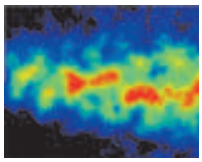
Chapter 10



Vision: The Eye

Overview

The human visual system is extraordinary in the quantity and quality of information it supplies about the world. A glance is sufficient to describe the location, size, shape, color, and texture of objects and, if the objects are moving, their direction and speed. Equally remarkable is the fact that visual information can be discerned over a wide range of stimulus intensities, from the faint light of stars at night to bright sunlight. The next two chapters describe the molecular, cellular, and higher-order mechanisms that allow us to see. The first steps in the process of seeing involve transmission and refraction of light by the optics of the eye, the transduction of light energy into electrical signals by photoreceptors, and the refinement of these signals by synaptic interactions within the neural circuits of the retina.

Anatomy of the Eye

The eye is a fluid-filled sphere enclosed by three layers of tissue (Figure 10.1). Only the innermost layer of the eye, the **retina**, contains neurons that are sensitive to light and are capable of transmitting visual signals to central targets. The immediately adjacent layer of tissue includes three distinct but continuous structures collectively referred to as the **uveal tract**. The largest component of the uveal tract is the **choroid**, which is composed of a rich capillary bed (important for nourishing the photoreceptors of the retina) as well as a high concentration of the light absorbing pigment melanin. Extending from the choroid near the front of the eye is the **ciliary body**, a ring of tissue that encircles the lens and consists of a muscular component that is important for adjusting the refractive power of the lens, and a vascular component (the so-called ciliary processes) that produces the fluid that fills the front of the eye. The most anterior component of the uveal tract is the **iris**, the colored portion of the eye that can be seen through the cornea. It contains two sets of muscles with opposing actions, which allow the size of the **pupil** (the opening in its center) to be adjusted under neural control. The **sclera** forms the outermost tissue layer of the eye and is composed of a tough white fibrous tissue. At the front of the eye, however, this opaque outer layer is transformed into the **cornea**, a specialized transparent tissue that permits light rays to enter the eye.

Beyond the cornea, light rays pass through two distinct fluid environments before striking the retina. In the **anterior chamber**, just behind the cornea and in front of the lens, lies **aqueous humor**, a clear, watery liquid that supplies nutrients to both of these structures. Aqueous humor is produced by the ciliary processes in the **posterior chamber** (the region between

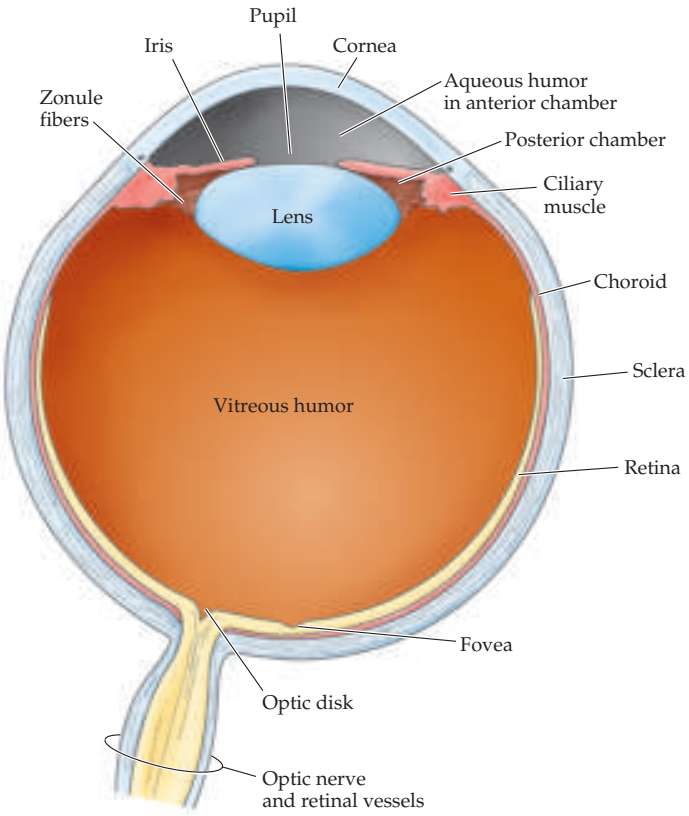


Figure 10.1 Anatomy of the human eye.

