stimulus level after a positive response (+) or a negative response followed by a positive one (−,+).

The 70.7% and 29.3% transformed up–down strategies are illustrated in Fig. 7.9. As for the simple up–down method, each strategy would be continued through six to eight reversals, and the average of the peaks and valleys would be taken as the target level. Since these two points are on the rising portion of the psychometric function and are equidistant from 50%, a reasonably good estimate of the 50% point can be obtained by averaging the levels from 70.7% and 29.3%. To increase

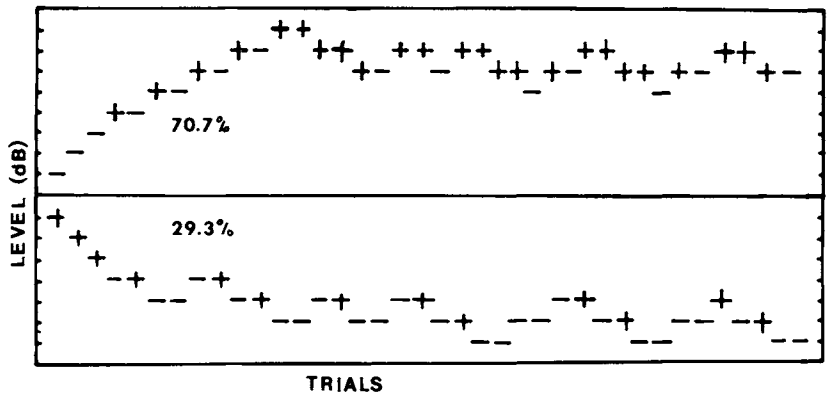

Figure 7.9 Transformed up–down strategies converging upon the 70.7% and 29.3% points of the psychometric function.

efficiency, one might start with a large step size, and then halve it in the target range for increased precision.

Other target points can be converged upon by using various sequences of positive and negative responses; and different sequences may be used to converge upon the same target points [4,13]. In addition, transformed up–down procedures are particularly applicable to testing speech discrimination functions [5,13,14].

A good approach to minimizing biases is to interleaf different test strategies [15]. In other words, two points on the psychometric function are converged upon during the same test session by switching in an approximately random manner between their two test strategies. For example, two reversals on the 29.3% strategy might be followed by a few reversals on the 70.7% strategy, then the experimenter would return to where he left off on the 29.3% sequence, and so forth. Such interleaving can also be applied to other psychoacoustic methods.

The transformed up–down procedure is made relatively simple to administer through the use of control charts [16]. These charts have the sequence of events for each strategy printed on them, so that all the experimenter needs to do is mark down whether the response was positive or negative, and otherwise simply "follow the dotted lines." Control charts also simplify tremendously the process of interleaving test strategies, by indicating whether the experimenter should stay on the current task or switch to the other one. Of course, greater flexibility and ease of measurement is made possible when the procedure is automated [17].

SCALES OF MEASUREMENT

Scaling is an important area in psychoacoustics. Stevens [18–21] described four scales of measurement which are now accepted as classical. These are descriptively named (1) nominal, (2) ordinal, (3) interval, and (4) ratio scales.

Nominal scales are the least restrictive, in the sense that the observations are simply assigned to groups. This is the lowest order of scaling because the nominal label does not tell us anything about the relationship among the groups other than that they are different with respect to some parameter. For example, the nominal scale "sex" enables us to separate people into two categories, "male" and "female." All we know is that the two categories are differentiable, and that we can count how many cases fall into each one. The same would apply to the number of subcompact cars made by different manufacturers. We know that there are so many Fords, Toyotas, etc.; but we have no idea of their relative attributes. A nominal scale, then, makes no assumptions of the order among the classes; thus, it is the least restrictive and least informative of the levels of scaling.

Ordinal scales imply that the observations have values which can be rank-ordered, so that one class is greater or less than another with respect to the parameter of interest. However, an ordinal scale does not tell us how far apart they are. Consider the relative quality of artistic reproductions. Painter A may produce a better reproduction of the Mona Lisa than Painter B, who in turn makes a better copy than Painter C, and so on. However, there may be one magnitude of distance between A and B, a second distance between B and C, and still a third distance between C and D. An ordinal scale thus gives the rank-order of the categories ($A > B > C \ldots$), but does not specify the distances between them. Whereas the nominal scale allows us to express the *mode* of the data (which category contains more cases than any other), ordinal scales permit the use of the *median* (the value with the same number of observations above and below it). Sometimes the nature of the categories enables some of them to be rank-ordered, but not others. This constitutes a *partially ordered scale* [22], which lies between the nominal and ordinal scales.

An *interval scale* specifies both the order among categories and the fixed distances among them. In other words, the distance between any two successive categories is equal to the distance between any other successive pair. Interval scales, however, do not imply a true zero reference point. Examples are temperature (in Celsius or Fahrenheit) and the

dates on a calendar. Since the distances between categories are equal, the central tendency of interval data may be expressed as a *mean* (average); however, interval data cannot be expressed as proportions (ratios) of one another because a true zero point is not assumed. It is also possible to rank the categories in such a way that there is an ordering of the distances between them. For example, the distances between successive categories may become progressively longer:

A-B--C---D----E-----F------G . . .

This is an *ordered metric scale* [22]. An ordered metric scale actually falls between the definitions of ordinal and interval scales, but may often be treated as an interval scale [23].

Ratio scales include all the properties of interval scales as well as an inherent zero point. The zero point permits values to be expressed as ratios, and hence the use of decibels in the expression of relationships. As the most restrictive level, ratio scales give the most information about the data and their interrelationships. Examples are length, time intervals, and temperature (in Kelvins), as well as loudness (sone) and pitch (mel) scales.

We might also think of three classes of scaling procedures [20,24]. *Discriminability* (or *confusion*) scales are generated by asking the subject to discriminate small differences among stimuli. For example, the subject might be required to detect small differences in intensity between otherwise equivalent tones. The smallest discriminable difference is the jnd or DL. The relationship between the magnitude of a sound (intensity) and the psychoacoustic response to it (loudness) may be indirectly inferred from the jnd's.

The other two scaling procedures ask the subject to make direct estimates of perceptual differences between stimuli. *Category* (or *partition*) scales are generated when the subject's task is to divide a range of stimuli (e.g. an intensity or frequency range) into equally spaced or sized categories. *Magnitude* (or *ratio*) scales are obtained when the subject is asked to estimate ratio (or proportional) relationships among the stimuli. For example, the subject may be asked to adjust the frequency of one tone so that it sounds twice as high in pitch as another tone. The measurement approaches in the following sections may be thought of in these terms. In general, the classical psychophysical methods discussed earlier in the chapter are discriminability scales, while those that follow are direct scaling procedures of the ratio type.

Direct Scaling

As mentioned above, direct scaling procedures ask the subject to establish a relationship between a standard stimulus and a comparison stimu-

lus. In other words, the subject must specify a perceptual continuum that corresponds to a physical continuum. Two types of continua may be defined [20]. Prothetic continua, such as loudness, have the characteristic of *amount* (or intensity). They are additive in that the excitation due to an increase in stimulus level is added to the excitation caused by the intensity which was already present. On the other hand, pitch has the characteristic of *kind* and azimuth has the characteristic of *location*. These are metathetic continua, and are substantive rather than additive. In other words, a change in the pitch corresponds to a substitution of one excitation pattern, as it were, for another.

Ratio Estimation and Production

In ratio estimation the subject is presented with two stimuli differing in terms of some parameter, and is asked to express the subjective magnitude of one stimulus as a ratio of the other. Subjective values are thus scaled as a function of the physical magnitudes. Suppose two 1000 Hz tones with different intensities are presented to a subject, who must judge the loudness of the second tone as a ratio of the first. He might report that the intensity of the second tone sounds one-half, one-quarter, twice, or five times as loud as the first tone.

Ratio production, or fractionalization, is the opposite of ratio estimation in that the subject's task is to adjust the magnitude of a variable stimulus so that it sounds like a particular ratio (or fraction) of the magnitude of a standard stimulus. For example, the subject might adjust the intensity of a comparison tone so that it sounds half as loud as the standard, twice as loud, etc. Fractionalization has been used in the development of scales relating loudness to intensity [25] and pitch to frequency [26,27].

Magnitude Estimation and Production

In magnitude estimation the subject assigns to physical intensities numbers that correspond to their subjective magnitudes. This may be done in two ways [28]. In the first method, the subject is given a standard stimulus and is told that its intensity has a particular value (modulus). He is then presented with other intensities and must assign numbers to these which are ratios of the modulus. Consider a loudness scaling experiment in which the subject compares the loudness of variable tones to a standard tone of 80 dB. If the 80 dB standard is called 10 (modulus) then a magnitude estimate of 1 would be assigned to the intensity 1/10 as loud, 60 would be assigned to the one which is 6 times as loud, etc. The relationship between these magnitude estimates and intensity is shown by the circles in Fig. 7.10.

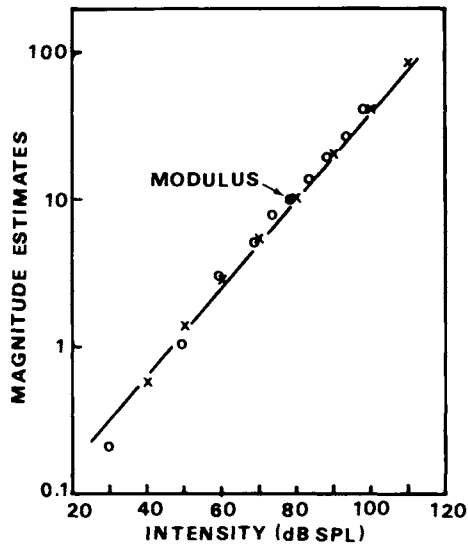

Figure 7.10 Magnitude estimation of loudness with a modulus (circles) and without a modulus (crosses) as a function of intensity. (Data from Stevens [28].)

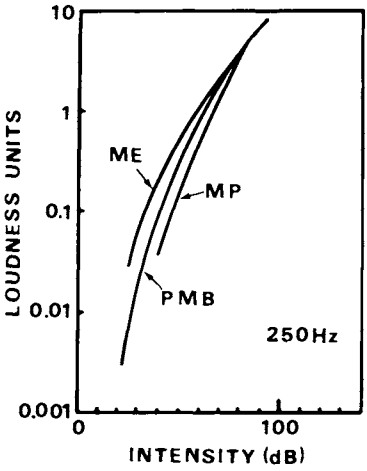

Figure 7.11 Bias effects in magnitude estimation (ME) and magnitude production (MP) are minimized by geometrical averaging in the method of psychological magnitude balance (PMB). (Adapted from Hellman and Zwislocki [30]. Permission *J. Acoust. Soc. Amer.*)

An alternative approach is to omit the modulus. Here, the subject is presented with a series of stimuli and is asked to assign numbers to them reflecting their subjective levels. The results of such an experiment are shown by the x's in Fig. 7.10. As the figure shows, magnitude estimates obtained with and without a modulus result in similar findings.

The reverse of magnitude estimation is magnitude production. In this approach the subject is presented with numbers and must adjust the magnitude of the stimulus to correspond to the numbers.

Subject bias causes magnitude estimation and production to yield somewhat different results, especially at high and low stimulus levels. Specifically, subjects tend not to assign extreme values in magnitude estimation, or to make extreme level adjustments in magnitude production. These bias effects are in opposite directions, so that the "real" function lies somewhat between the ones obtained from magnitude estimations and productions. This is illustrated in Fig. 7 .11 by the divergence of the magnitude estimation (ME) and magnitude production (MP) functions. The unbiased function may be obtained by using the method of psychological magnitude balance suggested by Hellman and Zwislocki [29,30]. This method involves taking the geometric mean of the magnitude estimates and productions along the intensity axis or the loudness axis. The result is shown by the curve PMB in Fig. 7.11.

REFERENCES

1. D. E. Robinson and C. S. Watson, Psychophysical methods in modern psychoacoustics, in *Foundations of Modern Auditory Theory* (J. V. Tobias, Ed.), Vol. 2, Academic, New York, 1973, pp. 99–131.
2. R. Carhart and J. Jerger, Preferred method for clinical determination of pure-tone thresholds, *J. Speech Hearing Dis. 24*, 330–345 (1959).
3. G. B. Wetherill and H. Levitt, Sequential estimation of points on a psychometric function, *Br. J. Math. Stat. Psych. 18*, 1–10 (1965).
4. H. Levitt, Transformed up–down methods in psychoacoustics, *J. Acoust. Soc. Amer. 49*, 467–477 (1971).
5. D. L. Bode and R. Carhart, Measurements of articulation functions using adaptive test procedures, *IEEE Trans. Audiol. Electroacoust. AU-21*, 196–201 (1973).
6. G. Bekesy, *Experiments in Hearing*, McGraw-Hill, New York, 1960.
7. W. J. Dixon and A. M. Mood, A method for obtaining and analyzing sensitivity data, *J. Amer. Stat. Ass. 43*, 109–126 (1948).
8. G. B. Wetherill, Sequential estimation of quantal responses, *J. Roy. Stat. Soc. 25*, 1–48 (1963).
9. M. M. Taylor and C. D. Creelman, PEST: Efficient estimates on probability functions, *J. Acoust. Soc. Amer. 41*, 782–787 (1967).

10. R. A. Campbell, Detection of a noise signal of varying duration, *J. Acoust. Soc. Amer. 35*, 1732–1737 (1962).
11. R. A. Campbell and S. A. Counter, Temporal energy integration and periodicity pitch, *J. Acoust. Soc. Amer. 45*, 691–693 (1969).
12. C. D. Creelman and M. M. Taylor, Some pitfalls in adaptive testing: Comments on "Temporal integration and periodicity pitch," *J. Acoust. Soc. Amer. 46*, 1581–1582 (1969).
13. H. Levitt and L. R. Rabiner, Use of a sequential strategy in intelligibility testing, *J. Acoust. Soc. Amer. 42*, 609–612 (1967).
14. D. L. Bode and R. Carhart, Stability and accuracy of adaptive tests of speech discrimination, *J. Acoust. Soc. Amer. 56*, 963–970 (1974).
15. H. Levitt, Testing for sequential dependencies, *J. Acoust. Soc. Amer. 43*, 65–69 (1968).
16. H. Levitt and M. Treisman, Control charts for sequential testing, *Psychometrika 34*, 509–518 (1969).
17. H. Levitt and D. E. Bock, Sequential programmer for psychophysical testing, *J. Acoust. Soc. Amer. 42*, 911–913 (1967).
18. S. S. Stevens, Mathematics, measurement, and psychophysics, in *Handbook of Experimental Psychology* (S. S. Stevens, Ed.), Wiley, New York, 1951.
19. S. S. Stevens, Problems and methods in psychophysics, *Psych. Bull. 55*, 177–196 (1958).
20. S. S. Stevens, The psychophysics of sensory function, in *Sensory Communication* (W. A. Rosenblith, Ed.), MIT Press, Cambridge, Mass., (1961).
21. S. S. Stevens, *Psychophysics*, Wiley, New York, 1975.
22. C. H. Coomb, Theory and methods of measurement, in *Research Methods in the Behavioral Sciences* (L. Festinger and D. Katz, Eds.), Holt, Rinehart, and Winston, New York, 1953, pp. 471–535.
23. R. P. Abelson and J. W. Tukey, Efficient conversion of non-metric information into metric information, in *The Quantitative Analysis of Social Problems* (E. R. Tufte, Ed.), Addison Wesley, Reading, Mass., (1959), pp. 407–417.
24. S. S. Stevens, Ratio scales, partition scales and confusion scales, in *Psychological Scaling: Theory and Applications* (H. Gulliksen and S. Messick, Eds.), Wiley, New York, 1960.
25. S. S. Stevens, A scale for the measurement of a psychological magnitude: Loudness, *Psych. Rev. 43*, 405–416 (1936).
26. S. S. Stevens, J. Volkmann, and E. B. Newman, A scale for the measurement of the psychological magnitude pitch, *J. Acoust. Soc. Amer. 8*, 185–190 (1937).
27. S. S. Stevens and J. Volkmann, The relation of pitch to frequency: A revised scale, *Amer. J. Psych. 53*, 329–353 (1940).
28. S. S. Stevens, The direct estimation of sensory magnitudes — loudness, *Amer. J. Psych. 69*, 1–25 (1956).
29. R. P. Hellman and J. Zwislocki, Monaural loudness function of a 1,000-cps tone and interaural summation, *J. Acoust. Soc. Amer. 35*, 856–865 (1963).
30. R. P. Hellman and J. Zwislocki, Loudness summation at low sound frequencies, *J. Acoust. Soc. Amer. 43*, 60–63 (1968).

8
Theory of Signal Detection

The previous chapter addressed itself to classical and modern psychophysical methods and the direct rating of sensory magnitudes, with respect to hearing. It left essentially unresolved, however, the problem of how to effectively separate sensitivity from response proclivity. In this chapter, we shall approach this problem from the standpoint of the theory of signal detection.

FACTORS AFFECTING RESPONSES

The theory of signal detection [1–3] provides the best approach to separating the effects of sensitivity from those of response bias. We might think of the theory of signal detection (TSD) as asking "what led to a 'yes' (or 'no') decision?" as opposed to "what did the subject hear (or not hear)?"

Suppose a subject is asked to say "yes" when he hears a tone during a test trial and "no" when a tone is not heard. A large number of trials are used for each of several stimulus levels, and half of those at each level are "catch trials" during which signals are not actually presented. There are thus four possible outcomes for each test trial. Two of these are correct:

The signal is present and the subject says "yes" (a hit).
The signal is absent and the subject says "no" (a correct rejection).

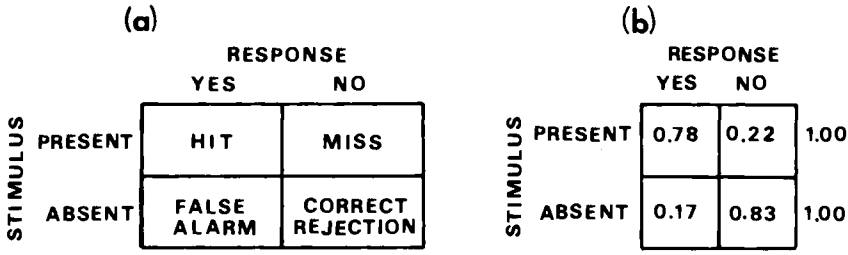

Figure 8.1 (a) Stimulus–response table showing the four possible outcomes for any trial. (b) Hypothetical results (in proportions) for 100 stimulus trials and 100 catch trials at a particular test level.

The other two alternatives are wrong:

The signal is present but the subject says "no" (a miss).
The signal is absent but the subject says "yes" (a false alarm).

A convenient way to show these possible outcomes is in the form of a stimulus–response matrix (Fig. 8.1).

The stimulus–response table is generally used to summarize the results of all trials at a particular test level; there would thus be such a table for each stimulus level used in an experiment. For example, Fig. 8.1b shows the results of 100 trials containing a signal and 100 catch trials. The subject responded to 78 of the signal trials (so that the probability of a hit was 0.78), did not respond to 22 signals (the probability of a miss is 0.22), said "yes" for 17 out of 100 catch trials (the probability of a false alarm is 0.17), and said "no" for the remaining absent-stimulus trials (the probability of a correct rejection is 0.83). One is tempted to say that the percent correct at this stimulus level is 78% (the hit rate), but the fact that the subject also responded 17 times when there was no stimulus present tells us that even the 78% correct includes some degree of chance success or guessing. One way to account for this error is to use the proportion of false alarms as an estimate of the overall guessing rate, and to correct the hit rate accordingly. The traditional formula to correct the hit rate for chance success is

$$p(\text{hit})_{\text{corrected}} = \frac{p(\text{hit}) - p(\text{false alarm})}{1 - p(\text{false alarm})}$$

In other words, the probability p of a hit corrected for chance success is obtained by dividing the difference between the hit rate and the false alarm rate by one* minus the false alarm rate. Thus, for this example

$$p(\text{hit})_{\text{corrected}} = \frac{0.78 - 0.17}{1.0 - 0.17} = \frac{0.61}{0.83} = 0.735$$

The original 78% correct thus falls to 73.5% when we account for the proportion of the "yes" responses due to chance.

Correcting for chance success is surely an improvement over approaches that do not account for guessing, but it still does not really separate the effects of auditory factors (sensitivity) and nonauditory factors. In essence, this process highlights the importance of nonauditory factors in determining the response, because the very fact that the subject said "yes" to catch trials and "no" to stimulus trials indicates that his *decision* to respond was affected by more than just sensitivity to the stimulus. The TSD is concerned with the factors that enter into this decision.

Let us, at least for the moment, drop the assumption that there is some clear-cut threshold that separates audibility from inaudibility, and replace it with the following assumptions of TSD. First, assume that there is always some degree of noise present. This may be noise in the environment, instrumentation noise, or noise due to the subject's moving around and fidgeting. Even if all of these noises were miraculously removed, there would still remain the subject's unavoidable physiological noises (heartbeat, pulse, breathing, blood rushing through vessels, stomach gurgles, etc.). Indeed, the noise is itself often presented as part of the experiment. For example, the task may be to detect a tone in the presence of a noise. Since there is always noise, which is by nature random, we also assume that the stimulation of the auditory system varies continuously. Finally, we shall assume that all of the stimulation occurs (or is at least measurable) along a single continuum. In other words, the subject must decide whether the stimulation in the auditory system (e.g. energy) is due to noise alone (N) or to signal-plus-noise (SN). This process may be represented by distributions on a "decision axis" like the one shown in Fig. 8.2. The subject's response is then a decision between "yes" ("I hear the signal as well as the noise") and "no" ("I hear the noise alone").

*Recall that the total probability of all catch trials is 1.0, so that 1 − p (false alarm) is the same as the probability of a correct rejection.

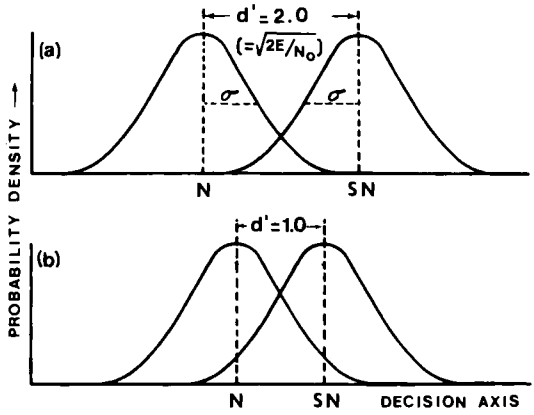

Figure 8.2 Separation between the noise (N) and signal-plus-noise (SN) curves determines the value of d'.

The N and SN distributions in Fig. 8.2 show the probability functions of noise alone (N) and signal-plus-noise (SN). We might think of these curves as showing the chances (or likelihood) of there being respectively a noise alone or a signal-plus-noise during a particular test trial. Obviously, there must always be more energy in SN than in N, due to the presence of the signal. The separation between the N and SN curves thus becomes a measure of sensitivity. This is an unbiased measure because the separation between the two curves is not affected by the subject's criteria for responding (biases). The separation is determined solely by the energy in the signals and the sensitivity of the auditory system. This separation is measured in terms of a parameter d'. The value of d' is equal to the difference between the means ($\bar{X}$) of the N and SN distributions divided by their standard deviation σ*

$$d' = \frac{\bar{X}_{SN} - \bar{X}_{N}}{\sigma}$$

Comparing Figs. 8.2a and b, we see that the greater the separation between N and SN distributions, the larger the value of d'. This value does not change even when different experimental methods are used [4].

*It is assumed that SN and N are normally distributed with equal variances. Since σ is the square root of the variance, and the variances of SN and N are assumed to be equal, then only one value of σ need be shown. The value of d' is equal to the square root of twice the energy in the signal (2E) divided by the noise power (N_0) in a band that is one cycle wide [5], or: $d' = (2E - N_0)^{1/2}$. Tables of d' are available in the literature [6]. However, since the standard deviation of SN is actually larger than that of N in some cases, a corrected value of d' may be a more valid measure [7].

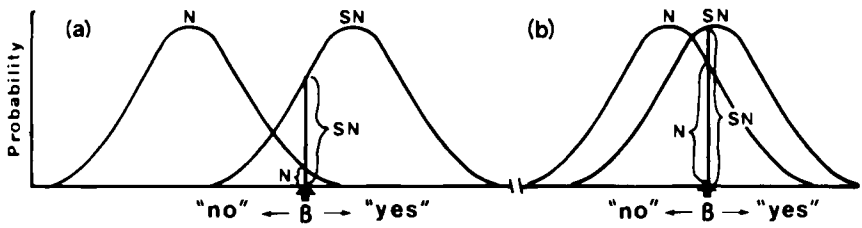

Figure 8.3 Effects of the criterion point β. See text.

How, then, does a subject decide whether to say "yes" or "no" for a given separation between the N and SN curves? Consider the N and SN distributions in Fig. 8.3, and note the point along the decision axis labeled β. This point is the subject's *criterion* for responding. Whenever the energy is greater than the criterion (to the right of β) the subject will say "yes." The subject will say "no" if the energy is less than the criterion (to the left of β). The value (or placement) of β depends on several factors which we will examine next.

The first factor affecting β is expressed by the question "what is the probabiity that there is a noise alone compared to the probability that there is a signal-plus-noise for a given β?" For any point along the decision axis, this question is the same as comparing the height of the N curve with the height of the SN curve (Fig. 8.3). The value of this ratio is due to the amount of overlap between the N and SN distributions, and to what the subject knows about the relative chances of a signal actually being presented. Comparison of Figs. 8.3a and b shows how overlapping of the N and SN functions changes this ratio. At any point, the heights of the two curves become closer as the separation between them decreases from that in Fig. 8.3a to that in Fib. 8.3b. An "ideal observer" (which is actually a mathematical concept rather than a real subject) would place the criterion point β at the ratio which minimizes the chances of error; that is, at the point at which misses and false alarms are minimized. However, the placement of the criterion point will also be adjusted somewhat by what the subject knows about the chances of occurrence of a noise alone versus a signal-plus noise. Let us now address ourselves to this factor.

Up to this point, it has been assumed that N and SN will be presented on a fifty-fifty basis. However, if the subject knows that a signal will actually be presented one-third of the time, then he will of course adjust his criterion β accordingly. In other words, he will adopt a stricter criterion. Alternatively, if the subject knows that a signal will occur more often than the noise alone, then he will relax his criterion for respond-

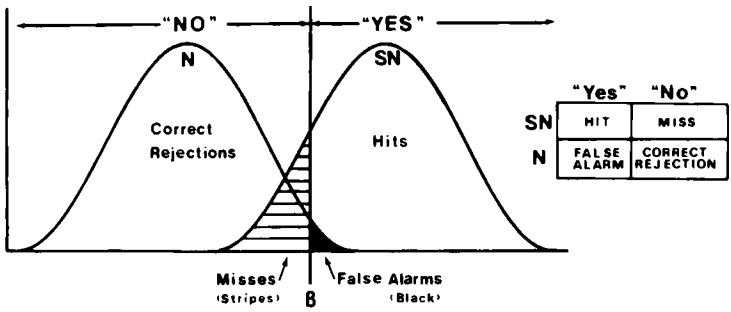

Figure 8.4 Outcomes of a "yes" or "no" response based upon a particular criterion β. Stimulus–response table is shown to the right.

ing, adjusting for the greater chances of the signal actually being presented. An "ideal observer" always knows these probabilities; a real subject is often, but not always, told what they are.

The last factor that we will discuss which affects the final value of the criterion β has to do with how much a correct response is worth and how much a wrong response will cost. We are therefore concerned with the chance of an error associated with a particular criterion for responding. These chances are shown in Fig. 8.4. The subject will say "no" whenever the actual presentation falls to the left of the criterion β, and will say "yes" when the presentation is to the right of β. Since the N and SN curves overlap, there will be both "yes" and "no" decisions for a certain proportion of both signal and no-signal presentations. With the criterion point β placed as shown in the figure, most of the "yes" decisions will be in response to SN; i.e., so that there actually was a signal when the subject says "yes" (a hit). However, a certain percentage of the N trials will fall to the right of β, so that the subject will say "yes" even though there actually was no signal presented (a false alarm). A stimulus–response table similar to the one in Fig. 8.1a is shown next to the N and SN distributions in Fig. 8.4 to illustrate how the two curves and the criterion relate to the possible outcomes of an experiment.

Now, suppose that a subject is told that it is imperative that he never miss a signal. He would thus move the criterion point toward the left to increase the hit rate; however, this shift would also have the effect of increasing the number of false alarms. The result would occur because moving β toward the left increases the proportions of *both* the N and SN curves that are in the "yes" region. On the other hand, if the subject is advised that a false alarm is the worst possible error, then the criterion point would be shifted toward the right, minimizing false alarms. Of course, this shift would also increase the number of misses, because a larger portion of the SN curve would now be in the "no" region.

Instead of telling the subject that one or another type of response is more (or less) important, the subject might be given a nickel for each correct response, lose three cents for a false alarm, etc. This, too, would cause the subject to adjust the criterion point so as to maximize the payoff associated with his responses. In effect, then, a set of values are attached to the responses so that each correct response has a *value* and each erroneous response has a *cost*.

An optimum criterion point (optimum β) is based upon the probabilities of noise alone (p_N) and of signal-plus-noise (p_{SN}), combined with the payoff resulting from the costs and values of each response. The payoff is the net result of the values of hits (V_H) and correct rejections (V_{CR}) and of the costs of misses (C_M) and false alarms (C_{FA}). In other words

$$\text{optimum } \beta = \left(\frac{(p_N)}{(p_{SN})}\right)\left(\frac{V_{CR} - C_{FA}}{V_H - C_M}\right)$$

The decision criterion is an attempt to maximize the payoff associated with the task. However, the subject in the real world is either not aware of all factors, or not able to use them as efficiently as the mathematically ideal observer. Therefore, the actual performance observed in an experiment generally falls short of what would have resulted had the subject been an ideal observer.

In summary, two types of information are obtained from the subject's responses in a TSD paradigm. One of these, d′, is a measure of sensitivity which is determined strictly by the separation between the N and SN distributions and by the ability of the auditory system to make use of this separation. The other measure is the subject's criterion for responding, which does not affect the actual measure of sensitivity.

How can we show all of this information at the same time in a meaningful manner? Consider the effects of several different response criteria for the same value of d′. These criteria may be obtained by changing the directions given to the subject, or by changing the payoff scheme. Another way would be to have the subject rank the degree of certainty with which he makes each yes/no decision (see the discussion of TSD methods, below, for the rationale of this approach).

For a given amount of sensitivity (i.e., a given value of d′), different criteria will result in different proportions of hits and false alarms. This result is shown for two arbitrarily selected values of d′ in Fig. 8.5. We may plot the proportions of hits versus false alarms for each criterion point, as in the center of Fig. 8.5. Such a graph is called a receiver-operating characteristic (ROC) curve. Notice that the ROC curve allows both the effects of sensitivity and response criterion to be illustrated at the same time. Sensitivity is shown by the distance of the ROC curve

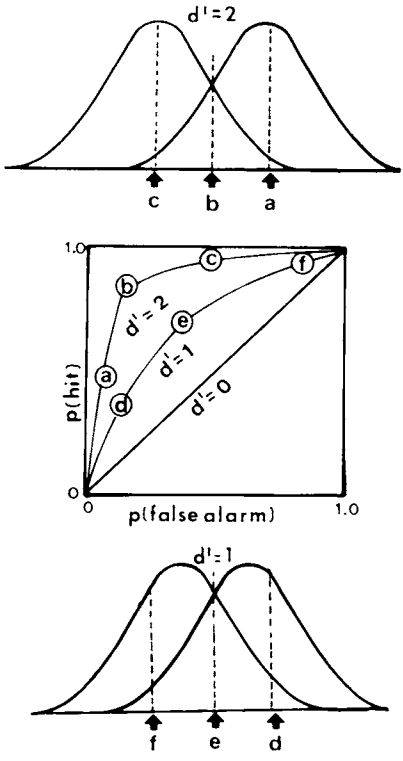

Figure 8.5 Sensitivity is shown by the distance of the ROC curve from the diagonal (center). Different response criteria (arrows) correspond to various points along the same ROC curve (indicated by letters).

from the diagonal (at which d′ = 0), or by the area under the ROC curve. On the other hand, the response criterion is indicated by the particular point along the ROC curve.

PSYCHOPHYSICAL METHODS IN TSD

Yes/No Methods

This discussion of the theory of signal detection has dealt primarily with the yes/no method. To recapitulate: The subject is presented with a large number of trials for each stimulus level, and a large proportion of these are actually catch trials. For each trial the subject says "yes" when a signal is detected and "no" when a signal is not detected. Fundamentally, then, the yes/no method in TSD is somewhat akin to the classical method of constant stimuli, although there are obvious differences in the number of trials, the large proportion of catch trials, and the manner of data analysis.

As in the classical methods, the TSD experiment is easily modified for use in a study of differential sensitivity. In this case, two signals are presented in a trial and the subject's task is to say "yes" ("they are different") or "no" ("they are not different").

The yes/no method is actually a subcategory of a larger class of experiments in which each trial contains one of two alternative signals. In this general case, the subject's task is to indicate which of two possible signals was present during the trial. For example, in the differential sensitivity experiment mentioned in the last paragraph, the decisions "same" versus "different"; or else the subject might be asked to decide between two alternative speech sounds (e.g. /p/ and /b/) while some parameter is varied. In this light the yes/no method might be thought of as a single-interval forced-choice experiment. In other words, the subject is presented with a stimulus interval during each trial, and is required to choose between signal-plus-noise (one of the alternatives) and noise alone (the other alternative).

Two- (and Multiple-) Interval Forced-Choice Methods

Just as the subject may be asked to choose between two alternatives in a single-interval trial, he might also be asked to decide *which* of two successive intervals contained a signal. In this case, the trial consists of two intervals, A and B, which are presented one after the other. One of the intervals (SN) contains the signal and the other (N) does not. The subject must indicate whether the signal as presented in interval A or in interval B.

In multiple-interval forced-choice experiments, several (e.g. four) intervals are included in each trial, among which the subject must choose the one that contained the signal.

Confidence Rating Methods

Recall that various points along the same ROC curve represent different response criteria with the same sensitivity d′. We might think of the response criterion as a reflection of how much confidence a subject has in his decision. In other words, a strict criterion means that the subject must have a great deal of confidence in his decision that the signal is present before he is willing to say "yes." In this case the criterion point β is pushed toward the right along the decision axis. Alternatively, a lax criterion means that the subject does not require as much confidence in his "yes" decision, which moves the criterion point toward the left.

We might apply this relationship between the confidence in the decision and the criterion point by asking the subject to rate how much confidence he has in each of his responses. For example, the subject might be instructed to rate a "yes" response as "five" when he is absolutely positive that there was a signal, and "four" when he thinks there was a signal. A rating of "three" would mean "I'm not sure whether there was a signal or no signal." "Two" would indicate that there probably was no signal present, and a rating of "one" would suggest that the subject is positive that a signal was not presented. This procedure is the same as adopting a series of criterion points located successively from right to left along the decision axis. Thus, the use of confidence ratings enables the experimenter to obtain several points along the ROC curve simultaneously. This approach results in data which are comparable to those obtained by the previously discussed direct methods [8].

IMPLICATIONS OF TSD

The theory of signal detection has importance in psychoacoustics because its application allows the experimenter to ferret out the effects of sensitivity and response criterion. Furthermore, TSD lends itself to experimental confirmation, and can be used to test theories and their underlying assumptions. A key application of TSD has been the testing of the classical concept of threshold as an absolute boundary separating sensation from no sensation. It is implicit in this discussion that such a concept of a clear-cut threshold is not supported by TSD. However, the more general concept of threshold remains unresolved. Threshold theory is beyond the scope of this text. The student is therefore referred to the very informative discussions in papers by Swets [9] and Krantz [10].

REFERENCES

1. J. A. Swets, Ed., *Signal Detection and Recognition by Human Observers*, Wiley, New York, 1964.
2. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, Krieger, New York, 1974.
3. J. P. Egan, *Signal Detection Theory and ROC Analysis*, Academic, New York, 1975.
4. J. A. Swets, Indices of signal detectability obtained with various psychophysical procedures, *J. Acoust. Soc. Amer. 31*, 511–513 (1959).
5. J. A. Swets, W. P. Tanner, Jr., and T. G. Birdsall, Decision processes in perception, *Psych. Rev. 68*, 301–340 (1961).

6. P. B. Elliot, Tables of d′, in *Signal Detection and Recognition by Human Observers* (J. A. Swets, Ed.), Wiley, New York, 1964, pp. 651–684.
7. L. H. Theodore, A neglected parameter: Some comments on "A table for calculation of d′ and β," *Psych. Bull. 78*, 260–261 (1972).
8. J. P. Egan, A. I. Schulman, and G. Z. Greenberg, Operating characteristics determined by binary decisions and by ratings, *J. Acoust. Soc. Amer. 31*, 768–773 (1959).
9. J. A. Swets, Is there a sensory threshold? *Science 134*, 168–177 (1961).
10. D. H. Krantz, Threshold theories of signal detection, *Psych. Rev. 76*, 308–324 (1969).

9
Auditory Sensitivity

The ear's extremely wide range of sensitivity is one of the most striking aspects of audition. The preceding chapters emphasized that hearing measurements are affected by psychophysical methods and other non-auditory factors; nevertheless, a reliable picture of auditory sensitivity has been provided by research over the years. Briefly, the ear is sensitive to a range of intensities from about 0 dB SPL (which is an amplitude of vibration of about the size of a hydrogen molecule) to roughly 140 dB (at which pain and damage to the auditory mechanism ensue). This dynamic range of approximately 140 dB corresponds to a pressure ratio of 10^7:1. In other words, the most intense sound pressure that is bearable is on the order of 10 million times as great as the softest one that is perceivable under optimum listening conditions. In terms of frequency, humans can hear tones as low as 2 Hz (although roughly 20 Hz is required for a perception of "tonality") and as high as about 20,000 Hz. Even the most avid hi-fi/stereo enthusiast must be impressed by the range over which the ear is responsive. Furthermore, the auditory system is capable of resolving remarkably small temporal differences.

The frequency and intensity sensitivities of the ear interact, affecting each other to a greater or lesser degree. In addition, when the duration of a sound is less than about half of a second, it affects both frequency and intensity sensitivity. Longer durations may be thought of as being infinitely long as far as auditory sensitivity is concerned.

Finally, the ear is able to discriminate small differences in a wide range of stimuli, i.e., it has remarkable differential sensitivity — the

ability to detect very small differences between similar sounds. This ability applies to all three parameters: intensity, frequency, and duration.

So much for sweeping generalizations. Let us now look at the details.

ABSOLUTE SENSITIVITY

Minimum Audible Levels

The problem of absolute sensitivity is essentially one of describing how much sound intensity is necessary for a typical, normally hearing person to just detect the presence of a stimulus. We must realize at the outset that these values are actually measures of central tendencies (means, medians, and/or modes) which describe a group of ostensibly normal subjects. In addition, it is essential to know how and where the minimum audible sound intensity is measured.

Two fundamental methods have been used to measure the intensity of a minimum audible stimulus [1]. The first involves testing a subject's thresholds through earphones, and then actually monitoring the sound pressures in the ear canal (between the earphone and eardrum) that correspond to these thresholds. This procedure yields a measure of minimum audible *pressure* (MAP). The alternative approach is to seat the subject in a sound field and test his thresholds for sounds presented through a loudspeaker. The subject then leaves the sound field and the threshold intensity is measured with a microphone placed where his head had been. This method measures the minimum audible *field* (MAF). It is important to dichotomize between the MAP and MAF methods, because they result in different threshold values. This discrepancy has been one of the most elusive problems in psychoacoustics.

Ostensibly, MAP refers to the sound pressure at the eardrum. This quantity is monitored by placing a probe tube in the subject's ear canal. The probe tube passes through the earphone enclosure and leads to a microphone, which measures the sound pressure at the tip of the probe tube. Since it is difficult to place the probe right at the drum (as well as potentially painful and dangerous), the probe is generally located somewhere in the ear canal, as close to the drum as is practicable.

Minimum audible pressures are often stated in terms of the sound pressure generated by an earphone in a standardized 6 cc metal cavity (coupler), which approximates the volume under an earphone on the subject's ear. Such coupler pressures form the reference levels used in audiometric standards (see below). Coupler-referred MAP values are more appropriately called MAPC to distinguish them from the probe tube MAP data obtained from actual ear canals [2].

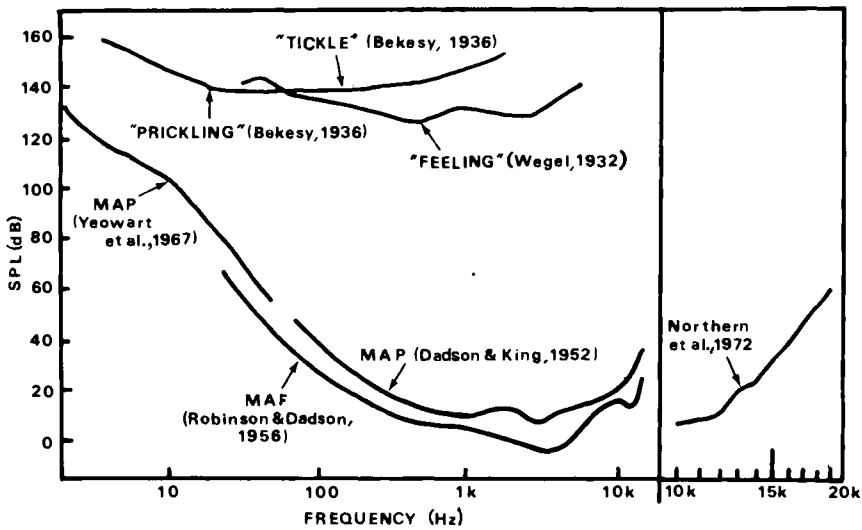

Figure 9.1 Monaural MAP after Dadson and King [3] and Yeowart et al. [5]. High frequency MAP after Northern et al. [6] is shown on the expanded scale to the right for clarity. Upper limits of "usable" hearing after Bekesy [7] and Wegel [8].

Sivian and White [1] reviewed the minimum audibility data available as of 1933, and reported the results of their own classic MAP and MAF experiments. Their work was essentially confirmed by Dadson and King [3] and by Robinson and Dadson [4], whose data are shown in the lower portion of Fig. 9.1. These curves show monaural MAP and binaural MAF (from a loudspeaker located directly in front of the subject, i.e., at 0° azimuth) as a function of frequency. Monaural MAP values extending to very low frequencies are also shown. The MAP values for frequencies between 10,000 and 18,000 Hz are shown in the figure on an expanded frequency scale. As these MAP and MAF curves clearly show, human hearing is most sensitive between about 2000 and 5000 Hz, and reasonably good sensitivity is maintained in the 100 Hz–10 kHz range. Absolute sensitivity becomes poorer above and below these frequencies.

An intriguing phenomenon, demonstrated in Fig. 9.1, is that the MAF curve falls consistently below the MAP curve. In other words, a lower intensity is needed to reach threshold in a sound field (MAF) than under earphones (MAP). This fact was first demonstrated by Sivian and White [1]; and the discrepancy of 6–10 dB is called the "missing 6 dB" [9]. Sivian and White proposed that the MAP/MAF discrepancy might be due to physiological noises picked up by the ear when it is covered by an earphone. These physiological noises would partially mask (see Chap.

10) the signal presented by the earphone, so that more sound pressure would be needed to achieve MAP than for the unmasked MAF. While this explanation accounts for part of the missing 6 dB problem, it falls short of accounting for the whole difference.

Several more recent studies [2,10–15] have formed the basis for resolving the MAP/MAF difference. This explanation has been presented in a cohesive manner by Killion [2]. To begin with, recall from Chap. 3 that diffraction and ear canal resonace enhance the pressure of a free field signal reaching the eardrum [16]. Thus, a corrected version of the international standard reference MAF curve* [17] may be converted to eardrum pressure by applying Shaw's [16] free field/eardrum pressure data. Since binaural thresholds are somewhat better than monaural ones (see Chap. 13), a correction is also made to account for the advantage of binaural MAF over monaural MAP. By accounting for differences between real-ear (MAP) and coupler (MAPC) values, the effects of impedance changes and ear canal distortion due to the placement of the earphone, and the effects of physiological noises, the MAP/MAF discrepancy is essentially resolved.

Upper Limits of Hearing

Just as we may conceive of the minimum audibility (MAP and MAF) curves as the lower limit of hearing sensitivity, the upper limits of hearing may be thought of as the sound pressure levels (SPLs) which result in discomfort. These levels are also referred to as the threshold of feeling, tickle, touch, tolerance, or pain. In general, these sensations occur at SPLs of about 120 dB or more, depending upon the laboratory and how the sensation is defined (discomfort, pain, etc.). However, this is not to imply that a pain threshold of about 140 dB SPL must be reached in order to define a subject's tolerance for intense sounds. Uncomfortable loudness has been associated with SPLs of more than 100 dB [18–20]; and Silverman et al. [21] reported discomfort to occur at about 120 dB SPL.

Standard Zero Reference Levels

One might now ask what constitutes a reasonable conception of normal hearing sensitivity for the population as a whole. That is, how much SPL does the average person who is apparently free of pathology need to detect a particular sound? The answer permits standardization of audiometric instruments so that we may quantify hearing loss relative to "what most people can hear."

*The correction accounts for an apparent error in the low-frequency MAF levels [2].

Table 9.1 Standard Reference Levels for 0 dB HL for WE 705A and TDH-39 Earphones

	Reference level (dB SPL)	
Frequency (Hz)	Western Electric WE 705A	Telephonics TDH-39
125	45.5	45
250	24.5	25.5
500	11	11.5
1000	6.5	7
1500	6.5	6.5
2000	8.5	9
3000	7.5	10
4000	9	9.5
6000	8	15.5
8000	9.5	13

Prior to 1964 several countries had their own standards for normal hearing and audiometer calibration based upon locally obtained data. For example, the 1954 British Standard [22] was based upon one group of studies [3,23], whereas the 1951 American Standard [24] reflected other findings [25,26]. Unfortunately, these standards differed by about 10 dB; and the American Standard was actually too lax at 250 and 500 Hz, and too strict at 4000 Hz [27]. This situation was rectified in 1964 with the issuance of Recommendation R389 by the International Organization for Standardization (ISO) [28]. This standard is generally referred to as ISO-1964. It was based upon a round-robin of loudness-balance and threshold experiments involving American, British, French, Russian, and West German laboratories, and as a result equivalent reference SPLs were obtained for the earphones used by each country [29].

Table 9.1 shows these SPLs for the Western Electric (WE) 705A earphones used in the ISO study. These reference levels were subsequently incorporated into the American National Standards Institute (ANSI) S3.6-1969 *American National Standard Specifications for Audiometers* [30]. Also shown in Table 9.1 are the equivalent zero reference levels for the Telephonics TDH-39 earphone [30], which is more commonly used in clinical audiometers. These are MAPC values since they represent the equivalent SPLs, measured in a 6 cc coupler, that correspond to normal threshold levels.

Hearing Level

Because each of the SPLs in Table 9.1 corresponds to minimum audibility, we may think of them as all representing the same *hearing*

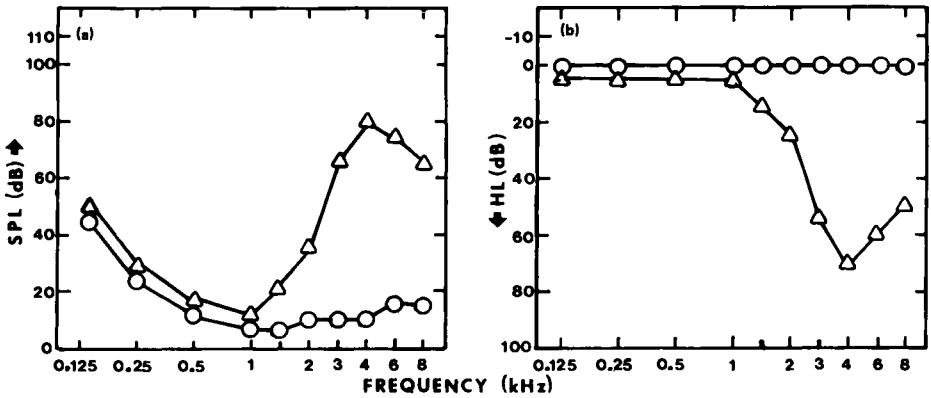

Figure 9.2 Audiograms showing normal hearing (circles) and a hearing loss in the high frequencies (triangles), expressed in (a) dB SPL and (b) dB HL. Note that intensity is shown downward on the clinical audiogram in dB HL.

level. Thus, each zero reference SPL may also be referred to as 0 dB hearing level (0 dB HL) for its respective frequency. For example, the zero reference level for a 1000 Hz tone (TDH-39 earphones) is 7 dB SPL, so that 0 dB HL corresponds to 7 dB SPL at 1000 Hz. At 250 Hz, more sound pressure is required to reach the normal zero reference threshold level, so that 0 dB HL equals 25.5 dB SPL at this frequency. The relationship between SPL and HL is exemplified in Fig. 9.2. Figure 9.2a shows the minimum audible zero reference threshold values in dB SPL as a function of frequency. As in Fig. 9.1, intensity increases upward on the y-axis. Figure 9.2b shows the same information in dB HL. Notice that the minimum audible values (0 dB HL) all lie along a straight line in terms of hearing level. In other words, the HL scale calls each zero reference SPL value "O dB HL," so that thresholds can be measured in comparison to a straight line rather than a curved one.

The graph in Fig. 9.2b is the conventional audiogram used in clinical audiology. (Actually, the term "audiogram" may legitimately be used to describe any graph of auditory sensitivity as a function of frequency.) By convention, increasing intensity (which indicates a hearing loss) is read downward on the y-axis when thresholds are plotted in dB HL.

Now, suppose that we measure the thresholds of a person whose cochlea has been partially damaged by excessive noise exposure. This kind of trauma often appears as a hearing loss in the higher frequencies. The triangles in Fig. 9.2 show the impaired threshold in terms of both SPL and HL. The difference in dB between the impaired thresholds and the reference values (circles) is the amount of hearing loss at that fre-

quency. For example, our hypothetical patient has a threshold of 5 dB HL at 1000 Hz. This means that he requires 5 dB HL to just detect the tone, as opposed to only 0 dB HL for a normal person. In SPL, this corresponds to a threshold of 12 dB (i.e., the 7 dB SPL zero reference level for 0 dB HL plus the 5 dB hearing loss. Had the threshold been 40 dB HL, the corresponding value would have been 47 dB SPL. Similarly, the 70 dB HL threshold at 4000 Hz is equivalent to 79.5 dB SPL (70 dB over the 9.5 dB SPL zero reference level). As one might expect, audiometers are calibrated to dB HL values by measuring the output of the earphone in SPL and then converting to HL by subtracting the zero reference in Table 9.1.

Effects of Duration

Thus far we have been considering tones lasting for about a second or more. From the standpoint of audition, such durations may be viewed as infinite. Auditory sensitivity is altered, however, for durations much shorter than a second. Extremely short durations, on the order of 10 msec or less, result in transients that spread energy across the frequency range. These transients will confound the result of an experiment if they are audible [31,32], so that special care is needed in the design and interpretation of studies using short durations.

The perception of tonality appears to change in a somewhat orderly manner as the duration of a very short tone burst is increased [33,34]. A click is heard at the shortest durations, then a click with tonal qualities (click pitch) at slightly longer durations. For frequencies below about 1000 Hz, a tonal pitch is perceived when the duration of the tone burst is long enough for the subject to hear several cycles (periods) of the tone. Thus, the duration threshold for tonality decreases from about 60 msec at 50 Hz to approximately 15 msec at 500 Hz. Above 1000 Hz, the threshold for tonality is essentially constant, and is on the order of about 10 msec.

Absolute sensitivity decreases when the duration of a stimulus becomes much shorter than a second; and the nature of this phenomenon reveals an interesting property of the auditory system. Although the exact results of individual experiments differ, two observations are routinely encountered [35–40]. First, for durations up to roughly 200–300 msec, a tenfold (decade) change in duration can offset an intensity change on the order of about 10 dB. In other words, reducing the duration of a tone burst at threshold from 200 msec to 20 msec (a decade reduction) reduces sensitivity to the degree that the intensity must be increased by 10 dB to reattain threshold. Alternatively, the threshold intensity decreases by about 10 dB when the duration of a tone burst is

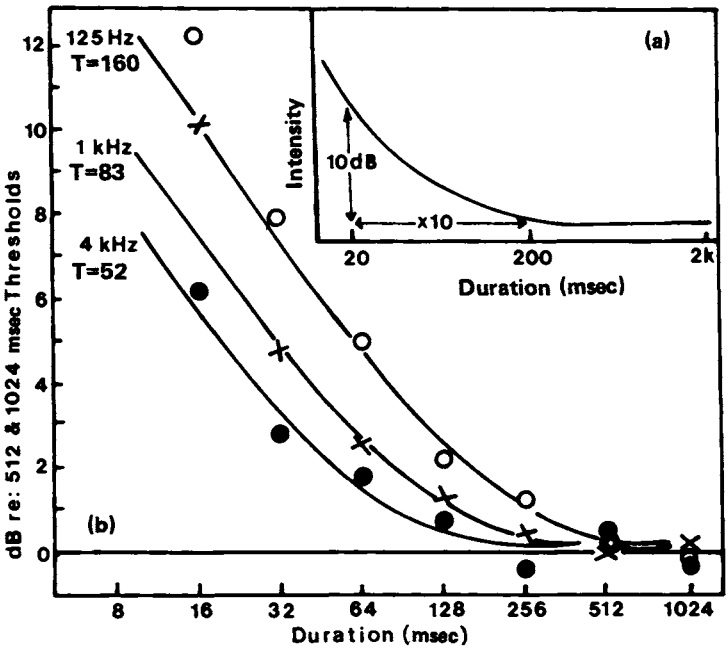

Figure 9.3 (a) Idealized temporal integration function showing that a decade change in duration is offset by an intensity change of about 10 dB up to 200–300 msec. Threshold remains essentially constant at longer durations. (b) Effect of frequency on temporal integration. (Adapted from Watson and Gengel [40]. Permission *J. Acoust. Soc. Amer.*)

increased from 20 msec to 200 msec. Second, durations longer than about 1/3 sec are treated by the ear as though they are infinitely long. That is, increasing or decreasing durations which are longer than approximately 300 msec does not change the threshold level. These observations are shown in idealized form in Fig. 9.3a.

The phenomenon under discussion is called temporal integration (or summation). It demonstrates that the ear operates as an energy detector which samples the amount of energy present within a certain time frame (or window). A certain amount of energy is needed within this time window for the threshold to be attached. This energy may be obtained by using a higher intensity for less time or a lower intensity for more time. The ear integrates energy over time *within* an integration time frame of roughly 200 msec. This interval might also be viewed as a period during which energy may be stored, and can be measured as a time consultant τ [36]. Energy available for longer periods of time is

not integrated with the energy inside the time window. This additional energy thus does not contribute to the detection of the sound, so that the threshold does not change durations longer than 200 msec. Photographers might think of this situation as analogous to the interaction of f-stop (intensity) and shutter speed (duration) in summating the light energy for a certain film speed (integration time): The lens opening and shutter speed may be traded against one another as long as the same amount of light is concentrated upon the film.

Figure 9.3b shows the effect of frequency upon temporal integration at threshold. Thresholds for shorter durations are shown relative to the threshold levels obtained for 512 msec, which are represented by the horizontal line. Notice that although temporal integration occurs for all frequencies shown, the functions become flatter (i.e., the time constant τ for integration becomes shorter) as frequency increases from 250 to 4000 Hz.

Temporal integration is observed at higher levels as well as at absolute threshold. Temporal summation of loudness is discussed in Chap. 11, and Chap. 3 covers this topic with respect to the acoustic reflex.

DIFFERENTIAL SENSITIVITY

Having examined the lower and upper bounds of hearing, we may now ask what is the smallest perceivable difference between two sounds. This quantity is called either the difference limen (DL) or the just noticeable difference (jnd); these terms will be used interchangeably in this text. The DL is the smallest perceivable difference in dB between two intensities (ΔI) or the smallest perceivable change in Hz between two frequencies (Δf). We may think of the jnd in two ways. One is as the absolute difference between two sounds, and the other is as the relative difference between them. The latter is obtained by dividing the absolute DL by the value of the starting level. Thus, if the starting level I is 1000 units and the DL ΔI is 50 units, then the relative DL $\Delta I/I$ is $50/1000 = 0.05$. This ratio, $\Delta I/I$, is called the Weber fraction.

A point about absolute versus relative DLs should be clarified before proceeding. The frequency DL Δf is an absolute difference in Hz, as opposed to the relative frequency DL obtained by dividing Δf by the starting frequency f. Suppose it is necessary to change a 1000 Hz tone (f) by a 3.6 Hz Δf in order for a particular subject to just detect the frequency difference. His absolute frequency DL is thus 3.6 Hz, whereas his relative DL is 0.0036. The situation is different, however, for the intensity DL, because we measure ΔI in dB. Since decibels are actually ratios, ΔI in dB is really a relative value. (This is why ΔI and I were

expressed as "units" in the above example.) For continuity, we shall discuss differential sensitivity for intensity in terms of the Weber fraction $\Delta I/I$.*

An important concept in psychophysics is Weber's law, which states that the value of $\Delta I/I$ (the Weber fraction) is a constant (k) regardless of the stimulus level, or

$$\Delta I/I = k$$

A classic illustration of this law is the number of candles one must add to a number of candles which are already lit in order to perceive a difference in the amount of light [41]. If 10 candles are lit, then only one more will produce a jnd of light (DL = 1). However, if there are originally 100 candles then 10 must be added to result in a perceptible difference; and to notice an increase in the light provided by 1000 candles, 100 must be added! Thus, the absolute value of the DL increases from 1 to 100, whereas the Weber fraction k has remained constant at 0.1 (since 1/10 = 10/100 = 100/1000 = 0.1), illustrating Weber's law.

Intensity Discrimination

In 1928, Riesz [42] reported on the differential sensitivity for intensity over a wide range of frequencies (35 Hz to 10 kHz) and sensation levels (0–100 dB). Riesz also overcame a major problem of previous experiments [43]: The abrupt switching on and off of the stimuli in these experiments had been marked by audible transient noises. These noises occur because abrupt switching causes energy to spread to other frequencies than the one being tested, and as a result it is unclear whether the subject is responding to the stimulus or to the audible transient. Riesz avoided this problem by using beating tones. When two tones of slightly different frequencies (e.g. 1000 and 1003 Hz) are presented simultaneously, the resulting tone will fluctuate in intensity (beat) at a rate equal to the difference in frequency between the two original tones. Thus, combining a 1000 Hz tone with a 1003 Hz tone results in three beats per second (beat frequency of 3 Hz). This method, which is essentially the same as amplitude modulation, produces no switching transients. Riesz found that DLs were optimized when the modulation rate was 3 beats/sec.

*Both $\Delta I/I$ and ΔI in dB are commonly encountered in the literature. They are related by the formula ΔI in dB = 10 log(1 + $\Delta I/I$). Thus, for $\Delta I/I$ = 1.0, ΔI in dB is 10 log(1 + 1) = 10 log 2, or 3 dB. When $\Delta I/I$ is 0.5, ΔI = 10 log(1 + 0.5) = 1.76 dB.

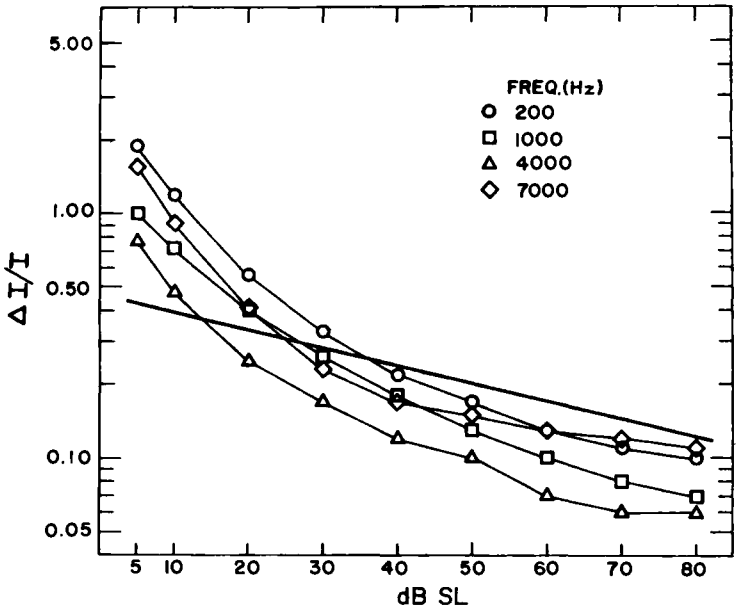

Figure 9.4 The Weber fraction ΔI/I as a function of SL. Riesz's [42] data are shown by the curves with symbols; the straight line shows the results of Jesteadt et al. (From Jesteadt, Wier, and Green [44]. Permission *J. Acoust. Soc. Amer.*)

Riesz's subjects adjusted the amplitude of one of the two beating tones until the beats became minimally audible. The intensity difference between the two tones was then taken as the measure of the DL. Figure 9.4 shows the results (ΔI/I as a function of sensation level) for several frequencies. Weber's law predicts that the data should fall along a horizontal line, indicating that ΔI/I remains constant at all stimulus levels (SLs). Contrary to Weber's law, Riesz found that ΔI/I decreased with increasing intensity, especially at low SLs. The curves became flatter at moderate and high SLs (with the Weber fraction approximating 0.3), but did not quite achieve the constant value of ΔI/I dictated by Weber's law. This discrepancy has come to be known as the "near miss" to Weber's law, and has been corroborated by subsequent studies using pulsed tones* [44–48]. Several examples are illustrated in Fig. 9.5.

The curves in Fig. 9.5 show that, while the trend is in the same direction across studies, the precise values are certainly quite variable. This variability reflects differences in methodology, stimuli, etc. Many of

*Modern electronics can produce pulsed tones without switching transients, which lend themselves to contemporary psychophysical methods (see Chaps. 7 and 8).

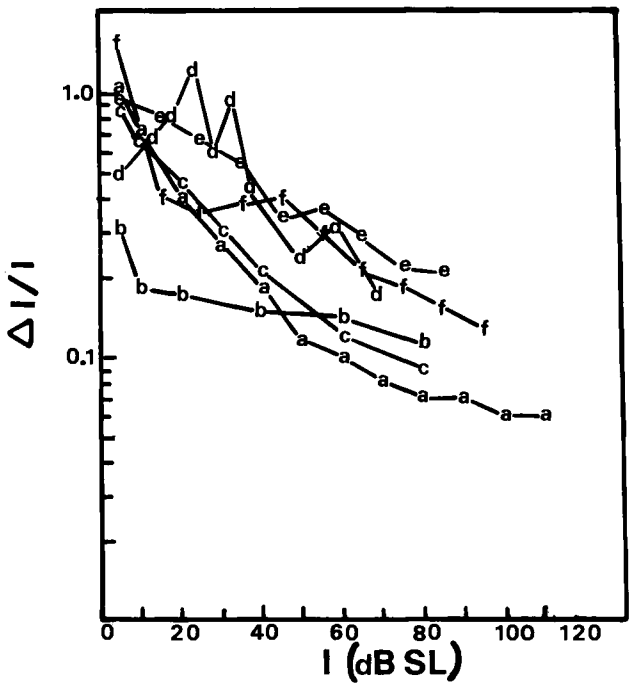

Figure 9.5 The Weber fraction as a function of SL (for 1000 Hz) from several studies: (a) Riesz [42]; (b) and (c) Harris [49]; (d) McGill and Goldberg [45]; (e) Campbell and Lasky [50]; (f) Luce and Green [51]. (Adapted from Luce and Green [51]. Permission *J. Acoust. Soc. Amer.*)

these factors have been enumerated for a large number of studies by Rabinowitz et al. [52]. They combined and summarized the results for differential sensitivity at 1000 Hz. Their results are shown in Fig. 9.6 in adapted form. Note that Weber's law holds for SLs between 10 and 40 dB (segment b in Fig. 9.6), although differential sensitivity changes with sensation level above and below this range (segments a and c). The horizontal dashed line is the function predicted by Weber's law.

Riesz [42], using the amplitude modulation method, also reported that the Weber fraction is frequency dependent: He found that $\Delta I/I$ became smaller as frequency increased from 35 Hz up to about 1000 Hz. The Weber fraction remained more or less constant for the frequencies above this, at least for SLs above 20 dB. This result has not been confirmed by subsequent studies [44,49,53,54]. In what was probably the most extensive study of the intensity DL since Riesz's classic paper, Jesteadt et al. [44] reported on $\Delta I/I$ for pulsed pure tones. They demonstrated that $\Delta I/I$ does not change with frequency, so that a straight line may be used to show $\Delta I/I$ as a function of sensation level. This result is

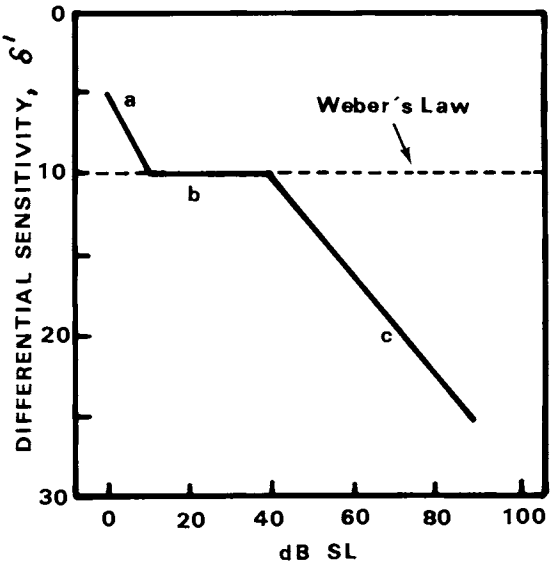

Figure 9.6 Averaged differential sensitivity normalized at 40 dB SL; see text. (δ' is related to the DL for intensity as follows: $\delta'(I) = 10/10 \log(1 + \Delta I/I)$, when $10 \log(\Delta I/I)$ produces a d' of 1.0). (Modified after Rabinowitz et al. [52].)

shown by the straight line in Fig. 9.4. Jesteadt et al. also demonstrated that the function relating $\Delta I/I$ to SL holds even for sensation levels as low as 5 dB.

In summary, differential sensitivity for intensity ($\Delta I/I$) becomes more acute with increasing sensation level in a manner which is a near miss to the constant predicted by Weber's law. Furthermore, recent research has shown that the Weber fraction remains constant across most of the audible frequency range.

Frequency Discrimination

The early work [43] on differential frequency sensitivity, like that on intensity discrimination was plagued by transient noise problems associated with the presentation of the stimuli. Shower and Biddulph [55] circumvented this problem by using frequency-modulated (FM) tones as the stimuli. In other words, the test tone was varied continuously in frequency at a rate of twice per second. The subject's task was to detect the presence of a modulated tone as opposed to a steady tone. The DL was taken as the smallest difference in frequency that produced a perceptible modulation of the original tone. Since Shower and Biddulph's

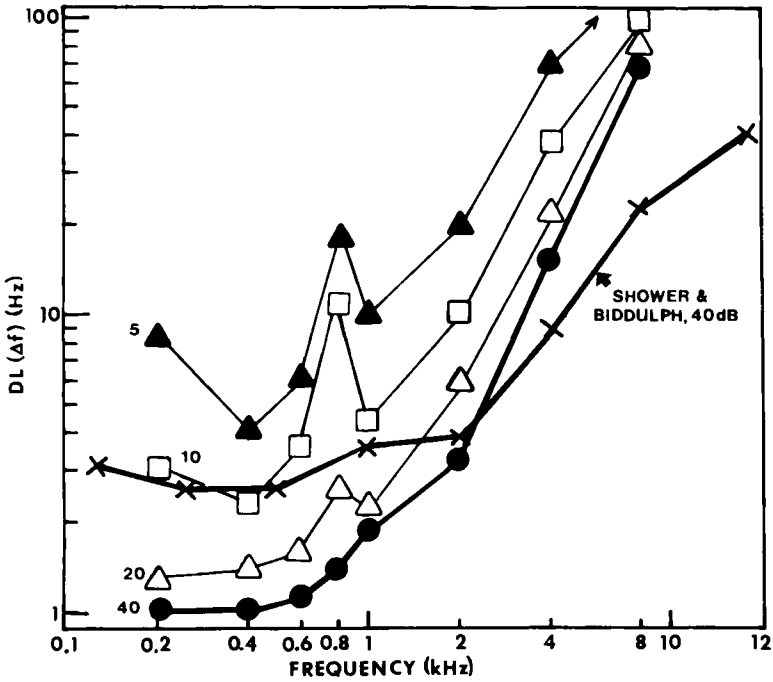

Figure 9.7 Frequency DL as a function of frequency at SLs of 5 dB (filled triangles), 10 dB (open squares), 20 dB (open triangles), and 40 dB (filled circles); based on data of Wier et al. [61]. Shower and Biddulph's [55] FM data at 40 dB SL (X's) are shown for comparison.

classic study included a very wide range of frequencies (62–11,700 Hz) and sensation levels (5–80 dB), it has remained the most widely cited study of differential frequency sensitivity for many years. However, subsequent studies using pulsed tones have generally resulted in better (smaller) DLs at low frequencies and poorer (larger) DLs at higher frequencies than were found with the FM tones [56–62]. We shall return to this point below. The most likely reason for the discrepancy is that frequency modulation results in a stimulus with a complex spectrum, so that we really cannot be sure what serves as the basis for the subject's responses.

Wier et al. [61] recently reported an extensive frequency-discrimination study using pulsed pure tones from 200 to 8000 Hz at sensation levels between 5 and 80 dB. They took the DL to be the smallest frequency difference Δf that the subject could detect 71% of the time. Figure 9.7 shows some of their results at four SLs. The important observations are that the frequency DL Δf becomes larger as frequency in-

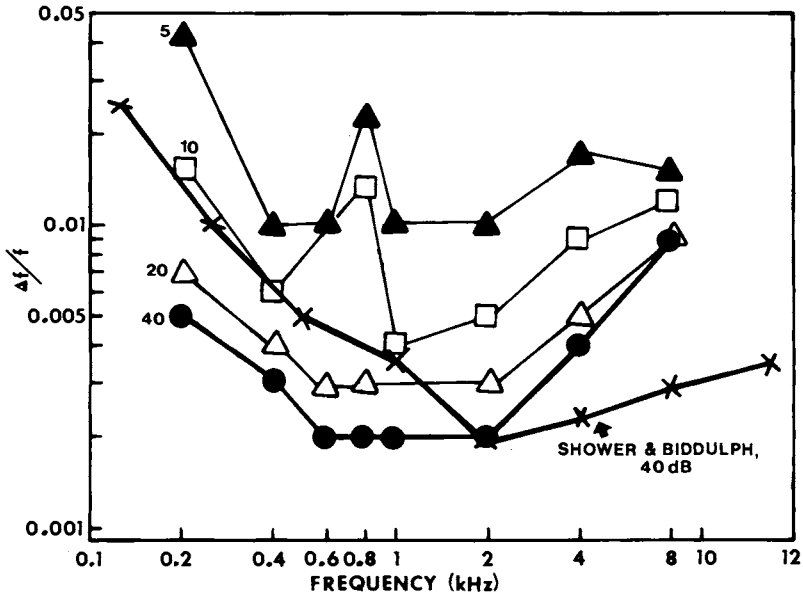

Figure 9.8 Differential sensitivity Δf/f as a function of frequency and SL. Symbols same as in Fig. 9.7.

creases, and that Δf becomes smaller as sensation level increases. The best (smallest) values of Δf — on the order of 1 Hz — occur for low frequencies presented at about 40 dB SL or more. The DL increases substantially above about 1000 Hz, so that Δf at 40 dB SL is roughly 16 Hz at 4000 Hz and 68 Hz by 8000 Hz. The figure also shows that Δf is not simply a monotonic function of frequency; it does not *always* get larger as frequency increases. We see a departure from a monotonically rising function between 200 and 400 Hz, and there are rather dramatic peaks in the curves in the vicinity of 800 Hz. Unfortunately, the origin of these peaks is unclear. We also find that sensation level is relatively more important at low frequencies than at high ones, where the curves tend to converge.

Figure 9.7 also shows the data obtained by Shower and Biddulph for FM tones at 40 dB SL (X's). Comparison of this curve with the one obtained by Wier et al. at the same SL (filled circles) demonstrates that Δf is larger for FM tones at low frequencies, smaller for FM tones at high frequencies, and about the same for the two kinds of stimuli at around 2 kHz. Other studies using pulsed tones at sensation levels between 30 and 50 dB agree with the findings of Wier et al. [56–60,62].*

*Nordmark's [59] data are in agreement with those of Wier et al. [61] when the latter are corrected for differences in experimental methodology [63].

Differential sensitivity is shown as the Weber fraction $\Delta f/f$ in Fig. 9.8. Again, the data obtained by Wier et al. are shown along with Shower and Biddulph's data at 40 dB SL for comparison. Notice that $\Delta f/f$ improves (becomes smaller) as SL increases, so that the Weber fraction approximates 0.002 for the mid-frequencies at 40 dB SL. The value of $\Delta f/f$ is relatively constant for moderate sensation levels between about 600 and 2000 Hz, but becomes larger at higher and lower frequencies. In summary, then, $\Delta f/f$ is a somewhat complex function of both frequency and intensity, unlike $\Delta I/I$ (Fig. 9.4), which appears to depend upon SL alone [44].

Temporal Discrimination

There are several interesting aspects of temporal discrimination which we shall touch upon. The importance of being able to make fine temporal discriminations should become obvious when one realizes that speech is made up of signals that change rapidly over time.

We might first ask what is the shortest period of time over which the ear can discriminate two signals. This period could be measured by asking the subject to discriminate between signals that are exactly the same except for a phase difference. Green [64–66] has referred to this quantity as temporal auditory acuity or minimum integration time. The latter phrase suggests that the ear's most sensitive temporal discriminations also provide an estimate of the shortest time period within which the ear can integrate energy. We could think of this time period as the "low end" of the scale for temporal integration, as discussed earlier in the chapter. Experiments on temporal auditory acuity have been done using clicks [67], special digital signals called Huffman sequences* [65,66,68], and tone bursts [66]. The essential finding has been that the ear can make temporal discriminations as small as about 2 msec. This value is essentially independent of frequency (although performance improves above 1000 Hz), and is not especially affected by intensity once a certain level has been attained.

The nature of the listener's task is extremely important in measuring temporal discrimination; different values are obtained when different methods are employed. Suppose, for example, that a subject is presented with two signals ("high" and "low") in rapid succession. The subject will be able to detect a difference as small as 2 msec between the onsets of the two signals if we ask him to tell whether there was one signal or two successive ones. On the other hand, he will require on the order of 20 msec in order to say which one came first. The former result

*A Huffman sequence is obtained by directing digital pulses through a digital all-pass filter composed of overlapping high- and low-pass sections. See Huffman [69].

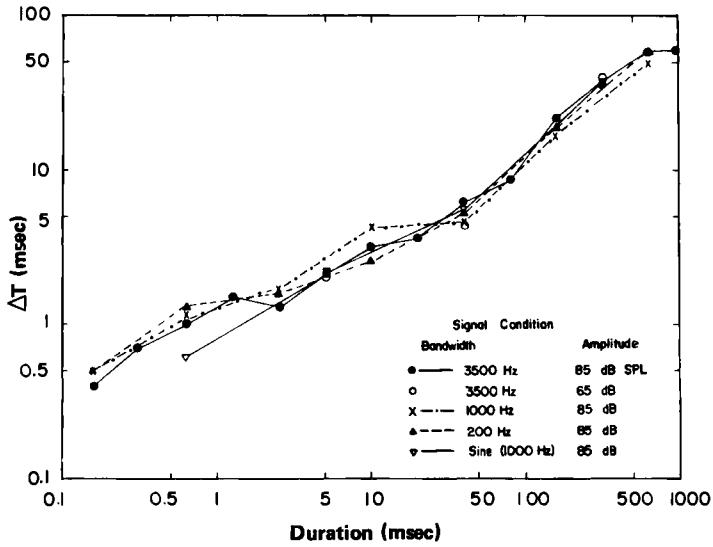

Figure 9.9 Values of ΔT as a function of duration from 0.16 to 960 msec. (From Abel [74]. Permission *J. Acoust. Soc. Amer.*)

may be viewed as an estimate of auditory successiveness (versus simultaneity), while the latter is a measure of perceived temporal order [70,71]. Highly trained subjects can identify a rapid sequence of three tone bursts for durations as small as 2–7 msec [72]. Divenyi and Hirsh [72] pointed out that auditory discrimination performance is substantially affected by at least four factors: (1) the number of stimuli in the sequence of items to be discriminated, (2) how the sequences are presented (separate presentations or continuously), (3) the kind of task which the subject must perform, and (4) the amount of training the subject has had.

We have seen that ΔI depends upon intensity and Δf upon frequency. Differential sensitivity for duration has also been investigated, with the general finding that the DL for duration (ΔT) becomes smaller as overall duration decreases [73,74]. Abel [74] studied ΔT for durations between 0.16 and 960 msec, using various bandwidths of noise from 200 to 300 Hz wide and also 1000 Hz tone bursts. She presented subjects with two intervals, one containing a standard stimulus duration T and the other containing a slightly longer duration T + ΔT. The subject listened to the two intervals (which were presented randomly), and indicated the one with the longer-duration signal. The smallest time difference correctly detected 75% of the time was taken as the DL for duration ΔT. As Fig. 9.9 shows, ΔT decreases from about 50 msec at

durations of 960 msec to on the order of 0.5 for durations of less than 0.5 msec. Differential sensitivity in terms of the Weber fraction $\Delta T/T$ is not a constant, but changes with duration so that $\Delta T/T$ is about 1.0 at 0.5–1 msec, roughly 0.3 at 10 msec, and approximately 0.1 from 50 to 500 msec. The results were essentially independent of bandwidth and intensity.

REFERENCES

1. L. J. Sivian and S. D. White, On minimal audible sound fields, *J. Acoust. Soc. Amer. 4*, 288–321 (1933).
2. M. C. Killion, Revised estimate of minimum audible pressure: Where is the "missing 6 dB"? *J. Acoust. Soc. Amer. 63*, 1501–1508 (1978).
3. R. S. Dadson and J. H. King, A determination of the normal threshold of hearing and its relation to the standardization of audiometers, *J. Laryng. Otol. 46*, 366–378 (1978).
4. D. W. Robinson and R. S. Dadson, A re-determination of the equal loudness relations for pure tones, *Brit. J. Appl. Phys. 7*, 166–181 (1956).
5. N. S. Yoewart, M. Bryan, and W. Tempest, The monaural MAP threshold of hearing at frequencies from 1.5 to 100 c/s, *J. Sound Vib. 6*, 335–342 (1967).
6. J. L. Northern, M. P. Downs, W. Rudmose, A. Glorig, and J. Fletcher, Recommended high-frequency audiometric threshold levels (8000–18000 Hz), *J. Acoust. Soc. Amer. 52*, 585–595 (1972).
7. G. Bekesy, *Experiments in Hearing*, McGraw-Hill, New York, 1960.
8. R. L. Wegel, Physical data and physiology of excitation of the auditory nerve, *Ann. Otol. 41*, 740–779 (1932).
9. W. A. Munson and F. M. Wiener, In search of the missing 6 dB, *J. Acoust. Soc. Amer. 24*, 498–501 (1952).
10. W. Rudmose, On the lack of agreement between earphone pressures and loudspeaker pressures for loudness balances at low frequencies, *J. Acoust. Soc. Amer. 35*, S1906 (1963).
11. E. Villchur, Free-field calibration of earphones, *J. Acoust. Soc. Amer. 46*, 1527–1534 (1969)
12. E. Villchur, Audiometer-earphone mounting to improve intersubject and cushion-fit reliability, *J. Acoust. Soc. Amer. 48*, 1387–1396 (1970).
13. E. Villchur, Comparison between objective and threshold-shift methods of measuring real-ear attenuation of external sound by earphones, *J. Acoust. Soc. Amer. 51*, 663–664 (1972).
14. D. W. Morgan and D. D. Dirks, Loudness discomfort level under earphone and in the free field: The effect of calibration, *J. Acoust. Soc. Amer. 56*, 172–178 (1974).
15. R. W. Stream and D. D. Dirks, Effects of loudspeaker on difference between earphone and free-field thresholds (MAP and MAF), *J. Speech Hearing Res. 17*, 549–568 (1974).

16. E. A. G. Shaw, Transformation of sound pressure level from the free field to the eardrum in the horizontal plane, *J. Acoust. Soc. Amer. 56,* 1848–1861 (1974).
17. International Organization for Standards, R226-1961, Normal equal-loudness contours for pure tones and normal threshold of hearing under free-field listening conditions, 1961.
18. J. D. Hood, Observations upon the relationship of loudness discomfort level and auditory fatigue to sound-pressure-level and sensation level, *J. Acoust. Soc. Amer. 44,* 959–964 (1968).
19. J. D. Hood and J. P. Poole, Tolerable limit of loudness: Its clinical and physiological significance, *J. Acoust. Soc. Amer. 40,* 47–53 (1966).
20. J. D. Hood and J. P. Poole, Investigations on hearing upon the upper physiological limit of normal hearing, *Intl. Aud. 9,* 250–255 (1970).
21. S. R. Silverman, C. E. Harrison, and H. S. Lane, Tolerance for pure tones and speech in normal and hard-of-hearing ears, OSRD Rept. 6303, Central Institute for the Deaf, Missouri (1946).
22. British Standards Institute, Standard 2497, The normal threshold of hearing for pure tones by earphone listening, 1954.
23. L. J. Wheeler and E. D. D. Dickson, The determination of the threshold of hearing, *J. Laryng. Otol. 46,* 379–395 (1952).
24. American Standards Association, ASA-Z24.5-1951, Specifications for audiometers for general diagnostic purposes, 1951.
25. U.S. Public Health Service, National Health Survey: Preliminary Reports, *Hearing Study Series Bulletins* 1–7, 1935–1936.
26. J. C. Steinberg, H. C. Montgomery, and M. B. Gardner, Results of the World's Fair hearing tests, *J. Acoust. Soc. Amer. 12,* 291–301 (1940).
27. H. Davis and F. W. Krantz, The international standard reference zero for pure-tone audiometers and its relation to the evaluation of impairment of hearing, *J. Speech Hearing Res. 7,* 7–16 (1964).
28. International Organization for Standards, Recommendation R389, Standard reference zero for the calibration of pure-tone audiometers, 1964.
29. P. G. Weissler, International standard reference for audiometers, *J. Acoust. Soc. Amer. 44,* 264–275 (1968).
30. American National Standards Institute, ANSI-S3.6-1969(R1973), American national standard specifications for audiometers, 1969, 1973.
31. H. N. Wright, Audibility of switching transients, *J. Acoust. Soc. Amer. 32,* 138 (1960).
32. H. N. Wright, An artifact in the measurement of temporal summation at the threshold of hearing, *J. Speech Hearing Dis. 32,* 354–359 (1967).
33. W. Bürck, P. Kotowski, and H. Lichte, Die Lautstärke von Knacken. Geräuchen und Tönen, *Elek. Nachr.-Tech. 12,* 278–288 (1935).
34. J. M. Doughty and W. R. Garner, Pitch characteristics of short tones. I. Two kinds of pitch threshold, *J. Exp. Psych. 37,* 351–365 (1947).
35. J. W. Hughes, The threshold of audition for short periods of stimulation, *Proc. Roy. Soc. B133,* 486–490 (1946).
36. R. Plomp and M. A. Bouman, Relation between hearing threshold and duration for pure tones, *J. Acoust. Soc. Amer. 31,* 749–758 (1959).

37. J. Zwislocki, Theory of auditory summation, *J. Acoust. Soc. Amer. 32*, 1046–1060 (1960).
38. A. M. Small, J. F. Brandt, and P. G. Cox, Loudness as a function of signal duration, *J. Acoust. Soc. Amer. 34*, 513–514 (1962).
39. R. S. Campbell and S. A. Counter, Temporal integration and periodicity pitch, *J. Acoust. Soc. Amer. 45*, 691–693 (1969).
40. C. S. Watson and R. W. Gengel, Signal duration and signal frequency in relation to auditory sensitivity, *J. Acoust. Soc. Amer. 46*, 989–997 (1969).
41. I. J. Hirsh, *The Measurement of Hearing*, McGraw-Hill, New York, 1952.
42. R. R. Riesz, Differential intensity sensitivity of the ear for pure tones, *Physiol. Rev. 31*, 867–875 (1928).
43. V. O. Knudsen, The sensibility of the ear to small differences in intensity and frequency, *Physiol. Rev. 21*, 84–103 (1923).
44. W. Jesteadt, C. C. Wier, and D. M. Green, Intensity discrimination as a function of frequency and sensation level, *J. Acoust. Soc. Amer. 61*, 169–177 (1977).
45. W. J. McGill and J. P. Goldberg, Pure-tone intensity discrimination as energy detection, *J. Acoust. Soc. Amer. 44*, 576–581 (1968).
46. W. J. McGill and J. P. Goldberg, A study of the near-miss involving Weber's law and pure tone intensity discrimination, *Percept. Psychophys. 4*, 105–109 (1968).
47. N. F. Viemeister, Intensity discrimination of pulsed sinusoids: The effects of filtered noise, *J. Acoust. Soc. Amer. 51*, 1265–1269 (1972).
48. B. C. J. Moore and D. H. Raab, Pure-tone intensity discrimination: Some experiments relating to the "near-miss" to Weber's law, *J. Acoust. Soc. Amer. 55*, 1049–1054 (1974).
49. J. D. Harris, Loudness Discrimination, *J. Speech Hearing Dis.*, Mono. Suppl. *11* (1963).
50. R. A. Campbell and E. Z. Lasky, Masker level and sinusoidal signal detection, *J. Acoust. Soc. Amer. 42*, 972–976 (1967).
51. R. D. Luce and D. M. Green, Neural coding and physiological discrimination data, *J. Acoust. Soc. Amer. 56*, 1554–1564 (1974).
52. W. M. Rabinowitz, J. S. Lim, L. D. Braida, and N. I. Durlach, Intensity perception: VI. Summary of recent data on deviations from Weber's law for 1000-Hz tone pulses, *J. Acoust. Soc Amer. 59*, 1506–1509 (1976).
53. P. N. Schacknow and D. H. Raab, Intensity discrimination of tone bursts and the form of the Weber function, *Percept. Psychophys. 14*, 449–450 (1973).
54. M. J. Penner, B. Leshowitz, E. Cudahy, and G. Richard, Intensity discrimination for pulsed sinusoids of various frequencies, *Percept. Psychophys. 15*, 568–570 (1974).
55. E. G. Shower and R. Biddulph, Differential pitch sensitivity of the ear, *J. Acoust. Soc. Amer. 3*, 275–287 (1931).
56. J. D. Harris, Pitch discrimination, *J. Acoust Soc. Amer. 24*, 750–755 (1952).
57. W. A. Rosenblith and K. N. Stevens, On the DL for frequency, *J. Acoust. Soc. Amer. 25*, 980–985 (1953).
58. G. B. Henning, Frequency discrimination in noise, *J. Acoust. Soc. Amer. 41*, 774–777 (1967).

59. J. O. Nordmark, Mechanisms of frequency discrimination, *J. Acoust. Soc. Amer. 44*, 1533–1540 (1968).
60. B. C. J. Moore, Frequency difference limens for short duration tones, *J. Acoust. Soc. Amer. 54*, 610–619 (1973).
61. C. C. Wier, W. Jesteadt, and D. M. Green, Frequency discrimination as a function of frequency and sensation level, *J. Acoust. Soc. Amer. 61*, 178–184 (1977).
62. W. Jesteadt and C. C. Wier, Comparison of monaural and binaural discrimination of intensity and frequency, *J. Acoust. Soc. Amer. 61*, 1599–1603 (1977).
63. C. C. Wier, W. Jesteadt, and D. M. Green, A comparison of method-of-adjustment and forced-choice procedures in frequency discrimination, *Percept. Psychophys. 19*, 75–79 (1976).
64. D. M. Green, Temporal auditory acuity, *Psych. Rev. 78*, 540–551 (1971).
65. D. M. Green, Temporal integration time, in *Basic Mechanisms in Hearing* (A. R. Møller, Ed.), Academic, New York, 1973, pp. 829–846.
66. D. M. Green, Temporal acuity as a function of frequency, *J. Acoust. Soc. Amer. 54*, 373–379 (1973).
67. D. A. Ronken, Monaural detection of a phase difference between clicks, *J. Acoust. Soc. Amer. 47*, 1091–1099 (1970).
68. J. Patterson and D. M. Green, Discrimination of transient signals having identical energy spectra, *J. Acoust. Soc. Amer. 48*, 894–905 (1970).
69. D. A. Huffman, The generation of impulse-equivalent pulse trains, *Trans IEEE IT8*, S10–S16 (1962).
70. I. J. Hirsh, Auditory perception of temporal order, *J. Acoust. Soc. Amer. 31*, 759–767 (1959).
71. I. J. Hirsh and C. E. Sherrick, Jr., Perceived order in different sense modalities, *J. Exp. Psych. 62*, 423–432 (1961).
72. P. L. Divenyi and I. J. Hirsh, Identification of temporal order in three-tone sequences, *J. Acoust. Soc. Amer. 56*, 144–151 (1974).
73. A. M. Small, Jr. and R. A. Campbell, Temporal differential sensitivity for auditory stimuli, *Amer. J. Psych. 75*, 401–410 (1962).
*74. S. M. Abel, Duration discrimination of noise and tone bursts, *J. Acoust. Soc. Amer. 51*, 1219–1223 (1972).

10
Masking

The last chapter dealt with auditory sensitivity. This one is concerned with how sensitivity for one sound is affected by the presence of another sound.

Suppose that the threshold for a sound A is found to be 40 dB SPL. A second sound B is then presented and the threshold of A is measured again, but this time in the presence of sound B. We now find that sound A must be presented at, say, 52 dB in order to be detected. In other words, sound A has a threshold of 40 dB when measured in quiet, but of 52 dB when determined in the presence of sound B. This increase in the threshold of one sound in the presence of another is called masking [1]. Our definition of masking may be expanded to include the reduction in loudness that can occur when a second sound is presented, a process referred to as partial masking [2,3].

We may use the word "masking" to denote either the threshold shift per se or the amount (in dB) by which the threshold is raised. Thus, sound A has been *masked* by sound B; and the amount of *masking* due to the presence of B is equal to 52 dB − 40 dB, or 12 dB. In this case, 40 dB is the *unmasked threshold* and 52 dB is the *masked threshold,* and 12 dB is the *amount of masking.* We may also say that since B causes 12 dB of masking to occur for A, the *effective level* of B is 12 dB. Finally, we may adopt the convention of calling sound B the *masker,* and sound A the *maskee,* test signal, or probe signal. As will become obvious, masking not only tells us about how one sound affects another, but also provides insight into the frequency-resolving power of the ear. This is the case because the mask-

ing pattern to a large extent reflects the excitation pattern along the basilar membrane. In Chap. 13 we shall see how masking is modified under certain conditions of binaural hearing.