the lens and the iris) and flows into the anterior chamber through the pupil. The amount of fluid produced by the ciliary processes is substantial: it is estimated that the entire volume of fluid in the anterior chamber is replaced 12 times a day. Thus the rates of a aqueous humor production must be balanced by comparable rates of drainage from the anterior chamber in order to ensure a constant intraocular pressure. A specialized meshwork of cells that lies at the junction of the iris and the cornea (a region called the **limbus**) is responsible for aqueous drainage. Failure of adequate drainage results in a disorder known as **glaucoma**, in which high levels of intraocular pressure can reduce the blood supply to the eye and eventually damage retinal neurons.

The space between the back of the lens and the surface of the retina is filled with a thick, gelatinous substance called the **vitreous humor**, which accounts for about 80% of the volume of the eye. In addition to maintaining the shape of the eye, the vitreous humor contains phagocytic cells that remove blood and other debris that might otherwise interfere with light transmission. The housekeeping abilities of the vitreous humor are limited, however, as a large number of middle-aged and elderly individuals with vitreal "floaters" will attest. Floaters are collections of debris too large for phagocytic consumption that therefore remain to cast annoying shadows on the retina; they typically arise when the aging vitreous membrane pulls away from the overly long eyeball of myopic individuals (Box A).

The Formation of Images on the Retina

Normal vision requires that the optical media of the eye be transparent, and both the **cornea** and the **lens** are remarkable examples of tissue specializations that achieve a level of transparency that rivals that found in inorganic materials such as glass. Not surprisingly, alterations in the composition of the cornea or the lens can significantly reduce their transparency and have serious consequences for visual perception. Indeed, **cataracts** (opacities in the lens) account for roughly half the cases of blindness in the world, and almost everyone over the age of 70 will experience some loss of transparency in the lens that ultimately degrades the quality of visual experience. Fortunately, there are successful surgical treatments for cataracts that can restore vision in most cases. Furthermore, the recognition that a major factor in the production of cataracts is exposure to ultraviolet (UV) solar radiation has heightened public awareness of the need to protect the lens (and the retina) by reducing UV exposure through the use of sunglasses.

Beyond efficiently transmitting light energy, the primary function of the optical components of the eye is to achieve a focused image on the surface of the retina. The cornea and the lens are primarily responsible for the refraction (bending) of light that is necessary for formation of focused images on the photoreceptors of the retina (Figure 10.2). The cornea contributes most of the necessary refraction, as can be appreciated by considering the hazy, out-of-focus images experienced when swimming underwater. Water, unlike air, has a refractive index close to that of the cornea; as a result, immersion in water virtually eliminates the refraction that normally occurs at the air/cornea interface; thus the image is no longer focused on the retina. The lens has considerably less refractive power than the cornea; however, the refraction supplied by the lens is adjustable, allowing objects at various distances from the observer to be brought into sharp focus**.**

Dynamic changes in the refractive power of the lens are referred to as **accommodation**. When viewing distant objects, the lens is made relatively thin and flat and has the least refractive power. For near vision, the lens becomes thicker and rounder and has the most refractive power (see Figure 10.2). These changes result from the activity of the **ciliary muscle** that surrounds the lens. The lens is held in place by radially arranged connective tissue bands (called zonule fibers) that are attached to the ciliary muscle. The shape of the lens is thus determined by two opposing forces: the elasticity of the lens, which tends to keep it rounded up (removed from the eye, the lens

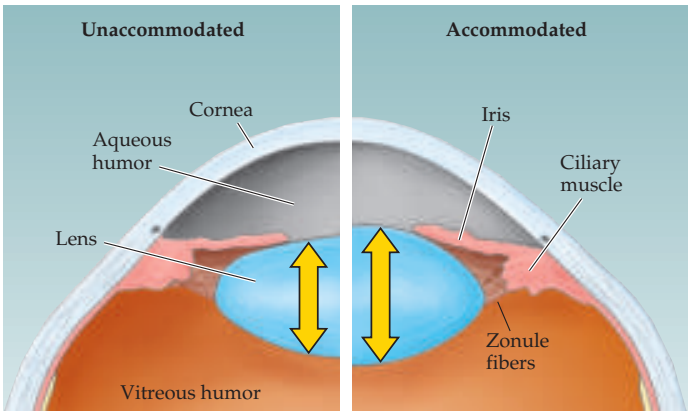


Figure 10.2 Diagram showing the anterior part of the human eye in the unaccommodated (left) and accommodated (right) state. Accommodation for focusing on near objects involves the contraction of the ciliary muscle, which reduces the tension in the zonule fibers and allows the elasticity of the lens to increase its curvature.

Box A

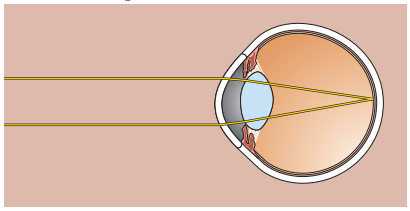
Myopia and Other Refractive Errors

Optical discrepancies among the various components of the eye cause a majority of the human population to have some form of refractive error, called *ametropia*. People who are unable to bring distant objects into clear focus are said to be nearsighted, or myopic (Figures A and B). *Myopia* can be caused by the corneal surface being too curved, or by the eyeball being too long. In either case, with the lens as flat as it can be, the image of distant objects focuses in front of, rather than on, the retina. People who are unable to focus on near objects are said to be farsighted, or hyperopic. *Hyperopia* can be caused by the eyeball being too short or the refracting system too weak (Figure C). Even with the lens in its most rounded-up state, the image is out of focus on the retinal surface (focusing at some point behind it). Both myopia and hyperopia are correctable by appropriate lenses—concave (minus) and convex (plus), respectively—or by the increasingly popular technique of corneal surgery.

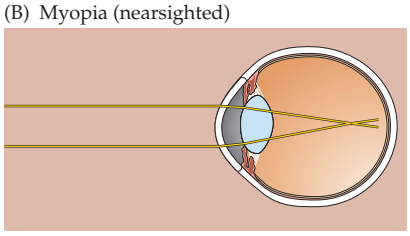
Myopia, or nearsightedness, is by far the most common ametropia; an estimated 50% of the population in the United States is affected. Given the large number of people who need glasses, contact lenses, or surgery to correct this refractive error, one naturally wonders how nearsighted people coped before spectacles were invented only a few centuries ago. From what is now known about myopia, most people's vision may have been considerably better in ancient times. The basis for this assertion is the surprising finding that the growth of the eyeball is strongly influenced by focused light falling on the retina. This phenomenon was first described in 1977 by Torsten Wiesel and Elio Raviola at Harvard Medical School, who studied monkeys reared with their lids sutured (the same approach used to demonstrate the effects of visual deprivation on cortical connections in the visual system; see Chapter 23), a procedure that deprives the eye of focused retinal images. They found that animals growing to maturity under these conditions show an elongation of the eyeball. The effect of focused light deprivation appears to be a local one, since the abnormal growth of the eye occurs in experimental animals even if the optic nerve is cut. Indeed, if only a portion of the retinal surface is deprived of focused light, then only that region of the eyeball grows abnormally.