The basic masking experiment is really quite straightforward. First, the unmasked threshold of the test stimulus is determined and recorded. This unmasked threshold becomes the baseline. Next, the masker is presented to the subject at a fixed level. The test stimulus is then presented to the subject and its level is adjusted (by whatever psychoacoustic method is being used) until its threshold is determined in the presence of the masker. This level is the masked threshold. As just described, the amount of masking is simply the difference in decibels between this masked threshold and the previously determined unmasked (baseline) threshold. This procedure may then be repeated for all parameters of the test stimulus and masker. An alternative procedure is to present the test stimulus at a fixed level and then to vary the masker level until the stimulus is just audible (or just masked).

MONAURAL MASKING

The masking produced by a particular sound is largely dependent upon its intensity and spectrum. Let us begin with pure tones, which have the narrowest spectra. As early as 1894, Mayer [4] had reported that, while low-frequency tones effectively mask higher frequencies, higher frequencies are not good maskers of lower frequencies. Masking, then, is not necessarily a symmetric phenomenon. This spread of masking to frequencies higher than that of the masker has been repeatedly demonstrated for tonal maskers [5–8]. We must therefore focus our attention not only upon the amount of masking, but also upon the frequencies at which masking occurs.

Figure 10.1 shows a series of masking patterns (sometimes called masking audiograms) obtained by Ehmer [6]. Each panel shows the amount of masking produced by a given pure tone masker presented at different intensities. In other words, each curve shows as a function of maskee frequency the masked thresholds of the maskee which are produced by a given masker presented at a given intensity. Masker frequency is indicated in each frame and masker level is shown near each curve. Several observations may be made from these masking patterns. First, the strongest masking occurs in the immediate vicinity of the masker frequency; the amount of masking tapers with distance from this "center" frequency. Second, masking increases as the intensity of the masker is raised.

The third observation deals with how the masking pattern depends upon the intensity and frequency of the masker. Concentrate for the moment upon the masking pattern produced by the 1000 Hz masker.

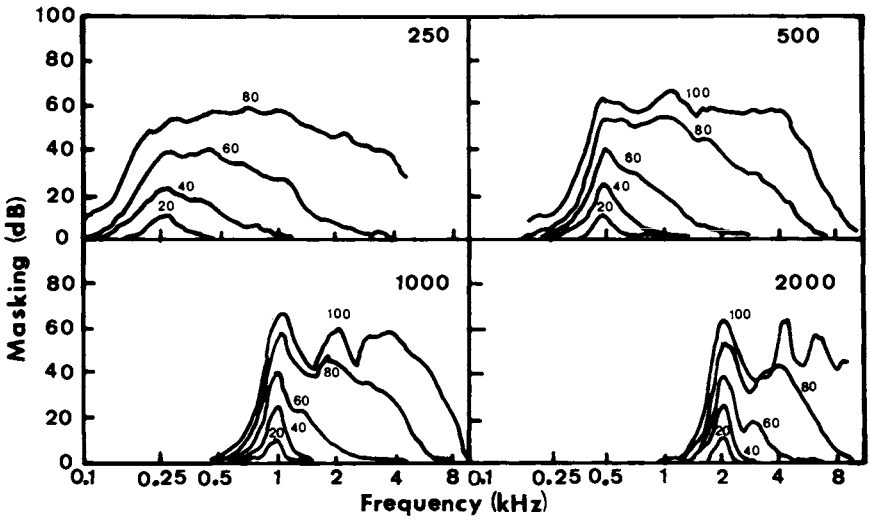

Figure 10.1 Masking patterns produced by various pure tone maskers (masker frequency indicated in each frame). Numbers on curves indicate masker level. (Adapted from Ehmer [6]. Permission *J. Acoust. Soc. Amer.*)

Note that the masking is quite symmetric around the masker frequency for relatively low masker levels (20 and 40 dB). However, the masking patterns become asymmetrically wider with increasing masker intensity, with the greatest masking occurring for tones higher than the masker frequency, but with very little masking at lower frequencies. Thus, as masker intensity is raised, there is considerable spread of masking upward in frequency but minimal downward spread. Note too that there are peaks in some of the masking patterns corresponding roughly to the harmonic of the masker frequency. Actually, however, these peaks are probably not due to aural harmonics (see Chap. 12) because they do not correspond precisely to multiples of the masker [6,7]. Small [7] found that these peaks occurred when the masker frequency was about 0.85 times the test tone frequency.

Finally, notice that the masking patterns are very wide for low-frequency maskers, and are considerably more restricted for high-frequency maskers. In other words, high-frequency maskers are only effective over a relatively narrow frequency range in the vicinity of the masker frequency, but low frequencies tend to be effective maskers over a very wide range of frequencies.

These masking patterns reflect the activity along the basilar membrane. Recall from Chap. 4 that the traveling wave envelope has a gradually increasing amplitude along its basal (high-frequency) slope,

reaches a peak, and then decays rapidly with a steep apical (low-frequency) slope. It is thus expected that higher (more basal) frequencies would be most affected by the displacement pattern caused by lower-frequency stimuli. In addition, the high-frequency traveling wave peaks and "decays away" fairly close to the basal turn, so that its masking effect would be more restricted. Lower frequencies, on the other hand, produce basilar membrane displacements along most of the partition. In addition, the excitation pattern becomes wider as signal level increaes. Recall from Chap. 5 that the tuning curves of the auditory nerve fibers reflect this basilar membrane activity. Recent masking experiments by Zwicker [9] and Zwicker and Schorn [10] employed test tones (maskees) presented at fixed frequencies and intensities as an analogue of the neural response area. The masker tone was then varied in level as a function of frequency, and the resulting masking patterns appeared as psychoacoustic tuning curves quite similar to the neural tuning curves shown in Chap. 5.

Although a great deal of information about masking has been derived from studies using tonal maskers, difficulties become readily apparent when both masker and test stimulus are tones. As we shall see in greater detail in Chap. 12, two tones differing by only a few cycles per second (e.g. 1000 and 1003 Hz) will produce audible fluctuations heard as "beats." Therefore, when the masker and test tones are very close in frequency, one cannot be sure whether the subject has responded to the beats or to the test tone. These audible beats result in notches at the peaks of the masking patterns [5] when masker and maskee are close in frequency. The situation is further complicated because combination tones result when two pure tones are presented together (Chap. 12). These problems may be eliminated by replacing the tonal maskers with narrow bands of noise centered around given frequencies [11–13]. The results of such narrow-band masking experiments have essentially confirmed the masking patterns generated in the tonal masking studies.

We have seen that upward spread of masking is the rule as masker level is increased. However, a very interesting phenomenon appears when the stimulus level is quite high, for example, at spectrum levels* of

*"Spectrum level" refers to the power in a one-cycle-wide band. In other words, spectrum level is level per cycle. It may be computed by subtracting 10 times the log of the bandwidth from the overall power in the band. Thus:

$$dB_{\text{spectrum level}} = dB_{\text{overall}} - 10 \log(\text{bandwidth}).$$

If bandwidth is 10 kHz and overall power is 95 dB, then the spectrum level is 95 − 10 log 10,000, or 95 − 40 = 55 dB.

about 60 to 80 dB. Bilger and Hirsh [14] found that threshold shifts (masking) also occurred at *low* frequencies when higher-frequency noise bands (maskers) were presented at intense levels. This is called "remote masking" because the threshold shifts occur at frequencies below and remote from the masker. In general, the amount of remote masking increases when the bandwidth of the masking noise is widened or its spectrum level is raised [15]. We might expect that remote masking is the result of the low-frequency effect of the acoustic reflex (see Chap. 3) since it occurs as higher masker levels. However, remote masking has been reported to occur even in the absence of the acoustic reflex [16]. Thus, although there is support for some contribution of the reflex to remote masking [17], remote masking is most likely due primarily to envelope detection of distortion products generated within the cochlea at high masker intensities [18–20]. (See Chap. 4 for a discussion of cochlear distortion.)

It is apparent from Fig. 10.1 that masking increases as the level of the masker is raised. We may now ask how the amount of masking relates to the intensity of the masker. In other words, how much of a threshold shift results when the masker level is raised by a given amount? This question was addressed in the classical studies of Fletcher [21] and Hawkins and Stevens [22]. Since the essential findings of the two studies agreed, let us concentrate upon the data reported by Hawkins and Stevens in 1950. They measured the threshold shifts for pure tones (and for speech) produced by various levels of a white noise masker. (It should be pointed out that although "white" noise connotes equal energy at all frequencies, the actual spectrum reaching the subject is shaped by the frequency response of the earphone or loudspeaker used to present the signal. Therefore, the exact masking patterns produced by a white noise depend upon the transducer employed, as well as on bandwidth effects that will be discussed in the next section.)

Figure 10.2 shows Hawkins and Stevens' data as masked threshold contours. These curves show the masked thresholds produced at each frequency by a white noise presented at various spectrum levels. The curves have been idealized in that the actual results were modified to reflect the masking produced by a true white noise. The actual data were a bit more irregular, with peaks in the curves at around 7 kHz, reflecting the effects of the earphone used. The bottom contour is simply the unmasked threshold curve. The essential fidning is that these curves are parallel and spaced at approximately 10 dB intervals, which is also the interval between the masker levels. This result suggests that a 10 dB increase in masker level produces a 10 dB increase in masked threshold; a point which will become clearer soon.

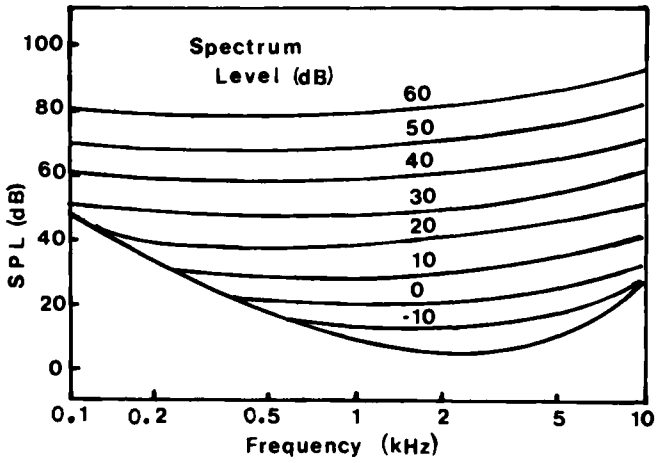

Figure 10.2 Masking contours showing masking as a function of frequency for various spectrum levels of an idealized white noise. Bottom curve is threshold in quiet. (Adapted from Hawkins and Stevens [22]. Permission *J. Acoust. Soc. Amer.*)

The actual amount of masking may be obtained by subtracting the unmasked threshold (in quiet) from the masked threshold. For example, the amount of masking produced at 1000 Hz by a white noise with a spectrum level of 40 dB is found by subtracting the 1000 Hz threshold in quiet (about 7 dB SPL) from that in the presence of the 40 dB noise spectrum level (roughly 58 dB). Thus, the amount of masking is $58 - 7 = 51$ dB in this example. Furthermore, since the masked thresholds are curved rather than flat, the white noise is not equally effective at all frequencies. We might therefore express the masking noise in terms of its effective level at each frequency. We may now show the amount of masking as a function of the effective level of the masking noise (Fig. 10.3). As Fig. 10.3 shows, once the masker attains an effective level, the amount of masking is a linear function of masker level. That is, a 10 dB increase in masker level results in a corresponding 10 dB increase in the masked threshold of the test signal. Hawkins and Stevens demonstrated that this linear relationship between masking and masker level is independent of frequency (as shown in the figure), and that it applies to speech stimuli as well as to pure tones.

CRITICAL BANDS AND RATIOS

Since a tone may be masked by another tone or by a narrow band of noise as well as by white noise, it is reasonable to ask how much of the

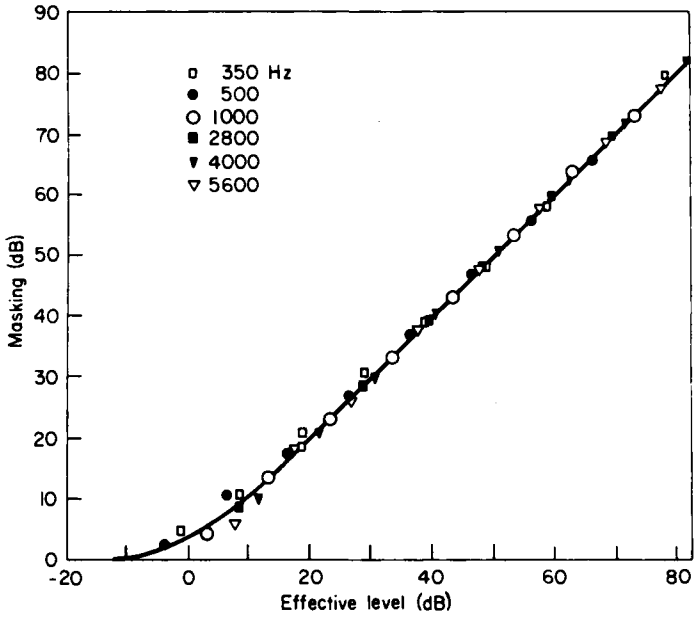

Figure 10.3 Masking produced at various frequencies as a function of the effective level of the masker. (Adapted from Hawkins and Stevens [22]. Permission *J. Acoust. Soc. Amer.*)

white noise actually contributes to the masking of a tone. Otherwise stated, does the entire bandwidth of the white noise contribute to the masking of a given tone, or is there a certain limited ("critical") bandwidth around the tone that alone results in masking? Fletcher [23] attacked this problem by finding for tones the masked thresholds produced by various bandwidths of noise centered around the test tones. He held the spectrum level constant, and found that the masked threshold of a tone increased as the bandwidth of the masking noise was widened. However, once the noise band reached a certain critical bandwidth, further widening of the band did not result in any more masking of the tone. Thus, Fletcher demonstrated that only a certain critical bandwidth within the white noise actually contributes to the masking of a tone at the center of that band, a finding which has been repeatedly confirmed [24–28].

This finding is easily understood if we think of the critical bandwidth as a filter. More and more of the energy in the white noise will be

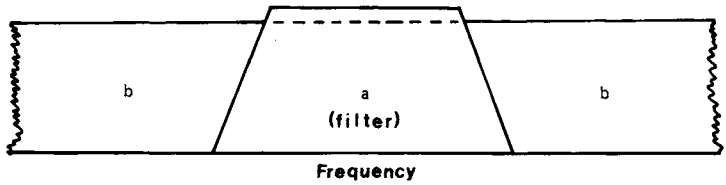

Figure 10.4 Energy within the filter (a) contributes to the masking of the tone at the center, whereas energy outside of the filter (b) does not contribute to the masking (see text).

made available by the filter as the filter's bandwidth is widened. On the other hand, energy present in the white noise that lies above and below the upper and lower cutoff frequencies of the filter is "wasted" from the standpoint of the filter (Fig. 10.4). Now, if this filter defines the critical bandwidth that contributes to the masking of a tone at the center of the band, then it is easy to see how only that portion of the noise that is inside the filter will be useful in masking the tone. Adding to the noise band beyond the limits of this filter (the areas labeled "b" in Fig. 10.4) will not add any more masking, although it will cause the noise to sound louder (see Chap. 11).

Fletcher [23] hypothesized that the signal power (S) would be equal to the noise power (N_0) located within the critical bandwidth (CB) when the tone was at its masked threshold: $S = CB \cdot N_0$. Thus, the critical band would be equal to the ratio of the signal power to the noise power; or $CB = S/N_0$. In decibels, this corresponds to $dB_S - dB_{N_0}$. Hawkins and Stevens [22] found that the masked threshold of a 1000 Hz tone was approximately 58 dB in the presence of a white noise whose spectrum level was 40 dB. The resulting estimate of the critical band is therefore 58 dB − 40 dB = 18 dB, which corresponds to a bandwidth of 63.1 Hz. This estimate of the critical band is shown by the X in Fig. 10.5. Notice that this indirect estimate of the critical bandwidth based upon the power ratio of signal to noise is actually quite a bit narrower than the other more direct estimates of the critical band shown in the figure. For this reason, the indirect estimates based upon Fletcher's formula are referred to as *critical ratios,* as opposed to the *critical bands* obtained by other, direct means. Good correspondence to the critical band is obtained when the critical ratio is multiplied by a factor of 2.5 [30]. This correspondence is demonstrated by the open circles in Fig. 10.5, which are the values of the critical ratios multiplied by 2.5 (based upon

Hawkins and Stevens' [22] data).* Note that there is good agreement with Greenwood's [26] masking data, as well as with the critical bands directly derived from loudness studies (see Chap. 11).

Figure 10.5 indicates that the critical band becomes wider as the center frequency increases. Scharf [29] has provided a table of critical bandwith estimates based upon the available data. Examples are a critical bandwidth of 100 Hz for a center frequency of 250 Hz, a 160 Hz band for 1000 Hz, and a 700 Hz band for 4000 Hz. Actually, one should be careful not to conceive of a series of discrete critical bands laid as it were end to end, but rather of a bandwidth around any particular frequency that defines the phenomenon we have discussed with respect to that frequency. (One should remember in this context Scharf's [29, p. 159] elegant definition: "the critical band is that bandwidth at which subjective responses rather abruptly change.") We should thus think of critical bandwidths as overlapping filters rather than as discrete, contiguous filters.

It would appear that the original concept [23] of the critical bandwidth as defining an internal auditory filter is fundamentally correct. Its location is more than likely peripheral, with critical bandwidths probably corresponding to 1–2 mm distances along the human cochlear partition [29]. Thus, the critical band presents itself as a fundamental concept in the frequency-analysis capability of the cochlea, the physiological aspects of which are discussed in Chap. 4. The shape of this internal filter is still unsubstantiated; however, curves by Patterson [32] account for a good deal of the available data. For detailed coverage of the critical band concept, the interested reader is referred to the work of Scharf [29] and Bilger [31].

*Bilger [31] proposed that the listener performs an intensity discrimination between the noise power in the critical band and the combined power of the noise plus signal at the masked threshold; as a result the critical ratio is equated to the familiar Weber fraction (Chap. 9):

$$\frac{S}{CB \cdot N} = \frac{\Delta I}{I}$$

This equation is solved for critical bandwidth by multiplying S/N by the reciprocal of the Weber fraction

$$CB = \frac{S}{N} \cdot \frac{I}{\Delta I}$$

Since the critical ratio is multiplied by 2.5 to obtain the critical band, this leads to a Weber fraction of 1/2.5 = 0.4, or a difference limen of 1.46 dB, a value that is in reasonable agreement with intensity DL data.

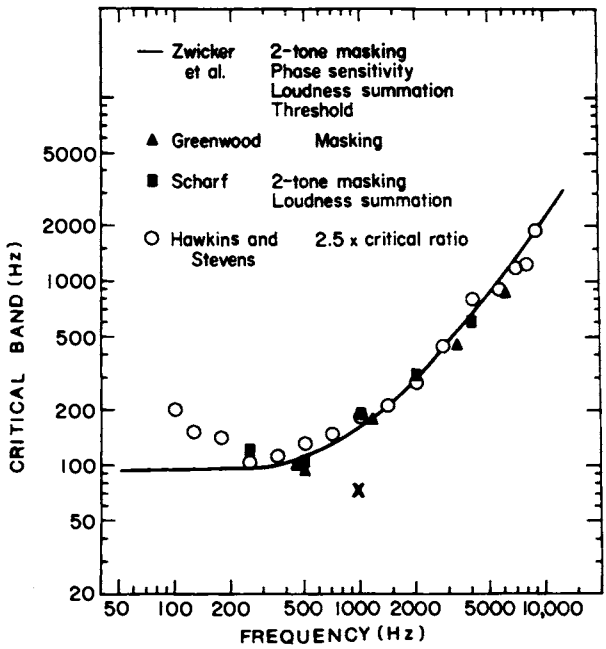

Figure 10.5 Critical bandwidth as a function of center frequency for various studies. The X is the critical ratio estimate of Hawkins and Stevens [22] for a 1000 Hz tone. [Adapted from Scharf [29], Critical bands, in *Foundations of Modern Auditory Theory* (J. V . Tobias, Ed.), Vol. 1, ©1970, Academic Press, New York.]

CENTRAL MASKING

The typical arrangement of a masking experiment involves presenting both the masker and the test stimulus to the same ear. Up to now, we have been discussing this ipsilateral type of masking. Another approach is to present the masker to one ear and the test signal to the opposite ear. Raising the intensity of the masker will eventually cause the masker to become audible in the other ear, in which case it will mask the test stimulus (a process known as cross-hearing or contralateralization of the masker). This is actually a case of ipsilateral masking as well, because it is the amount of masker that crosses the head, so to speak, that causes the masking of the signal. However, it has been demonstrated that a masker presented to one ear can cause a threshold shift for a signal at the other ear even when the masker level is too low for it to cross over to the signal ear [33–39]. This contralateral effect of the masker is most likely due to

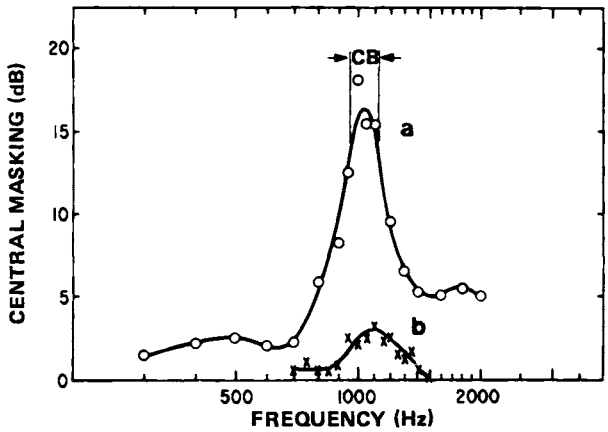

Figure 10.6 Central masking produced by a 1000 Hz tonal masker at 60 dB SL for an individual subject. Curve a is for a masker and maskee pulsed on and off together; curve b is for a continuously on masker and a pulsed signal. (Adapted from Zwislocki et al. [39]. Permission *J. Acoust. Soc. Amer.*)

an interaction of the masker and test signal within the central nervous system, probably at the level of the superior olivary complex where bilateral representation is available [40].

Central masking is in some ways similar to, yet in other ways quite different from, the monaural (direct, ipsilateral) masking discussed earlier. In general, the amount of threshold shift produced by central masking is far less than by monaural masking; and more central masking occurs for higher-frequency tones than for low. The amount of masking is greatest at masker onset and decays to a steady-state value within about 200 msec. Of particular interest is the frequency dependence of central masking. The greatest amount of central masking occurs when the masker and test tones are close together in frequency. This frequency dependence is shown rather clearly in Fig. 10.6, in which the masker is a 1000 Hz tone presented at a sensation level of 60 dB to the opposite ear. Note that the most masking occurs in a small range of frequencies around the masker frequency. This frequency range is quite close to critical bandwidth (CB). As the figure also shows, more central masking results when the masker and test tones are pulsed on and off together (curve a) rather than when the masker is continuously on and the signal is pulsed in the other ear (curve b). This is a finding common to most central masking experiments, although the amount of masking produced by a given masker level varies among studies and between subjects in the same study. Furthermore, central masking increases as

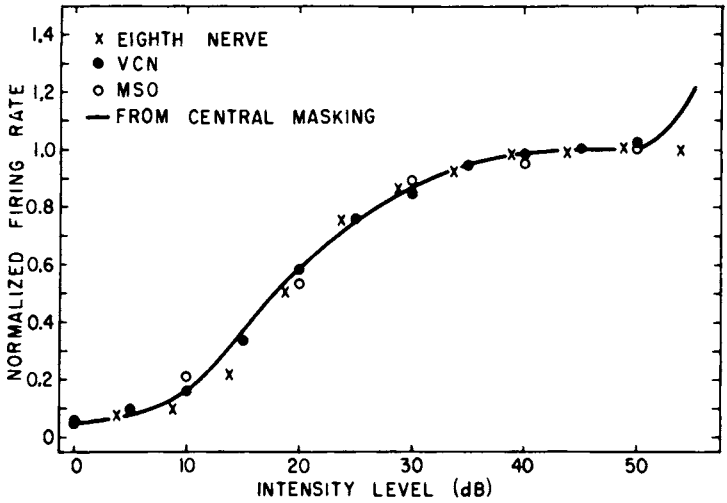

Figure 10.7 Relationship between actually obtained firing rates in the auditory (eighth) nerve, ventral cochlear nucleus (VCN), and medial superior olive (MSO), and rates predicted from central masking data. [From Zwislocki [41], In search of physiological correlates of psychoacoustic characteristics, in *Basic Mechanisms in Hearing* (A. R. Møller, Ed.), ©1973, Academic Press, New York.]

the level of the masker is raised only for the pulsed masker/pulsed signal arrangement, whereas the amount of masking produced by the continuous masker/pulsed signal paradigm remains between about 1 and 2 dB regardless of masker level.

An excellent review of the relationships between the psychophysical and electrophysiological correlates of central masking may be found in two papers by Zwislocki [40,41]. An example is shown in Fig. 10.7, which demonstrates the firing rates of neurons at various levels of the lower auditory nervous system (see Chaps. 5 and 6), and those predicted on the basis of central masking data. With few exceptions, the agreement shown in the figure for the intensity parameter also holds for frequency and time. Thus, central masking is shown to be related to activity in the lower auditory pathways.

TEMPORAL MASKING

In defining masking as the threshold shift for one sound due to the presence of another sound, we have been considering situations in which

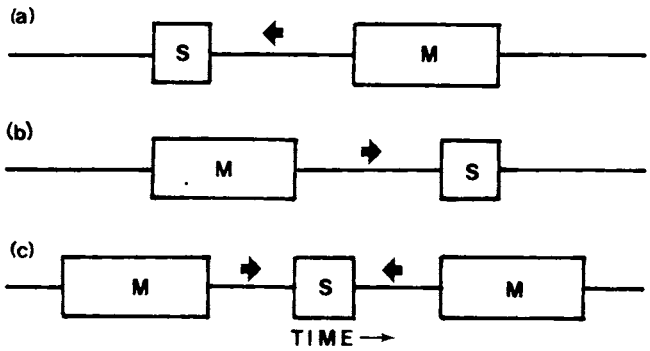

Figure 10.8 Temporal masking paradigms: (a) In backward masking the masker (M) follows the signal (S). (b) In forward masking the masker precedes the signal. (c) Combined forward and backward masking. The heavy arrows show the direction of the masking effect.

the masker and test signal occur simultaneously. Let us now extend our viewpoint to include masking that occurs when the test signal and masker do not overlap in time: temporal masking. This phenomenon may be better understood with reference to the diagrams in Fig. 10.8. In Fig. 10.8a the signal (probe) is presented and terminated, and then the masker is presented after a brief time delay following signal offset. Masking occurs in spite of the fact that the signal and masker are not presented together. This arrangement is called backward masking because the masker is preceded by the signal; i.e., the masking effect occurs backward in time (as shown by the arrow in the figure). Forward masking (Fig. 10.8b) is just the opposite. Here, the masker is presented first and then the signal is turned on after an interval following masker offset. As the arrow shows, the masking of the signal now occurs forward in time.

Interest in temporal masking became prevalent with the studies reported by Samoilova [42,43] and Chistovich and Ivanova [44] in the 1950s, although backward and forward masking had been reported upon earlier [45,46]. In the United States, interest began with studies by Pickett [47] and Elliott [48–50]. The basic experiment follows the schemes in Fig. 10.8. The amount of masking of the test signal produced by the following (a) or preceding (b) masker is determined while various parameters of the probe and masker are manipulated. These parameters may be the time interval between probe and masker, masker level, masker duration, etc.

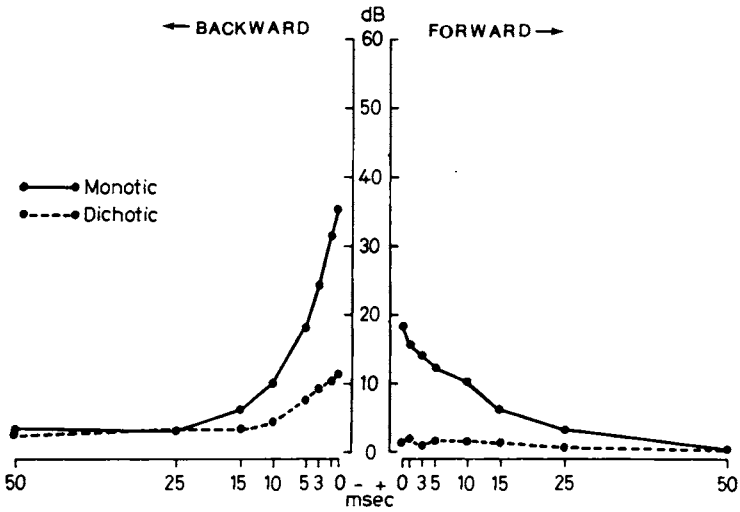

Figure 10.9 Temporal masking in decibels as a function of the interval between signal and masker. (Signal: 10 msec, 1000 Hz tone bursts; masker: 50 msec broad-band noise bursts at 70 dB SPL). (Adapted from Elliott [48]. Permission *J. Acoust. Soc. Amer.*)

Figure 10.9 shows some results which are typical of temporal masking data. The ordinate is the amount of masking produced by 50-msec noise bursts presented at 70 dB SPL for a test signal of 1000 Hz lasting 10 msec. The abscissa is the time interval between masker and test signal for the backward and forward masking paradigms. Finally, the solid lines show the amount of masking produced when the masker and probe signal are presented to the same ear (monotically), and the dotted lines reveal the masking that results when the noise goes to one ear and the tone to the other (dichotic masking).

At least three observations may be made from Fig. 10.9. First, backward masking is more effective than forward masking. In other words, more masking occurs when the masker follows the signal by a brief time interval than when the signal follows the masker. Second, more masking occurs when both signal and masker are delivered to the same ear (monotically) than when the probe is presented to one ear and the masker to the other (dichotically). Finally, the closer in time the signal and masker are presented, the more masking results. Conversely, less masking occurs as the time gap between probe and masker widens.

Although Fig. 10.9 shows data for masker–probe separations of up to 50 msec, Elliott [51] has demonstrated significant backward masking

for temporal gaps as wide as 100 msec. One would expect to find that temporal masking would increase as the level of the masker is increased. However, the linear increase of masked threshold as a function of masker level seen for simultaneous masking, discussed earlier, is not observed for temporal masking [43,48,52]. That is, increasing the masker level by 10 dB may result in an additional threshold shift on the order of only 3 dB.

Returning to backward masking as a function of the delay between probe and masker, notice in Fig. 10.9 that the amount of masking decreases dramatically as the interval increaes from 0 to 15 msec. For longer delays, the amount of backward masking decreases very little as the interval is lengthened further. The same situation occurs for the forward masking curve, but to a lesser degree. The steep and shallow segments of the monotic temporal masking curves suggest that two different mechanisms might be active [53]. The steeper segments within approximately 15–20 msec of the masker are probably due to overlapping in time between the traveling wave patterns of the signal and masker; i.e., to a cochlear event. This conclusion is supported by other findings [49,54], as well. It would appear that neural events are responsible for the shallower temporal masking curves at longer delays between the probe and masker. Furthermore, the dichotic temporal masking must rely upon central processing, since the signal and masker are presented to different ears.

The duration of the masker influences the amount of forward masking, but this does not appear to occur for backward masking [51]. That is, more forward masking is produced by a masker 200 msec in duration than by a 25 msec masker. Interestingly, more masking occurs when backward and forward masking are combined than would result if the individual contributions of backward and forward masking were simply added together [55–57]. Such a result is obtained by placing the probe signal between the two maskers, as shown in Fig. 10.8c. Such findings suggest that forward and backward masking may well depend upon different underlying mechanisms.

As we might expect, temporal masking depends upon the frequency relationship between the probe signal and the masker [51,58] just as we have seen to occur for simultaneous masking. In other words, more masking occurs when the signal and masker are close together in frequency than when they are far apart.

The underlying mechanisms of temporal masking are not fully resolved. We have already seen that overlapping in time of the cochlear displacement patterns is a reasonable explanation at very short masker–signal delays. Evidence for the role of central processing in temporal

masking has been pointed out in the literature [48,59,60]. Among other factors, forward masking may be due to a persistence of the masker's representation in the auditory neural channel, and/or to a temporary reduction in the absolute sensitivity of the stimulated units resulting from stimulation by the masker. It may be that backward masking occurs because the masker is presented before the signal has been fully processed within the auditory nervous system, thus interfering with the process of its perception. In other words, the occurrence of the masker might be thought of as overriding the in-process processing of the signal.

POSTSTIMULATORY FATIGUE (TTS)

It is not uncommon to experience a period of decreased hearing sensitivity, which lasts for some time, after being exposed to high sound intensities, for example after leaving a discotheque. This temporary shift in auditory threshold may last as long as roughly 16 hours, improving gradually. The phenomenon is quite descriptively called poststimulatory fatigue or temporary threshold shift (TTS).

The nature of the TTS is in clear contrast to the threshold shifts associated with simultaneous masking discussed previously in this chapter. Poststimulatory fatigue appears to be a manifestation of temporary changes in the hair cells as a result of exposure to the fatiguing stimulus. As one might expect, excessive and/or long-standing exposures may result in permanent threshold shifts, reflecting pathologic changes or destruction of the hair cells and their associated structures. From the practical standpoint, the amount of TTS produced by exposure to a given fatiguing stimulus has been used as a predictor of individual susceptibility to permanent threshold shifts ("damage risk"). However, this approach is not unchallenged. Since space permits only brief coverge of TTS, the reader is referred to other sources [61–64] for excellent reviews of this topic and its related areas.

Except for levels below approximately 70–75 dB SPL, which do not appear to produce any TTS, the amount of threshold shift increases as the stimulus intensity is raised [65–67]. Furthermore, it appears that higher-frequency stimuli result in more TTS than do lower frequency fatiguers [68]. For a given intensity, the amount of TTS will increase with duration [67,69]. In other words, the TTS will become worse as exposure time is increased. Furthermore, the rate at which TTS increases is proportional to the logarithm of the exposure time.

The amount of TTS is smaller for intermittent sounds than for continuous stimulation [69,70]; it appears to be related to the total or

average time that the fatiguer is "on" during the course of stimulation [62].

The frequency range over which TTS occurs becomes wider as the stimulus level is raised, and this affect is asymmetric in that the higher frequencies (up to roughly 4000–6000 Hz) are the most severely affected. The TTS reaches a maximum at a higher frequency than the fatiguer frequency, generally about one-half to one octave above (though higher) [61,62,71].

The course of recovery from TTS may be measured at various times after the fatiguing tone has been turned off. The course is rather complicated within about 2 min following the offset of the fatiguer [72,73]: It is nonmonotonic, with a "bounce" occurring in the vicinity of the 2 min point. This bounce is a reversal of the recovery, after which the TTS begins to decrease again with time. For longer recovery times, the amount of TTS tends to decrease from the value at the 2 min point (generally called TTS_2) at a rate that is proportional to the logarithm of the time since exposure [74].

REFERENCES

1. American National Standards Institute, Standard S1.1-1960, Acoustical terminology, 1960.
2. M. F. Meyer, Masking: Why restrict it to the threshold level? *J. Acoust. Soc. Amer. 31*, 243 (1959).
3. B. Scharf, Partial masking, *Acoustica 14*, 16–23 (1964).
4. A. M. Mayer, Researches in acoustics, *Philos. Mag. 37*, 259–288 (1894).
5. R. L. Wegel and C. E. Lane, The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear, *Physiol. Rev. 23*, 266–285 (1924).
6. R. H. Ehmer, Masking patterns of tones, *J. Acoust. Soc. Amer. 31*, 1115–1120 (1959).
7. A. M. Small, Pure tone masking, *J. Acoust. Soc. Amer. 31*, 1619–1625 (1959).
8. A. Finck, Low-frequency pure tone masking, *J. Acoust. Soc. Amer. 33*, 1140–1141 (1961).
9. E. Zwicker, On a psychoacoustical equivalent of tuning curves, in *Facts and Models in Hearing* (E. Zwicker and E. Tehrhardt, Eds.), Springer, Berlin, 1974, pp. 95–99.
10. E. Zwicker and K. Schorn, Psychoacoustical tuning curves in audiology, *Audiology 17*, 120–140 (1978).
11. J. P. Egan and H. W. Hake, On the masking pattern of a simple auditory stimulus, *J. Acoust. Soc. Amer. 22*, 622–630 (1950).
12. R. H. Ehmer, Masking by tones vs. noise bands, *J. Acoust. Soc. Amer. 31*, 1253–1256 (1959).

13. D. D. Greenwood, Auditory masking and the critical band, *J. Acoust. Soc. Amer. 33*, 484–501 (1961).
14. R. C. Bilger and I. J. Hirsh, Masking of tone by bands of noise, *J. Acoust. Soc. Amer. 28*, 623–630 (1956).
15. R. C. Bilger, Intensive determinants of remote masking, *J. Acoust. Soc. Amer. 30*, 817–824 (1958).
16. R. C. Bilger, Remote masking in the absence of intra-aural muscles, *J. Acoust. Soc. Amer. 39*, 103–108 (1966).
17. J. L. Shapely, Reduction in the loudness of a 250-cycle tone in one ear following the introduction of a thermal noise in the opposite ear, *Proc. Iowa Acad. Sci. 61*, 417–422 (1954).
18. W. Spieth, Downward spread of masking, *J. Acoust. Soc. Amer. 29*, 502–505 (1957).
19. B. H. Deatherage, H. Davis, and D. H. Eldredge, Physiological evidence for the masking of low frequencies by high, *J. Acoust. Soc. Amer. 29*, 132–137 (1957).
20. B. H. Deatherage, H. Davis, and D. H. Eldredge, Remote masking in selected frequency regions, *J. Acoust. Soc. Amer. 29*, 512–514 (1957).
21. H. Fletcher, Relation between loudness and masking, *J. Acoust. Soc. Amer. 9*, 1–10 (1937).
22. J. E. Hawkins and S. S. Stevens, The masking of pure tones and of speech by white noise, *J. Acoust. Soc. Amer. 22*, 6–13 (1950).
23. H. Fletcher, Auditory patterns, *J. Acoust. Soc. Amer. 12*, 47–65 (1940).
24. T. H. Schaefer, R. S. Gales, C. A. Shewaker, and P.O. Thompson, The frequency selectivity of the ear as determined by masking experiments, *J. Acoust. Soc. Amer. 22*, 490–497 (1950).
25. P. M. Hamilton, Noise masked thresholds as a function of tonal duration and masking noise band width, *J. Acoust. Soc. Amer. 29*, 506–511 (1957).
26. D. D. Greenwood, Auditory masking and the critical band, *J. Acoust. Soc. Amer. 33*, 484–502 (1961).
27. J. A. Swets, D. M. Green, and W. P. Tanner, On the width of the critical bands, *J. Acoust. Soc. Amer. 34*, 108–113 (1962).
28. C. E. Bos and E. deBoer, Masking and discrimination, *J. Acoust. Soc. Amer. 39*, 708–715 (1966).
29. B. Scharf, Critical bands, in *Foundations of Modern Auditory Theory* (J. V. Tobias, Ed.), Vol. 1, Academic, New York, 1970, pp. 157–202.
30. E. Zwicker, G. Flottrop, and S. S. Stevens, Critical bandwidth in loudness summation, *J. Acoust. Soc. Amer. 29*, 548–557 (1957).
31. R. C. Bilger, A revised critical-band hypothesis, in *Hearing and Davis: Essays Honoring Hallowell Davis* (S. K. Hirsh, D. H. Eldredge, Wash. Univ. Press, St. Louis, 1976, pp. 191–198.
32. R. D. Patterson, Auditory filter shape, *J. Acoust. Soc. Amer. 55*, 802–809 (1974).
33. R. Chocolle, La sensibilité auditive différentielle d'intensité en présence d'un son contralatéral de même fréquence. *Acoustica 7*, 75–83 (1957).
34. J. G. Ingham, Variations in cross-making with frequency, *J. Exp. Psych. 58*, 199–205 (1959).

35. C. E. Sherrick and P. L. Mangabeira-Albernaz, Auditory threshold shifts produced by simultaneous pulsed contralateral stimuli, *J. Acoust. Soc. Amer. 33*, 1381–1385 (1961).
36. D. D. Dirks and C. Malmquist, Changes in bone-conduction thresholds produced by masking of the non-test ear, *J. Speech Hearing Res. 7*, 271–287 (1964).
37. D. D. Dirks and J. C. Norris, Shifts in auditory thresholds produced by pulsed and continuous contralateral masking, *J. Acoust. Soc. Amer. 37*, 631–637 (1966).
38. J. Zwislocki, E. N. Damianopoulus, E. Buining, and J. Glantz, Central masking: Some steady-state and transient effects, *Percept. Psychophys 2*, 59–64 (1967).
39. J. Zwislocki, E. Buining, and J. Glantz, Frequency distribution of central masking, *J. Acoust. Soc. Amer. 43*, 1267–1271 (1968).
40. J. Zwislocki, A theory of central masking and its partial validation, *J. Acoust. Soc. Amer. 52*, 644–659 (1972).
41. J. Zwislocki, In search of physiological correlates of psychoacoustic characteristics, in *Basic Mechanisms in Hearing (A. R. Møller, Ed.), Academic, New York, 1973, pp. 787–808.*
42. I. K. Samoilova, Masking effect of a strong sound stimulus on a previously applied weak one, *Biofizika 1*, 79–97 (1956).
43. I. K. Samoilova, Masking of short tone signals as a function of the time between masked and masking sounds, *Biofizika 4*, 550–558 (1959).
44. L. A. Chistovich and V. A. Ivanova, Mutual masking of short sound pulses, *Biofizika 4*, 170–180 (1959).
45. H. Werner, Studies on contour. I. Qualitative analyses. *Amer. J. Psych. 47*, 40–64 (1935).
46. E. Lüscher and J. Zwislocki, Adaptation of the ear to sound stimuli, *J. Acoust. Soc. Amer. 21*, 135–139 (1949).
47. J. M. Pickett, Backward masking, *J. Acoust. Soc. Amer. 31*, 1613–1615 (1959).
48. L. L. Elliott, Backward masking: Monotic and dichotic conditions, *J. Acoust. Soc. Amer. 34*, 1108–1115 (1962).
49. L. L. Elliott, Backward and forward masking of probe tones of different frequencies, *J. Acoust. Soc. Amer. 34*, 1116–1117 (1962).
50. L. L. Elliott, Backward masking: Different durations of the masker stimulus, *J. Acoust. Soc. Amer. 36*, 393 (1964).
51. L. L. Elliott, Development of auditory narrow-band frequency contours, *J. Acoust. Soc. Amer. 42*, 143–153 (1967).
52. H. Babkoff and S. Sutton, Monaural temporal masking of transients, *J. Acoust. Soc. Amer. 44*, 1373–1378 (1968).
53. H. Duifhuis, Consequences of peripheral frequency selectivity for nonsimultaneous masking, *J. Acoust. Soc. Amer. 54*, 1471–1488 (1973).
54. J. H. Patterson, Additivity of forward and backward masking as a function of signal frequency, *J. Acoust. Soc. Amer. 50*, 1123–1125 (1971).
55. I. Pollack, Interaction of forward and backward masking, *J. Aud. Res. 4*, 63–67 (1964).

56. L. L. Elliott, Masking of tones before, during, and after brief silent periods in noise, *J. Acoust. Soc. Amer. 45*, 1277–1279 (1969).
57. R. H. Wilson and R. Carhart, Forward and backward masking: Interactions and additivity, *J. Acoust. Soc. Amer. 49*, 1254–1263 (1971).
58. H. N. Wright, Backward masking for tones in narrow-band noise, *J. Acoust. Soc. Amer. 36*, 2217–2221 (1964).
59. D. H. Raab, Forward and backward masking between acoustic clicks, *J. Acoust. Soc. Amer. 33*, 137–139 (1961).
60. B. H. Deatherage and T. R. Evans, Binaural masking: Backward, forward and simultaneous effects, *J. Acoust. Soc. Amer. 46*, 362–371 (1969).
61. D. N. Elliott and W. Fraser, Fatigue and adaptation, in *Foundations of Modern Auditory Theory* (J. V. Tobias, Ed.), Vol. 1, Academic, New York, 1970, pp. 115–155.
62. W. D. Ward, Adaptation and fatigue, in *Modern Developments In Audiology* (J. Jerger, Ed.), 2nd Ed., Academic, New York, 1973, pp. 301–344.
63. K. D. Kryter, Impairment to hearing from exposure to noise, *J. Acoust. Soc. Amer. 53*, 1211–1234 (1973). [Also: Comments by A. Cohen, H. Davis, B. Lempert, and W. D. Ward, and reply by K. D. Kryter, *J. Acoust. Soc. Amer. 53*, 1235–1252 (1973)].
64. D. Henderson, R. P. Hamernik, D. S. Dosanjh, and J. H. Mills, Eds., *Effects of Noise on Hearing*, Raven, New York, 1976).
65. I. J. Hirsh and R. C. Bilger, Auditory-threshold recovery after exposures to pure tones, *J. Acoust. Soc. Amer. 27*, 1186–1194 (1955).
66. W. D. Ward, A. Glorig, and D. L. Sklar, Temporary threshold shift from octave-band noise: Applications to damage-risk criteria, *J. Acoust. Soc. Amer. 31*, 522–528 (1959).
67. J. H. Mills, R. W. Gengel, C. S. Watson, and J. D. Miller, Temporary changes in the auditory system due to exposure to noise for one or two days, *J. Acoust. Soc. Amer. 48*, 524–530 (1970).
68. W. D. Ward, Auditory fatigue and masking, in *Modern Developments in Audiology* (J. Jerger, Ed.), Academic, New York, 1963, pp. 241–286.
69. W. D. Ward, A. Glorig, and D. L. Sklar, Dependence of temporary threshold shift at 4kc on intensity and time, *J. Acoust. Soc. Amer. 30*, 944–954 (1958).
70. W. D. Ward, A. Glorig, and D. L. Sklar, Temporary threshold shift produced by intermittent exposure to noise, *J. Acoust. Soc. Amer. 31*, 791–794 (1959).
71. W. D. Ward, Damage-risk criteria for line spectra, *J. Acoust. Soc. Amer. 34*, 1610–1619 (1962).
72. I. J. Hirsh and W. D. Ward, Recovery of the auditory threshold after strong acoustic stimulation, *J. Acoust. Soc. Amer. 24*, 131–141 (1952).
73. I. J. Hirsh and R. C. Bilger, Auditory-threshold recovery after exposures to pure tones, *J. Acoust. Soc. Amer. 27*, 1186–1194 (1955).
74. W. D. Ward, A. Glorig, and D. L. Sklar, Relation between recovery from temporary threshold shift and duration of exposure, *J. Acoust. Soc. Amer. 31*, 600–602 (1959).