Although the mechanism of light-mediated control of eye growth is not fully understood, many experts now believe that the prevalence of myopia is due to some aspect of modern civilization—perhaps learning to read and write at an early age—that interferes with the normal feedback control of vision on eye development, leading to abnormal elongation of the eyeball. A corollary of this hypothesis is that if children (or, more likely, their parents) wanted to improve their vision, they might be able to do so by practicing far vision to counterbalance the near work "overload." Practically, of

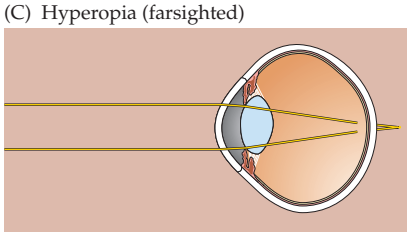
(A) Emmetropia (normal)



(B) Myopia (nearsighted)



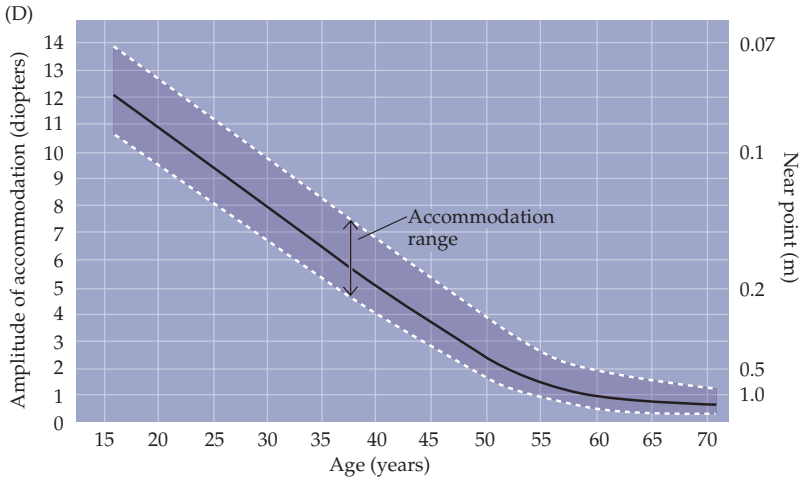
(C) Hyperopia (farsighted)



Refractive errors. (A) In the normal eye, with ciliary muscles relaxed, an image of a distant object is focused on the retina. (B) In myopia, light rays are focused in front of the retina. (C) In hyperopia, images are focused at a point beyond the retina.

becomes spheroidal), and the tension exerted by the zonule fibers, which tends to flatten it. When viewing distant objects, the force from the zonule fibers is greater than the elasticity of the lens, and the lens assumes the flatter shape appropriate for distance viewing. Focusing on closer objects requires relaxing the tension in the zonule fibers, allowing the inherent elasticity of the lens to increase its curvature. This relaxation is accomplished by the sphincter-like contraction of the ciliary muscle. Because the ciliary muscle forms a ring around the lens, when the muscle contracts, the attachment points of the zonule fibers move toward the central axis of the eye, thus

(D) Changes in the ability of the lens to round up (accommodate) with age. The graph also shows how the near point (the closest point to the eye that can be brought into focus) changes. Accommodation, which is an optical measurement of the refractive power of the lens, is given in diopters. (After Westheimer, 1974.)



course, most people would probably choose wearing glasses or contacts or having corneal surgery rather than indulging in the onerous daily practice that would presumably be required. Not everyone agrees, however, that such a remedy would be effective, and a number of investigators (and drug companies) are exploring the possibility of pharmacological intervention during the period of childhood when abnormal eye growth is presumed to occur. In any event, it is a remarkable fact that deprivation of focused light on the retina causes a compensatory growth of the eye and that this feedback loop is so easily perturbed.

Even people with normal (emmetropic) vision as young adults eventually experience difficulty focusing on near objects. One of the many consequences of aging is that the lens loses its elasticity; as a result, the maximum curvature the lens can achieve when the ciliary muscle contracts is gradually reduced. The near point (the closest point that can be brought into clear focus) thus recedes, and objects (such as this book) must be farther and farther away from the eye in order to focus them on the retina. At some point, usually during early middle age, the accommodative ability of the eye is so reduced that near vision tasks like reading become difficult or impossible (Figure D). This condition is referred to as presbyopia, and can be corrected by convex lenses for near-vision tasks, or by bifocal lenses if myopia is also present (which requires a negative correction). Bifocal correction presents a particular problem for those who prefer contact lenses. Because contact lenses float on the surface of the cornea, having the distance correction above and the near correction below (as in conventional bifocal glasses) doesn't work (although "omnifocal" contact lenses have recently been used with some success). A surprisingly effective solution to this problem for some contact lens wearers has been to put a near correcting lens in one eye and a distance correcting lens in the other! The success of this approach is another testament to the remarkable ability of the visual system to adjust to a wide variety of unusual demands.