11
Loudness

The intensity of a sound refers to its physical magnitude, which may be expressed in such terms as its power or pressure. Turning up the "volume" control on a stereo amplifier thus increases the intensity of the music coming out of the loudspeakers. This intensity is easily measured by placing the microphone of a sound-level meter near the loudspeaker. The perception of intensity is called loudness; generally speaking, low intensities are perceived as "soft" and high intensities as "loud." In other words, intensity is the physical parameter of the stimulus and loudness is the percept associated with that parameter. However, intensity and loudness are not one and the same; and although increasing intensity is generally associated with increasing loudness, there is not a simple one-to-one correspondence between the two. Furthermore, loudness is also affected by factors other than intensity. For example, it is a common experience to find that loudness changes when the "bass" and "treble" controls of a stereo amplifier are adjusted, even though the volume control itself is untouched.*

LOUDNESS LEVEL

We may begin our discussion of loudness by asking whether the same amount of intensity results in the same amount of loudness for tones of

*Bass and treble are the relative contributions of the lower and higher frequency ranges, respectively. Thus, raising the bass emphasizes the low frequencies, and raising the treble emphasizes the high.

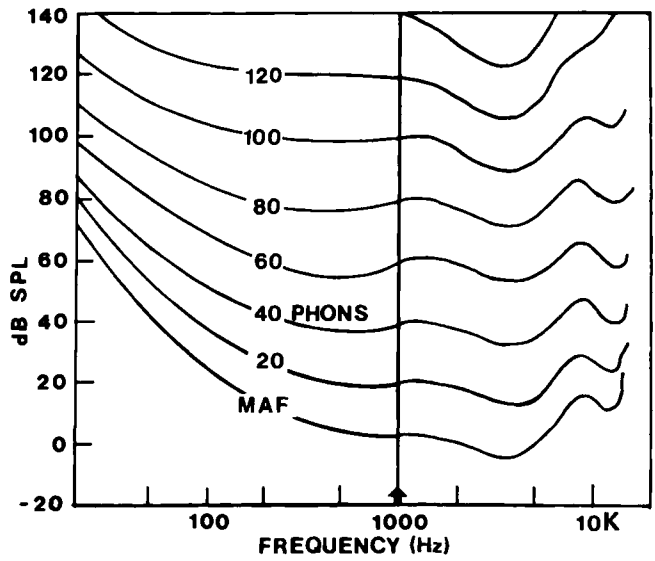

Figure 11.1 Equal loudness contours (phon curves) adapted from Robinson and Dadson [2]. (Permission *British J. Appl. Phys.* Copyright 1956 by The Institute of Physics.)

different frequencies. For example, does a 100 Hz tone at 40 dB SPL have the same loudness as a 1000 Hz tone also presented at 40 dB? The answer is no. However, a more useful question is to ask how much intensity is needed in order for tones of different frequencies to sound equally loud. These values may be appropriately called equal loudness levels.

Although the exact procedures differ, the fundamental approach for determining equal loudness levels is quite simple. One tone is presented at a fixed intensity level, and serves as the reference tone for the experiment. The other tone is then varied in level until its loudness is judged equal to that of the reference tone. (The traditional reference tone has been 1000 Hz; but Stevens [1] more recently suggested the use of 3150 Hz, where threshold sensitivity is most acute.) A third frequency tone may then be balanced with the reference tone; then a fourth, a fifth, and so on. The result is a list of sound pressure levels at various frequencies, all of which sound equal in loudness to the reference tone. We can then draw a curve showing these equally loud intensity levels as a function of frequency. If the experiment is repeated for different reference tone intensities, the result is a series of contours like the ones in Fig. 11.1.

The contour labeled "40 phons" shows the intensities needed at each frequency for a tone to sound equal in loudness to a 1000 Hz reference tone presented at 40 dB SPL. Thus, any sound which is equal in loudness to a 1000 Hz tone at 40 dB has a loudness level of 40 phons. A tone which is as loud as a 1000 Hz tone at 50 dB has a loudness level of 50 phons; one that is as loud as a 1000 Hz tone at 80 dB has a loudness level of 80 phons, etc. We may now define the phon as the unit of loudness level. All sounds that are equal in phons have the same loudness level even though their physical magnitudes may be different. Since we are expressing loudness level in phons relative to the level of a 1000 Hz tone, phons and dB SPL are necessarily equal at this frequency.

The earliest equal loudness data were reported by Kingsbury [3] in 1927. However, the first well-accepted phon curves were published in 1933 by Fletcher and Munson [4], and as a result, equal loudness contours have come to be called "Fletcher–Munson curves." Subsequently, extensive equal loudness contours were also published by Churcher and King [5] and by Robinson and Dadson [2]. There are some differences among the curves given by the various studies, although the basic configurations are reasonably similar. Robinson and Dadson's curves appear to be the most accurate, and are the ones shown in Fig. 11.1. Equal loudness contours have also been reported for narrow bands of noise [6].

At low loudness levels the phon curves are quite similar in shape to the minimum audible field (MAF) curve. Thus, more intensity is needed to achieve equal loudness for lower frequencies than for higher ones. However, notice that the phon curves tend to flatten for higher loudness levels, indicating that the lower frequencies grow in loudness at a faster rate than the higher frequencies, overcoming, so to speak, their disadvantage at near-threshold levels. This effect can be experienced in a simple, at-home experiment, We begin by playing a record album at a moderate level, with the bass and treble controls set so that the music is as "natural sounding" as possible. If we decrease the volume to a much softer level, the music will also sound as though the bass was decreased, demonstrating the de-emphasis of the low (bass) frequencies at lower loudness levels. If we raise the volume to a quite loud level, then the music will sound as though the bass was turned up as well. This "boomy" sound reflects the faster rate of growth for the lower frequencies with increasing loudness levels.

Since the same sound pressure level (SPL) will be associated with different loudness levels as a function of frequency, it would be convenient to have a frequency weighting network which could be applied to the wide-band sounds encountered in the environment. Such a

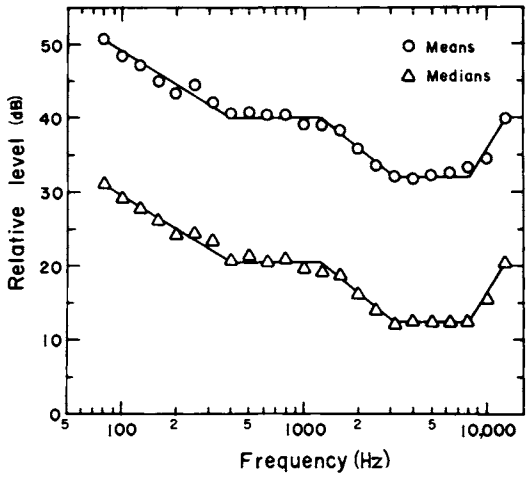

Figure 11.2 Combined equal perceived magnitude (loudness) contours. Circles show mean data across studies and triangles show medians. (From Stevens [1]. Permission *J. Acoust. Soc. Amer.*)

weighting function would facilitate calculating the loudness of such sounds as highway noise, sonic booms, etc. This has been done to some extent in the construction of electronic weighting networks for sound level meters (see Peterson and Gross [7] for a complete discussion). These networks are rough approximations to various phon curves. For example, the A-weighting network approximates the general shape of the 40 phon curve by de-emphasizing the low frequencies and more efficiently passing the high. The B-weighting network roughly corresponds to the 70 phon loudness level; and the C-weighting network is designed to mimic the essentially flat response of the ear at high loudness levels.

More precise calculation methods have been proposed by Stevens [1,8,9], Zwicker [10], and Kryter [11,12]. We will not discuss these methods except to say that the basic approach is to assign weights to various frequency bands in the sound being measured, and then to combine these weights in order to arrive at the predicted loudness of the sound. Stevens [1] combined the results of 25 equal loudness (and "equal noisiness") contours across a wide variety of studies in order to arrive at the representative contours shown in Fig. 11.2. These combined equal perceived loudness contours appear to represent the consensus of the existing data as of this writing, and serve as the basis for Stevens' loudness calculation procedure ("Mark VII") [1].

The use of loudness levels represents a significant improvement over such vague concepts as "more intense sounds are louder." However, the phon itself does not provide a direct measure of loudness, per se. We must still seek an answer to the question of how the loudness percept is related to the level of the physical stimulus.

LOUDNESS SCALING

Loudness scales show how the loudness percept is related to the intensity of the sound stimulus. Since we are interested not only in the loudness of a particular sound, but also in how much louder one sound is than another, the relationship between loudness and intensity is best determined with direct ratio scaling techniques (see Chap. 7).* This approach was pioneered by Stevens [15,17,19–22], whose earliest attempts to define a ratio scale of loudness used the fractionalization method [19]. Stevens later adopted the use of magnitude estimation and magnitude production [21], and subsequent work has almost exclusively employed these techniques alone or in combination [23–29].

The unit of loudness is the sone [19]; one sone is the loudness of a 1000 Hz tone presented at 40 dB SPL. Since SPL in decibels and loudness level in phons are equivalent at 1000 Hz, we may also define one *sone* as the *loudness* corresponding to a *loudness level* of 40 *phons*. We may therefore express loudness in sones as a function of loudness level in phons [13] as well as as a function of stimulus intensity. (In this context, recall that Stevens [1] proposed that the reference frequency be moved to about 3150 Hz. This change would lower the reference for one sone to 32 dB SPL at 3150 Hz, as was confirmed by Hellman [29], who measured the loudness functions of 1000 and 3000 Hz tones.) Since loudness level does not vary with frequency (i.e., 40 phons represents the same loudness level at any frequency even though the SPLs are different), we can ignore frequency to at least some extent when assigning loudness in sones to a tone, as long as sones are expressed as a function of phons. As previously mentioned, methods that calculate loudness for complex sounds first weight the frequency bands comprising the complex sound, using what is essentially a phon curve, as in Fig. 11.2. Loudness in sones is then calculated from these weighted bands.

*The intensity difference limen (DL), or just noticeable difference (jnd), has also been proposed as a basis for loudness scaling, as has been the partitioning of the audible intensity range into equal loudness categories. However, the consensus of data supports ratio scaling. See Robinson's [13] review and study, and also Gardner [14] and Stevens [15], for summaries of the controversy. Also see Marks [16], Stevens [17] and Gescheider [18] for informative treatments within the general context of psychophysics.

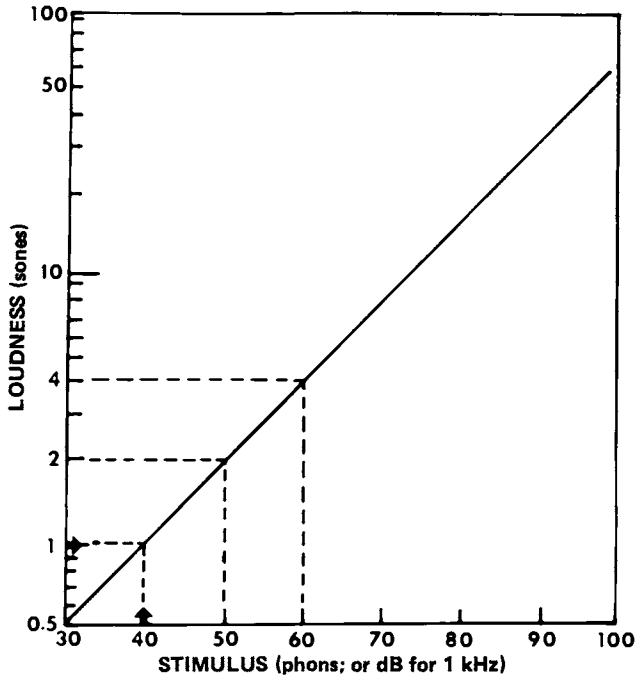

Figure 11.3 Idealized sone scale showing a doubling of loudness for a 10 dB (phon) stimulus increase.

The sone scale is easily understood. Having assigned a value of one sone to the reference sound, we assign a loudness of two sones to the intensity that sounds twice as loud as the reference, 0.5 sones to the level that sounds half as loud, etc. The resulting function appears as shown in Fig. 11.3. Notice that loudness in sones is a straight line when plotted as a function of level on log–log coordinates. This fact is extremely important, because a straight line on a log–log plot indicates a power function. In other words, loudness L may be expressed as a power e of the stimulus level I according to the formula

$$L = kI^e$$

where k is a constant. This is a case of Stevens' law [30], which states that sensation grows as a power of stimulus level. The exponent indicates the rate at which the sensation grows with stimulus magnitude. Thus, an exponent of 1.0 would mean that the magnitude of the sensation increases at the same rate as the stimulus level. Exponents less than 1.0 show that the sensation grows at a slower rate than stimulus level

(examples are loudness and brightness); whereas exponents greater than 1.0 indicated that the perceived magnitude grows faster than the physical level (examples are electric shock and heaviness). Conveniently, the exponent also corresponds to the slope of the function. (The constant k depends upon the units used to express the magnitudes.)

Early studies resulted in a median exponent of 0.6 for loudness as a function of sound pressure level, so that a 10 dB level increase would correspond to a doubling of loudness [20]; and Robinson [13] reported that loudness increased twofold with a change in loudness level of 10 phons. On the other hand, J. C. Stevens and Guirao [25] reported much faster loudness growth, with exponents averaging 0.77. However, the consensus places the exponent of the loudness function on the order of about 0.67 [1,31–33]. This slope of two-thirds leads to a doubling of loudness with an increase in sound pressure level of 9 dB, and is incorporated into the procedure for calculating loudness proposed by Stevens [1].

CRITICAL BANDS AND LOUDNESS OF COMPLEX SOUNDS

The critical band concept was introduced with respect to masking in the last chapter. As we shall see, loudness also bears an intimate relationship to the critical bandwidth; and loudness experiments provide a direct estimate of the width of the critical band. As Scharf [34] pointed out, it is convenient to think of the critical band as the bandwidth where abrupt changes occur. Consider the following experiment with this concept in mind.

Suppose pairs of simultaneous tones are presented to a subject, both tones always at the same fixed level. The first pair of tones presented are very close together in frequency; and the subject compares their loudness to the loudness of a standard tone. The frequency difference between the two tones is then increased, and the resulting loudness is again compared to the standard. We find that the loudness of the two tones stays about the same as long as the tones are separated by less than the critical bandwidth, but that there is a dramatic increase in loudness when the components are more than a critical bandwidth apart. The open circles in Fig. 11.4 show typical results for two subjects. In this figure the amount of loudness summation is shown as the level difference between the standard and comparison stimuli (ordinate) as a function of bandwidth (abscissa). Notice that the loudness of the two-tone complex stays essentially the same for frequency separations smaller than the critical bandwidth (roughly 200 Hz in this example); whereas loudness increases when the frequency difference is greater than the width of the critical band.

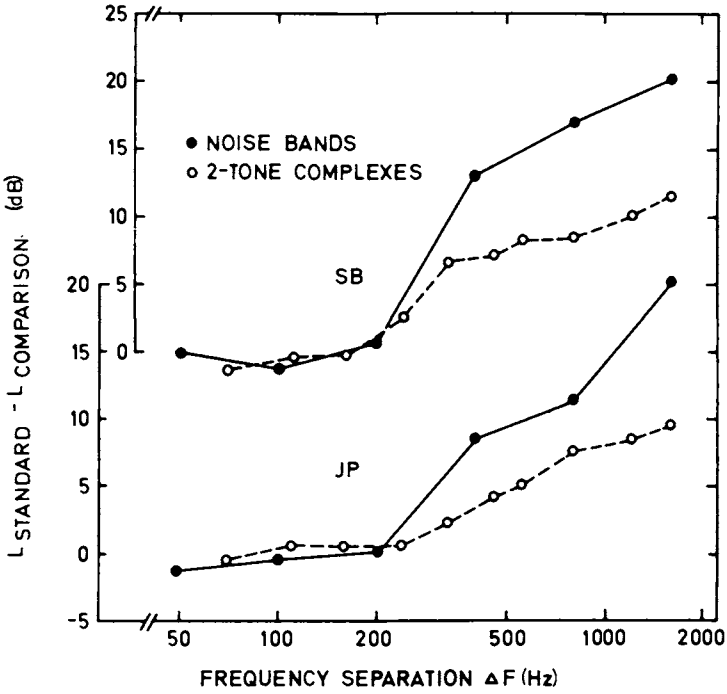

Figure 11.4 Effect of critical bandwidth upon loudness summation for a two-tone complex (open circles) and bands of noise (filled circles) for two subjects (SB and JP). Test level was 65 dB SPL and center frequency was 1000 Hz (see text). (From Florentine et al. [35]. Permission *J. Acoust. Soc. Amer.*)

The fact that loudness remains essentially the same for bandwidths (or frequency separations) smaller than the critical band, but increases when the critical band is exceeded, has been demonstrated for two-tone and multitone complexes, and also for bands of noise [35–39]. This loudness summation effect is minimal at near-threshold levels, and the greatest loudness increases occur for moderate signal levels [36,38,39]. As Fig. 11.5 shows, loudness summation becomes greater as the number of components of a multitone complex is increased, with the most loudness summation occurring for bands of noise wider than the critical band (35). This relation is shown in Fig. 11.4, in which the same loudness results from both two-tone complexes (open circles) and noise bands (filled circles) narrower than the critical band, but much greater loudness summation results for the noise when the critical bandwidth is exceeded.

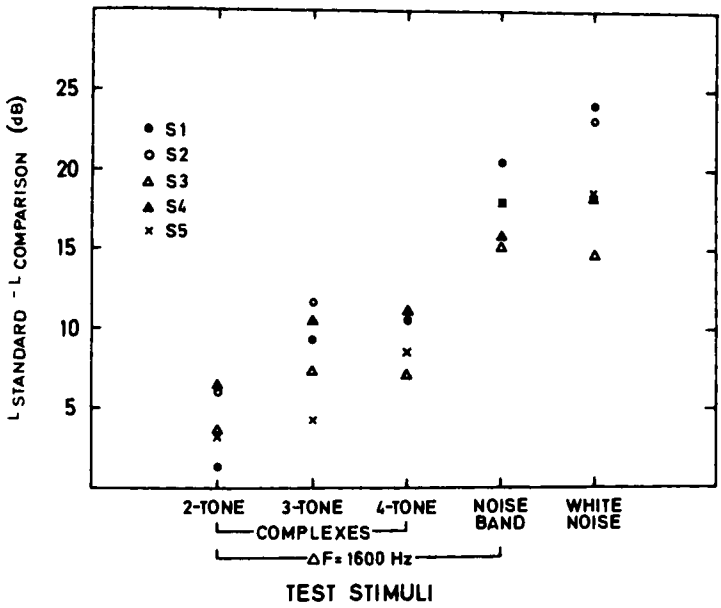

Figure 11.5 Loudness summation for tonal complexes and for noise (symbols show data for individual subjects). (From Florentine et al. [35]. Permission *J. Acoust. Soc. Amer.*)

These findings, by Florentine et al. [35], contradict earlier observations by Scharf [40], who found that loudness summation was unaffected by the number of components in a tonal complex. Florentine et al. attributed the discrepancy to variability which may have been introduced into Scharf's results by pooling data across subjects and by not testing at 65 dB SPL, where summation is most apparent. The small effect associated with the number of components would have been obscured by the large degree of variability caused by these factors.

TEMPORAL INTEGRATION OF LOUDNESS

Temporal integration (summation) at threshold was discussed in Chap. 10, where we found that sensitivity improves as signal duration increases up to about 200–300 msec, after which thresholds remain essentially constant. Temporal integration was also covered with respect to the acoustic reflex in Chap. 3. A similar phenomenon is also observed for loudness [41–48]. Increasing the duration of a very brief signal at a given level above threshold will, within the same general time frame as in the cases previously discussed, cause it to sound louder.

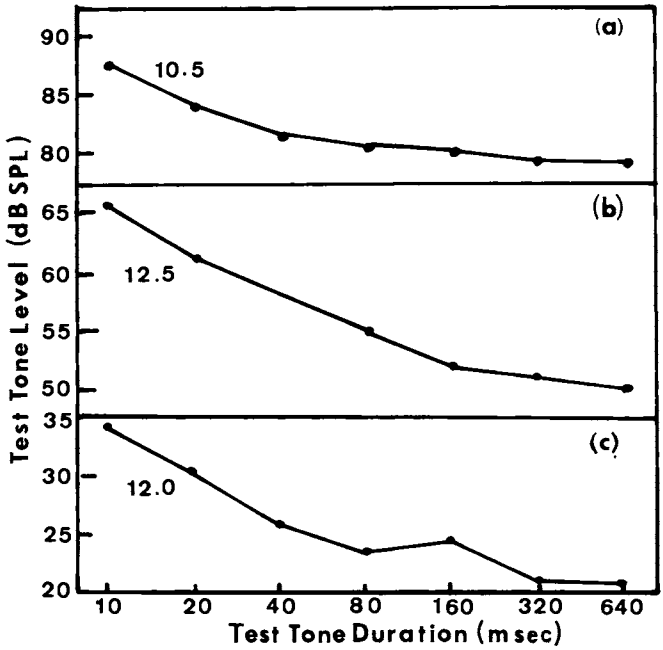

Figure 11.6 Temporal integration of loudness at 1000 Hz based upon loudness balances to a 500 msec tone at (a) 20 dB SPL, and (b) 50 dB SPL and (c) 80 dB SPL. Slopes of the steeper portions (dB per decade duration change) are shown near each curve. Based upon data by Richards [48]. See text.

There are two basic techniques which may be used to study the temporal integration of loudness. One method is similar to that used in establishing phon curves. The subject is presented with a reference sound at a given intensity and is asked to adjust the level of a second sound until it is equal in loudness with the first one [41–43,48]. In such cases, one of the sounds is "infinitely" long (i.e., long enough so that we may be sure that temporal integration is maximal, say 1 sec), and the other is a brief tone burst (of a duration such as 10 msec, 20 msec, 50 msec, etc.). Either stimulus may be used as the reference while the other is adjusted, and the result is an equal loudness contour as a function of signal duration. The alternate method involves direct magnitude scaling [44,45,47] from which equal loudness curves can be derived.

Figure 11.6 shows representative curves for the temporal integration of loudness. These curves are based upon the findings of Richards [48]. In his experiment, test tones of various durations were balanced in

loudness to a 500 msec reference tone presented at either 20, 50, or 80 dB SPL. The ordinate shows the test tone levels (in dB SPL) needed to achieve equal loudness with the reference tone. This quantity is plotted as a function of test tone duration on the abscissa. Notice that loudness increases (less intensity is needed to achieve a loudness balance with the reference tone) as test tone duration increases. This increase in loudness is greater for increases in duration up to about 80 msec, and then tends to slow down. In other words, increases in duration from 10 msec to about 80 msec has a steeper loudness summation slope than increases in duration above 80 msec. However, Richards did find that there was still some degree of additional loudness integration at longer durations.

These data are basically typical of most findings on temporal integration of loudness. That is, there is an increase of loudness as duration is increased up to some "critical duration;" and loudness growth essentially stops (or slows down appreciably) with added duration. On the other hand, the critical duration is quite variable among studies, and has generally been reported to decrease as a function of sensation level [41,42], though not in every study [45]. In addition, the rate at which loudness has been found to increase with duration varies among studies. McFadden [47] found large differences also among individual subjects. Richards [48] fitted the steeper portions of the temporal integration functions with straight lines, and found that their slopes were on the order of 10–12 dB per decade change in duration. The mean values are shown on the figure, and agree well with other studies [42,45]. One might note at this point that the results of loudness integration experiments have been shown to be affected by the methodology used and the precise nature of the instructions given to the patients, and also by confusions on the part of subjects between loudness and duration of the signal [49].

LOUDNESS ADAPTATION

Loudness adaptation refers to the apparent decrease in the loudness of a signal that is continuously presented at a fixed level for a reasonably long period of time. In other words, the signal appears to become softer as time goes on even though the intensity is the same. Hood's [50] classic experiment demonstrates this phenomenon rather clearly. A 1000 Hz tone is presented to the subject's right ear at 80 dB. This adapting stimulus remains on continuously. At the start, a second 1000 Hz tone is presented to the left ear, and the subject adjusts the level of this second (comparison) tone to be equally loud to the adapting tone in the right ear

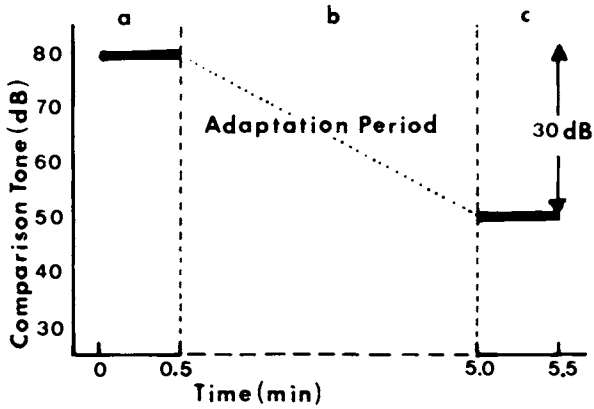

Figure 11.7 Loudness adaptation as shown by the level of the comparison stimulus. Modified from Hood [50].

(part a in Fig. 11.7). Thus, the level of the comparison tone is used as an indicator of the loudness of the adapting tone in the opposite ear. This first measurement represents the loudness prior to adaptation (the preadaptation balance).

The comparison tone is then turned off while the adapting tone continues to be applied to the right ear (adaptation period b in Fig. 11.7). After several minutes of adaptation, the comparison signal is reapplied to the opposite ear, and the subject readjusts it to be equally loud with the 80 dB adapting tone. This time, however, the comparison tone is adjusted by the subject to only 50 dB in order to achieve a loudness balance with the adaptor (segment c in Fig. 11.7), indicating that the loudness of the adapting tone has decreased by an amount comparable to a 30 dB drop in signal level. Thus, there has been 30 dB of adaptation due to the continuous presentation of the tone to the right ear. Since the loudness decrease occurs during stimulation, the phenomenon is also called *per*stimulatory adaptation. This phenomenon contrasts, of course, with the TTS described in the previous chapter, which constitutes *post*stimulatory fatigue.

The method just described for the measurement of adaptation may be considered a simultaneous homophonic loudness balance. In other words, the subject must perform a loudness balance between two tones of the same frequency that are presented, one to each ear, at the same time. Other studies employing this approach [51–53] have reported similar findings.

An problem inherent in these early studies of loudness adaptation was their use of simultaneously presented adapting and comparison

tones of the same frequency. Although the presumed task was a loudness balance between the ears, the experiments were confounded by the perception of a fused image due to the interaction of the identical stimuli at the two ears. As we shall see in Chap. 13, the relative intensities at the two ears will determine whether the fused image is perceived centralized within the skull or lateralized toward one side or the other. It is therefore reasonable to question whether the loudness adaptation observed in these studies was really due to an interaction between the ears, i.e., to a central process.

One way around the problem is to use a comparison tone having a different frequency than the adapting tone. This procedure would reduce the lateralization phenomenon because the stimuli would be different at the ears. In 1955, Egan [51] found no significant difference in the amount of loudness adaptation caused by this heterophonic method and the homophonic approach described above. Subsequently, however, Egan and Thwing [54] reported that loudness balances involving lateralization cues did in fact result in greater adaptation than techniques that kept the effects of lateralization to a minimum.

The consensus of more recent studies [55–60] is that loudness adaptation is reduced or absent when binaural interactions (lateralization cues) are minimized. This may be accomplished by using adapting and comparison tones of different frequencies (heterophonic loudness balances), or by a variety of other means. For example, Stokinger and colleagues [58,59] reduced or obliterated loudness adaptation by shortening the duration of the comparison tone and by delaying the onset of the comparison tone until after the adapting tone was removed. Both of these approaches reduced or removed the interaction of the two tone between the ears.

The implication of these more recent studies is that the loudness adaptation measured by simultaneous loudness balances with similar adapting and comparison tones is due to an interaction between the signals presented to the two ears. In other words, the effect is associated with some central process.

REFERENCES

1. S. S. Stevens, Perceived level of noise by Mark VII and decibels (E), *J. Acoust. Soc. Amer. 51,* 575–601 (1972).
2. D. W. Robinson and R. S. Dadson, A redetermination of the equal loudness relations for pure tones, *Brit. J. Appl. Phys. 7,* 166–181 (1956).
3. B. A. Kingsbury, A direct comparison of the loudness of pure tones, *Phys. Rev. 29,* 588–600 (1927).

4. H. Fletcher and W. A. Munson, Loudness, its definition, measurement, and calculation, *J. Acoust. Soc. Amer. 5*, 82–105 (1933).
5. B. G. Churcher and A. J. King, The performance of noise meters in terms of the primary standard, *J. Inst. Elec. Eng. 81*, 57–90 (1937).
6. I. Pollack, The loudness of bands of noise, *J. Acoust. Soc. Amer. 24*, 533–538 (1952).
7. A. P. G. Peterson and E. E. Gross, *Handbook of Noise Measurement*, 7th ed., General Radio, Concord, Mass. 1972.
8. S. S. Stevens, Calculation of the loudness of complex noise, *J. Acoust. Soc. Amer. 28*, 807–832 (1956).
9. S. S. Stevens, Procedure for calculating loudness: Mark VI, *J. Acoust. Soc. Amer. 33*, 1577–1585 (1961).
10. E. Zwicker, Über psychologische and methodische Grundlagen der Lautheit, *Acoustica 1*, 237–258 (1958).
11. K. D. Kryter, Concepts of perceived noisiness, their implementation and application, *J. Acoust. Soc. Amer. 43*, 344–361 (1968).
12. K. D. Kryter, *The Effects of Noise on Man.* Academic, New York, 1970.
13. D. W. Robinson, The subjective loudness scale, *Acoustica 7*, 217–233 (1957).
14. W. R. Gardner, Advantages of the discriminability criteria for a loudness scale, *J. Acoust. Soc. Amer. 30*, 1005–1012 (1958).
15. S. S. Stevens, On the validity of the loudness scale, *J. Acoust. Soc. Amer. 31*, 995–1003 (1959).
16. L. E. Marks, *Sensory Processes: The New Psychophysics*, Academic, New York, 1974.
17. S. S. Stevens, In *Psychophysics: Introduction to Its Perceptual, Neural and Social Prospects* (G. Stevens, Ed.), Wiley, New York, 1975.
18. G. A. Gescheider, *Psychophysics: Method and Theory*, Lawrence Erlbaum, New Jersey, 1976.
19. S. S. Stevens, A scale for the measurement of a psychological magnitude: Loudness. *Psych. Rev. 43*, 405–416 (1936).
20. S. S. Stevens, The measurement of loudness, *J. Acoust. Soc. Amer. 27*, 815–829 (1955).
21. S. S. Stevens, The direct estimation of sensory magnitudes—loudness, *Amer. J. Psychol. 69*, 1–25 (1956).
22. S. S. Stevens, Concerning the form of the loudness function, *J. Acoust. Soc. Amer. 29*, 603–606 (1957).
23. R. P. Hellman and J. Zwislocki, Monaural loudness function at 1000 cps and interaural summation, *J. Acoust. Soc. Amer. 35*, 856–865 (1963).
24. R. P. Hellman and J. Zwislocki, Loudness function of a 1000-cps tone in the presence of a masking noise, *J. Acoust. Soc. Amer. 36*, 1618–1627 (1964).
25. J. C. Stevens and M. Guirao, Individual loudness functions, *J. Acoust. Soc. Amer. 36*, 2210–2213 (1964).
26. S. S. Stevens and H. Greenbaum, Regression effect in psychophysical judgment, *Percept. Psychophysiol. 1*, 439–446 (1966).
27. R. P. Hellman and J. Zwislocki, Loudness determination at low sound frequencies, *J. Acoust. Soc. Amer. 43*, 60–64 (1968).

28. R. R. Rowley and G. A. Studebaker, Monaural loudness-intensity relationships for a 1000-Hz tone, *J. Acoust. Soc. Amer. 45*, 1186–1192 (1969).
29. R. P. Hellman, Growth of loudness at 1000 and 3000 Hz, *J. Acoust. Soc. Amer. 60*, 672–679 (1976).
30. S. S. Stevens, On the psychophysical law, *Psych. Rev. 54*, 153–181 (1957).
31. S. S. Stevens, Matching functions between loudness and ten other continua. *Percept. Psychophysiol. 1*, 5–8 (1966).
32. J. C. Stevens and S. S. Stevens, Brightness function: Effects of adaptation, *J. Opt. Soc. Amer. 53*, 375–385 (1963).
33. S. S. Stevens, On predicting exponents for cross-modality matches, *Percept. Psychophysiol. 4*, 251–256 (1969).
34. B. Scharf, Critical bands, in Foundations of Modern Auditory Theory (J. V. Tobias, Ed.), Vol. I, New York Academic, 1970, pp. 157–202.
35. M. Florentine, S. Buus, and P. Bonding, Loudness of complex sounds as a function of the standard stimulus and the number of components, *J. Acoust. Soc. Amer. 64*, 1036–1040 (1978).
36. E. Zwicker and J. Feldtkeller, Über die Lautstärke von gleichförmigen Geräuschen, *Acoustica 5*, 303–316 (1955).
37. J. Feldtkeller and E. Zwicker, Das Ohr als Nachrichtenempfänger, Hirzel, Stuttgart, 1956.
38. E. Zwicker, G. Flottrop, and S. S. Stevens, Critical bandwidth in loudness summation, *J. Acoust. Soc. Amer. 29*, 548–557 (1957).
39. B. Scharf, Critical bands and the loudness of complex sounds near threshold, *J. Acoust. Soc. Amer. 31*, 365–380 (1959).
40. B. Scharf, Loudness of complex sounds as a function of the number of components, *J. Acoust. Soc. Amer. 31*, 783–785 (1959).
41. G. A. Miller, The perception of short bursts of noise, *J. Acoust. Soc. Amer. 20*, 160–170 (1948).
42. A. M. Small, J. F. Brandt, and P. G. Cox, Loudness as a function of signal duration, *J. Acoust. Soc. Amer. 34*, 513–514 (1962).
43. C. D. Creelman, Detection, discrimination, and the loudness of short tones, *J. Acoust. Soc. Amer. 35*, 1201–1205 (1963).
44. E. Ekman, G. Berglund, and V. Berglund, Loudness as a function of duration of auditory stimulation, *Scand. J. Psych. 7*, 201–208 (1966).
45. J. C. Stevens and J. W. Hall, Brightness and loudness as a function of stimulus duration, *Percept. Psychophysiol. 1*, 319–327 (1966).
46. J. S Zwislocki, Temporal summation of loudness: An analysis, *J. Acoust. Soc. Amer. 46*, 431–441 (1969).
47. D. McFadden, Duration-intensity reciprocity for equal-loudness, *J. Acoust. Soc. Amer. 57*, 701–704 (1975).
48. A. M. Richards, Loudness perception for short-duration tones in masking noise, *J. Speech Hearing Res. 20*, 684–693 (1977).
49. S. D. G. Stephens, Methodological factors influencing loudness of short duration tones, *J. Sound Vib. 37*, 235–246 (1974).
50. J. D. Hood, Studies on auditory fatigue and adaptation, *Acta Otol.*, Suppl. 92 (1950).

51. J. P. Egan, Perstimulatory fatigue as measured by heterophonic loudness balances, *J. Acoust. Soc. Amer. 27*, 111–120 (1955).
52. H. N. Wright, Measurement of perstimulatory auditory adaptation, *J. Acoust. Soc. Amer. 32*, 1558–1567 (1960).
53. A. M. Small and F. D. Minifie, Effect of matching time on perstimulatory adaptation, *J. Acoust. Soc. Amer. 33*, 1028–1033 (1961).
54. J. P. Egan and E. J. Thwing, Further studies on perstimulatory fatigue, *J. Acoust. Soc. Amer. 27*, 1225–1226 (1955).
55. T. E. Stokinger and G. A. Studebaker, Measurement of perstimulatory loudness adaptation, *J. Acoust. Soc. Amer. 44*, 250–256 (1968).
56. W. D. Fraser, J. W. Petty, and D. N. Elliott, Adaptation: Central or peripheral? *J. Acoust. Soc. Amer. 47*, 1016–1021 (1970).
57. J. W. Petty, W. D. Fraser, and D. N. Elliot, Adaptation and loudness decrement: A reconsideration, *J. Acoust. Soc. Amer. 47*, 1074–1082 (1970).
58. T. E. Stokinger, W. A. Cooper, and W. A. Meissner, Influence of binaural interaction on the measurement of perstimulatory loudness adaptation, *J. Acoust. Soc. Amer. 51*, 602–607 (1972).
59. T. E. Stokinger, W. A. Cooper, W. A. Meissner, and K. O. Jones, Intensity, frequency, and duration effects in the measurement of monaural perstimulatory loudness adaptation, *J. Acoust. Soc. Amer. 51*, 608–616 (1972).
60. D. E. Morgan and D. D. Dirks, Suprathreshold loudness adaptation, *J. Acoust. Soc. Amer. 53*, 1560–1564 (1973).

12
Pitch

In this chapter we shall deal with several attributes of sounds, grossly classified as "pitch." Like loudness, the word pitch denotes a perception with which we are all familiar. Yet, in spite of hundreds of years of study, the manner in which pitch is perceived remains enigmatic. For example, the issue of place versus temporal coding of frequency discussed in earlier chapters has been somewhat resolved: we know that both mechanisms are operative. However, the precise interaction of frequency and temporal coding in auditory frequency analysis remains largely unresolved. Furthermore, the perception of pitch, per se, appears to be multifaceted; and it may be that there are various kinds of pitch. Thus, we find that psychoacoustic pitch scales that have been confirmed by the rigors of laboratory replication are often at odds with musical scales that have withstood the tests of time and experience.

In light of these factors, the reader should not be surprised to find his fingers acting as placemarkers to other sections (particularly Chaps. 2, 4–6, 9, and 10) while perusing this chapter. The need to refer back to such phenomena as cochlear and neural coding, distortion, differential sensitivity, and masking should not be viewed as an inconvenience, but rather as exemplary of the pervasiveness of pitch in all aspects of auditory perception; a factor that makes it an exciting area in psychoacoustics.

PITCH SCALES

Stated in formal terms, pitch is the psychological correlate of frequency, such that high frequency tones are heard as being "high" in pitch and

low frequencies are associated with "low" pitches. We saw in Chap. 9 that not all changes in frequency are perceptible. Instead, a certain amount of frequency change is needed before the difference limen (DL) is reached. In other words, the frequency difference between two tones must be at least equal to the DL before they are heard as being different in pitch. Obviously, pitch and frequency are related, but not in a simple one-to-one manner.

How, then, is pitch related to frequency? This problem was tackled by Stevens et al. [1] using the fractionalization method discussed in Chap. 7. That is, their subjects were presented with a standard tone and were asked to adjust the frequency of a second tone until its pitch was one-half that of the standard. The result was a scale in which pitch is expressed as a function of frequency. This scale was revised by Stevens and Volkmann [2] in 1940. In this study, the subjects were required to adjust the frequencies of five tones within a certain frequency range until they were separated by equal pitch intervals. In other words, their task was to make the distance in pitch between tones A and B equal to that between B and C, which in turn was to be the same as that between C and D, and so on. Stevens and Volkmann also repeated the earlier fractionalization experiment, except that they now gave their subjects a 40 Hz tone that was arbitrarily assigned a pitch of zero. (The rationale will become clear momentarily.) Pitch scales have also been developed using the method of magnitude estimation [3,4].

The various methods result in slightly different pitch scales. Keeping this fact in mind, we will examine one scale [2] as an example of the relationship between frequency and pitch.

Figure 12.1 shows the revised pitch scale reported by Stevens and Volkmann [2]. This scale is a good fit to the data they obtained using equal pitch distances and fractionalization. Frequency is shown along the abscissa and pitch on the ordinate. As shown, pitch is expressed in units called *mels* [1,2]. By convention, 1000 mels is the pitch of a 1000 Hz tone presented at 40 phons. This association is shown by the lines labeled a in the figure. The frequency that sounds twice as high as 1000 mels has a pitch of 2000 mels, while 500 mels is half the pitch of 1000 mels, and so on. Thus, the lines labeled b in the figure indicate that 3000 Hz is 2000 mels. Note that tripling the frequency from 1000 to 3000 Hz only doubles the pitch, from 1000 to 2000 mels. Furthermore, even a cursory look at Fig. 12.1 reveals that the audible frequency range of about 20,000 Hz is focused down to a pitch range of only about 3500 mels. Nevertheless, Stevens and Volkmann demonstrated rather good correspondence between pitch in mels and distance along the basilar membrane, both quantities expressed as a function of frequency. Scharf [5] pointed out that approximately 150 mels corresponds to the typical critical bandwidth (see Chaps. 10 and 11).

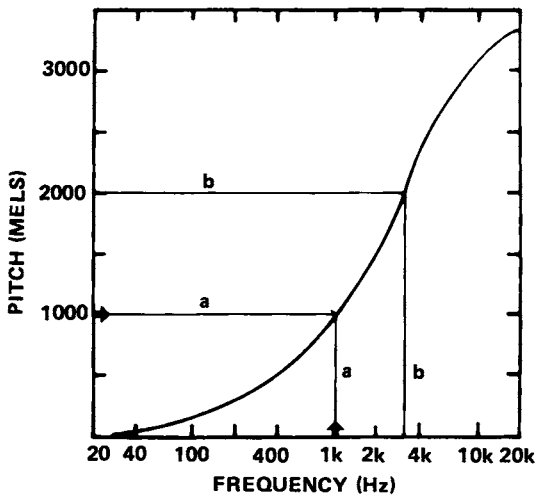

Figure 12.1 The revised pitch (mel) scale (see text). (Adapted from Stevens and Volkmann [2]. Permission *Amer. J. Psych.*)

Stevens and Volkmann employed both direct estimates of the pitch remaining below 40 Hz and extrapolations to arrive at 20 Hz as being the lowest perceptible pitch. (The interval between 25 and 40 Hz was equal in pitch to that between 40 and 52 Hz; hence, the assignment of "zero pitch" to 40 Hz in the fractionalization method used in their study.) This estimate is consistent with Bekesy's [6] observation that the lowest frequency yielding a sensation of pitch is approximately 20 Hz.

It is interesting to ask how the psychoacoustic pitch scale (in mels) relates to the musical pitch scales which divide the frequency range into subjective intervals such as octaves (1:2 relationships), fifths (2:3 relationships), etc. For informative reviews of musical perception, see Littler [7] and Ward [8]. A pervasive problem has been that the psychoacoustic and musical pitch scales are often at odds with one another. For example, the musical interval between 100 and 200 Hz corresponds to one octave, as does the interval from 1000 to 2000 Hz. On the other hand, the interval in mels between 100 and 200 Hz is different from the interval between 1000 and 2000 Hz (see Fig. 12.1) Obviously, however, we cannot arbitrarily dismiss one scale in favor of the other simply because they disagree. The evidence for both approaches is firmly defensible. What, then, might be the source of the differences between the psychoacoustic and musical pitch scales?

Insight into this problem was provided by Terhardt [9], who suggested a dichotomy between the *spectral pitch* of pure tones and the *virtual pitch* of tonal complexes. This concept is probably best aproached

Figure 12.2 Visual analogy for the perception of virtual pitch.

by comparing psychoacoustic and musical consonance. Consonance simply means that when two sounds are presented together they result in a pleasant perception; in contrast, sounds that appear unpleasant together are dissonant. It has been shown [10–12] that consonance for pairs of pure tones depends upon the difference between the two frequencies ($f_2 - f_1$). If the tones are relatively close in frequency there will be a sensation of "roughness" due to rapid beats between the two tones. (Beats will be discussed later in the chapter. For now, they may be thought of as a waxing and waning in the perception of two simultaneously presented tones that differ slightly in frequency.) This roughness is perceived as being unpleasant or dissonant. If the tone are far enough apart that there is no roughness, then the two tones are consonant. Thus, psychoacoustic consonance relies upon the distance in frequency between two tones. In music, however, consonance is produced by frequency intervals which are small integer ratios (e.g., 1:2, 2:3, 3:4). A complex ratio will be judged dissonant even if there is no sensation of roughness.

Let us return now to the difference between spectral and virtual pitch. While both quantities are of course related to the spectrum of the stimuli, virtual pitch goes beyond the frequency analysis to include the perception of an auditory gestalt calling upon what has been previously learned. We draw upon learned cues in the recognition of virtual pitch in the same way that we use them in Fig. 12.2 to visually synthesize a complete image of the word "hearing" from the incomplete contours. Musical pitch intervals depend upon a knowledge of the intervals between the lowest six to eight harmonics of a complex tone. This knowledge is, in turn, learned in the course of one's ongoing processing of speech. Hence, we might consider the perception of virtual pitch to be a matter of recognizing learned pitch intervals which are familiar to the pitch "processor" in the central auditory nervous system.

PITCH AND INTENSITY

Is the pitch of a tone changed by varying its intensity? To study this problem, Stevens [13] asked subjects to adjust the intensity of a tone until it had the same pitch as a standard tone of slightly different fre-

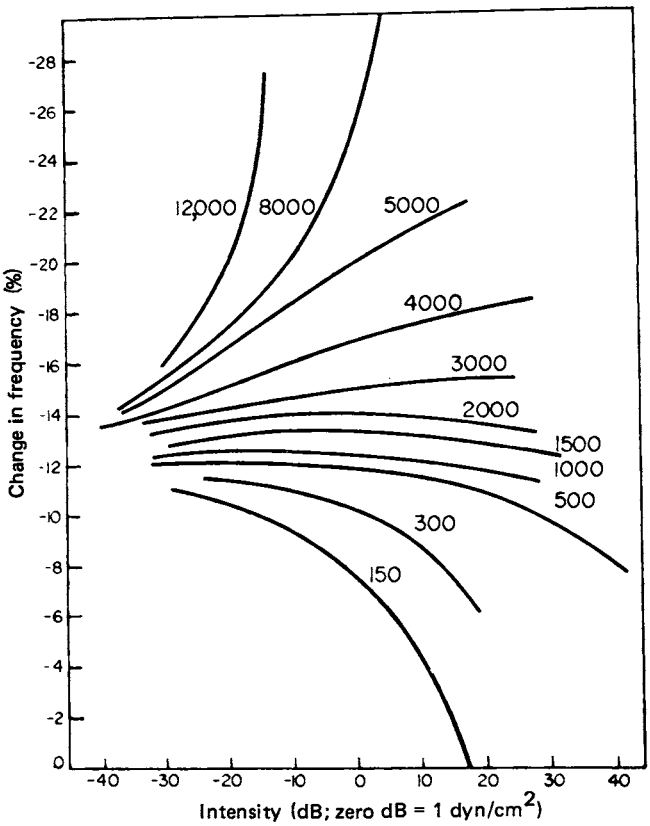

Figure 12.3 Equal pitch contours for a single subject (see text). (Adapted from Stevens [13]. Permission *J. Acoust. Soc. Amer.*)

quency. The results obtained for one subject who was a "good responder" are summarized in Fig. 12.3. As the figure shows, increasing the intensity of the tone increased its pitch for frequencies 3000 Hz and above, but lowered its pitch for frequencies 1000 Hz and below. The pitch stayed essentially constant as intensity was varied for tones between 1000 and 3000 Hz. Snow [14] reported that pitch changed substantially as intensity was increased for tones below about 300 Hz, but that there was little or no change between approximately 1000 and 2000 Hz.

Although the data in Fig. 12.3 are frequently cited as representative of how intensity affects pitch, subsequent studies [15–17] have failed to corroborate the large pitch changes associated with intensity increases or the universality of such observations. For example, Ward [16] studied the effect of intensity on pitch at 250 Hz, using five musicians as subjects.

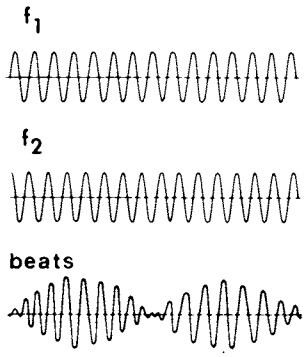

Figure 12.4 Tones of slightly different frequency, f_1 and f_2, result in beats which fluctuate (wax and wane) at a rate equal to the difference between them ($f_2 - f_1$). See text.

He found relatively large pitch changes with increasing intensity for only two subjects, whereas the other three showed essentially no change in pitch between 40 and 90 dB SPL. Furthermore, Cohen [17] found no significant differences in pitch as intensity was increased for low frequencies, and pitch changes of only 2% or less for high frequencies. It thus appears that one set of "representative" curves cannot be used to show how pitch is affected by intensity, and that changes that do occur are rather small.

BEATS, HARMONICS, AND COMBINATION TONES

We have seen in several contexts that a pure tone stimulus will result in a region of maximal displacement along the basilar membrane according to the place principle. Now, suppose that a second tone is added whose frequency f_2 is slightly different from that of the first sinusoid (f_1), as in Fig. 12.4. If the frequency difference between the two tones ($f_2 - f_1$) is small (say, 3 Hz), then the two resulting excitation patterns along the cochlear partition will overlap considerably, so that the two stimuli will be indistinguishable. However, the small frequency difference between the two tones will cause them to be in phase and out of phase cyclically in a manner that repeats itself at a rate equal to the frequency difference $f_2 - f_1$. Thus, a combination of a 1000 Hz tone and a 1003 Hz tone will be heard as a 1000 Hz tone that waxes and wanes in level (beats) at a rate of three times per second. This perception of aural beats therefore reflects the limited frequency-resolving ability of the ear.

If the two tones are equal in level, then the resulting beats will alternate between maxima that are twice the level of the original tones and minima that are inaudible due to complete out-of-phase cancellation. Such beats are aptly called "best beats." Tones which differ in level result in smaller maxima and incomplete cancellation. As one would expect, the closer are the levels of the two tones, the louder the beats will sound.

As the frequency difference between the two tones widens, the beats become faster. These rapid amplitude fluctuations are perceived as "roughness" rather than as discernible beats, as discussed earlier in the chapter. Further widening of the frequency separation results in the perception of the two original tones, in addition to which a variety of combination tones may be heard. Combination tones, as well as aural harmonics, are the result of nonlinear distortion in the ear.

A simple example demonstrates how nonlinear distortions produce outputs which differ from the inputs. Consider two levers, one rigid and the other springy. The rigid lever represents a linear system. If one moves one arm of this lever up and down sinusoidally (the input) then the opposite arm will also move sinusoidally (the output). On the other hand, the springy lever has a nonlinear response. A sinusoidal input to one arm will cause the other arm to move up and down, but there will also be superimposed overshoots and undershoots in the motion, due to the "bounce" of the springy lever arms. Thus, the responding arm of the lever will move with a variety of superimposed frequencies (distortion products) even though the stimulus is being applied sinusoidally. In other words, the distortion products are those components at the output of the system that were not present at the input.

The simplest auditory distortion products are aural harmonics. As their name implies, these are distortion products which have frequencies that are multiples of the stimulus frequency. For example, a stimulus frequency f_1, when presented at a high enough level, will result in aural harmonics whose frequencies correspond to $2f_1$, $3f_1$, etc. Therefore, a 500 Hz primary tone (f_1) will result in aural harmonics that are multiples of 500 Hz (1000 Hz, 1500 Hz, etc.).

If two primary tones f_1 and f_2 are presented together, nonlinear distortion will result in the production of various combination tones due to the interactions among the primaries and the harmonics of these tones. For convenience, we will call the lower-frequency primary tone f_1 and the higher one f_2. There are several frequently encountered combination tones that we shall touch upon. See Boring [18] and Plomp [19] for interesting historical perspectives, and Goldstein et al. [20] for an excellent review of the compatibility among physiological and psychoacoustic findings on combination tones.

It is necessary to devise methods that enable us to quantify the aspects of combination tones. A rather classical method takes advantage of the phenomenon of aural beats discussed above. Recall that best beats occur when the beating tones are of equal amplitudes. Generally stated, this technique involves the presentation of a "probe" tone at a frequency close enough to the combination tone of interest so that beats will occur between the combination and probe tones. Characteristics of the combination tone are inferred by varying the amplitude of the probe until the subject reports hearing best beats (i.e., maximal amplitude variations). The best beats method, however, has been the subject of serious controversy [21–24], and has been largely replaced by a cancellation technique [25–27]. The cancellation method also employs a probe tone; but in this method, instead of asking the subject to detect best beats, the probe tone is presented at the frequency of the combination tone, and its phase and amplitude are adjusted until the combination tone is cancelled. Cancellation occurs when the probe tone is equal in amplitude and opposite in phase to the combination tone. The characteristics of the combination tone may then be inferred from those of the probe tone that cancels it.

Techniques such as these have resulted in various observations about the nature of combination tones. We will only mention a few. The simplest combination tones result, of course, from adding or subtracting the two primary tones. The former is the summation tone $f_1 + f_2$. For primaries of 800 Hz (f_1) and 1000 Hz (f_2), the summation tone would be 1800 Hz. We will say little about the summation tone except to point out that it is quite weak and not always audible. On the other hand, the difference tone $f_2 - f_1$ is a significant combination tone which is frequently encountered [19]. For the above-mentioned 800 and 1000 Hz primaries, the difference tone would be 200 Hz. The difference tone is heard only when the primary tones are presented well above threshold. Plomp [19] found, despite wide differences among his subjects, that the primaries had to exceed approximately 50 dB sensation level in order for the difference tone to be detected.

The cubic difference tone ($2f_1 - f_2$) is another significant and frequently encountered combination tone [19]. For our 800 and 1000 Hz primary tones, the resulting cubic difference tone is $2(800) - 1000 = 600$ Hz. The particularly interesting aspect of the cubic difference tone is that it is audible even when the primaries are presented at low sensation levels. For example, Smoorenburg [28] demonstrated that $2f_1 - f_2$ is detectable when the primaries are only 15–20 dB above threshold, although he did find variations among subjects.

When the primary tones exceed 1000 Hz, the level of the difference tone $f_2 - f_1$ tends to be rather low (approximately 50 dB below the level of the primaries); in contrast, the difference tone may be as little as 10 dB below the primaries when they are presented below 1000 Hz [25,26,29,30]. On the other hand, the cubic difference tone $2f_1 - f_2$ appears to be limited to frequencies below the lower primary f_1, and its level increases as the ratio f_2/f_1 becomes smaller [26–30]. Furthermore, the cubic difference tone has been shown to be within approximately 20 dB of the primaries when the frequency ratio of f_2 and f_1 is on the order of 1.2 [28,30].

An interesting attribute of combination tones is their stimulus-like nature [20]. In other words, the combination tones themselves interact with primary (stimulus) tones as well as with other combination tones to generate beats and higher-order (secondary) combination tones, such as $3f_1 - 2f_2$ and $4f_1 - 3f_2$. Goldstein et al. [20] have recently shown that such secondary combination tones have properties similar to those of combination tones generated by the primaries. Interestingly, Goldstein et al. demonstrated that secondary combination tones generated by the interaction of the cubic difference tone $2f_1 - f_2$ and a third primary (stimulus) tone were obliterated when $2f_1 - f_2$ was cancelled. This result, of course, supports the concept that combination tones actually exist as distinct entities within the cochlea once they have been generated by nonlinear processes. The exact nature(s) of these distortion process(es) have yet to be unquestionably established.

COMPLEX TONES AND PERIODICITY PITCH

We generally think of the pitch of a sound as being dependent upon the frequencies at which energy is physically present. Thus, high-frequency energy would be associated only with high pitch, as predicted by the place principle. However, we have also seen (Chap. 5) that auditory neurons fire in synchrony with the period of a tone for frequencies up to about 1100 Hz, and that statistically significant phase locking occurs for frequencies as high as approximately 5000 Hz, regardless of the neuron's characteristic frequency. On this basis, we should not be surprised to find that pitch perception may be based upon temporal factors (periodicity) as well as upon the coding of frequency by place (see also Chap. 2).

Pitch perception based upon the periodicity of the stimulus waveform is descriptively called periodicity pitch. For reasons that will become readily apparent, it is also known as residue pitch, low pitch, and

repitition pitch, and by other such terms. Periodicity pitch was first described in 1841 by Seebeck [31]. His observations of a temporally based pitch were inconsistent with the popular resonance theory, and the furor that ensued has been described by Schouten [32], who reintroduced the phenomenon in 1940.

The manner in which periodicity results in a perception of pitch was shown rather clearly in an experiment by Thurlow and Small [34]. They found that if a high-frequency tone is interrupted periodically, then the subject will perceive a pitch corresponding to the frequency whose period is equal to the interruption rate. Thus, if the high-frequency tone is interrupted every 10 msec (the period of a 100 Hz tone), then subjects will match the pitch of the interrupted high-frequency tone to that of a 100 Hz tone. Perhaps the best-known example of periodicity pitch is the perception of the "missing fundamental" [31,32]. Consider a complex periodic tone containing energy only above 1200 Hz, spaced as shown in Fig. 12.5a. This power spectrum shows energy at 1200 Hz, 1400 Hz, 1600 Hz, etc., all of which are higher harmonics of 200 Hz. Thus, there is no energy available to stimulate the more apical area of the basilar membrane, which would respond to the 200 Hz fundamental frequency. Nevertheless, subjects presented with this complex tone will match its pitch to that of a 200 Hz tone. In other words, the pitch of this complex periodic tone corresponds to its fundamental frequency (200 Hz) even though no energy is physically present at that frequency. It appears that the auditory system is responding to the period of the complex periodic tone (5 msec), which results in a pitch perception based upon that period (1/5 msec = 200 Hz). If the components were separated by 100 Hz (i.e., 2000, 2100, 2200, 2300 Hz), then the waveform would have a period of 10 msec, corresponding to 100 Hz. With a large enough number of harmonics, periodicity pitches can be perceived as high as roughly 1400 Hz [35,36].

We might now ask which harmonics are most important in the pitch perception of complex tones. The results of several experiments [35,37,38] suggest that the first three to five harmonics provide this information. It is understandable, in view of the nature of the critical bandwidth, that the higher harmonics will contribute less than the lower ones. Recall that the critical band widens as frequency increases. Thus, the lower harmonics will fall more than a critical bandwidth apart, so that they will be perceived separately from one another. However, the higher harmonics will fall within a critical bandwidth of each other, so that they will not contribute separately to the pitch of the complex tone.

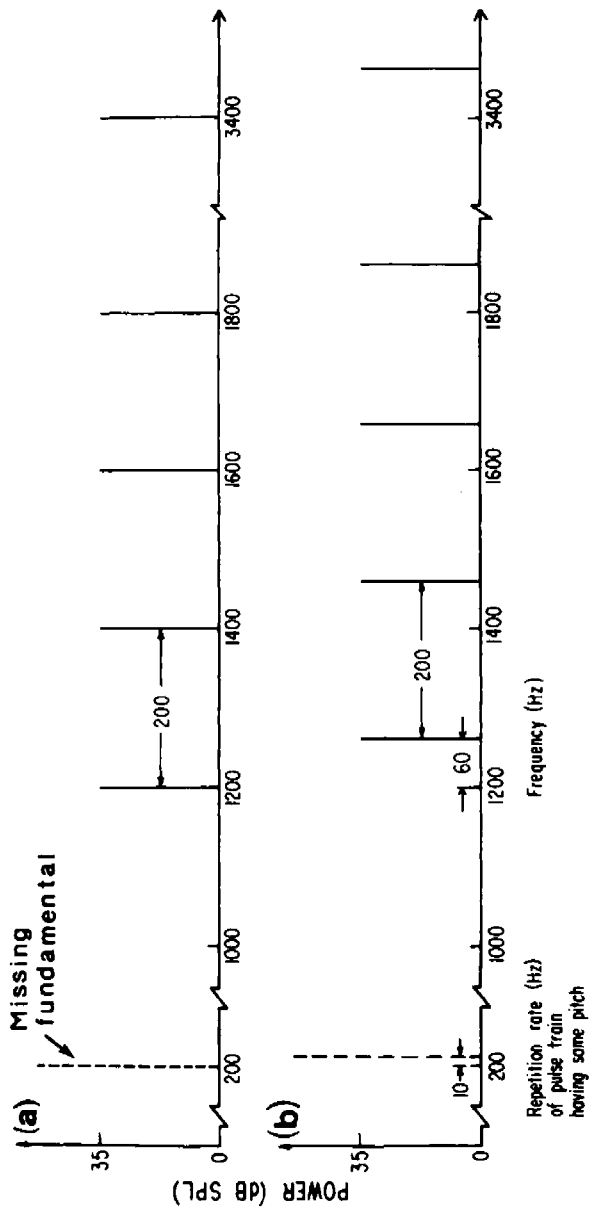

Figure 12.5 Power spectra showing (a) the missing fundamental, and (b) pitch shift. (Adapted from Patterson [33]. Permission *J. Acoust. Soc. Amer.*)

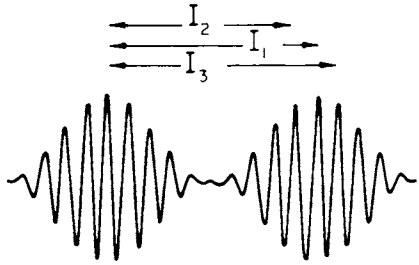

Figure 12.6 Comparison of the interval I_1 between equivalent peaks on repetitions of a waveform. Pitch ambiguity results when the pitch-extracting mechanism compares the peak of one repetition to peaks not equivalent to it on the next repetition of the waveform (I_2 and I_3). (From Schouten et al., [39]. Permission *J. Acoust. Soc. Amer.*)

An especially interesting observation is the pitch shift of the missing fundamental [32,33,39–43]. Suppose we have a complex periodic tone composed of exact harmonics of 200 Hz, as in Fig. 12.5a (1200 Hz, 1400 Hz, 1600 Hz, etc.). The resulting missing fundamental has a pitch of 200 Hz. Interestingly, if the frequencies making up the complex are increased to, say, 1260 Hz, 1460 Hz, 1660 Hz, etc. (Fig. 12.5b), then there will be a slight increase (shift) in the pitch of the missing fundamental, even though these frequencies are still all multiples of 200 Hz. In the first case, the harmonics are exact multiples of 200 Hz, so that there is exactly 5 msec between equivalent peaks (Fig. 12.6) in the fine structure of the repeated waveform. The upward shift of the harmonics by 60 Hz results in a slightly shorter interval between the equivalent peaks as the waveform is repeated, so that the pitch shifts upwards a bit. Ambiguities of pitch would result when the nearly but not exactly equivalent peaks are compared, as in the figure.

This fine structure or "peak picking" explanation was suggested by Schouten and colleagues [32,39], and by deBoer [40], among others. A problem, however, is that a mechanism that picks out aspects of the fine structure requires that phase changes must alter the pitch of the complex waveform. Unfortunately, there is considerable disagreement in the literature [33,41–47] as to whether the phase of a complex tone affects its pitch or timbre. It probably does not. (We may define timbre as those aspects of a complex sound's spectral content that enable us to differentiate, say, between middle C played on a violin, a piano, and a French horn. Plomp [48] defines timbre as the sensory attribute that enables one to judge differences between complex sounds having the same pitch, loudness, and duration.)

In light of this conclusion, Whightman [47] has suggested a mathematical model explaining these aspects of pitch on the basis of pattern recognition. Pattern recognition may be thought of as the extraction of those similar attributes of a stimulus which allow it to be recognized as a member of a class in spite of variations in details. For example, the letter "A" is recognized whether it is printed upper- or lowercase, in italics, or even in this author's illegible penmanship (see Fig. 12.2). In Whightman's model the waveform is first coarsely analyzed at a peripheral level, and the result is then subjected to a Fourier analysis. The resulting neural pattern is examined by a pitch extractor that looks for important features shared by stimuli having the same pitch. The details of the model are of course well beyond the current scope; the reader is referred to Plomp [48,49], Ritsma [50], and Whightman [42,47] for detailed reviews of such models.

Place theory would suggest that the missing fundamental is due to energy present at the fundamental frequency as a result of distortions. In other words, the difference tone $f_2 - f_1$ would be the same as the missing fundamental since, for example, 1100 − 1000 = 100 Hz. However, this supposition is quite unlikely because the missing fundamental differs from combination tones in several dramatic ways. For example, the missing fundamental is heard at levels as low as about 20 dB SPL [34,51] whereas difference tones are not heard until the primary tones are presented at levels of 60 dB SPL or more [6,19]. Furthermore, if a probe tone is presented to a subject at a frequency close to a difference tone (which is actually represented at a place along the basilar membrane), then aural beats are heard. However, beats do not occur when the probe is added to the missing fundamental [32].

Even stronger evidence against the supposition that the missing fundamental is the result of energy at the apex of the cochlea due to distortions (or other means) comes from masking studies [51–53]. These experiments demonstrated that masking of the frequency range containing the missing fundamental does not obliterate its audibility. In other words, real low-frequency tones and difference tones can be masked, but the missing fundamental cannot be. Of course, this result supports a neural basis for periodicity pitch perception.

The neural basis of periodicity pitch is highlighted by using dichotic stimuli — different stimuli presented to each ear. Houtsma and Goldstein [54] presented one harmonic to each ear and asked their subjects to identify melodies based upon the perception of missing fundamentals. If the missing fundamental were really a difference tone created in the cochlea, then the subjects would not hear it, because only one tone was available to each cochlea. The missing fundamental was perceived when the harmonics were presented separately to the two ears, indicating that

the phenomenon has to occur within the central auditory nervous system, since this is the only region where the harmonics are simultaneously represented. The reader should consult the recent works by Goldstein and associates [55,56] for interesting insights into the possible nature of a central pitch processor.

REFERENCES

1. S. S. Stevens, J. Volkmann, and E. B. Newman, A scale for the measurement of the psychological magnitude pitch, *J. Acoust. Soc. Amer. 8*, 185–190 (1937).
2. S. S. Stevens and J. Volkmann, The relation of pitch to frequency: A revised scale, *Amer. J. Psych. 53*, 329–353 (1940).
3. J. Beck and W. A. Shaw, Magnitude estimations of pitch, *J. Acoust. Soc. Amer. 34*, 92–98 (1962).
4. J. Beck and W. A. Shaw, Single estimates of pitch magnitude, *J. Acoust. Soc. Amer. 35*, 1722–1724 (1963).
5. B. Scharf, Critical bands, in *Foundations of Modern Auditory Theory* (J. V. Tobias, Ed.), Vol. I, Academic, New York, 1970, pp. 157–202.
6. G. Bekesy, *Experiments in Hearing*, McGraw-Hill, New York, 1960.
7. T. S. Littler, *The Physics of the Ear*, Pergamon, London 1965.
8. W. D. Ward, Musical perception, in *Foundations of Modern Auditory Theory* (J. V. Tobias, Ed.), Vol. I, Academic, New York, 1970, pp. 405–447.
9. E. Terhardt, Pitch, consonance, and harmony, *J. Acoust. Soc. Amer. 55*, 1061–1069 (1974).
10. R. Plomp and W. J. M. Levelt, Tonal consonance and critical bandwidth, *J. Acoust. Soc. Amer. 38*, 548–560 (1965).
11. R. Plomp and H. J. M. Steeneken, Interference between two simple tones, *J. Acoust. Soc. Amer. 43*, 883–884 (1968).
12. A. Kameoka and M. Kuriyagawa, Consonance theory. Part I: Consonance of dyads, *J. Acoust. Soc. Amer. 45*, 1451–1459 (1969).
13. S. S. Stevens, The relation of pitch to intensity, *J. Acoust. Soc. Amer. 6*, 150–154 (1935).
14. W. B. Snow, Changes of pitch with loudness at low frequencies, *J. Acoust. Soc. Amer. 8*, 14–19 (1936).
15. C. T. Morgan, W. R. Garner, and R. Galambos, Pitch and intensity, *J. Acoust. Soc. Amer. 23*, 658–663 (1951).
16. W. D. Ward, The subjective octave and the pitch of pure tones, Ph.D. Dissertation, Harvard University, Cambridge, Mass., 1953.
17. A. Cohen, Further investigations of the effects of intensity upon the pitch of pure tones, *J. Acoust. Soc. Amer. 33*, 1363–1376 (1961).
18. E. G. Boring, *Sensation and Perception in the History of Experimental Psychology*, Appleton-Century, New York, 1942.
19. R. Plomp, Detectability threshold for combination tones, *J. Acoust. Soc. Amer. 37*, 1110–1123 (1965).

20. J. L. Goldstein, G. Buchsbaum, and M. Furst, Compatability between psychophysical and physiological measurements of aural combination tones, *J. Acoust. Soc. Amer. 63,* 474–485 (1978).
21. M. Lawrence and P. J. Yantis, Onset and growth of aural harmonics in the overloaded ear, *J. Acoust. Soc. Amer. 28,* 852–858 (1956).
22. M. F. Meyer, Aural harmonics are ficticious, *J. Acoust. Soc. Amer. 29,* 749 (1957).
23. R. Chocolle and J. P. Legouix, On the inadequacy of the method of beats as a measure of aural harmonics, *J. Acoust. Soc. Amer. 29,* 749–750 (1957).
24. M. Lawrence and P. J. Yantis, In support of an "inadequate" method for detecting "ficticious" aural harmonics, *J. Acoust. Soc. Amer. 29,* 750–751 (1957).
25. E. Zwicker, Der Ungewöhnliche Amplitudengang der Nichtlinearen Verzerrungen des Ohres, *Acoustica 5,* 67–74 (1955).
26. J. L. Goldstein, Auditory nonlinearity, *J. Acoust. Soc. Amer. 41,* 676–689 (1967).
27. J. L. Hall, Nonmonotonic behavior of distortion product $2f_1 - f_2$: Psychophysical observations, *J. Acoust. Soc. Amer. 58,* 1046–1050 (1975).
28. G. F. Smoorenburg, Audibility region of combination tones, *J. Acoust. Soc. Amer. 52,* 603–614 (1972).
29. J. L. Hall, Auditory distortion products $f_2 - f_1$ and $2f_1 - f_2$, *J. Acoust. Soc. Amer. 51,* 1863–1871 (1972).
30. J. L. Hall, Monaural phase effect: Cancellation and reinforcement of distortion products $f_2 - f_1$ and $2f_1 - f_2$, J. Acoust. Soc. Amer. 51, 1872–1881 (1972).
31. A. Seebeck, Beohachtungen über einige Bedingungen der Entstehung von Tönen, *Ann. Phys. Chem. 53,* 417–436 (1841).
32. J. F. Schouten, The residue, a new concept in subjective sound analysis, *Proc. Kon. Ned. Akad. 43,* 356–365 (1940).
33. R. D. Patterson, The effects of relative phase and the number of components on residue pitch, *J. Acoust. Soc. Amer. 53,* 1565–1572 (1973).
34. W. R. Thurlow and A. M. Small, Pitch perception of certain periodic auditory stimuli, *J. Acoust. Soc. Amer. 27,* 132–137 (1955).
35. R. Plomp, Pitch of complex tones, *J. Acoust. Soc. Amer. 41,* 1526–1533 (1967).
36. B. C. J. Moore, Some experiments relating to the perception of complex tones, *Quart. J. Exp. Psych. 25,* 451–475 (1973).
37. R. J. Ritsma, Frequencies dominant in the perception of the pitch of complex tones, *J. Acoust. Soc. Amer. 42,* 191–198 (1967).
38. F. A. Bilsen and R. J. Ritsma, Repetition pitch mediated by temporal fine structure at dominant spectral regions, *Acoustica 19,* 114–116 (1967).
39. J. F. Schouten, R. Ritsma, and B. Cardozo, Pitch of the residue, *J. Acoust. Soc. Amer. 34,* 1418–1424 (1962).
40. E. deBoer, On the residue in hearing, Ph.D. Dissertation, University of Amsterdam, Amsterdam, 1956.
41. G. Smoorenburg, Pitch perception of two-frequency stimuli, *J. Acoust. Soc. Amer. 48,* 924–942 (1971).

42. F. L. Whightman, Pitch and stimulus fine structure, *J. Acoust. Soc. Amer. 54*, 397–406 (1973).
43. T. J. F. Buunen, J. M. Festen, F. A. Bilsen, and G. van den Brink, Phase effects in a three-component signal, *J. Acoust. Soc. Amer. 55*, 297–303 (1974).
44. R. C. Mathes and R. Miller, Phase effects in monaural perception, *J. Acoust. Soc. Amer. 19*, 780–797 (1947).
45. J. C. R. Licklider, Influence of phase coherence upon the pitch of complex periodic sounds, *J. Acoustic Soc. Amer. 27*, 996 (1957).
46. R. Ritsma and F. Engel, Pitch of frequency-modulated signals, *J. Acoust. Soc. Amer. 36*, 1637–1644 (1964).
47. F. L. Whightman, The pattern-transformation model of pitch, *J. Acoust. Soc. Amer. 54*, 407–416 (1973).
48. R. Plomp, Timbre as a multidimensional attribute of complex tones, in *Frequency Analysis and Periodicity Detection in Hearing* (R. Plomp and G. F. Smoorenburg, Eds.), Sijtthoff, Leiden 1970, pp. 397–414.
49. R. Plomp, Auditory psychophysics, *Ann. Rev. Psych. 26*, 207–232 (1975).
50. R. Ritsma, Periodicity detection, in *Frequency Analysis and Periodicity Detection in Hearing* (R. Plomp and G. F. Smoorenburg, Eds.), Sijtthoff, Leiden 1970, pp. 250–266.
51. A. M. Small and R. A. Campbell, Masking of pulsed tones by bands of noise, *J. Acoust. Soc. Am. 33*, 1570–1576 (1961).
52. J. C. R. Licklider, Periodicity pitch and place pitch, *J. Acoust. Soc. Amer. 26*, 945 (1954).
53. R. D. Patterson, Noise masking of a change in residue pitch, *J. Acoust. Soc. Amer. 45*, 1520–1524 (1969).
54. A. J. M. Houtsma and J. L. Goldstein, The central origin of the pitch of complex tones: Evidence from musical interval recognition, *J. Acoust. Soc. Amer. 51*, 520–529 (1972).
55. J. L. Goldstein, A. Gerson, P. Srulovicz, and M. Furst, Verification of an optimal probabilistic basis of aural processing in pitch of complex tones, *J. Acoust. Soc. Amer. 63*, 486–497 (1978).
56. A. Gerson and J. L. Goldstein, Evidence for a general template in central optimal processing for pitch of complex tones, *J. Acoust. Soc. Amer. 63*, 498–510 (1978).

13
Binaural Hearing

We shall now examine several aspects of binaural hearing — i.e., of hearing with two ears rather than one. We shall see that binaural hearing offers a number of important advantages over monaural hearing which have obvious implications for daily living. The binaural processing of signals clearly involves the neural interaction of inputs from the two ears (see Chaps. 5 and 6). Furthermore, the importance of relatively low-frequency signal processing which we have already seen operating in pitch perception (Chap. 12) is a recurring theme in binaural hearing.

BINAURAL SUMMATION

Although Sivian and White [1] did not find significant differences between the minimal audible field (MAF) for the better ear and binaural MAF, subsequent studies demonstrated that the intensity needed to reach threshold is lower when listening with two ears than with one. The essential finding is that, if one first corrects for any difference in monaural threshold between the two ears, so that they are equal in terms of sensation level (SL), then the binaural threshold will be approximately 3 dB better (lower) than the monaural thresholds [2–5]. For example, to correct for the difference between a monaural threshold of 11 dB in the right ear and one of 16 dB in the left, the binaural stimulus would be presented 5 dB higher in the left ear. The resulting binaural threshold would be about 3 dB below these equated monaural thresholds. Hirsh [6] refers to this threshold advantage that occurs when listening with two ears as "binaural summation at threshold." (Recall that 3 dB is a doubl-

ing of power, 10 log 2/1 = 3 dB; so that "summation" implies a 2:1 advantage.) Similar binaural advantages have been demonstrated when the stimulus is white noise [7] or speech [2,3].

Loudness is also enhanced by binaural hearing. Based upon loudness level measurements, Fletcher and Munson [8] concluded that a stimulus presented at a given SPL will sound twice as loud binaurally as monaurally. "Binaural summation of loudness" [6] was shown as a function of SL by Caussé and Chavasse [9], who performed loudness balances between binaurally and monaurally presented tones. At SLs close to threshold, they found that a binaural tone had to be about 3 dB lower in intensity than a monaural tone in order to produce the same sensation of loudness. This binaural advantage increased gradually with SL so that equal loudness was produced by a binaural tone 6 dB softer than the monaural stimulus at about 35 dB SL. This difference remained essentially constant at approximately 6 dB for higher sensation levels.

Perfect binaural summation means that a sound is twice as loud binaurally as it is monaurally. That loudness summation actually occurs at the two ears was questioned by Reynolds and Stevens [10]. They found that the ratio of binaural to monaural loudness increased as a power function of stimulus level (with an exponent of approximately 0.066). Thus, although both binaural and monaural loudness increased as power functions of SPL, the rate of binaural loudness growth was steeper (exponent of 0.6) than of monaural growth (exponent 0.54). Reynolds and Stevens attributed this binaural advantage, or at least part of it, to processes in the central nervous system.

It thus appears that the nature of binaural summation is somewhat uncertain, particularly the problem of whether there is actually binaural *summation* of loudness, per se. For example, some findings have suggested that a tone is twice as loud binaurally as it is monaurally [11], whereas others have reported less than perfect binaural summation [12]. Marks [13] recently reported on the binaural summation of loudness for 100, 400, and 1000 Hz tones using magnitude estimation (and also loudness matches for corroboration). His findings are summarized in Fig. 13.1. The circles and squares show the loudness estimates for the left and right ears, respectively. The dotted lines show what the binaural estimates should be if summation is perfect. Notice that the actual binaural loudness estimates (shown by the triangles) fall almost exactly along the predicted functions. This indicates essentially perfect binaural summation at each frequency.

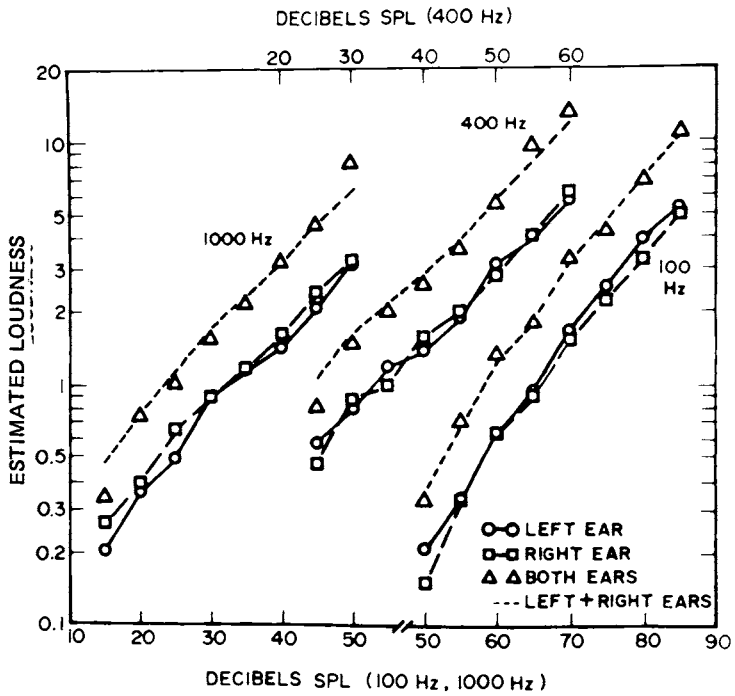

Figure 13.1 Loudness magnitude estimates for each ear and binaurally at 100, 400, and 1000 Hz. The dotted lines are predicted values for perfect summation. See text. (From Marks [13]. Permission *J. Acoust. Soc. Amer.*)

DIFFERENTIAL SENSITIVITY

Various studies suggest that differential sensitivity for both intensity [14–17] and frequency [17–19] is better binaurally than when listening with only one ear. A problem, however, has been that the small differences detected between monaural and binaural difference limens (DLs) may have been the result of loudness summation. Pickler and Harris [19] highlighted this problem. They found that the frequency DL was better binaurally than monaurally at low SLs. Recall from Chap. 9 that the effect of intensity upon differential sensitivity is greatest at low SLs, and that binaural hearing enhances sensitivity (or loudness) by roughly 3–6 dB. Thus, the smaller binaural DL may be due to summation rather than to some binaural mechanism for discrimination. To test this idea, Pickler and Harris adjusted the binaural signal level to account for the loudness

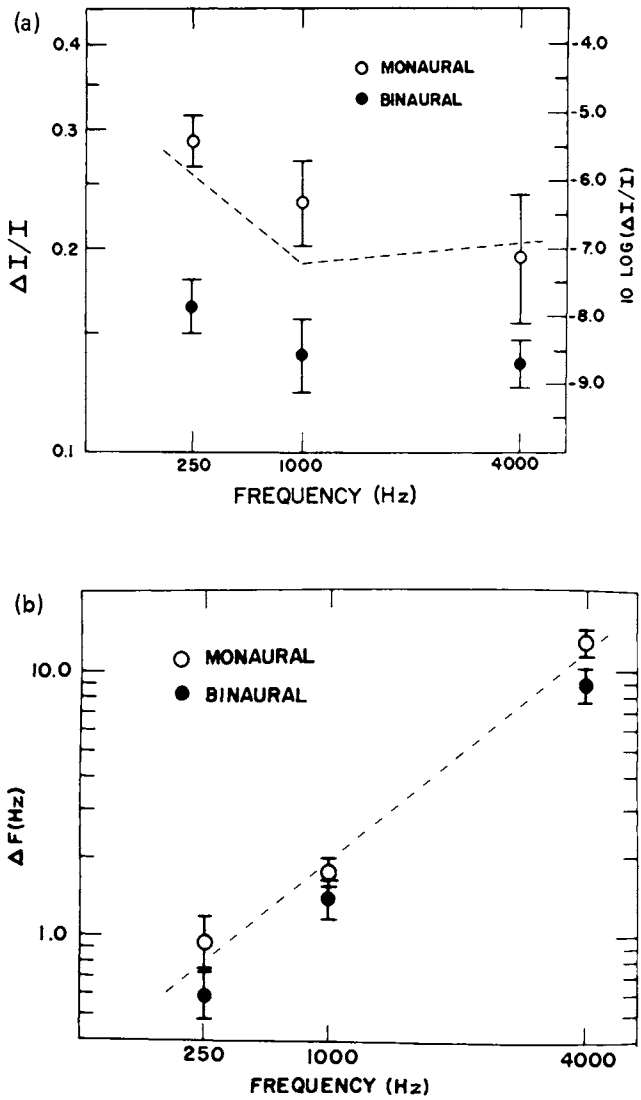

Figure 13.2 (a) Mean values of binaural and monaural ΔI/I. Dotted line shows predicted monaural values from Jesteadt et al. [20]. (b) Mean binaural and monaural values of Δf. Dotted line shows predicted monaural DLs from Wier et al. [21]. (From Jesteadt et al. [17]. Permission *J. Acoust. Soc. Amer.*)

advantage, and also tested DLs at a high level where differential sensitivity should not be affected by intensity. In both cases the difference between monaural and binaural DLs disappeared. It is thus unclear whether the binaural DL is smaller than it is monaurally, or whether the difference just reflects a level difference.

This enigma was essentially resolved in a recent study by Jesteadt et al. [17]. They obtained intensity and frequency DLs at 70 dB SPL for 250, 1000, and 4000 Hz tones using a two-interval forced-choice method. Their results are shown in Fig. 13.2. Note that binaural differential sensitivity is uniformly better (the DL is smaller) than monaural, and that the difference is largely the same regardless of frequency. The ratio of the monaural to the binaural DL is on the order of 1.65 for intensity and 1.44 for frequency. The binaural–monaural differences obtained by Jesteadt et al. are not attributable to a loudness advantage for binaural hearing because a difference of about 30 dB would have been needed to produce the observed binaural DL advantages [17,18]; and binaural summation is equivalent to only about 3–6 dB.

BINAURAL FUSION AND BEATS

Even though the sounds of daily life reach the two ears somewhat differently in terms of time, intensity, and spectrum, we still perceive a single image. As Cherry [22] points out, we perceive one world with two ears. More precisely, the similar but nonidentical signals reaching the two ears are fused into a single, coherent image (gestalt). This process is called binaural fusion.

Binaural fusion experiments require earphone listening because this allows us to precisely control the stimuli presented to the two ears, as well as how these signals are related. Generally, the experimenter is looking for a combination of stimuli that results in a fused image lateralized to the center (midline) of the head. The essential finding is that, although completely dissimilar signals are not fused, the auditory system does achieve binaural fusion as long as the signals presented to the two ears are similar in some way (23–27). The low frequencies, below roughly 1500 Hz, appear to be the most important. Thus, if each ear is presented with a 300 Hz tone at the same time, the subject will perceive a fused image in the center of his head.

A second example will demonstrate an important property of binaural fusion. If two different high-frequency tones are presented one to each ear, they will be heard as two separate signals. However, if a single low-frequency tone is superimposed upon both high frequencies so that they are caused to modulate at the frequency of the low tone, the listener will report a fused image [26]. This result shows that the audi-

tory system uses the low-frequency envelopes of the complex signals (their macrostructures) for fusion even though the details of the signals (their microstructures) are different. Fusion of speech can be shown to occur, for example, when only the high-frequency components of the speech waveform are directed to one ear and only the lows are presented to the other [27]. Even though neither ear alone receives enough of the speech signal for identification, the resulting fused image is readily identified.

The binaural fusion mechanism has been described in terms of a mathematical model by Cherry and Sayers [24,25]. The details of the model are beyond the current scope; however, its basis is that the central auditory nervous system carries out a running cross-correlation between the inputs to the two ears. In other words, the signals entering the ears are viewed as statistical events, and the fusion mechanism operates by looking for commonalities between the inputs coming from the two ears on an ongoing basis.

A very interesting phenomenon occurs when one tone is presented to the right ear and a second tone of slightly different frequency is presented to the left. The result is the perception of beats (see Chap. 12) in the fused image. Recall that beats occur when one combines two tones slightly different in frequency because phase differences between the tones result in alternating increases and decreases in amplitude. The intriguing aspect of binaural beats is that they occur even though the two signals are acoustically completely isolated from one another. Obviously, binaural beats must result from some interaction between the neural codings of the signals from the two ears, taking place within the central nervous system. (Cells have been identified in the superior olive that are responsive to the envelope of binaural beats [28]. They are probably at least partially involved in subserving the perception of binaural beats.)

Binaural beats differ from monaural beats in several ways [29–31]. Whereas monaural beats can be heard for interacting tones across the audible frequency range, binaural beats are associated with the lower frequencies, and the best responses are for tones between about 300 and 600 Hz. Binaural beats can still be heard even if the frequency difference between the ears is relatively wide, although the perception of the image changes with frequency separation (see below). Furthermore, binaural beats can also be perceived if there is a substantial intensity difference between the ears, and even if one of the tones is presented at a level below the behavioral threshold for that ear. (Recall from Chap. 5 that phase locking to stimulus cycle occurs at the very lowest levels at which an auditory neuron responds.)

Licklider et al. [29] reported that perceptual differences occur as the frequency separation widens between the ears. When identical frequencies are presented to two ears, the listener hears a fused image. When the frequencies are 2–10 Hz apart, the subject reports loudness fluctuations, which give way to a perception of "roughness" when the frequency difference reaches about 20 Hz. As the frequency separation becomes wider and wider, the fused image appears first to split into two smooth tones, and these tones then migrate in perceived location to the respective ears.

DIRECTIONAL HEARING

How do we determine the direction of a sound source? Intuitively, we know that some sort of comparison betwen the two ears must be involved.* The question is to determine what cues are used. Although we are most often concerned with binaural listening in a sound field (stereophony), we must often resort to the use of earphones to precisely control experimental conditions, as described above. Interestingly, however, stereophonic and headphone listening result in different perceptions of space. Sounds heard in a sound field seem to be *localized* in the environment. However, sounds presented through a pair of earphones are perceived to come from within the head, and their source appears to be *lateralized* along a plane between the two ears. Thus, identical sounds impinging at the same time upon the two ears appear to come from directly in front of (or behind) the listener, unless he is listening through earphones, in which case the source seems to be in the center of the head. This difference between extracranial localization and intracranial lateralization is easily experienced by comparing the way the same stereo record sounds through speakers and through headphones.

The arrangement in Fig. 13.3a shows the directional cues available in stereophonic listening. The signal from the speaker, which is off to the right, must follow a longer path to the far (left) ear than to the near (right) ear. As Fig. 13.3b shows, low frequencies have wavelengths that are longer than the path around the head, so that they "bend around" the head to the far ear (diffraction). However, higher frequencies (Fig. 13.3c) have wavelengths smaller than the head, so that they are "blocked" in the path to the far (left) ear. This "head shadow" causes a reduction in the intensity of the signal at the far ear. We would thus expect lower frequencies to be localized on the basis of time or phase dif-

*See Chap. 3 for a discussion of monaural and midline localization cues.

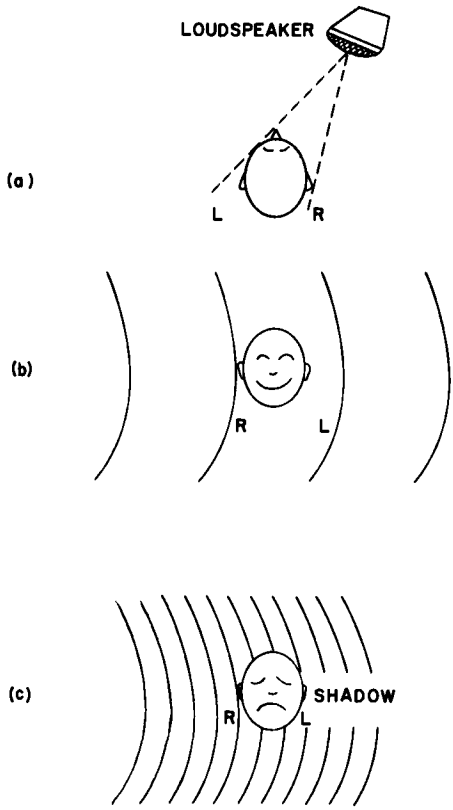

Figure 13.3 (a) Relationship between a loudspeaker and the two ears. (b) Low frequencies bend around the head due to their large wavelengths λ. (c) High frequencies have wavelengths smaller than head diameter so that an acoustic shadow results at the far ear.

ferences between the ears as the signal bends around the head; whereas localization of high frequencies would depend upon intensity differences due to the head shadow. (Recall Fig. 3.3, which shows the effects of azimuth angle on the sound pressure at the eardrum.)

Phase differences between the two ears are useful localization cues as long as the wavelength is larger than the distance the tone must travel from the near (right) ear to the far (left) ear. Assuming that this path around the head is 22–23 cm (roughly 8.75 in.) long, then it will take approximately 660 μsec for the sound to get from the near ear to the far ear, corresponding to a frequency of about 1500 Hz. For much longer wavelengths (i.e., lower frequencies), there will be a significant phase dif-

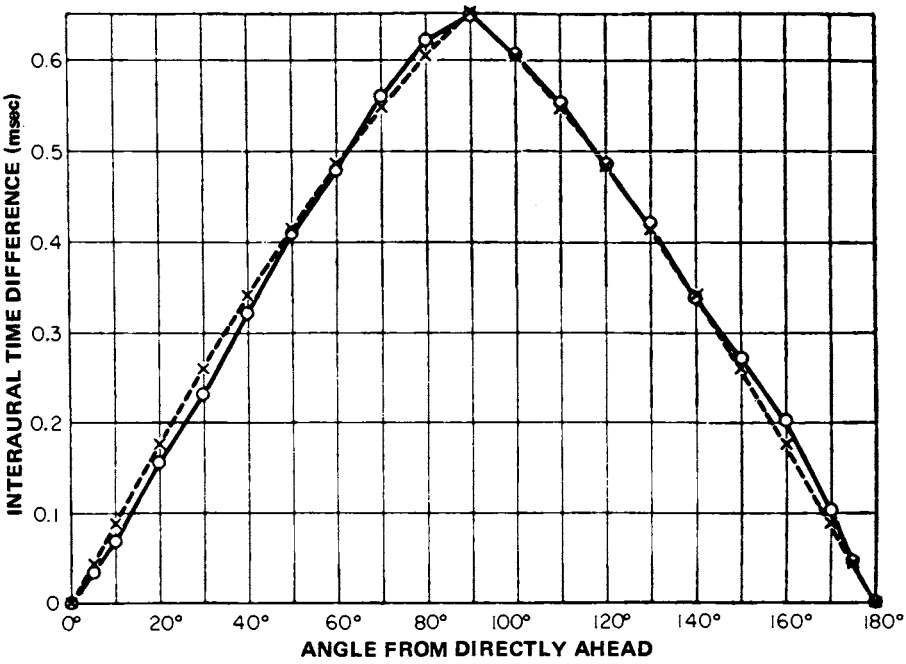

Figure 13.4 Interaural time differences for different loudspeaker azimuths. Circles are measurements of Feddersen et al. [32]. Crosses are calculated differences based on a solid sphere [33]. See text. (From Feddersen et al. [32]. Permission *J. Acoust. Soc. Amer.*)

ference between the ears which will provide an unequivocal localization cue. However, as the frequency approaches 1500 Hz, the phase difference becomes smaller and smaller, so that the phase discrepancy between the ears becomes quite ambiguous. These ambiguities in the phase cue should result in localization errors. At about 1500 Hz, where the path between the ears and the wavelength are equal, there will be no phase difference between the two ears (except between the times of arrival of the first wavefront). Thus, the localization mechanism will be unable to tell whether the signal is coming from the right or the left. Further increases in frequency will cause the tonal wavelength to be shorter than the size of the head. The resulting head shadow will produce intensity differences between the ears which then form the available localization cues.

Feddersen et al. [32] measured the interaural time and intensity differences for human heads as functions of angle around the head (azimuth) and frequency. The circles in Fig. 13.4 show their results for

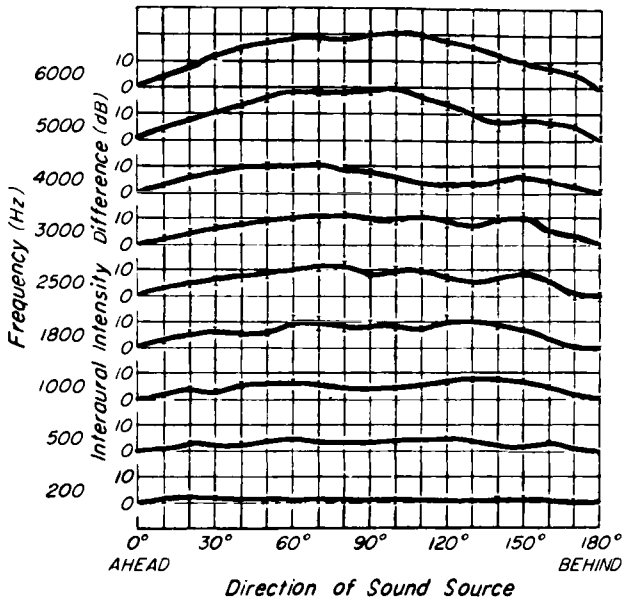

Figure 13.5 Interaural intensity differences as a function of frequency and azimuth (see text). (From Feddersen et al. [32]. Permission *J. Acoust. Soc. Amer.*)

interaural time differences. Notice that there is no difference in the arrival time at the two ears when the signals (clicks) come from directly in front of the listener (0°) or directly behind (180°) since the ears are equidistant from the sound source in both cases. Interaural arrival time differences develop as the loudspeaker is moved around the head, bringing it closer to one ear than the other. The time difference between the ears increases to a maximum when the loudspeaker is directly in front of one ear (90° azimuth), where the distance (and therefore the time delay) between the ears is greatest. As the speaker continues to move around the head, the interaural time difference (ITD) becomes smaller, reaching zero when the speaker is directly behind the head (180° azimuth), where once again the ears are equidistant from the sound source. Notice that the maximum ITD measured at 90° azimuth is about 660 μsec, which is what we expected based upon the previous discussion. These actual measurements correspond closely to what is predicted by calculations of interaural time delay based upon a solid sphere having the same diameter as the head (crosses in Fig. 13.4) [33].

Figure 13.5 shows the interaural intensity differences (IIDs) obtained by Feddersen et al. There were no intensity differences between the ears when the sound source was directly in front of (0°) or behind

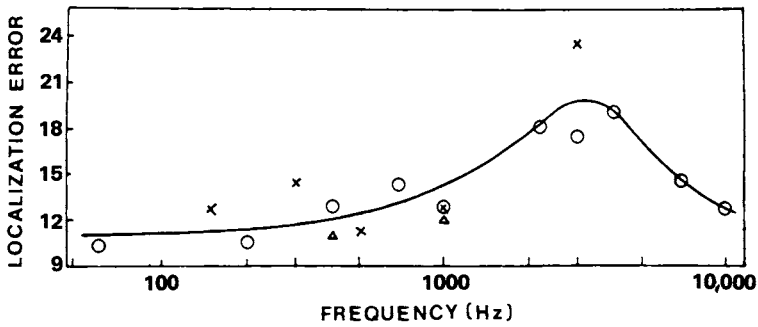

Figure 13.6 Localization error as a function of frequency. (Adapted from Stevens and Newman [35]. Permission *Amer. J. Psych.*)

(180°) the head. However, IIDs occurred when the loudspeaker was closer to one ear than the other, with a maximum when it was directly in front of one ear (90°). Note in particular that the IIDs were negligible at 200 Hz, and increased with frequency, reaching as much as 20 dB at 6000 Hz.

The importance of IIDs in high-frequency localization and of ITDs in low-frequency directionality was first demonstrated in 1907 by Lord Rayliegh [34]. However, the classic paper by Stevens and Newman [35] probably represents the first modern rigorous experimentation on directional hearing. Since echoes and reverberation (see below) introduce major complications into the localization of a sound source, Stevens and Newman sat their subjects in a chair elevated about 12 feet above the roof of the Harvard Biological Laboratories building. This made the test environment as anechoic (echo-free) as possible, by essentially removing all reflecting surfaces except for the roof below. The sound source was a loudspeaker mounted on a boom arm extended 12 feet from the listener (6 feet for low frequencies). Rotating the boom allowed the speaker to be placed at any azimuth angle. The loudspeaker was moved to the subject's right in 15° steps from 0° to 180°, and the task was to listen for the signal and report its apparent direction. We have already seen that interaural time and intensity differences are nil at 0° and 180° azimuth. This fact resulted in constant front–back reversals. Stevens and Newman therefore considered front–back reversals as correct responses; and the location of the sound source was judged relative to 0° or 180°, whichever was closer. In other words, since front and back were virtually indistinguishable, 30° right (30° off center from 0°) and 150° (30° off center from 180°) were considered to be equivalent responses.

Stevens and Newman's findings are shown as a function of frequency in Fig. 13.6. Localizations were most accurate below 1000 Hz and

above 4000 Hz, with the greatest errors between about 2000 and 4000 Hz. In terms of azimuth, the smallest errors were on the order of 4.6°, and occurred when the sound sources were off to one side. Furthermore, Stevens and Newman observed that noiselike sounds were better localized than tones. They attributed this result to quality (spectral) differences as well as to IIDs for the high-frequency energy in these noises.

Sandel et al [36] asked subjects to localize sounds in an anechoic room. Their heads were held immobile to preclude localization cues from head movements. Loudspeakers were placed at 0° azimuth and at 40° right and left. The stimuli were presented from one or two speakers. Two novel aspects of this study are notable. First, the use of two speakers permitted the generation of "phantom" sound sources between the speakers or off to one side, depending upon the phases of the signals. Second, an "acoustic pointer" was used to show the perceived location of the sound source. That is, the subject controlled the location of a speaker that rotated on a boom around his head. A noise from this speaker alternated with the presentations of the test tones. The subject indicated the perceived location of the sound source by placing the noise loudspeaker (pointer) at the same angle around his head.

There were systematic localization errors between about 1500 and 5000 Hz. When these errors were eliminated from the data, Sandel et al. found that ITDs accounted for the localization of tones below about 1500 Hz, but that the only available cues at high frequencies were IIDs. There were many random errors at 1500 Hz, showing that localization cues are ambiguous around this frequency. These results essentially confirm those of Stevens and Newman.

Another way to examine localization is to find the smallest difference in location between two sound sources that results in a different perceived location. Since the two sound sources are viewed relative to the head, this is the same as asking what is the smallest angle (or difference in azimuth) that a listener can discriminate. Mills [37–39] studied this phenomenon in depth, and called it the minimal audible angle (MAA). Specifically, he tested the MAA as a function of frequency when the sound sources were located in front of the subject (0°), and when they were 30°, 60°, and 75° off to the side. He found that the MAA was smallest (best) for the frequencies below about 1500 Hz and above approximately 2000 Hz, and was largest (poorest) between these frequencies. This result reflects the ambiguity of localization cues in the vicinity of 1500 Hz, thus confirming the previously mentioned findings. Mills also found that MAAs were most acute (1–2°) when the sound sources were directly in front of the head, and that they increased dramatically to very high values when the sources were at the side of the

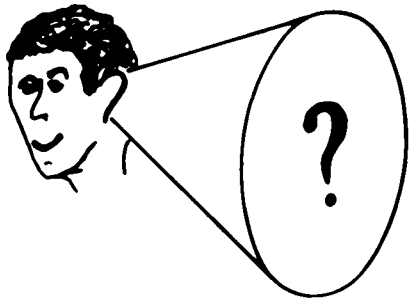

Figure 13.7 Cone of confusion (see text). (Modified after Mills [39].)

head. This result occurs because small changes in location in front of the head result in large interaural differences (especially ITDs). However, when the sources are off to one side of the head (facing one ear), the interaural differences remain largely the same in spite of reasonably large changes in angle between the loudspeakers. We might thus conceive of a "cone of confusion" (Fig. 13.7) to one side of the head, within which the interaural differences do not vary when the sound sources change location [39]. This image demonstrates the importance of head movements in localization, since these movements keep changing the position of the cone of confusion — the zone of ambiguity — thereby minimizing its detrimental effect.

Since earphones allow us to precisely control and manipulate the signals presented to the ears, lateralization experiments using them clarify and expand upon what we know about directional cues. Before discussing these studies, it is wise to point out a distinction between time and phase differences. Recall that low frequences need a longer time period t than high frequencies for completion of one full cycle (period is the reciprocal of frequency, $t = 1/f$). We are concerned with the effects of delaying the onset of a stimulus at one ear relative to the other. If we impose the same delay upon three different frequencies, the delay will correspond to different phases of their cycles (Fig. 13.8). In other words, the higher the frequency, the greater the phase difference due to a given time delay. Thus, if we impose a time delay upon a broad-band noise or click, there will be more of a phase difference for the higher-frequency components of the signal than for the low.

Several studies [40–42] have examined the effects of interaural time (phase) and intensity cues upon lateralization. While exact procedures vary, the general approach is to present two stimuli to the subject that differ with respect to interaural time (phase) or intensity, and to deter-

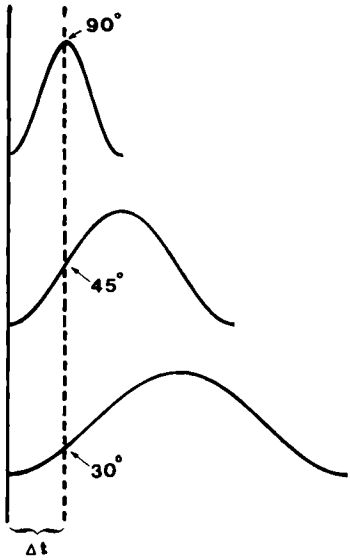

Figure 13.8 Relation between time delay Δt and phase shift for different frequencies.

mine whether this interaural disparity results in a perceptible change in lateralization. The findings essentially agree with the localization data. That is, ITDs are most important up to about 1300 Hz, and IIDs take over as the primary lateralization cue for higher frequencies.

Yost [43] performed a particularly interesting lateralization experiment analogous to Mills' MAA studies. He presented subjects with two stimuli. The first included a particular interaural time (actually phase) difference Θ. This difference, of course, resulted in a lateralization toward one side of the head analogous to the azimuth position of the stimuli in Mills' work. The second stimulus was the same except that the phase difference between the ears was larger by a slight amount $\Delta\Theta$. Thus, it was $\Theta + \Delta\Theta$. The subjects had to detect the value of $\Delta\Theta$ by discriminating between the two stimuli (Θ versus $\Theta + \Delta\Theta$). We might think of the phase difference Θ as analogous to azimuth, and of $\Delta\Theta$ as the MAA. For any value of Θ, the smaller the value of $\Delta\Theta$ needed for a change in apparent lateralization, the better the discrimination of interaural phase. Yost's findings are summarized in Fig. 13.9. Note that $\Delta\Theta$ is smallest (best) when Θ is 0° or 360°. These values of Θ are midline lateralizations because 0° and 360° correspond to zero phase disparity between the ears. Thus, the most acute interaural phase discriminations are made at the midline. On the other hand, interaural phase discrimination was poorest ($\Delta\Theta$ was largest) when Θ was 180°, i.e., when the signals were lateralized directly to one side. These findings are consistent with the MAA data.

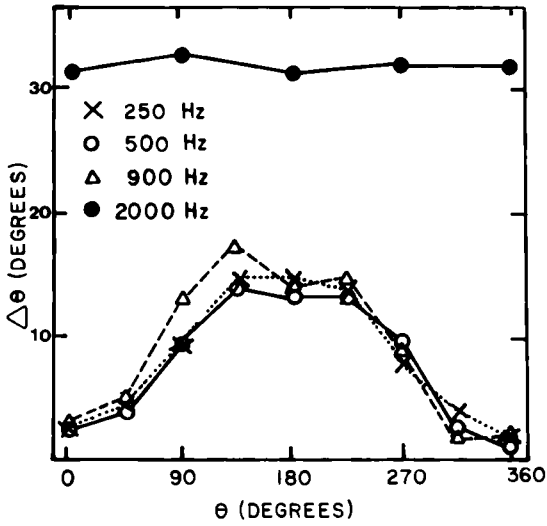

Figure 13.9 Changes in interaural phase ($\Delta\Theta$) needed to detect a difference in lateralization from a standard (reference) phase difference Θ. (Adapted from Yost [43]. Permission *J. Acoust. Soc. Amer.*)

Fig. 13.9 also shows that $\Delta\Theta$ is essentially the same for all frequencies up to 900 Hz. Interaural phase discrimination was significantly poorer at 2000 Hz, where it was constant at about 30°. Interaural phase had no effect at all for 4000 Hz (not shown on the graph). Thus, interaural phase was shown to be an important cue at low frequencies but unimportant for high, in a manner consistent with the sound field data.

Recall that Stevens and Newman [35] found better localization accuracy for noiselike sounds than for tones. Since noises contain considerable transient information, this suggests that transient sounds provide excellent directional cues. If so, then we would expect improved lateralization as more transient cues are made available to the listener. In other words, very short-term events (e.g., clicks), which provide a transient cue only at their onset and offset, should require larger ITDs for lateralization than longer events (e.g., noises), which provide ongoing transient disparities between the ears. Klumpp and Eady [40] found that the ITD required to achieve 75% correct lateralizations was 28 μsec for a single click. The ITDs were reduced to 11 μsec for a 2 sec burst containing 30 clicks in rapid succession, and to 9–10 μsec for a 1.4 sec noise. Tobias and Zerlin [44] found smaller ITDs as the duration of a noise burst was increased from 10 msec to 700 msec. The ITD was only 6 μsec for noise bursts that lasted 700 msec or more. (Tobias [45] pointed out that this is

puzzling, because synaptic delays average about 1000 μsec.) Furthermore, Tobias and Schubert [46] found that very small ongoing ITDs could offset much larger onset ITDs. Thus, lateralization acuity improves substantially when ongoing transient ITDs are provided by a noise or a click train.

Klumpp and Eady [40] found that ITDs of about 10 μsec resulted in correct lateralizations of noises that contained low frequencies. However, high-frequency noise bands required substantially larger ITDs (44 μsec for 2400–3400 Hz; 62 μsec for 3056–3344 Hz). Yost et al. [47] found that the lateralization of clicks was unaffected by removing their high frequencies by filtering or masking. Removing the low-frequency components, however, substantially interfered with the correct lateralization of the clicks.

Transient energy is also associated with the onset and offset of tone bursts. Yost [48] found that this transient energy enabled his subjects to lateralize a high-frequency (4050 Hz) tone burst on the basis of an ITD (71 μsec). Low-pass filtering did not detract from this performance; but the 4050 Hz tone burst could not be lateralized when the low-frequency segments of the associated transient were filtered out. The experiments show the importance of low-frequency information in the lateralization of wide-band and transient signals. It is interesting to note at this point that ITDs can subserve lateralizations of high-frequency tones if they are modulated by a lower frequency. Henning [49] demonstrated that a 3900 Hz tone can be lateralized on the basis of ITDs if it is modulated at 300 Hz. In this case, it appears that the lateralization is based upon the 300 Hz envelope rather than the 3900 Hz carrier tone.

TIME–INTENSITY TRADES

Having seen that both interaural time and intensity differences serve as lateralization cues, we might ask whether there is a trading relationship between them. In other words, can an ITD favoring one ear be offset by an IID favoring the other? This area has been extensively studied because of its theoretical implications.

Individual studies have varied considerably in technique. However, the basic experiment is to vary the interaural intensity of a signal in order to offset a given ITD, or vice versa. The subject's task is generally to adjust the IID (or ITD) until a midline image is perceived. A standard comparison signal presented without interaural differences may be used as a midline marker to assist the subject in achieving the lateralization.

Time–intensity trading studies are marked by inconsistent results between subjects. This fact, plus the nature of the findings, draws us to conclude that lateralization is actually a complex process rather than a

simple interaction of time and intensity cues. Harris [50] performed time–intensity trades for high- and low-pass filtered clicks. Instead of a consistent trading ratio, he observed a ratio of 25 μsec/dB for clicks containing energy below 1500 Hz, and of 90 μsec/dB for clicks containing above 1500 Hz. In another study, Moushegian and Jeffress [51] found that the trading ratio for a 500 Hz tone was about 20–25 μsec/dB for two subjects but only on the order of 2.5 μsec/dB for another subject.

A series of experiments by Jeffress, Hafter, and colleagues [52–55] has contributed substantially to our understanding of the lateralization mechanism. For a variety of stimuli, these workers found that subjects often reported two lateralized images instead of just one. Furthermore, the two images are associated with different trading ratios. One image is particularly dependent upon ITDs (especially for frequencies below 1500 Hz), but is essentially unaffected by IIDs. This is the "time image." The other is an "intensity image" that is responsive to both IIDs and ITDs at all audible frequencies. Typical trading ratios (click stimuli) are on the order of 2–35 μsec/dB for the time image and 85–150 μsec/dB for the intensity image [53]. It may be that the large intersubject differences reported by Harris [50] and by Moushegian and Jeffress [51] were due to responses to the time image by some subjects and to the intensity image by others.

A rather straightforward mechanism was proposed by Jeffress and McFadden [54] for the two lateralization images. In this model, the time image depends upon a comparison of the signals at the two ears on a cycle-by-cycle basis — i.e., it depends on an interaural comparison of the fine structures of the signals. The intensity image, on the other hand, depends upon a comparison of the envelopes (gross structures) at the two ears. In this case, the amplitudes of the peaks of the waveform envelopes serve as the interaural intensity cues, and the ITDs between envelope peaks are the timing cues. We have already seen that neurons in the auditory nerve respond to both stimulus intensity and cycle (Chap.5), and that brainstem neurons are responsive to ITDs and/or IIDs (Chap. 6). This situation provides a neurophysiological basis for the two mechanisms.

PRECEDENCE EFFECT, LOCALIZATION, AND REVERBERATION

Consider two apparently unrelated situations. The first involves listening to a monaural record album or a news broadcast through both speakers of a home stereo system. Thus, identical signals are coming from both speakers. Sitting equidistant from the speakers causes us to perceive a phantom sound source between them. However, sitting closer

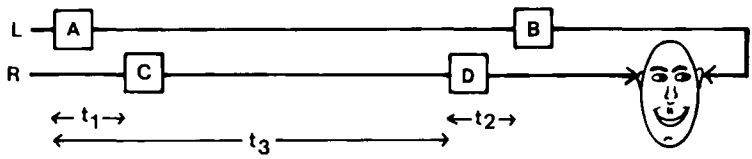

Figure 13.10 The precedence effect (see text).

to one speaker (so that the signal reaches our ears sooner from that direction than the other) gives us the impression that all of the sound is coming from the closer speaker. This occurs even though the other speaker is still on.

The second situation involves listening to someone talking in a hard-walled room. The sounds reaching our ears will be composed of the direct sound from the talker's lips plus reflections of these sounds from the walls. Because they take an indirect route (via the walls), the reflected sounds reach our ears later and from different directions than the direct sound. Nevertheless, we localize the sound source based on the direction of the direct sound, not the reflections. In addition, although the reflections will "color" the quality of what we hear, we perceive only the earlier-arriving direct sound. (This applies to reflections arriving within a certain time after the direct sound. Longer-lasting reflections are perceived as reverberation, and discrete reflections arriving a certain time after the direct sound are heard as echoes. We shall discuss this further below.)

These situations illustrate the precedence effect, also known as the Haas phenomenon and the principle of the first wavefront. As a broad generality, the precedence effect states that within a certain time frame, the earlier-arriving signal (wavefront) will dominate over a later-arriving signal (e.g., echo) in determining what we hear. This effect was first described in detail by Wallach et al. [56] and Haas [57,58] in 1949, although it had been known for some time earlier [59].

The classic experiment by Wallach et al [56] illustrates the precedence effect. Figure 13.10 shows two clicks (A and B) delivered to the left ear and two others (C and D) presented to the right ear. The entire set of clicks was presented within 2 msec, so that they were perceived as a single fused image. Notice that there are two opposing ITDs at the onset and offset of the composite signal. The onset disparity t_1 favors the left ear, whereas the offset time difference t_2 favors the right. The question is which of these opposing ITDs is more important in lateralizing the fused image. The onset disparity t_1 was set to a certain interval so that

the click would be lateralized left, and then the offset disparity t_2 was adjusted to give a centered image. If t_1 and t_2 were equally important directional cues, then a centered image would occur when the two delays were equal ($t_1 = t_2$). However, Wallach et al. found that a much larger offset delay t_2 was needed to overcome a small onset delay t_1. For example, t_2 had to be about 400 μsec to counteract an opposing t_1 of about 50 μsec. This result demonstrates the importance of the first arriving signal in determining the location of the sound source. The precedence effect was maintained even if the level of the second click pair was as much as 15 dB greater than the first.

The findings just described were obtained as long as the interval t_3 between the click pairs was from 2 to 40 msec, depending on the stimuli. Longer durations of t_3 caused the subject to hear two separate signals, one at each ear.

Haas [57,58] presented speech stimuli from two loudspeakers. If the signal from one speaker was delayed up to 35 msec, then the stimulus was perceived to come only from the leading loudspeaker. Longer delays caused the listener to detect the presence of the second (delayed) sound, although the signal was still localized toward the leading side. Delays longer than about 50 msec caused the listener to hear a distinct echo from the second loudspeaker. Thus, these findings correspond to the effect of the interval t_3 in the study of Wallach et al.

The precedence effect serves us not only in sound source localization, but also by suppressing the effects of reflections arriving soon after the direct sound. This occurs because the reflections are integrated with the direct sound into a single fused image, at least for delays up to about 30 msec. These reflections would otherwise interfere with our perception of the direct sound, and particularly with speech intelligibility. This was demonstrated by Lochner and Burger [60], who found that speech discrimination was unaffected by reflections arriving up to 30 msec after the direct sound. However, later-arriving reflections were not integrated with the direct sounds, and resulted in reduced intelligibility. Similarly, Nabelek and Robinette [61] found that speech recognition was reduced by single echoes arriving 20 msec or more after the direct stimulus. Reflections that arrive beyond the time when the precedence effect is operative result in distortions of the speech signal because they tend to mask the direct signal [61–64]. Gelfand and Silman [64] found that the speech sound confusions associated with small-room reverberation were similar to those associated with masking or filtering of the speech signal. This result confirms that the reflections distort the speech signal via a masking mechanism.

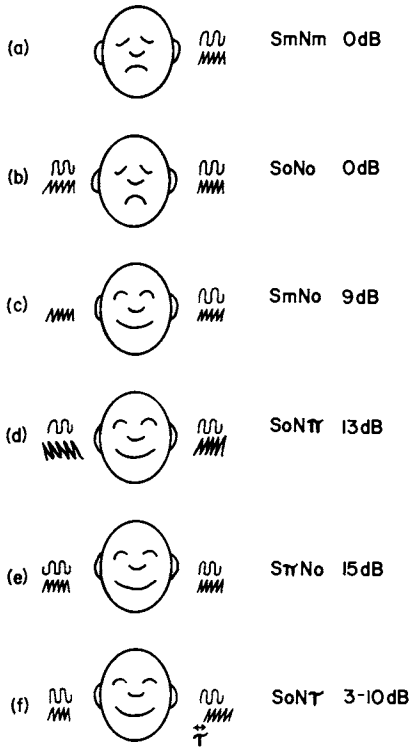

Figure 13.11 Masking level differences (MLDs) for various conditions.

MASKING LEVEL DIFFERENCES

The term "masking level difference" (MLD) may be a bit confusing at first glance. Obviously it refers to some sort of difference in masking. Let us see how. Consider a typical masking experiment (Chap. 10) in which a signal S is barely masked by a noise N. This can be done in one ear (monotically), as in Fig. 13.11a, or by presenting an *identical* signal and noise to both ears (diotically), as in Fig. 13.11b. Identical stimuli are obtained by simply directing the output of the same signal and noise sources to both earphones (in phase). For brevity and clarity, we will adopt a shorthand to show the relationships among the stimuli and ears. The letter "m" will denote a monotic stimulus and "o" will refer to a diotic stimulus. Thus, SmNm indicates that the signal and noise are presented to one ear, and SoNo means that the same signal and the same noise are simultaneously presented to both ears. Either of these conditions can be used as our starting point.

Suppose that we add an identical noise to the unstimulated ear of Fig. 13.11a, so that the signal is still monotic but the noise is now diotic

(SmNo), as in Fig. 13.11c. Oddly enough, the previously masked signal now becomes audible again! Starting this time from the masked situation in Fig. 13.11b (SoNo), we can make the signal audible again by reversing the phase of (inverting) the noise between the ears (Fib. 13.11d) or by reversing the phase of the signal between the ears (Fig. 13.11e). The phase reversal is indicated by the Greek letter π, since the stimuli are now 180° (or one radian, π) out of phase between the ears. (The phase reversal is accomplished by simply reversing the + and − poles at one of the earphones.) These new conditions are thus called SoNπ and SπNo, respectively. Note that the binaural advantage occurs only when the stimuli are in some way *different* at the two ears (dichotic). These fascinating observations were first reported in 1948 by Hirsh [65] for tonal signals and by Licklider [66] for speech.

We may now define the MLD as the difference (advantage) in masked threshold between dichotically presented stimuli and signals that are presented monotically (or diotically). It is not surprising to find that the MLD is also referred to as binaural unmasking, as binaural release from masking, or as the binaural masking level difference (BMLD). We shall express the magnitude of the MLD as the difference in dB between a particular dichotic arrangement and either the SmNm or SoNo conditions. Other MLD conditions are discussed below.

The size of the MLD varies from as large as about 15 dB for the SπNo condition [67] to as little as 0 dB, depending upon a variety of parameters. Typical MLD magnitudes associated with various dichotic arrangements [68] are shown in Fig. 13.11. It is a universal finding that the MLD becomes larger as the spectrum level of the masking noise is increased, especially when the noise is presented to both ears (No) at the same level [65,69–72].

The largest MLDs are obtained when either the signal (SπNo) or the noise (SoNπ) is opposite in phase at the two ears. The large MLDs obtained from these antiphasic conditions have been known since it was first described [65] in 1948, and have been repeatedly confirmed [73,74]. Recall from Chap. 5 that the firing patterns of auditory nerve fibers are phase-locked to the stimulus, particularly at low frequencies. Thus, the large MLDs associated wtih antiphasic conditions may be related to this phase-locking in the neural coding of the stimuli [67]. Furthermore, since the degree of phase-locking is greatest at low frequencies, and decreases as frequency becomes higher, we would expect the size of the MLD to be related to stimulus frequency as well.

Figure 13.12 shows the relationship between MLD size and stimulus frequency for a variety of studies, as summarized by Durlach [75]. As expected, the MLD is largest for low frequencies — about 15 dB for 250 Hz — and decreases for higher frequencies until a constant of about 3 dB is maintained about 1500–2000 Hz. (Look at the individual data

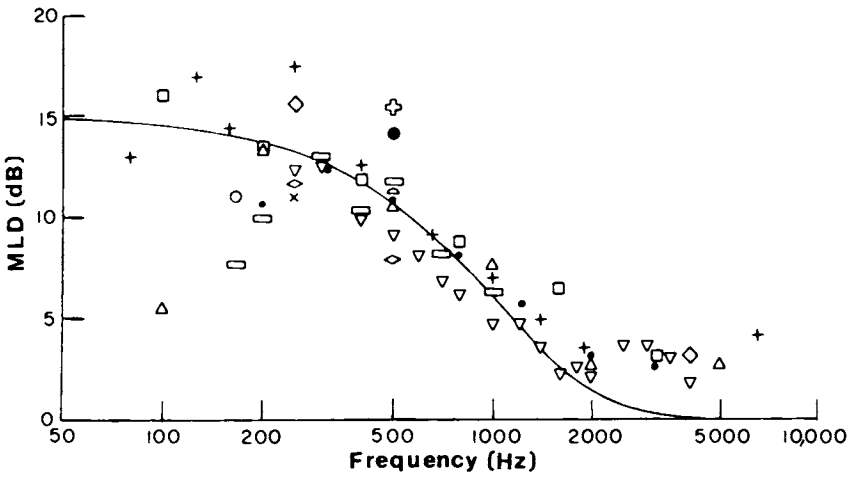

Figure 13.12 Magnitude of the MLD ($S\pi No - SoNo$) as a function of frequency for many studies. (Adapted from Durlach [75]. Binaural signal detection: Equalization and cancellation theory. In J. V. Tobias (ed.), *Foundations of Modern Auditory Theory, II,* ©1972, Academic Press, New York.)

points in Fig. 13.12 rather than at the smooth line, which is a statistical approximation of the actual results.) It is a bit intriguing to observe that the MLD (at least for $S\pi No$) does not fall to zero above 1500 Hz, since we have seen that phase information is essentially useless for high frequencies. However, these findings have been confirmed repeatedly [65,76–80]. One is drawn to recall in this context that statistically significant phase-locking is maintained as high as 5000 Hz (Chap. 5).

There is very good agreement about the size of the MLD for the frequencies above 250 Hz. At lower frequencies there is a great deal of variation in the MLD sizes reported by different studies [65,71,76,78,81]. This is shown in Fig. 13.12 by the substantial spread among the data points for the lower frequencies. Much of this variation in the lower frequencies may be explained on the basis of differences in noise level. In particular, Dolan [71] showed that the MLDs at 150 and 300 Hz increase with the spectrum level of the masker, attaining a value of approximately 15 dB when the noise spectrum level is 50 dB or more. Thus, the MLD becomes rather stable once moderate levels of presentation are reached.

We have been assuming that the noises at the two ears (SoNo) are derived from the same noise source, insuring that the waveforms are exactly the same at both ears. Another way to indicate identical wave-

forms is to say that the noises are perfectly correlated. Had we used two separate noise generators, then the noises would no longer be perfectly correlated. We would then say that the noises are uncorrelated (Nu). Robinson and Jeffress [82] added noises from the same (correlated) and different (uncorrelated) generators to study how noise correlation affects the size of the MLD. They found that the MLD resulting from uncorrelated noises is on the order of 3–4 dB, and that the MLD becomes larger as the degree of correlation decreases. The MLD resulting from uncorrelated noise may contribute to our ability to overcome the effects of reverberation. This relation was demonstrated by Koenig et al. [83], who found that room reverberation decorrelates the noise reaching the two ears, resulting in an MLD of about 3 dB.

Since only a certain critical bandwidth contributes to the masking of a tone (Chap. 10), it is not surprising that the MLD also depends upon the critical band [84,85] around a tone. In fact, the MLD actually increases as the noise band narrows [86,87]. As it turns out, a very narrow band of noise looks much like a sinusoid that is being slowly modulated in frequency and amplitude. If we present such a narrow-band noise to both ears and delay the wavefront at one ear relative to the other, then the degree to which the noises are correlated will change periodically as a function of the interaural time delay.

With this in mind, consider the arrangement in Fig. 13.11f. The noises presented to the two ears are from the same generator, but the noise is delayed at one ear relative to the other ($N\tau$). The interaural time delay decorrelates the noises in a manner dependent upon the time delay. This situation ($SoN\tau$) results in MLDs which are maximal when the time delay corresponds to half-periods of the signal, and minimal when the time delays correspond to the period of the signal [81,88]. Figure 13.13 shows examples at several frequencies. The effect is clearest at 500 Hz. The period of 500 Hz is 2 msec, and the half-period is thus 1 msec. As the figure shows, the MLDs are largest at multiples of the half-period (in the vicinity of 1 msec and 3 msec for 500 Hz) and are smallest at multiples of the full period (about 2 msec and 4 msec for 500 Hz). Also notice that successive peaks tend to become smaller and smaller.

Licklider [66] reported MLDs for speech about the same time that the phenomenon was described for tones [65]. Interestingly enough, the unmasking of speech is associated with the MLDs for pure tones within the spectral range critical for speech perception [90–92]. This was shown quite clearly in a study by Levitt and Rabiner [91] which used monosyllabic words as the signal S and white noise as the masker N. The subjects were asked to indicate whether the test words were detectible in

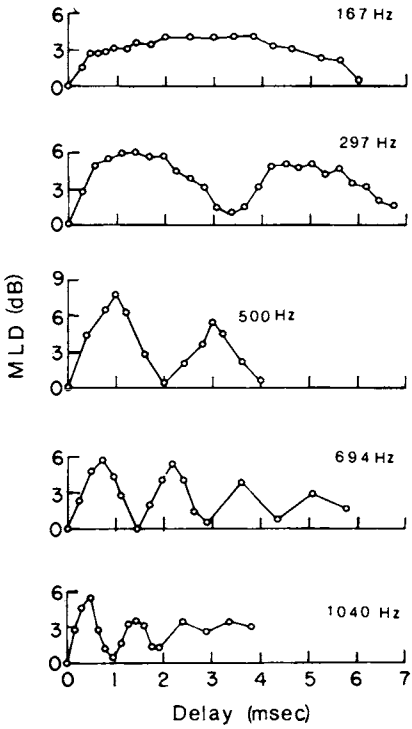

Figure 13.13 The MLD as a function of interaural time delay. (Adapted from Rabiner et al. [81]. Permission *J. Acoust. Soc. Amer.*)

the presence of the noise while different parts of the speech spectrum were reversed in phase between the ears. They found MLDs ($S\pi No$) on the order of 13 dB when the frequencies below 500 Hz were reversed in phase, indicating that the MLD for speech detection is primarily determined by the lower frequencies in the speech spectrum.

The MLD for speech detection obviously occurs at a minimal level of intelligibility. That is, a signal whose intelligibility is zero (see Chap. 14) may still be barely detectable. Increasing the presentation level of the words would result in higher intelligibility, i.e., a larger proportion of the words would be correctly repeated. The level at which half of the words are correctly repeated may be called the 50% intelligibility level; the 100% intelligibility level is the test level at which all of the words are repeated correctly. We may now note an interesting aspect of MLDs for speech: The speech MLD is quite large at near-detection levels. However, MLDs for speech are smallest at higher presentation levels where overall intelligibility is good [68,89–92].

Levitt and Rabiner [91,93] suggested the term "binaural intelligibility level difference" (BILD or ILD) to indicate the amount of unmasking for speech intelligibility. The ILD is the difference between the levels at which a particular percentage of the test words are correctly repeated for a dichotic condition and for SoNo. Levitt and Rabiner [91] found that the ILD was on the order of only 3–6 dB for an intelligibility level of 50%. Thus, the release from masking increases from about 3 dB to 13 dB as the intelligibility level goes down toward bare detection. At the lowest intelligibility level, where one can detect but not repeat the words, the ILD and MLD for speech are synonymous. In this light, we may consider the MLD as the limiting case of the ILD [93].

Although the speech MLD depends on frequencies below 500 Hz, Levitt and Rabiner [91] found that the entire speech spectrum makes a significant contribution to the ILD. In a later paper [93], they presented a numerical method for predicting the ILD. The procedure assumes that the SπNo condition reduces the noise level in a manner that depends upon frequency and S/N ratio. That is, the lower frequencies are given greater relative importance at low S/N ratios where overall intelligibility is poor. This technique makes predictions which are in close agreement with the empirical data [66,89,91].

Some work has also been done on MLDs for differential sensitivity and loudness [94,95]. We shall not go into detail in these areas, except to point out that MLDs for loudness and discrimination of intensity occur mainly at near-detection levels, and become insignificant well above threshold. This situation, of course, is analogous to what we have seen for ILDs as opposed to MLDs for speech signals.

MLD Models

Several models [75,76,78,96–99] try to explain the mechanism of MLDs. These models do not account for all known facets of MLDs, but they do provide good explanations of the phenomenon and actually allow one to make predictions about the results of experiments as yet undone. We will take a brief look at the two most established models, the Webster–Jeffress lateralization theory [76,96] and Durlach's equalization–cancellation model [75,78].

The Webster–Jeffress theory attributes MLDs to interaural phase and intensity differences. Recall that only a certain critical band contributes to the masking of a tone, and that this limited-bandwidth concept also applies to MLDs. Basically, the lateralization model compares the test tone to the narrow band of frequencies in the noise that contribute to its masking. Changing phase between the ears (SπNo) results in a time-of-arrival difference at a central mechanism, which provides

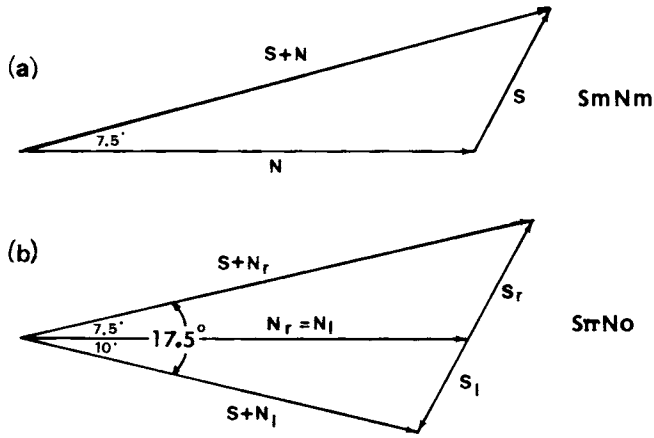

Figure 13.14 Vector diagrams for (a) SmNm and (b) SπNo (see text). (Adapted from Jeffress [96], Binaural signal detection: Vector theory, *Foundations of Modern Auditory Theory* in (J. V. Tobias, Ed.), Vol. 2, ©1972, Academic Press, New York.)

the detection cue. The vector diagrams in Fig. 13.14 show how this system might operate for SπNo versus SmNm. Figure 13.14a shows the interaction between the noise N and signal S amplitudes as vectors 7.5° apart. The resulting S + N amplitude is too small for detection. Fig. 13.14b shows the situation when the signal is reversed in phase at the two ears (SπNo). Now, the right and left noise amplitudes are equal ($N_r = N_l$), but the signal vectors for the two ears (S_r and S_l) point in opposite directions due to the reversed phase. Thus, the signal in the right ear leads the noise in phase by 7.5° while in the left ear it lags in phase by 10°; so that the phase difference between the resulting S + N at the two ears is 17.5°. In other words, S + N at the right ear is now 17.5° ahead of that in the left. If the signal is a 500 Hz tone, this lag corresponds to a time-of-arrival advantage for the right S + N of 97 μsec. In addition, the lengths of the S + N vectors indicates an amplitude advantage which is also available as a detection cue, since the right S + N is now substantially longer (more intense) than the left. That is, the SπNo condition causes the combined S + N from the right ear to reach a central detection mechanism sooner and with greater amplitude. This causes a lateralization of the result (to the right in this example) which the model uses as the basis for detection of the signal.

Durlach's equalization–cancellation (EC) model is shown schematically in Fig. 13.15. The stimuli pass through critical band filters at the

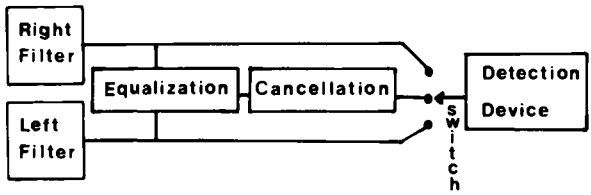

Figure 13.15 A simplified block diagram of Durlach's equalization-cancellation (EC) model.

two ears, and then follow both monaural and binaural routes to a detection device, which decides whether the signal is present. The detection device switches between the three possible channels (two monaural and one binaural), and will use the channel with the best S/N ratio as the basis for a response. The monaural channels go straight to the detection mechanism. The binaural channel, however, includes two special stages.

In the first stage, the inputs from the two ears are adjusted to be equal in amplitude (the equalization step). Then the inputs from the two ears are subtracted one from the other in the cancellation step.* Of course, if the signal and noise were identical in both ears (SoNo), then the entire binaural signal would be cancelled. In this case the detection device would choose between the monaural inputs so that no MLD resulted. However, for the SπNo condition, the subtraction cancels the in-phase noise and actually enhances the out-of-phase signal. Thus, if the EC model works perfectly, the S/N ratio will be improved infinitely. In reality, the mechanism operates less than perfectly, so that cancellation is not complete. This imperfection is due to atypical stimuli which necessitate unusual types of equalization, or to random jitter in the process, which causes the cancellation mechanism to receive inputs that are imperfectly equalized.

*Recall the rules for algebraic subtraction: $(+1) - (+1) = 0$, as for the in-phase noise; $(+1) - (-1) = +2$, as for the out-of-phase signal.

REFERENCES

1. L. J. Sivian and S. D. White, On minimal audible fields, *J. Acoust. Soc. Amer. 4*, 288–321 (1933).
2. J. Keys, Binaural versus monaural hearing, *J. Acoust. Soc. Amer. 19*, 629–631 (1947).
3. W. A. Shaw, E. B. Newman, and I. J. Hirsh, The difference between monaural and binaural thresholds, *J. Exp. Psycho. 37*, 229–242 (1947).

4. R. Caussé and P. Chavasse, Récherches sur les seuil de l'audition binauriculaire comparé au seuil monauriculaire en fonction de la fréquence, *Comp. R. Soc. Biol. 135*, 1272–1275 (1941).
5. R. Caussé and P. Chavasse, Différence entre le seuil de l'audition binauriculaire et le seuil monauriculaire de la fréquence, *Comp. R. Soc. Biol. 136*, 301 (1942).
6. I. J. Hirsh, Binaural summation: A century of investigation, *Psych. Bull. 45*, 193–206 (1948).
7. I. Pollack, Monaural and binaural threshold sensitivity for tones and white noise, *J. Acoust. Soc. Amer. 20*, 52–58 (1948).
8. H. Fletcher and W. Munson, Loudness: Its definition, measurement and calculation, *J. Acoust Soc. Amer. 5*, 82–108 (1933).
9. R. Caussé and P. Chavasse, Différence entre l'ecoute binauriculaire et monauriculaire por la perception des intensités supraliminaires. *Comp. R. Soc. Biol. 139*, 405 (1942).
10. G. S. Reynolds and S. S. Stevens, Binaural summation of loudness, *J. Acoust. Soc. Amer. 32*, 1337–1344 (1960).
11. R. P. Hellman and J. Zwislocki, Monaural loudness function at 1000 cps, and interaural summation, *J. Acoust. Soc. Amer. 35*, 856–865 (1963).
12. B. Scharf and D. Fishken, Binaural summation of loudness: Reconsidered, *J. Exp. Psych. 86*, 374–379 (1970).
13. L. E. Marks, Binaural summation of loudness of pure tones, *J. Acoust. Soc. Amer. 64*, 107–113 (1978).
14. B. G. Churcher, A. J. King, and H. Davies, The minimal perceptible change of intensity of a pure tone, *Phil. Mag. 18*, 927–939 (1934).
15. J. D. Harris, Loudness discrimination, *J. Speech Hearing Dis.*, Monograph Suppl 11 (1963).
16. R. C. Rowland and J. V. Tobias, Interaural intensity difference limens, *J. Speech Hearing Res. 10*, 745–756 (1967).
17. W. Jesteadt, C. C. Wier, and D. M. Green, Comparison of monaural and binaural discrimination of intensity and frequency, *J. Acoust. Soc. Amer. 61*, 1599–1603 (1977).
18. E. G. Shower and R. Biddulph, Differential pitch sensitivity of the ear, *J. Acoust. Soc. Amer. 3*, 275–287 (1931).
19. A. G. Pickler and J. D. Harris, Channels of reception in pitch discrimination, *J. Acoust. Soc. Amer. 27*, 124–131 (1955).
20. W. Jesteadt, C. C. Wier, and D. M. Green, Intensity discrimination as a function of frequency and sensation level, *J. Acoust. Soc. Amer. 61*, 169–177 (1977).
21. C. C. Wier, W. Jesteadt, and D. M. Green, Frequency discrimination as a function of frequency and sensation level, *J. Acoust. Soc. Amer. 61*, 178–184 (1977).
22. C. Cherry, Two ears — but one world, in *Sensory Communication* (W. A. Rosenblith, ed.), MIT Press, Cambridge, Mass., 1961, pp. 99–117.
23. E. C. Cherry, Some experiments upon the recognition of speech with one and two ears, *J. Acoust. Soc. Amer. 25*, 975–979 (1953).

24. E. C. Cherry and B. McA. Sayers, "Human cross-'correlator' " — A technique for measuring certain parameters of speech perception, *J. Acoust. Soc. Amer. 28*, 889–895 (1956).
25. B. McA. Sayers and E. C. Cherry, Mechanism of binaural fusion in the hearing of speech, *J. Acoust. Soc. Amer. 29*, 973–987 (1957).
26. D. M. Leakey, B. McA. Sayers, and E. C. Cherry, Binaural fuson of low- and high frequency sounds, *J. Acoust. Soc. Amer. 30*, 222–223 (1958).
27. D. E. Broadbent, A note on binaural fusion, *Quart. J. Exp. Psych. 7*, 46–47 (1955).
28. J. S. Wernick and A. Starr, Electrophysiological correlates of binaural beats in superior-olivary complex of cat, *J. Acoust. Soc. Amer. 40*, 1276 (1966).
29. J. C. R. Licklider, J. C. Webster, and J. M. Hedlun, On the frequency limits of binaural beats, *J. Acoust. Soc. Amer. 22*, 468–473 (1950).
30. J. V. Tobias, Application of a "relative" procedure to the binaural-beat problem, *J. Acoust. Soc. Amer. 35*, 1442–1447 (1963).
31. J. J. Groen, Super- and subliminate binaural beats, *Acta Otol. 57*, 224–230 (1964).
32. W. E. Feddersen, T. T. Sandel, D. C. Teas, and L. A. Jeffress, Localization of high-frequency tones, *J. Acoust. Soc. Amer. 29*, 988–991 (1957).
33. R. S. Woodworth, *Experimental Psychology*, Holt, Rinehart, and Winston, New York, 1938.
34. Lord Rayliegh, Our perception of sound duration, *Phil. Mag. 13*, 214–232 (1907).
35. S. S. Stevens and E. B. Newman, The localization of actual sources of sound, *Amer. J. Psycho. 48*, 297–306 (1936).
36. T. T. Sandel, D. C. Teas, W. E. Feddersen, and L. A. Jeffress, Localization of sound from single and paired sources, *J. Acoust. Soc. Amer. 27*, 842–852 (1955).
37. A. W. Mills, On the minimal audible angle, *J. Acoust. Soc. Amer. 30*, 237–246 (1958).
38. A. W. Mills, Auditory perception of spatial relations, *Proc. Int. Cong. Tech. Blind.*, Vol. 2, Am. Found. Blind, New York, 1963.
39. A. W. Mills, Auditory localization, In *Foundations of Modern Auditory Theory* (J. V. Tobias, Ed.), Vol. 2, Academic, New York, 1972, pp. 301–348.
40. R. G. Klumpp and H. R. Eady, Some measurements of interaural time difference thresholds, *J. Acoust. Soc. Amer. 28*, 859–860 (1956).
41. J. Zwislocki and R. S. Feldman, Just noticeable differences in dichotic phase, *J. Acoust. Soc. Amer. 28*, 860–864 (1956).
42. A. W. Mills, Lateralization of high-frequency tones, *J. Acoust. Soc. Amer. 32*, 132–134 (1960).
43. W. A. Yost, Discriminations of interaural phase differences, *J. Acoust. Soc. Amer. 55*, 1299–1303 (1974).
44. J. V. Tobias and S. Zerlin, Lateralization threshold as a function of stimulus duration, *J. Acoust. Soc. Amer. 31*, 1591–1594 (1959).
45. J. V. Tobias, Curious binaural phenomena, in *Foundations of Modern Auditory Theory* (J. V. Tobias, Ed.), Vol. 2, Academic, New York, 1972, pp. 463–486.

46. J. V. Tobias and E. D. Schubert, Effective onset duration of auditory stimuli, *J. Acoust. Soc. Amer. 31*, 1595–1605 (1959).
47. W. A. Yost, F. L. Whightman, and D. M. Green, Lateralization of filtered clicks, *J. Acoust. Soc. Amer., 50*, 1526–1531 (1971).
48. W. A. Yost, Lateralization of pulsed sinusoids based on interaural onset, ongoing, and offset temporal differences, *J. Acoust. Soc. Amer. 61*, 190–194 (1977).
49. G. B. Henning, Detectability of interaural delay in high-frequency complex waveforms, *J. Acoust. Soc. Amer. 55*, 84–90 (1974).
50. G. G. Harris, Binaural interactions of impulsive stimuli and pure tones, *J. Acoust. Soc. Amer. 32*, 685–692 (1960).
51. G. Moushegian and L. A. Jeffress, Role of interaural time and intensity differences in the lateralization of low-frequency tones, *J. Acoust. Soc. Amer. 31*, 1441–1445 (1959).
52. R. A. Whitworth and L. A. Jeffress, Time vs. intensity on the localization of tones. *J. Acoust. Soc. Amer. 33*, 925–929 (1961).
53. E. R. Hafter and L. A. Jeffress, Two-image lateralization of tones and clicks, *J. Acoust. Soc. Amer. 44*, 563–569 (1968).
54. L. A. Jeffress and D. McFadden, Differences of interaural phase and level in detection and lateralization, *J. Acoust. Soc. Amer. 49*, 1169–1179 (1971).
55. E. R. Hafter and S. C. Carrier, Binaural interaction in low-frequency stimuli: The inability to trade time and intensity completely, *J. Acoust. Soc. Amer. 51*, 1852–1862 (1972).
56. H. Wallach, E. B. Newman, and M. R. Rosenzweig, The precedence effect in sound localization, *Amer. J. Psych. 62*, 315–336 (1949).
57. H. Haas, The influence of a single echo on the audibility of speech, *Library Com. 363*, Dept. Sci. Indust. Rest., Garston, Watford, England, 1949.
58. H. Haas, Über den Einfluss eines Einfachechos aud die Hörsamkeit von Sprache, *Acoustica 1*, 49–58 (1951). [English translation: *J. Audiol. Eng. Soc. 20*, 146–159 (1972).]
59. M. B. Gardner, Historical background of the Haas and/or precedence effect, *J. Acoust. Soc. Amer. 43*, 1243–1248 (1968).
60. J. P. A. Lochner and J. F. Burger, The influence of reflections on auditorium acoustics, *J. Sound Vib. 1*, 426–454 (1964).
61. A. K. Nabelek and L. Robinette, Influence of the precedence effect on word identification by normal hearing and hearing-impaired subjects, *J. Acoust. Soc. Amer. 63*, 187–194 (1978).
62. R. H. Bolt and A. D. MacDonaldd, Theory of speech masking by reverberation, *J. Acoust. Soc. Amer. 21*, 577–580 (1949).
63. H. Kurtović, The influence of reflected sound upon speech intelligibility, *Acoustica*, 33, 32–39 (1975).
64. S. A. Gelfand and S. Silman, Effects of small room reverberation upon the recognition of some consonant features, *J. Acoust. Soc. Amer. 66*, 22–29 (1979).
65. I. J. Hirsh, The influence of interaural phase on interaural summation and inhibition, *J. Acoust. Soc. Amer. 20*, 536–544 (1948).

66. J. C. R. Licklider, The influence of interaural phase relations upon the masking of speech by white noise, *J. Acoust. Soc. Amer. 20*, 150–159 (1948).
67. D. M. Green and G. R. Henning, Audition, *Ann. Rev. Psych. 20*, 105–128 (1969).
68. D. M. Green and W. A. Yost, Binaural analysis, in *Handbook of Sensory Physiology* (W. D. Keidel and W. D. Neff, Eds.), Vol. V/2: *Auditory System*, Springer-Verlag, 1975, pp. 461–480.
69. H. C. Blodgett, L. A. Jeffress, and R. H. Whitworth, Effect of noise at one ear on the masking threshold for tones at the other, *J. Acoust. Soc. Amer. 34*, 979–981 (1962).
70. T. R. Dolan and D. E. Robinson, An explanation of masking-level differences that result from interaural intensive disparities of noise, *J. Acoust. Soc. Amer. 42*, 977–981 (1967).
71. T. R. Dolan, Effects of masker spectrum level on masking-level differences at low signal frequencies, *J. Acoust. Soc. Amer. 44*, 1507–1512 (1968).
72. D. McFadden, Masking-level differences determined with and without interaural disparities in masking intensity, *J. Acoust. Soc. Amer. 44*, 212–223 (1968).
73. L. A. Jeffress, H. C. Blodgett, and B. H. Deatheredge, The masking of tones by white noise as a function of interaural phases of both components, *J. Acoust. Soc. Amer. 24*, 523–527 (1952).
74. H. S. Colburn and N. I. Durlach, Time-intensity relations in binaural unmasking, *J. Acoust. Soc. Amer. 38*, 93–103 (1965).
75. N. I. Durlach, Binaural signal detection: Equalization and cancellation theory, in *Foundations of Modern Auditory Theory* (J. V. Tobias, Ed.), Vol. 2, Academic, New York, 1972, pp. 369–462.
76. F. A. Webster, The influence of interaural phase on masked thresholds. I. The role of interaural time-duration, *J. Acoust. Soc. Amer. 23*, 452–462 (1951).
77. I. J. Hirsh and M. Burgeat, Binaural effects in remote masking, *J. Acoust Soc. Amer. 30*, 827–832 (1958).
78. N. I. Durlach, Equalization and cancellation theory of binaural masking-level differences, *J. Acoust. Soc. Amer. 35*, 1206–1218 (1963).
79. K. D. Schenkel, Über die Abhängigkeit der Mithörschwellen von der interauralen Phasenlage des Testchalls, *Acoustica 14*, 337–346 (1964).
80. D. M. Green, Interaural phase effects in the masking of signals of different durations, *J. Acoust. Soc. Amer. 39*, 720–724 (1966).
81. L. R. Rabiner, C. L. Lawrence, and N. I. Durlach, Further results on binaural unmasking and the EC model, *J. Acoust. Soc. Amer. 40*, 62–70 (1966).
82. D. E. Robinson and L. A. Jeffress, Effect of varying the interaural noise correlation on the detectability of tonal signals, *J. Acoust. Soc. Amer. 35*, 1947–1952 (1963).
83. A. H. Koenig, J. B. Allen, D. A. Berkley, and T. H. Curtis, Determination of masking-level differences in a reverberant environment, *J. Acoust. Soc. Amer. 61*, 1374–1376 (1977).

84. M. M. Sondhi and N. Guttman, Width of the spectrum effective in the binaural release of masking, *J. Acoust. Soc. Amer. 40*, 600–606 (1966).
85. B. E. Mulligan, M. J. Mulligan and J. F. Stonecypher, Critical band in binaural detection, *J. Acoust. Soc. Amer. 41*, 7–12 (1967).
86. P. J. Metz, G. von Bismark, and N. I. Durlach, I. Further results on binaural unmasking and the EC model. II. Noise bandwidth and interaural phase, *J. Acoust. Soc. Amer. 43*, 1085–1091 (1967).
87. F. L. Whightman, Detection of binaural tones as a function of masker bandwidth, *J. Acoust. Soc. Amer. 50*, 623–636 (1971).
88. T. L. Langford and L. A. Jeffress, Effect of noise cross-correlation on binaural signal detection, *J. Acoust. Soc. Amer. 36*, 1455–1458 (1964).
89. E. D. Schubert and M. C. Schultz, Some aspects of binaural signal selection, *J. Acoust. Soc. Amer. 34*, 844–849 (1962).
90. R. Carhart, T. Tillman, and K. Johnson, Release of masking for speech through interaural time delay, *J. Acoust. Soc. Amer. 42*, 124–138 (1967).
91. H. Levitt and L. R. Rabiner, Binaural release from masking for speech and gain in intelligibility, *J. Acoust. Soc. Amer. 42*, 601–608 (1967).
92. R. Carhart, T. Tillman, and P. Dallos, Unmasking for pure tones and spondees: Interaural phase and time disparities, *J. Speech Hearing Res. 11*, 722–734 (1968).
93. H. Levitt and L. R. Rabiner, Predicting binaural gain in intelligibility and release from masking for speech, *J. Acoust. Soc. Amer. 42*, 820–829 (1967).
94. G. B. Henning, Effect of interaural phase on frequency and amplitude discrimination, *J. Acoust. Soc. Amer. 54*, 1160–1178 (1973).
95. T. H. Townsend and D. P. Goldstein, Suprathreshold binaural unmasking, *J. Acoust. Soc. Amer. 51*, 621–624 (1972).
96. L. A. Jeffress, Binaural signal detection: Vectory theory, in *Foundations of Modern Auditory Theory* (J. V. Tobias, Ed.), Vol. 2, Academic, New York, 1972, pp. 349–368.
97. E. R. Hafter, W. T. Bourbon, A. S. Blocker, and A. Tucker, Direct comparison between lateralization and detection under antiphasic masking, *J. Acoust. Soc. Amer. 46*, 1452–1457 (1969).
98. E. R. Hafter and S. C. Carrier, Masking-level differences obtained with a pulsed tonal masker, *J. Acoust. Soc. Amer. 47*, 1041–1048 (1970).
99. E. Osman, A correlation model of binaural masking level differences, *J. Acoust. Soc. Amer. 50*, 1494–1511 (1971).

14
SPEECH PERCEPTION

Pure tones, clicks, and the like enable us to study specific aspects of audition in a precise and controllable manner. On the other hand, we communicate with each other by speech, which is composed of particularly complex and variable waveforms. A knowledge of how we perceive simpler sounds is the foundation upon which an understanding of speech perception must be built. As one might suppose, speech perception and intimately related areas constitute a voluminous subject encompassing far more than hearing science, per se. For this reason, more than a cursory coverage of the topic would be neither prudent nor profitable within the current context. The interested reader is referred to several references [1–4].

Speech perception and speech production are inherently interrelated. We must be able to speak what we can perceive, and we must have the ability to perceive the sounds that our speech mechanisms produce. Traditionally, the sounds of speech have been described in terms of the vocal and articulatory manipulations that produce them. We too shall begin with production. For the most part, our discussion will focus upon phonemes.

By a "phoneme" we mean a group of sounds that are classified as being the same by native speakers of a given language. Let us see what the "sameness" refers to. Consider the phoneme /p/* as it appears at the

*By convention, phonemes are enclosed between slashes and phonetic elements between brackets.

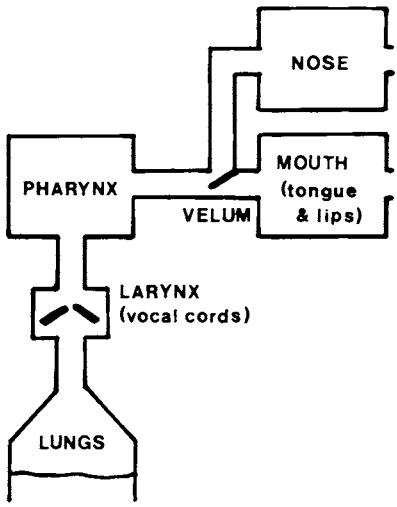

Figure 14.1 Schematic representation of the vocal tract.

beginning and end of the word "pipe." There are actually several differences between the two productions of /p/ in this word. For example, the initial /p/ is accompanied by a release of a puff of air (aspiration) whereas the final /p/ is not. In other words, the actual sounds are different (they are distinct phonetic elements). In spite of this, native speakers of English will classify both as belonging to the family designated as the /p/ phoneme. Such phonetically dissimilar members of the same phonemic class are called allophones of that phoneme. Consider a second example. The words "beet" and "bit" (/bit/ and /bɪt/, respectively) sound different to speakers of English but the same to speakers of French. This happens because the phonetic elements [i] and [ɪ] are different phonemes in English, but are allophones of the same phoneme in French. Since the Frenchman classifies [i] and [ɪ] as members of the same phonemic family, he hears them as being the same, just as English speakers hear the aspirated and unaspirated productions of /p/ to be the same.

This last example also demonstrates the second important characteristic of phonemes. Changing a phoneme changes the meaning of a word. Thus, /i/ and /ɪ/ are different phonemes in English, because replacing one for the other changes the meaning of at least some words. However, [i] and [ɪ] are not different phonemes in French (i.e., they are allophones), because replacing one for the other does not change the meaning of words. Implicit in the distinction of phonetic and phonemic elements is that even elementary speech sound classes are to some extent learned. All babies the world over produce the same wide range of

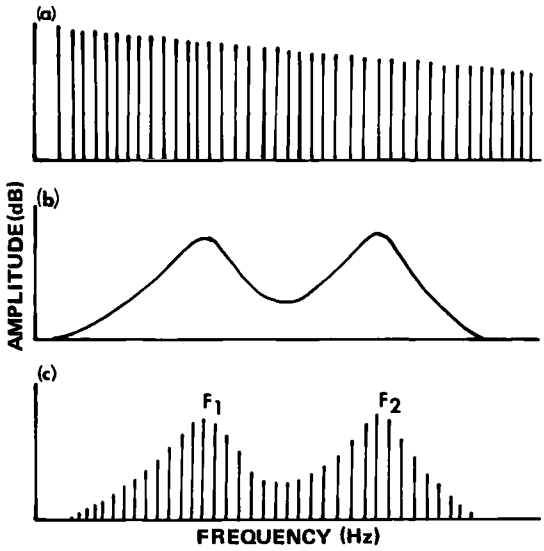

Figure 14.2 Idealized spectra showing that when the complex vocal waveform (a) is passed through the vocal tract filters (b) the resulting waveform (c) represents the acoustic characteristics of the vocal tract. Vocal tract resonances result in the formants (F_1, F_2).

sounds phonetically; it is through a process of learning that these phonetic elements become classified and grouped into families of phonemes that are used in the language of the community.

SPEECH SOUNDS: PRODUCTION AND PERCEPTION

Our discussion of speech sounds will be facilitated by reference to the simplified schematic diagram of the vocal tract in Fig. 14.1. The "power source" is the air in the lungs, which is directed up and out under the control of the respiratory musculature. Voiced sounds are produced when the vocal cords are vibrated (see Flanagan [5] and Stevens and House [6] for detailed discussions.) The result of this vibration is a periodic complex waveform made up of a fundamental frequency on the order of 100 Hz in males and 200 Hz in females, with as many as 40 harmonics of the fundamental represented in the waveform [5] (see Fig. 14.2a). This complex periodic waveform is then modified by the resonance characteristics of the vocal tract. In other words, the vocal tract constitutes a group of filters which are added together, and whose effect is to shape the spectrum of the waveform from the larynx. The resonance characteristics of the vocal tract (Fig. 14.2b) are thus reflected in the speech spectrum (Fig. 14.2c).

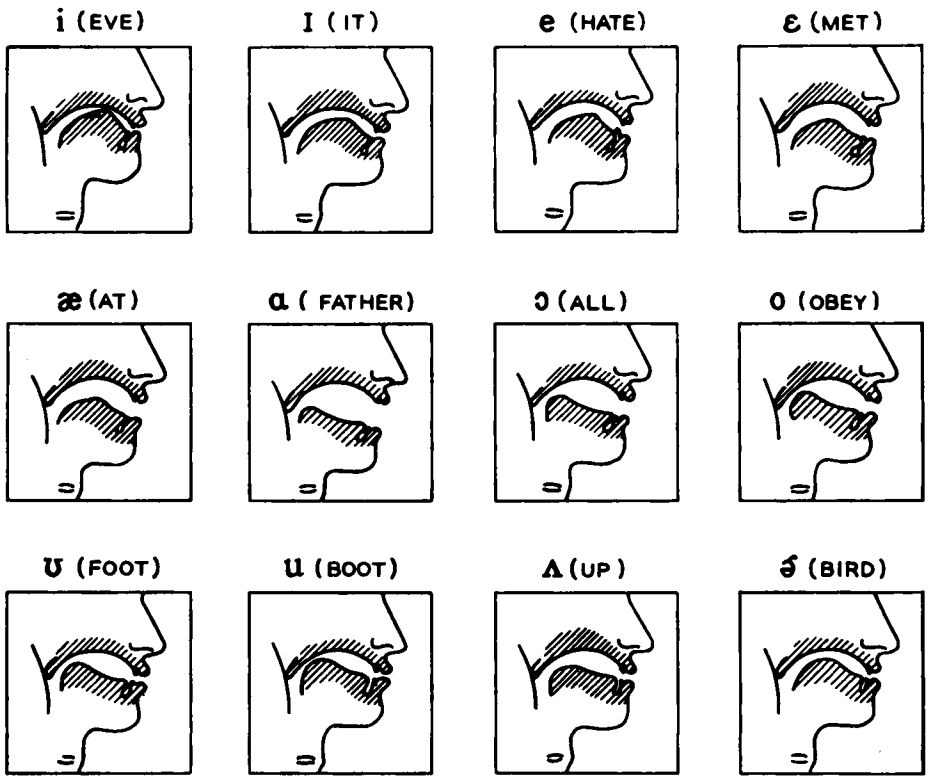

Figure 14.3 Articulation of various English vowels (see text). [From Flanagan [12], adapted from Potter et al. [7], *Visible Speech.* Copyright (1947) Bell Telephone Laboratories. Reprinted by permission.].

Voiceless (unvoiced) sounds are produced by opening the airway between the vocal folds so that they do not vibrate. Voiceless sounds are aperiodic and noiselike, being produced by turbulences due to partial or complete obstruction of the vocal tract.

Vowels

Speech sounds are generally classified broadly as vowels and consonants. Vowels are voiced sounds* whose spectral characteristics are determined by the size and shape of the vocal tract. Changing the shape of the vocal

*Certain exceptions are notable. Whispered speech is all voiceless; and vowels may also be voiceless in some contexts of voiceless consonants in connected discourse. Also, the nasal cavity is generally excluded by action of the velum, unless the vowel is in the environment of a nasal sound.

tract changes its filtering characteristics, which in turn changes the frequencies at which the speech signal is enhanced or de-emphasized (Fig. 14.2). Examples of how the vocal tract is shaped for various English vowels are shown in Fig. 14.3. As the figure shows, the vowel we hear is largely determined by the effects of tongue and lip position and movement. (Diphthongs such as /aɪ/ in "buy" and /oʊ/ in "toe" are heard when one vowel glides into another.) The resulting frequency regions where energy is concentrated are called formants, and are generally labeled starting from the lowest as F_1, F_2 etc. The lower formants (especially F_1 and F_2, as well as F_3) are primarily responsible for vowel recognition [8–10].

In general, the formant frequencies depend upon where and to what extent the vocal tract is constricted [9,11,12]. The locations and degrees of these constrictions control the sizes and locations of the volumes in the vocal tract (Fig. 14.3). For example, elevation of the back of the tongue results in a larger volume between this point of constriction and the lips than does elevation of the tongue tip. We may thus describe a vowel from front to back in terms of the location of tongue elevation, and from high to low in terms of the amount of tongue elevation. Lip rounding is another important factor. In English, front vowels (/i, ɪ, e, ɛ, æ/) are produced with retraction of the lips, while the lips are rounded when the back vowels (/u, ʊ, o, ɔ, a/) are formed. Rounding the front vowel /i/ as in "tea" while keeping the high-front tongue placement results in the French /y/ as in "*tu*."

The degree of tenseness associated with the muscle contractions is also a factor in vowel production and perception, as in the differentiation of the tense (i) ("peat") from the lax /ɪ/ ("pit"). Tense vowels are generally more intense and longer in duration than their lax counterparts.

The middle vowels (/ʌ, ə, ɜ, ɝ, ɚ/) are produced when tongue elevation is in the vicinity of the hard palate. These include the neutral vowel /ə/ associated mainly with unstressed syllables (eg., "*a*bout" and "s*u*pport").

Without going into great detail, the frequency of the first formant F_1 is largely dependent upon the size of the volume behind the tongue elevation, i.e., upon the larger of the vocal tract volumes. This volume must of course increase as the elevated part of the tongue moves forward. Thus, front tongue elevation produces a larger volume behind the point of constriction, which in turn is associated with lower F_1 frequencies; whereas back tongue elevations decrease the size of this volume, thereby raising the frequency of F_1. The frequency of F_2 depends largely upon the size of the volume in front of the point of tongue elevation, becoming higher when the cavity is made smaller (when the tongue is elevated

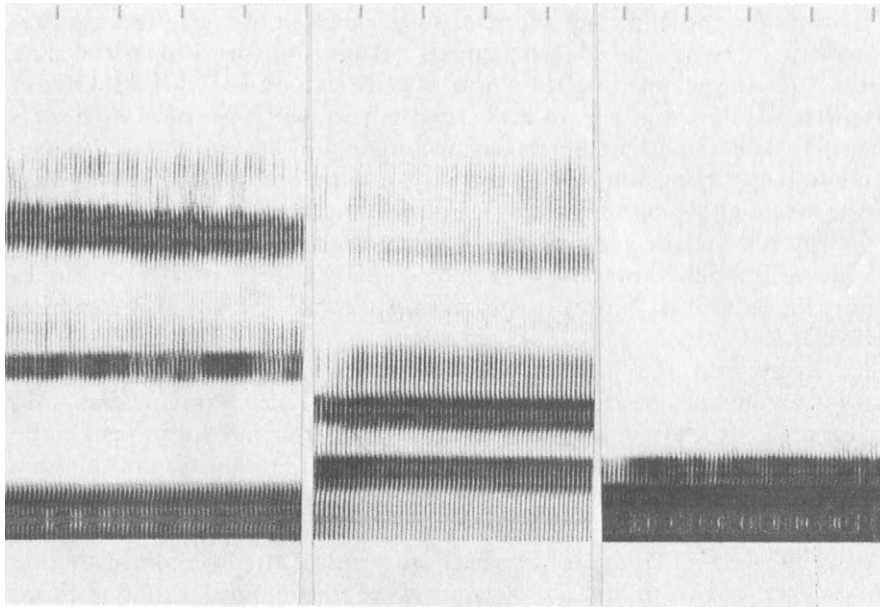

Figure 14.4 Spectrograms showing sustained production of the vowels /i/,/æ/ and /u/ (left to right). Timing marks along the top are 100 msec apart.

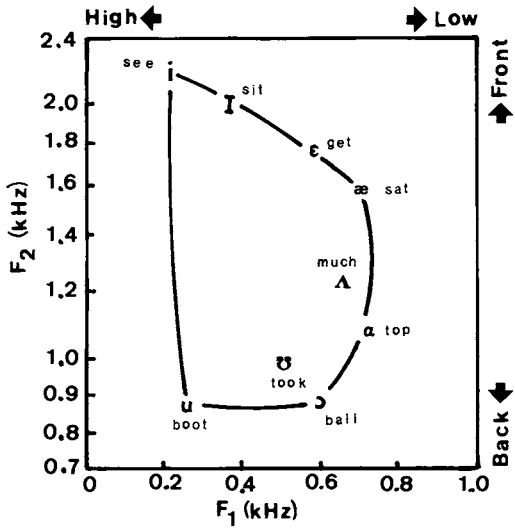

Figure 14.5 Approximate values of F_1 and F_2 for several vowels as they relate to tongue height and position. (Modified from Peterson and Barney [9].)

closer to the front of the mouth). Lip rounding lowers the first two formants by reducing the size of the mouth opening.

Figure 14.4 shows sustained productions of several vowels in the form of sound spectrograms [7,13]. Frequency is shown up the ordinate, time along the abscissa, and intensity as relative blackness. Thus, blacker areas represent frequencies with higher energy concentrations and lighter areas indicate frequency regions with less energy. The formants are indicated by frequency bands much darker than the rest of the spectrogram. These horizontal bands represent frequency regions containing concentrations of energy and thus reflect the resonance characteristics of the vocal tract. The vertical striations correspond to the period of the speaker's fundamental frequency.

The relationship between tongue position and the frequencies of F_1 and F_2 are shown for several vowels in Fig. 14.5. However, it should be noted that these formant frequencies are approximations. Formant center frequencies and bandwidths tend to become higher going from men to women to children, which reflects the effect of decreasing vocal tract length. Furthermore, they vary among studies and between subjects.

Given these wide variations it is doubtful that vowels are identified on the basis of their formant frequencies, per se. The relationships among the formants, as well as the environment of the vowel and its duration, provide important cues [14–16]. It has been suggested that, for each individual speaker, the listener adjusts the "target" values of the formants according to the utterances of that speaker [17,18].

Consonants

The consonants are produced by either a partial or complete obstruction somewhere along the vocal tract. The ensuing turbulence causes the sound to be quasi-periodic or aperiodic and noiselike. The consonants are differentiated on the basis of manner of articulation, place of articulation, and voicing; i.e., on how and where the obstruction of the vocal tract occurs and on whether there is vocal cord vibration. Table 14.1 shows the English consonants, arranged horizontally according to place of articulation and vertically by manner and voicing. Examples are given where the phonetic and orthographic symbols differ.

The stops, fricatives, and affricates may be either voiced or voiceless, whereas the nasals and semivowels are virtually always voiced. The nasal cavities are excluded from the production of all consonants except the nasals by elevation of the velum. Cross-sectional diagrams for the pro-

Table 14.1 Consonants of English

	Bilabial	Labiodental	Linguadental	Alveolar	Palatal	Velar	Glottal
Stops							
vl[a]	p			t		k	
vd[a]	b			d		g	
Fricatives							
vl	ʍ (*wh*ich)	f	θ (*th*ing)	s	ʃ (*sh*oot)		h
vd		v	ð (*th*is)	z	ʒ (beige)		
Affricates							
vl					tʃ (ca*tch*)		
vd					dʒ (do*dge*)		
Nasals[b]	m			n	ŋ (si*ng*)		
Semivowels[b]	w			l,r	j (*y*es)		

[a]vl, voiceless; vd, voiced.
[b]nasals and semivowels are voiced.

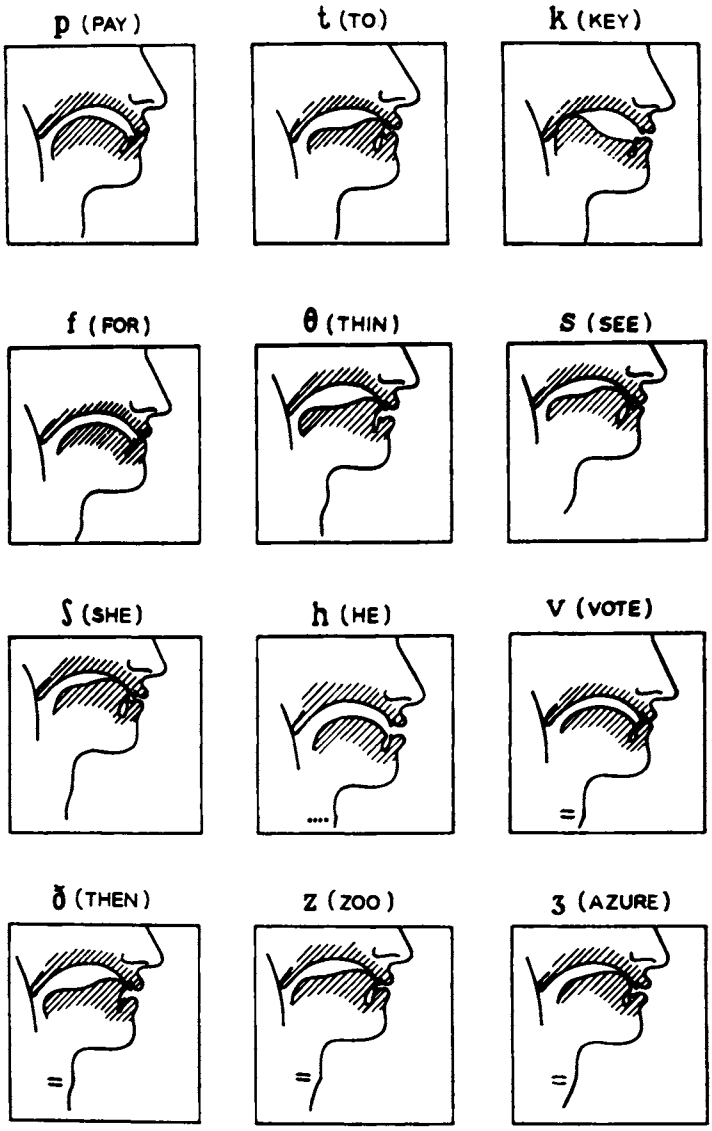

Figure 14.6 Articulation of some English stops and fricatives. (From Flanagan [12], adapted from Potter et al. [7], *Visible Speech.* Copyright (1947) Bell Telephone Laboratories. Reprinted by permission.)

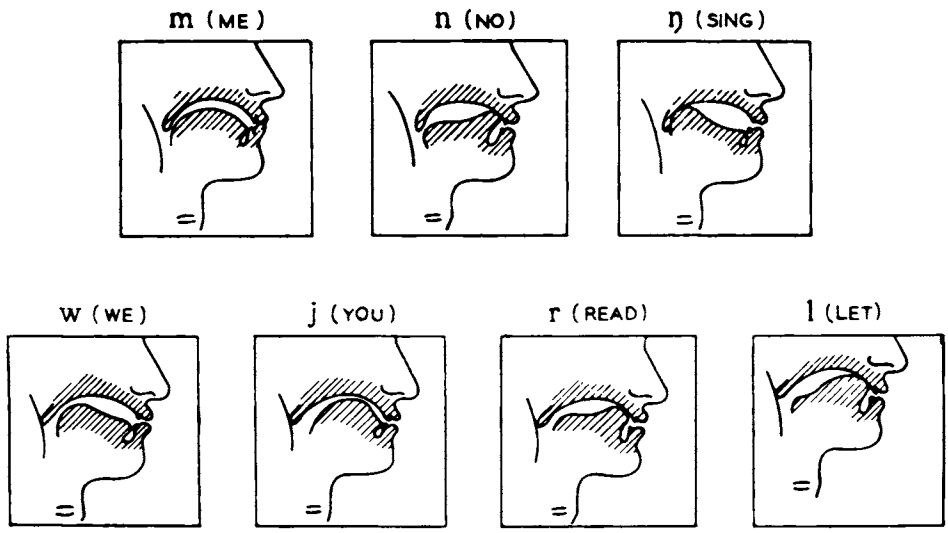

Figure 14.7 Articulation of some English nasals and semivowels (From Flanagan [12], adapted from Potter et al. [7], *Visible Speech.* Copyright (1947) Bell Telephone Laboratories. Reprinted by permission.)

duction of the stops, fricatives, and affricates are given in Fig. 14.6, and for the nasals and semivowels in Fig. 14.7. We shall briefly discuss the consonants in order of the manner of their articulation.

The stops (plosives) are produced by a transition (see below) from the preceding vowel, a silent period on the order of roughly 30 msec during which air pressure is impounded behind a complete obstruction somewhere along the vocal tract, a release (burst) of the built-up pressure, and finally a transition into the following vowel [19–22]. Of course, whether there is a transition from the preceding vowel and/or into the following vowel depends upon the environment of the stop consonant. Voiceless stops in the initial position are generally aspirated, or released with a puff of air. Initial voiced stops and all final stops tend not to be aspirated, although this does not apply always or to all speakers. The voiceless stops (/p,t,k/) and their voiced cognates (/b,d,g/) are articulated in the same way except for the presence or absence of voicing and/or aspiration (see Fig. 14.6). As a rule, the voiceless stops tend to have longer and stronger pressure buildups than do their voiced counterparts [23,24].

The six stops are produced at three locations. The bilabials (/p,b/) are produced by an obstruction at the lips; the alveolars (/t,d/) by the tongue tip against the upper gum ridge; and the velars (/k,g/) by the

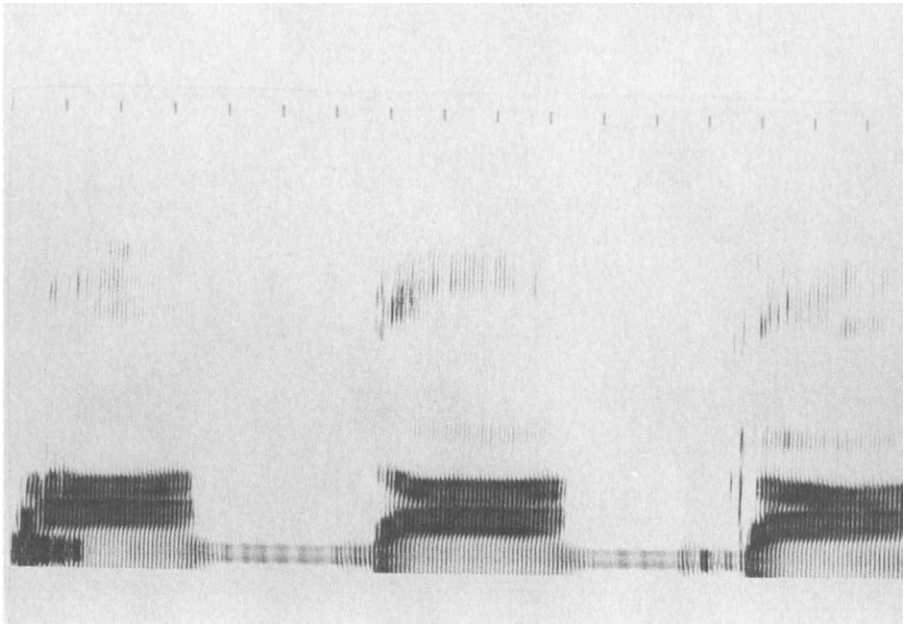

Figure 14.8 Spectrograms of /ba/, /da/ and /ga/ (left to right). Note second formant transitions.

tongue dorsum against the soft palate. Whether the sound is heard as voiced or voiceless is of course ultimately due to whether there is vocal cord vibration. However, cues differ according to the location of the stop in an utterance. The essential voicing cue for initial stops is voice onset time (VOT), which is simply the time delay between the onset of the stop burst and commencement of vocal cord vibration [25,26]. In general, voicing onset precedes or accompanies stop burst onset for voiced stops but lags behind the stop burst for voiceless stops. For final stops and those which occur medially within an utterance, the essential voicing cue appears to be the duration of the preceding vowel [27]. Longer vowel durations are associated with the perception that the following stop is voiced. Voiceless stops are also associated with longer closure durations [28], faster formant transitions [29], greater burst intensities [22], and somewhat higher fundamental frequencies [30], than voiced stops.

Place of articulation (bilabial versus alveolar versus velar) for the stops is largely cued by the second formant (F_2) transition of the associated vowel [31,32], along with some contribution from the F_3 transitions [33]. By a formant transition we simply mean a change with time of the

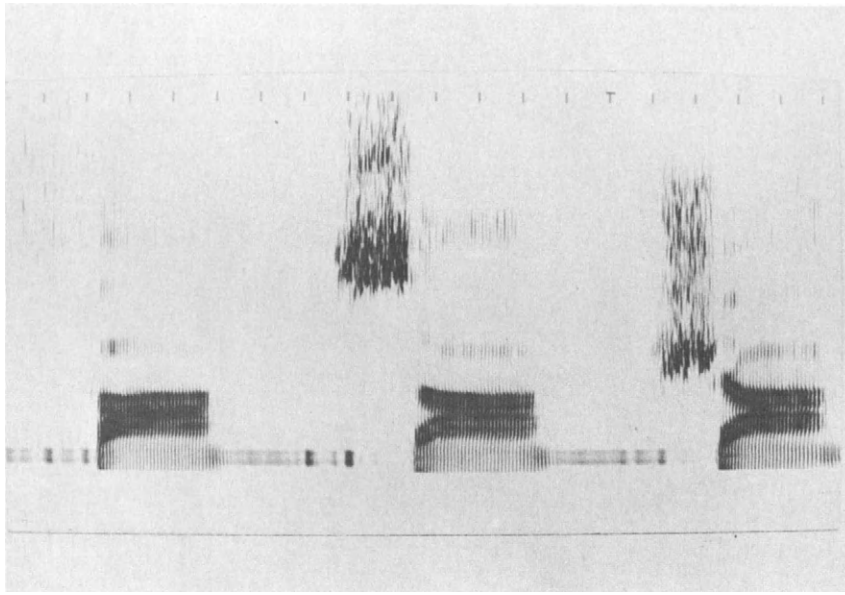

Figure 14.9 Spectrograms of /fa/, /sa/ and /ʃa/ (left to right).

formant frequency in going from the steady-state frequency of the vowel into the consonant (or vice versa). Formant transitions may be seen for several initial voiced stops in Fig. 14.8. The F_2 transitions point in the direction of approximately 700 Hz for bilabial stops, 1800 Hz for the alveolars, and 3000 Hz for the velars [31]. These directions relate to the location of vocal tract obstruction. That is, a larger volume is enclosed behind an obstruction at the lips (/p,b/) than at the alveolus (/t,d/) or velum (/k,g/), so that the resonant frequency associated with that volume is lower for more frontal obstructions. Moving the point of obstruction backward reduces the cavity volume and thus increases the resonant frequency. Further place information is provided by the frequency spectrum of the stop burst. Stop bursts tend to have concentrations of energy at relatively low frequencies (500–1500 Hz) for the bilabials, at high frequencies (about 4000 Hz and higher) for the alveolars, and at intermediate frequencies (between around 1500 and 4000 Hz) for the velars [22].

The fricatives are produced by a partial obstruction of the vocal tract so that the air coming through becomes turbulent. The nature of the fricatives has been well described [34–40], and examples are shown by the spectograms in Fig. 14.9. Fricatives are distinguished from other manners of articulation by the continuing nature of their turbulent

energy (generally lasting 100 msec or more); and vowels preceding fricatives tend to have greater power and duration, and somewhat longer fundamental frequencies, than vowels preceding stops. As is true for the stops, fricatives may be either voiced or voiceless, and voicing is similarly cued by VOT and by the nature of the preceding vowel. In addition, voiceless fricatives tend to have longer durations than their voiced counterparts.

Spectral differences largely account for place distinctions between the alveolar (/s,z/) and palatal (/ʃ,ʒ/) fricatives. The palatals have energy concentrations extending to lower frequencies (about 1500–7000 Hz) than do the alveolars (roughly 4000–7000 Hz), most likely because of the larger volume in front of the point of vocal tract obstruction for /ʃ,ʒ/ than for /s,z/. On the other hand, /θ,ð/ and /f,v/ are differentiated mainly on the basis of formant transitions. Due to resonation of the entire vocal tract above the glottis, /h/ possesses more low frequencies than the more anterior fricatives. The affricates are produced by the rapid release of their stop components into their fricative components.

The nasals are produced by opening the port to the nasal cavities at the velum (Fig. 14.7). They are semivowels in that they are voiced and have some of the formant characteristics of vowels. However, they differ from other sounds by the coupling of the nasal and oral cavities. The characteristics of nasals have been described by Fugimura [41] and others [19,42,43]. The coupling of the nasal cavities to the volume of oral cavity behind the point of obstruction (the velum for /ŋ/, alveolus for /n/, and lips for /m/ constitutes a side-branch resonator. This results in antiresonances at frequencies which become lower as the volume of the side branch (oral cavity) becomes larger. Thus, we find that antiresonances appear in the frequency regions of roughly 1000 Hz for /m/ (where the side branch is largest), 1700 Hz for /n/, and 3000 Hz for /ŋ/ (where the side branch is shortest). Furthermore, overall intensity is reduced, and for the first formant is lower than for the vowels, constituting a characteristic low-frequency nasal murmur. Place of articulation is cued by differences in formant structure among the nasal consonants and by formant transitions for adjacent vowels.

The semivowels are /w,j/ and /r,l/. The latter two are also called liquids. The semivowels have been described by O'Connor et al. [44], Lisker [45], and Fant [46]. The bilabial semivowel /w/ is produced initially in the same way as the vowel /u/, with a transition into the following vowel over the course of about 100 msec. For /j/, the transition into the following vowel is from /i/, and takes about the same amount of time as for /w/. The first formants of both /w/ and /j/ are on the order of 240 Hz, with the second formant transitions beginning below about 600 Hz for /w/ and above 2300 Hz for /j/. Furthermore, there appears to be rela-

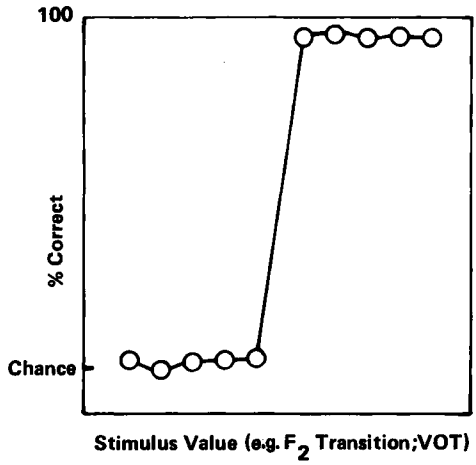

Figure 14.10 Idealized functions for a categorical perception experiment (see text).

tively low-frequency frication-like noise associated with /w/ and higher-frequency noise with /j/, due to the degree of vocal tract constriction in their production.

The liquid semivowels (/r,l/) have short-duration steady-state portions of up to about 50 msec, followed by transitions into the following vowel over the course of roughly 75 msec. The /r/ is produced with some degree of lip rounding, and /l/ is produced with the tongue tip at the upper gum ridge so that air is deflected laterally. The first formants of the liquids are on the order of 500 Hz for /r/ and 350 Hz for /l/, which is relatively high; and the first and second formant transitions are roughly similar for both liquids. The major difference appears to be associated with the presence of lip rounding for /r/ but not /l/. Since lip rounding causes the third formant to be lower in frequency, /r/ is associated with a rather dramatic third formant transition upward in frequency which is not seen for /l/, at least not for transitions into unrounded vowels. The opposite would occur for /r/ and /l/ before a rounded vowel.

Categorical Perception

Liberman et al. [47] prepared synthetic consonant–vowel (CV) monosyllables composed of two formants each. They asked their subjects to discriminate between pairs of these synthetic CVs as the second formant transition was varied, and obtained a finding which has had a profound effect upon the study of speech perception. Subjects' ability to discriminate between the two CVs in a pair was excellent when the consonants

were identifiable as different phonemes, whereas discrimination was poor when the consonants were identified as belonging to the same phonemic category. This observation of categorical perception has been confirmed by a variety of studies [25,48–60]. Typical (idealized) results are shown in Fig. 14.10.

The circles in Fig. 14.10 represent the percentage of correct discriminations between the CVs as a function of the difference between them along some continuum (such as the direction of F_2 transition or voice onset time). For small values along this continuum the subjects are essentially unable to discriminate between the two CVs beyond a chance rate. At a certain point, however, there is a dramatic jump in correct discrimination to near 100% performance. If we now ask the subject to identify each of the stimuli according to phonemic category, we find that the CVs identified as belonging to the same category are those which were poorly discriminated from one another, whereas those belonging to different categories are the ones that were easily discriminated.

This categorical perception of speech sounds was not found to occur when Liberman et al. [48] inverted the synthetic stimuli to form nonspeech materials. Instead, discrimination among the nonspeech sounds occurred along a continuum without the peaks and troughs seen in the categorical perception of speech stimuli. Using VOT as the parameter to be discriminated, Abramson and Lisker [25,49,50] demonstrated that categorical perception occurs across languages, generally with voicing time offsets on the order of approximately 20 msec. In addition, similar categorical discrimination for VOT was shown to occur in both infants and adults [51]. These findings suggested that discriminability depends upon some kind of phonetic (rather than acoustic) categorization of the stimuli, and have been proposed as evidence that there is an innate special mode for the perception of speech [52]. As we shall see, however, more recent findings rather powerfully support an acoustically based discriminability related to whether the stimuli are perceived as simultaneous versus successive, and upon their perceived order [53].

Several findings cast doubt upon a phonetic basis for categorical perception. First, categorical perception has been demonstrated to occur for a variety of nonspeech materials [53–55]. Second, three categorical boundaries along the VOT parameter have been found to be discriminated by Spanish and Kikuyu infants [56,57]. This result contradicts the phonetic basis because there is only one phonetic boundary for voicing in Spanish and none for the labial stops in Kikuyu. Most impressive, however, is the finding by Kuhl and Miller [58] that chinchillas can be trained to discriminate along the VOT parameter, and that the "phoneme" boundaries derived from the chinchilla data are excellent approximations of the human findings. The absence of a phonetic system is a reasonable assumption. Thus, one is drawn toward some sort

of psychoacoustic rather than phonetic discrimination as the basis for categorical perception.

Pisoni [53] found that onset-time disparities between two tones were categorically perceived at boundaries comparable to those for stop consonant VOT. Further, he demonstrated that judgments of whether the onsets of the two tones were simultaneous or successive lined up remarkably well with the categorical perception data for these stimuli. That is, two events were heard when the onsets of the two tones were 20 msec apart, whereas a single onset was heard when they were less than 20 msec apart. Since these discrepancies are the same as those needed to perceive the temporal order of two stimuli [59,60], Pisoni proposed that this psychoacoustic phenomenon may well be the basis of categorical perception.

POWER OF SPEECH SOUNDS

Several points about the power of speech sounds are noteworthy prior to a discussion of speech intelligibility. From the foregoing, we would expect to find most of the power of speech in the vowels; and since the vowels have a preponderance of low-frequency energy, we would expect the long-term speech spectrum to reflect this as well. This expectation is borne out by the literature [1]. The weakest sound in English is the voiceless fricative /θ/ and the strongest is the vowel /ɔ/ [1,61]. If /θ/ is assigned a power of one, then the relative power of /ɔ/ becomes 680 [1]. The relative power of the consonants range up to 80 for /ʃ/; are between 36 (/n/) and 73 (/ŋ/) for the nasals; are on the order of 100 for the semivowels; and range upward from 220 (/i/) for the vowels [1]. As one would expect, the more powerful sounds are detected and are more intelligible at lower intensities than are the weaker ones.

The long-term average spectrum of speech is shown in Fig. 14.11. This figure shows the general range of findings from several studies [62–64] in terms of relative level in dB. As expected, most of the energy in speech is found in the lower frequencies, particularly below about 1000 Hz. Intensity falls off as frequency increases above this range. It should be noted that the crosshatched region in the figure is the relative speech spectrum averaged over time for male and female speakers in several studies. The average overall sound pressure level of male speech tends to be on the order of 3 dB higher than for females. The spectrograms shown earlier in the chapter are examples of the short-term speech spectrum.

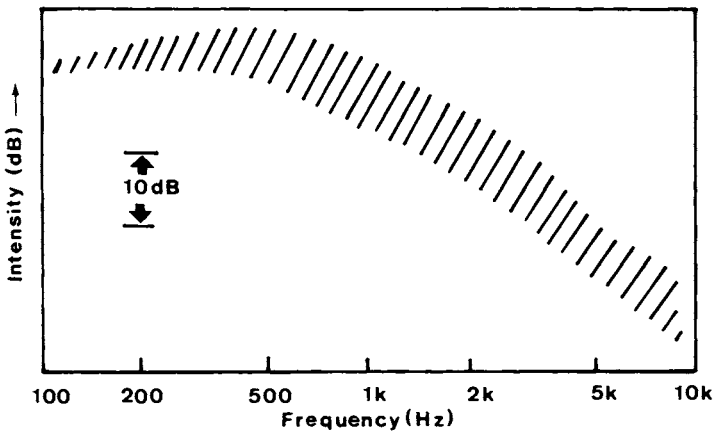

Figure 14.11 Long term average speech spectrum from several studies (see text).

SPEECH INTELLIGIBILITY AND DISTORTION

In general, speech intelligibility refers to how well the listener receives and comprehends the speech signal. The basic approach to studying speech intelligibility is quite simple and direct. The subject is presented with a series of stimuli (syllables, words, phrases, etc.), and is asked to identify what he has heard. The results are typically reported as the percent correct, which is called the articulation or discrimination score [65–67]. The approach may be further broken down into open-set methods requiring the subject to repeat (or write) what was heard without prior knowledge of the corpus of test items [67–69]; and close-set methods which provide a choice of response alternatives from which the subject must choose [70,71]. Both paradigms require the subject to perform a discrimination task; however, the results of open-set tests are often called discrimination scores while close-set results are generally referred to as recognition scores. These tests were originally devised in the development of telephone communication systems, and efforts were accelerated for obvious though unfortunate reasons during the second world war. The factors that contribute to speech intelligibility (or interfere with it) may be examined by obtaining articulation scores under various stimulus conditions and in the face of different kinds of distortions.

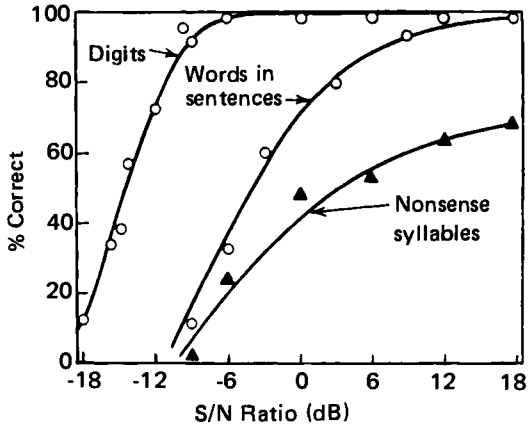

Figure 14.12 Psychometric functions showing the effects of test materials. [From Miller, Heise, and Lichten [73], The intelligibility of speech as a function of the context of the test materials, *J. Exp. Psych. 41*, 329–335. Copyright (1951) by the American Psychological Association. Reprinted by permission.]

Intensity and Materials

It is a general finding for all speech materials that intelligibility improves as the level of the signal is increased from barely audible levels up to higher levels, at which discrimination reaches a maximum [1,64,66]. In other words, there is a psychometric function (Chap. 7) for speech intelligibility as well as for the other psychoacoustic phenomena we have discussed. As a rule, discrimination performance becomes asymptotic when maximum intelligibility is reached for a given type of speech material; however, discrimination may actually decrease in some cases if intensity is raised to excessive levels.

Intelligibility depends upon the nature of the speech materials used, and a general rule is that performance improves as the speech stimuli become more redundant in any of a variety of ways [67,72–75]. For example, Miller et al. [73] asked their subjects to discriminate (1) words in sentences, (2) digits from 0 to 9, and (3) nonsense (meaningless) syllables. A psychometric function was generated for each type of test material, showing percent correct performance as a function of signal-to-noise (S/N) ratio (Fig. 14.12). Note that the digits were audible at softer levels (lower S/N ratios) than were the words in sentences, which were in turn more accurately perceived than the nonsense syllables. We observe this result in several ways. First, at each S/N ratio in the figure percent correct performance is best for the digits, less good for the words in sentences, and poorest for the syllables. Conversely, the subjects were able

to repeat 50% of the digits at a level 17 dB softer than that needed to reach 50% correct for the monosyllables. Second, a small increase in S/N ratio resulted in a substantial increase in digit discrimination but a much smaller improvement for the syllables, with improvement for word identification lying between these. Finally, notice that digit discrimination becomes asymptotic at 100% correct at low levels. On the other hand, words in sentences do not approach 100% intelligibility until the S/N ratio reaches 18 dB, and monosyllable discrimination fails to attain even 70% at the highest S/N ratio tested. Thus, we find that the most restricted (redundant) materials are more intelligible than the less redundant ones. In other words, we need only get a small part of the signal to tell one digit from another; whereas there is little other information for the subject to call upon if part of a nonsense syllable is unclear. (For example, if one hears "–en" and knows in advance that the test item is a digit, the test is inordinately easier than when the initial sound might be *any* of the 25 or so consonants of English.) Word redundancy falls between the exceptionally high redundancy of the digits and the rather minimal redundancy of the nonsense syllables. As expected, Miller and associates found that words were more intelligible in a sentence than than when presented alone, which also reflects redundancy afforded by the context of the test item.

In a related experiment, Miller et al. [73] obtained intelligibility measures for test vocabularies made up of 2, 4, 8, 16, 32 or 256 monosyllable words, as well as for an "unrestricted" vocabulary of approximately 1000 monosyllables. The results were similar to those just described. The fewer the alternatives (i.e., the more redundant or predictable the message), the better the discrimination performance. As the number of alternatives increases, greater intenstiy (a higher S/N ratio) was needed in order to obtain the same degree of intelligibility.

Frequency

How much information about the speech signal is contained in various frequency ranges? The answer to this question is not only important in describing the frequencies necessary to carry the speech signal (an important concept if communication channels are to be used with maximal efficiency), but also may enable us to predict intelligibility. Egan and Wiener [76] studied the effects upon syllable intelligibility of varying the bandwidth around 1500 Hz. They found that widening the bandwidth improved intelligibility, which reached 85% when a 3000 Hz bandwidth was available to the listener. Narrowing the passband resulted in progressively lower intelligibility; conversely, discrimination was improved by raising the level of the stimuli. That is, the narrower the band of fre-

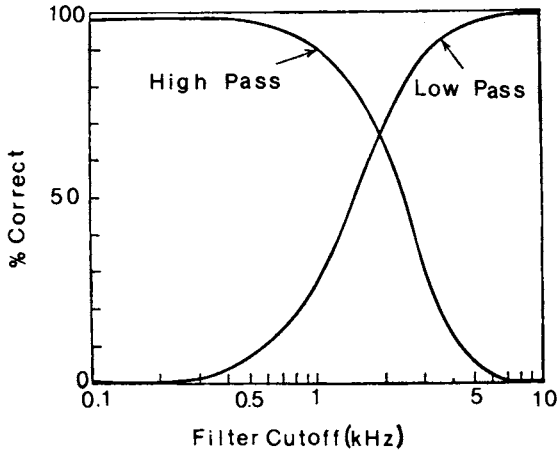

Figure 14.13 Syllable discrimination as a function of high- and low-pass filtering. (Adapted from French and Steinberg [64]. Permission *J. Acoust. Soc. Amer.*)

quencies, the higher the speech level must be in order to maintain the same degree of intelligibility.

French and Steinberg [64] determined the intelligibility of male and female speakers under varying conditions of low- and high-pass filtering. Discrimination was measured while filtering out the high frequencies above certain cutoff points (low-pass), and while filtering out the lows below various cutoffs (high-pass). Increasing amounts of either high- or low-pass filtering reduced intelligibility; and performance fell to nil when the available frequencies (the passband) were limited to those below about 200 Hz or above roughly 6000 Hz. As illustrated in Fig. 14.13, the high- and low-pass curves intersected at approximately 1900 Hz, where discrimination was about 68%. In other words, roughly equivalent contributions accounting for 68% intelligibility each were found for the frequencies above and below 1900 Hz. One must be careful, however, not to attach any magical significance to this frequency or percentage. For example, the crossover point dropped to about 1660 Hz for only male talkers. Furthermore, Miller and Nicely [77] showed that the crossover point depends upon what aspect of speech (feature) is examined. Their high- and low-pass curves intersected at 450 Hz for the identification of nasality, 500 Hz for voicing, 750 Hz for frication, and 1900 Hz for place of articulation.

The results of filtering studies have provided a basis for methods of predicting speech intelligibility — an articulation index. The articulation index was introduced by French and Steinberg [64], and has been

improved upon and expanded by others, particularly by Beranek [78] and by Kryter [79,80]. The actual calculations are beyond the current scope, and the interested reader is referred to the papers by Beranek [78] and Kryter [79] for details, and to Janssen [81] for an example of a simplified approach. The basic concept involves the use of 20 contiguous frequency bands, each of which contributes the same proportion (5%) to the overall intelligibility of the message. These bands are combined into a single number from 0 to 1.0 — the articulation index — which has been shown to be a reasonably good predictor of actual discrimination performance for a variety of speech materials [79,80].

Masking and Reverberation

The presentation of a noise has the effect of masking all or part of a speech signal. The general relationship between the effective level of the masker and the amount of masking for tones (Chap. 10) also holds true for the masking of speech by a broad-band noise [82]. That is, once the noise reaches an effective level, a given increment in noise level will result in an equivalent increase in speech threshold. Furthermore, this linear relationship between masker level and speech masking holds true for both the detection of the speech signal and intelligibility.

Recall from Chap. 10 that masking spreads upward in frequency, so that we would expect an intense low-frequency masker to be more effective in masking the speech signal than one whose energy is concentrated in the higher frequencies. This has been confirmed by Stevens et al. [83] and by Miller [84]. Miller also found that when noise bands were presented at lower intensities, the higher-frequency noise bands also reduced speech discrimination. This effect reflects the masking of consonant information concentrated in the higher frequencies.

Miller and Nicely [77] demonstrated that the effect of a wide-band noise upon speech intelligibility similar to that of low-pass filtering. This is expected, since a large proportion of the energy in the noise is concentrated in the higher frequencies. Both the noise and low-pass filtering resulted in rather systematic confusions among consonants, primarily affecting the correct identification of place of articulation. On the other hand, voicing and nasality, which rely heavily upon the lower frequencies, were minimally affected.

Reverberation is the persistence of acoustic energy in an enclosed space after the sound source has stopped; it is due to multiple reflections from the walls, ceiling, and floor of the enclosure (normally a room). It is a common experience that intelligibility decreases in a reverberant room, and this has been demonstrated experimentally as well [85–91].

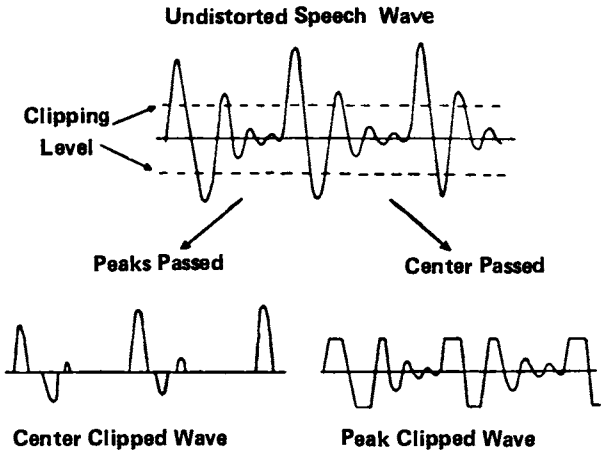

Figure 14.14 Effects of peak-clipping (right) and center clipping upon the speech waveform. "Clipping level" indicates the amplitude above (or below) which clipping occurs. (From *Language and Communication* by Miller [2]. Copyright © 1951 McGraw-Hill. Used with permission of McGraw-Hill Book Company.)

The amount of discrimination impairment becomes greater as the reverberation time* increases, particularly in small rooms where the reflections are "tightly packed" in time. Excellent reviews of the phenomenon may be found in Nabelek [89] and Nabelek and Robinette [90].

Reverberation appears to act as a masking noise in reducing speech intelligibility [85,91]. The reflected energy operates by overlapping the direct (original) speech signal so that perceptual cues are masked. Gelfand and Silman [91] demonstrated that reverberation results in rather systematic error patterns reminiscent of those associated with masking by noise and low-pass filtering, thus supporting the concept that reverberation acts as a masker. Their results also suggested that error patterns are affected by the distribution of phonemes in the language [92] as well as by the alternatives imposed by the test being used.

Amplitude Distortion

If the dynamic range of a system is exceeded, then there will be peak-clipping of the waveform. In other words, the peaks of the wave will be "cut off" as shown in Fig. 14.14. The resulting waveform approaches the

*Reverberation time is the time it takes for the reflections to decrease 60 dB after the sound source has been turned off.

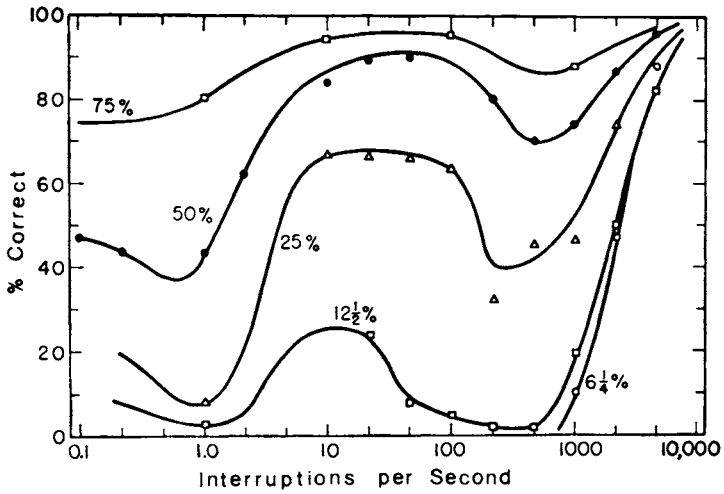

Figure 14.15 Discrimination as a function of interruption rate with speech-time fraction as the parameter. (Adapted from Miller and Licklider [96]. Permission *J. Acoust. Soc. Amer.*)

appearance of a square wave, as the figure clearly demonstrates. The effects of clipping have been studied in detail by Licklider and colleagues [93–95], and the essential though surprising finding is that peak-clipping does not result in any appreciable decrease in speech intelligibility even though the waveform is quite distorted. On the other hand, if the peaks are maintained but the center portion of the wave is removed (center clipping), then speech intelligibility quickly drops to nil.

Interruptions and Temporal Distortion

Miller and Licklider [96] studied the effect of rapid interruptions upon word intelligibility. They electronically interrupted the speech waveform at rates from 0.1–10,000 times per second, and with speech-time fractions between 6.25% and 75%. The speech-time fraction is simply the proportion of the time that the speech signal is actually on. Thus, a 50% speech-time fraction means that the speech signal was on and off for equal amounts of time, while 12.5% indicates that the speech signal was actually presented 12.5% of the time.

For the lowest interruption rate, the signal was alternately on and off for several seconds at a time. Thus, the discrimination score was roughly equal to the percent of the time that the signal was actually presented. The results for faster interruption rates are also shown in Fig. 14.15. When the speech-time fraction is 50%, performance was poorest when

the signal was interrupted about one time per second. This is an expected finding because we would expect intelligibility to be minimal when the interruption is long enough to overlap roughly the whole test word. At much faster interruption rates, the subjects get many "glimpses" of the test word. That is, assuming a word duration of 0.6 sec and five interruptions per second, there would be about three glimpses of each word, and so forth for higher interruption rates. (The dip in the function between about 200 and 2000 interruptions per second may be due to an interaction between the speech signal and the square wave that was used to modulate the speech signal to produce the interruptions.) Looking now at the remaining curves in Fig. 14.15, we find that the more the speech signal is actually on, the better is the discrimination; whereas when the speech-time fraction falls well below 50% intelligibility drops substantially at all interruption rates. Essentially similar findings have been reported in more recent work [97]. We see from such observations the remarkable facility with which the ear can "piece together" the speech signal, as well as the tremendous redundancy contained within the speech waveform.

Other forms of temporal distortion also substantially decrease speech intelligibility — particularly speeding (or time-compression) of the speech signal [98–100]. Space and scope preclude any detailed discussion except to note that intelligibility decreases progressively with speeding (or time compression) of the speech signal. An excellent review may be found in Beasley and Maki [101].

DICHOTIC LISTENING

Dichotic listening involves the presentation of one signal to the right ear and another to the left ear. For an exceptional detailed review, see Berlin and McNeil [102]. This approach was introduced to the study of speech perception by Broadbent [103,104] as a vehicle for the study of memory, and was extended to the study of hemispheric lateralization for speech by Kimura [105,106]. Kimura asked her subjects to respond to different digits presented to the two ears. The result was a small but significant advantage in the perception of the digits at the right ear — the right ear advantage (REA). On the other hand, there was a left ear advantage when musical material was presented dichotically [107]. Since the primary and most efficient pathways are from the right ear to the left cerebral hemisphere, and from the left ear to the right hemisphere, these findings have been interpreted as indicating a right-eared (left hemispheric) dominance for speech and a left-eared (right hemispheric) dominance for melodic material. This appears to be the case for the overwhelming majority (about 95%) of right-handed individuals, and for roughly 70% of left-handed people as well.

A major breakthrough in the study of dichotic listening was the use of CV syllables as stimuli by Shankweiler and Studdert-Kennedy [108], and others [109–112]. The basic experiment is similar to Kimura's, except that the dichotic digits are replaced by a much more controllable pair of dichotic CV syllables. Most often, the syllables /pa, ka, ta, ba, da, ga/ are used.

The CV studies confirmed and expanded the earlier digit observations. Specifically, Shankweiler and Studdert-Kennedy found a significant REA for the CVs but not for vowels. They further found that the REA was larger when the consonants differed in both place of articulation and voicing (e.g., /pa/ versus /ga/) than when the contrast was out of place (e.g., /pa/ versus /ta/) or voicing (e.g., /ta/ versus /da/) alone. Similarly, Studdert-Kennedy and Shankweiler [109] found that errors were less common when the dichotic pair had one feature (place or voicing) in common than when both place and voicing were different. These findings support the concept of a final linguistic (phonetic) processor within the left hemisphere since (1) they suggest that the neural representations from the two ears are recombined at one location, and (2) the features appear to be independently processed. On the other hand, auditory (nonphonetic) features do not appear to be independently processed [113,114].

The contention that there is a specialized phonetic processor in the left hemisphere, as well as a preliminary auditory processor in both hemispheres, has been supported by electrophysiological studies by Wood and colleagues [115,116]. They presented their subjects with dichotic CVs differing in one of two ways. A phonetic difference between the CVs was provided in the standard way (/ba/ versus /ga/). An auditory (nonphonetic) difference was produced by presenting the same CV syllable to the two ears, with a difference in fundamental frequency. In other words, the tasks were phonetic (/ba/–/fa/) and auditory (high pitch–low pitch). In addition to the dichotic listening task, they also obtained auditory evoked potentials (see Chap. 7) from both hemispheres of their subjects. The responses to the auditory and phonetic materials were significantly different over the left hemisphere, but not over the right; thus supporting the presence of a phonetic process occurring in the left hemisphere that differs from the auditory processing of the materials.

The acoustic parameters of the REA were studied by Cullen et al. [112]. They demonstrated that the REA was maintained until (1) the stimuli were presented to the left ear at a level 20 dB higher than the right, (2) the S/N ratio in the right ear was 12 dB poorer than in the left, or (3) the CVs presented to the right ear were filtered above 3000 Hz while the left ear received an unfiltered signal. Thus, the REA is maintained until the signal to the right ear is at quite a disadvantage relative

to the left. Interestingly, Cullen et al. demonstrated that when the right ear score decreased, the left ear score actually became proportionally better, so that the total percent correct (right plus left) was essentially constant. This suggests that there is a finite amount of information that can be handled at one time by the phonetic processor in the left hemisphere. The REA reflects the longer and noisier neural route (see Chap. 1) from the left ear to the left hemisphere processor for speech.

An interesting phenomenon occurs when the dichotic stimuli are presented with a time delay (lag) between the ears. If the CV delivered to one ear is delayed relative to the presentation of the CV to the other ear, then there is an advantage for the ear receiving the lagging stimulus, particularly for delays on the order of 30–60 msec [110,111]. This phenomenon is the "lag effect," and has been cited as evidence favoring the existence of a specialized phonetic processor. However, a lag effect has also been shown to occur for certain nonspeech (though speechlike) sounds, and there is evidence that the lag effect may rely upon the more general phenomenon of backward masking [117–120]. There is thus evidence implicating the perception of auditory as well as phonetic features in dichotic listening; an issue which remains to be fully resolved. In any case, the lag effect does appear to represent the interference with the perception of a stimulus by a following stimulus, and as such appears to be related to the time needed to complete the processing of the CV syllable [111].

REFERENCES

1. H. Fletcher, *Speech and Hearing in Communication, Van Nostrand, New York, 1953.*
2. *G. A. Miller, Language and Communication,* McGraw-Hill, New York, 1951.
3. I. Lehiste, *Readings in Acoustic Phonetics,* MIT Press, Cambridge, Mass., 1967.
4. M. E. Hawley, *Speech Intelligibility and Speech Recognition,* Dowden, Hutchinson, and Ross, Stroudsberg, Penn., 1977.
5. J. L. Flanagan, Some properties of the glottal sound source, *J. Speech Hearing Res. 1,* 99–116 (1958).
6. K. N. Stevens and A. S. House, An acoustical theory of vowel productions and some of its implications, *J. Speech Hearing Res. 4,* 302–320 (1961).
7. R. K. Potter, G. A. Kopp, and H. C. Green, *Visible Speech,* Van Nostrand, New York, 1947.
8. G. E. Peterson, The information-bearing elements of speech, *J. Acoust. Soc. Amer. 24,* 629–637 (1952).
9. G. E. Peterson and H. L. Barney, Control methods used in a study of the vowels, *J. Acoust. Soc. Amer. 24,* 175–184 (1952).

10. P. Delattre, A. M. Liberman, F. S. Cooper, and L. J. Gerstman, An experimental study of the acoustic determinants of vowel color; Observations on one- and two-formant vowels synthesized from spectrographic patterns, *Word 8*, 195–210 (1952).
11. K. N. Stevens and A. S. House, Development of a quantitative description of vowel articulation, *J. Acoust. Soc. Amer. 27*, 484–493 (1955).
12. J. L. Flanagan, *Speech Analysis, Synthesis and Perception*, Springer-Verlag, New York, 1972.
13. W. Koenig, H. K. Dunn, and L. Y. Lacy, The sound spectrograph, *J. Acoust. Soc. Amer. 17*, 19–49 (1946).
14. W. R. Tiffany, Vowel recognition as a function of duration, frequency modulation and phonetic context, *J. Speech Hearing Dis. 18*, 289–301 (1953).
15. K. N. Stevens and A. S. House, Pertubation of vowel articulations by consonantal context: An acoustical study, *J. Speech Hearing Res. 6*, 111–128 (1963).
16. B. Lindblom and M. Studdert-Kennedy, On the role of formant transitions in vowel recognition, *J. Acoust. Soc. Amer. 42*, 830–843 (1967).
17. P. Ladefoged and D. E. Broadbent, Information conveyed by vowels, *J. Acoust. Soc. Amer. 29*, 98–104 (1957).
18. P. Lieberman, On the evaluation of language: A unified view, *Cognition 2*, 59–94 (1973).
19. F. S. Cooper, P. C. Delattre, A. M. Liberman, J. M. Borst, and L. J. Gerstman, Some experiments on the perception of synthetic speech sounds, *J. Acoust. Soc. Amer. 24*, 597–617 (1952).
20. E. Fischer-Jorgensen, Acoustic analysis of stop consonants, *Misc. Phonetica 2*, 42–59 (1954).
21. A. M. Liberman, P. C. Delattre, L. J. Gerstman, and F. S. Cooper, Tempo of frequency as a cue for distinguishing classes of speech sounds, *J. Exp. Psych. 52*, 127–137 (1956).
22. M. Halle, G. W. Hughes, and J. P. A. Radley, Acoustic properties of stop consonants, *J. Acoust. Soc. Amer. 29*, 107–116 (1957).
23. D. J. Sharf, Duration of post-stress intervocalic stops preceding vowels, *Lang. Speech 5*, 26–30 (1962).
24. J. H. Arkebauer, T. J. Hixon, and J. C. Hardy, Peak intraoral air pressure during speech, *J. Speech Hearing Res. 10*, 196–208 (1967).
25. L. Lisker and A. S. Abramson, Cross-language study of voicing in initial stops: Acoustical measurements, *Word 20*, 384–422 (1964).
26. L. Lisker and A. S. Abramson, Some effects of context on voice onset time in English stops, *Lang. Speech 10*, 1–28 (1967).
27. L. J. Raphael, Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English, *J. Acoust. Soc. Amer. 51*, 1296–1303 (1972).
28. L. Lisker, Closure duration and the intervocalic voiced-voiceless distinction in English, *Language 33*, 42–49 (1957).
29. I. H. Slis, Articulatory measurements on voiced, voiceless and nasal consonants, *Phonetica 24*, 193–210 (1970).

30. M. P. Haggard, S. Ambler, M. Callow, Pitch as a voicing cue, *J. Acoust. Soc. Amer. 47*, 613–617 (1970).
31. A. M. Liberman, P. C. Delattre, F. S. Cooper, and L. J. Gerstman, The role of consonant-vowel transitions in the perception of the stop and nasal consonants, *Psych. Monogr. 68*, 1–13 (1954).
32. P. C. Delattre, A. M. Liberman, and F. S. Cooper, Acoustic loci and transitional cues for consonants, *J. Acoust. Soc. Amer. 27*, 769–773 (1955).
33. K. S. Harris, H. S. Hoffman, A. M. Liberman, P. C. Delattre, and F. S. Cooper, Effect of third formant transitions on the perception of the voiced stop consonants, *J. Acoust. Soc. Amer. 30*, 122–126 (1958).
34. A. S. House and G. Fairbanks, The influence of consonant environment upon the secondary characteristics of vowels, *J. Acoust. Soc. Amer. 25*, 105–113 (1953).
35. G. W. Hughes and M. Halle, Spectral properties of fricative consonants, *J. Acoust. Soc. Amer. 28*, 303–310 (1956).
36. L. J. Gerstman, Perceptual dimensions for the friction portions of certain speech sounds, Unpublished Ph.D. Dissertation, NY: New York University, New York, 1957.
37. K. S. Harris, Cues for the discrimination of American English Fricatives in spoken syllables, *Lang. Speech 1*, 1–7 (1958).
38. P. Strevens, Spectra of fricative noise in human speech, *Lang. Speech 3*, 32–49 (1960).
39. J. M. Heinz and K. N. Stevens, On the properties of voiceless fricative consonants, *J. Acoust. Soc. Amer. 33*, 589–596 (1961).
40. W. Jassem, The formants of fricative consonants, *Lang. Speech 8*, 1–16 (1965).
41. O. Fujimura, Analysis of nasal consonants, *J. Acoust. Soc. Amer. 34*, 1865–1875 (1962).
42. A. Malécot, Acoustic cues for nasal consonants, *Language 32*, 274–284 (1956).
43. A. S. House, Analog studies of nasal consonants, *J. Acoust. Soc. Amer. 22*, 190–204 (1957).
44. J. D. O'Connor, L. J. Gerstman, A. M. Liberman, P. C. Delattre, and F. S. Cooper, Acoustic cues for the perception of initial /w,j,r,l/ in English, *Word 13*, 24–43 (1957).
45. L. Lisker, Minimal cues for separating /w,r,l,y/ in intervocalic position, *Word 13*, 256–267 (1957).
46. G. Fant, *Acoustic Theory of Speech Perception*, Mouton, The Hague, 1970.
47. A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, The discrimination of speech sounds within and across phoneme boundaries, *J. Exp. Psych. 54*, 358–368 (1957).
48. A. M. Liberman, K. S. Harris, J. A. Kinney, and H. L. Lane, The discrimination of relative onset of certain speech and non-speech patterns, *J. Exp. Psych. 61*, 379–388 (1961).
49. A. S. Abramson and L. Lisker, Discriminability along the voicing continuum: Cross-language tests, in *Proceedings of the Sixth International Congress on Phonetic Science*, Academia, Prague, 1970, pp. 569–573.

50. L. Lisker and A. S. Abramson, The voicing dimension: Some experiments in comparative phonetics, *Proceedings of the Sixth International Congress on Phonetic Science*, Academia, Prague, 1970, pp. 563–567.
51. P. D. Eimas, P. Siqueland, P. Jusczyk, and J. Vigorito, Speech perception in infants, *Science 171*, 303–306 (1971).
52. A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, Perception of the speech code, *Psych. Rev. 74*, 431–461 (1967).
53. D. B. Pisoni, Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops, *J. Acoust. Soc. Amer. 61*, 1352–1361 (1977).
54. J. E. Cutting and B. S. Rosner, Categories and boundaries in speech and music, *Percept. Psychophysiol. 16*, 564–570 (1974).
55. J. D. Mitler, C. C. Wier, R. Pastore, W. J. Kelly, and R. J. Dooling, Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception, *J. Acoust. Soc. Amer. 60*, 410–417 (1976).
56. R. E. Lasky, A. Syrdal-Lasky, and R. E. Klein, VOT discrimination by four to six and a half month old infants from Spanish environments, *J. Exp. Child Psych. 20*, 213–225 (1975).
57. L. A. Streeter, Language perception of 2-month old infants shows effects of both innate mechanisms and experience, *Nature 259*, 39–41 (1976).
58. P. K. Kuhl and J. D. Miller, Speech perception by the chinchilla: Voice-voiceless distinctions in alveolar plosive consonants, *Science 190*, 69–72 (1975).
59. I. J. Hirsh and C. E. Sherrick, Perceived order in different sense modalities, *J. Exp. Psych. 62*, 423–432 (1961).
60. K. N. Stevens and D. H. Klatt, The role of formant transitions in the voiced-voiceless distinctions for stops, *J. Acoust. Soc. Amer. 55*, 653–659 (1974).
61. C. F. Sacia and C. J. Beck, The power of fundamental speech sounds, *Bell Sys. Tech. J. 5*, 393–403 (1926).
62. H. K. Dunn and S. D. White, Statistical measurements on conversational speech, *J. Acoust. Soc. Amer. 11*, 278–288 (1940).
63. H. W. Rudmose, et al., Effects of high altitude on the human voice, OSRD Rept. 3106, Harvard University, Cambridge, Mass., 1944.
64. N. R. French and G. C. Steinberg, Factors governing the intelligibility of speech, *J. Acoust. Soc. Amer. 19*, 90–114 (1947).
65. G. A. Campbell, Telephonic intelligibility, *Phil. Mag. 19*, 152–159 (1910).
66. H. Fletcher and J. C. Steinberg, Articulation testing methods, *Bell Sys. Tech. J. 8*, 848–852 (1929).
67. J. P. Egan, Articulation testing methods, *Laryngoscope 58*, 955–981 (1948).
68. I. J. Hirsh, H. Davis, S. R. Silverman, E. G. Reynolds, E. Eldert, and R. W. Benson, Development of materials for speech audiometry, *J. Speech Hearing Dis. 17*, 321–337 (1952).
69. G. E. Peterson and I. Lehiste, Revised CNC lists for auditory tests, *J. Speech Hearing Dis. 27*, 62–70 (1962).

70. G. Fairbanks, Test of phonemic differentiation: The rhyme test, *J. Acoust. Soc. Amer. 30,* 596–600 (1958).
71. A. House, C. Williams, H. Hecker, and A. Kryter, Articulation-testing methods: Consonantal differentiations with a close-response set, *J. Acoust. Soc. Amer. 37,* 158–166 (1965).
72. C. V. Hudgens, J. E. Hawkins, J. E. Karlin, and S. S. Stevens, The development of recorded auditory tests for measuring hearing loss for speech, *Laryngoscope 57,* 57–89 (1947).
73. G. A. Miller, G. A. Heise, and W. Lichten, The intelligibility of speech as a function of the context of the test materials, *J. Ex. Psych. 41,* 329–335 (1951).
74. M. R. Rosenzweig and L. Postman, Intelligibility as a function of frequency of usage, *J. Exp. Psych. 54,* 412–422 (1957).
75. L. Hochhouse and J. R. Antes, Speech identification and "knowing that you know," *Percept. Psychophysiol. 13,* 131–132 (1973).
76. J. P. Egan and F. M. Wiener, On the intelligibility of bands of speech in noise *J. Acoust. Soc. Amer. 18,* 435–441 (1946).
77. G. A. Miller and P. E. Nicely, An analysis of perceptual confusions among some English consonants, *J. Acoust. Soc. Amer. 27,* 338–352 (1955).
78. L. L. Beranek, The design of communication systems, *IRE Proc. 35,* 880–890 (1947).
79. K. D. Kryter, Methods for the calculation and use of the articulation index, *J. Acoust. Soc. Amer. 34,* 1689–1697 (1962).
80. K. D. Kryter, Validation of the articulation index, *J. Acoust. Soc. Amer. 34,* 1698–1702 (1962).
81. J. H. Jassen, A method for the calculation of the speech intelligibility under conditions of reverberation and noise, *Acoustica 7,* 305–310 (1957).
82. J. E. Hawkins and S. S. Stevens, The masking of pure tones and of speech by white noise, *J. Acoust. Soc. Amer. 22,* 6–13 (1950).
83. S. S. Stevens, J. Miller, and I. Truscott, The masking of speech by sine waves, square waves, and regular and modulated pulses, *J. Acoust. Soc. Amer. 18,* 418–424 (1946).
84. G. A. Miller, The masking of speech, *Psych. Bull. 44,* 105–129 (1947).
85. R. H. Bolt and A. D. MacDonald, Theory of speech masking by reverberation, *J. Acoust. Soc. Amer. 21,* 577–580 (1949).
86. A. K. Nabelek and J. M. Pickett, Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation and hearing aids, *J. Acoust. Soc. Amer. 56,* 628–639 (1974).
87. H. Kurtović, The influence of reflected sound upon speech intelligibility, *Acoustica 33,* 32–39 (1975).
88. S. A. Gelfand and I. Hochberg, Binaural and monaural speech discrimination under reverberation, *Audiology 15,* 72–84 (1976).
89. A. K. Nabelek, Reverberation effects for normal and hearing-impaired listeners, in *Hearing and Davis: Essays Honoring Hallowell Davis* (S. K. Hirsh, D. H. Eldredge, I. J. Hirsh, and S. R. Silverman, Eds.), Washington University Press, St. Louis, 1976, pp. 333–341.
90. A. K. Nabelek and L. N. Robinette, Reverberation as a parameter in clinical testing, *Audiology 17,* 239–259 (1978).

91. S. A. Gelfand and S. Silman, Effects of small room reverberation upon the recognition of some consonant features, *J. Acoust. Soc. Amer. 66*, 22–29 (1979).
92. P. B. Denes, On the statistics of spoken English, *J. Acoust. Soc. Amer. 35*, 892–904 (1963).
93. J. C. R. Licklider, Effects of amplitude distortion upon the intelligibility of speech, *J. Acoust. Soc. Amer. 18*, 429–434 (1946).
94. J. C. R. Licklider, D. Bindra, and I. Pollack, The intelligibility of rectangular speech waves, *Amer. J. Psych. 61*, 1–20 (1948).
95. J. C. R. Licklider and I. Pollack, Effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech, *J. Acoust. Soc. Amer. 20*, 42–51 (1948).
96. G. A. Miller and J. C. R. Licklider, The intelligibility of interrupted speech, *J. Acoust. Soc. Amer. 22*, 167–173 (1950).
97. G. L. Powers and C. Speaks, Intelligibility of temporally interrupted speech, *J. Acoust. Soc. Amer. 54*, 661–667 (1973).
98. C. Calearo and A. Lazzaroni, Speech intelligibility in relation to the speed of the message, *Laryngoscope 67*, 410–419 (1957).
99. G. Fairbanks and F. Kodman, Word intelligibility as a function of time compression, *J. Acoust. Soc. Amer. 29*, 636–641 (1957).
100. D. S. Beasley, S. Schwimmer, and W. F. Rintelmann, Intelligibility of time-compressed CNC monosyllables, *J. Speech Hearing Res. 15*, 340–350 (1972).
101. D. S. Beasley and J. E. Maki, Time- and frequency-altered speech, in *Handbook of Perception* (E. C. Carterette and M. P. Friedman, Eds.), Vol. 7: *Language and Speech*, Academic, New York, 1976, pp. 419–458.
102. C. I. Berlin and M. R. McNeil, Dichotic listening, in *Contemporary Issues in Experimental Phonetics* (N. J. Lass, Ed.), Academic, New York, 1976, pp. 327–387.
103. D. E. Broadbent, The role of auditory localization in attention and memory span, *J. Exp. Psych. 47*, 191–196 (1954).
104. D. E. Broadbent, Successive responses to simultaneous stimuli, *Quart J. Exp. Psych. 8*, 145–152 (1956).
105. D. Kimura, Cerebral dominance and the perception of verbal stimuli, *Canadian J. Psych 15*, 166–171 (1961).
106. D. Kimura, Functional asymmetry of the brain in dichotic listening, *Cortex 3*, 163–178 (1967).
107. D. Kimura, Left-right differences in the perception of melodies, *Quart. J. Exp. Psych. 16*, 355–358 (1964).
108. D. Shankweiler and M. Studdert-Kennedy, Identification of consonants and vowels presented to the left and right ears, *Quart. J. Exp. Psych. 19*, 59–63 (1967).
109. M. Studdert-Kennedy and D. Shankweiler, Hemispheric specialization for speech perception, *J. Acoust. Soc. Amer. 48*, 579–594 (1970).
110. M. Studdert-Kennedy, D. Shankweiler, and S. Schulman, Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables, *J. Acoust. Soc. Amer. 48*, 599–602 (1970).
111. C. I. Berlin, S. S. Lowe-Bell, J. K. Cullen, C. L. Thompson, and C. F. Loovis, Dichotic speech perception: An interpretation of right-ear advantage and temporal-offset effects, *J. Acoust. Soc. Amer. 53*, 699–709 (1973).

112. J. K. Cullen, C. L. Thompsen, L. F. Hughes, C. I. Berlin, and D. S. Samson, The effects of various acoustic parameters on performance in dichotic speech perception tasks, *Brain Lang. 1*, 307–322 (1974).
113. P. T. Smith, Feature-testing models and their application to perception and memory for speech, *Quart J. Exp. Psych. 25*, 511–534 (1973).
114. J. R. Sawusch and D. Pisoni, On the identification of place and voicing features in synthetic stop consonants, *J. Phonetics 2*, 181–194 (1974).
115. C. C. Wood, Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses, *J. Exp. Psych. 104*, 1–33 (1975).
116 C. C. Wood, W. R. Goff, and R. S. Day, Auditory evoked potentials during speech perception, *Science 173*, 1248–1251 (1971).
117. C. J. Darwin, Dichotic backward masking of complex sounds, *Quart. J. Exp. Psych. 23*, 386–392 (1971).
118. D. Pisoni and S. D. McNabb, Dichotic interactions and phonetic feature processing, *Brain Lang. 1*, 351–362 (1974).
119. P. J. Mirabile and R. J. Porter, Dichotic and monotic interaction between speech and nonspeech sounds at different stimulus-onset-asynchronies. Paper at 89th meeting of the Acoustical Society of America, Austin, Texas, 1975.
120. R. J. Porter, Effect of delayed channel on the perception of dichotically presented speech and nonspeech sounds, *J. Acoust. Soc. Amer. 58*, 884–892 (1975).

Author Index

Subject Index