References

Bock, G. and K. Widdows (1990) *Myopia and the Control of Eye Growth*. Ciba Foundation Symposium 155. Chichester: Wiley.

Coster, D. J. (1994) *Physics for Ophthalmologists*. Edinburgh: Churchill Livingston.

Kaufman, P. L. and A. Alm (eds.) (2002) *Adler's Physiology of the Eye: Clinical Application*, 10th Ed. St. Louis, MO: Mosby Year Book.

Sherman, S. M., T. T. Norton and V. A. Casagrande (1977) Myopia in the lid-sutured tree shrew. Brain Res. 124: 154–157.

Wallman, J., J. Turkel and J. Tractman (1978) Extreme myopia produced by modest changes in early visual experience. Science 201: 1249–1251.

Wiesel, T. N. and E. Raviola (1977) Myopia and eye enlargement after neonatal lid fusion in monkeys. Nature 266: 66–68.

reducing the tension on the lens. Unfortunately, changes in the shape of the lens are not always able to produce a focused image on the retina, in which case a sharp image can be focused only with the help of additional corrective lenses (see Box A).

Adjustments in the size of the pupil also contribute to the clarity of images formed on the retina. Like the images formed by other optical instruments, those generated by the eye are affected by spherical and chromatic aberrations, which tend to blur the retinal image. Since these aberrations are greatest for light rays that pass farthest from the center of the lens, narrow-

ing the pupil reduces both spherical and chromatic aberration, just as closing the iris diaphragm on a camera lens improves the sharpness of a photographic image. Reducing the size of the pupil also increases the depth of field—that is, the distance within which objects are seen without blurring. However, a small pupil also limits the amount of light that reaches the retina, and, under conditions of dim illumination, visual acuity becomes limited by the number of available photons rather than by optical aberrations. An adjustable pupil thus provides an effective means of reducing optical aberrations, while maximizing depth of field to the extent that different levels of illumination permit. The size of the pupil is controlled by innervation from both sympathetic and parasympathetic divisions of the visceral motor system, which are in turn modulated by several brainstem centers (see Chapters 19 and 20).

The Retina

Despite its peripheral location, the **retina** or neural portion of the eye, is actually part of the central nervous system. During development, the retina forms as an outpocketing of the diencephalon, called the optic vesicle, which undergoes invagination to form the optic cup (Figure 10.3; see also Chapter 21). The inner wall of the optic cup gives rise to the retina, while the outer wall gives rise to the **retinal pigment epithelium**. This epithelium is a thin melanin-containing structure that reduces backscattering of light that enters the eye; it also plays a critical role in the maintenance of photoreceptors, renewing photopigments and phagocytosing the photoreceptor disks, whose turnover at a high rate is essential to vision.

Consistent with its status as a full-fledged part of the central nervous system, the retina comprises complex neural circuitry that converts the graded electrical activity of photoreceptors into action potentials that travel to the brain via axons in the optic nerve. Although it has the same types of functional elements and neurotransmitters found in other parts of the central nervous system, the retina comprises fewer classes of neurons, and these are arranged in a manner that has been less difficult to unravel than the circuits in other areas of the brain. There are five types of neurons in the retina: **photoreceptors, bipolar cells, ganglion cells, horizontal cells**, and **amacrine cells**. The cell bodies and processes of these neurons are stacked in alternating layers, with the cell bodies located in the inner nuclear, outer nuclear, and ganglion cell layers, and the processes and synaptic contacts located in the inner plexiform and outer plexiform layers (Figure 10.4). A direct three-

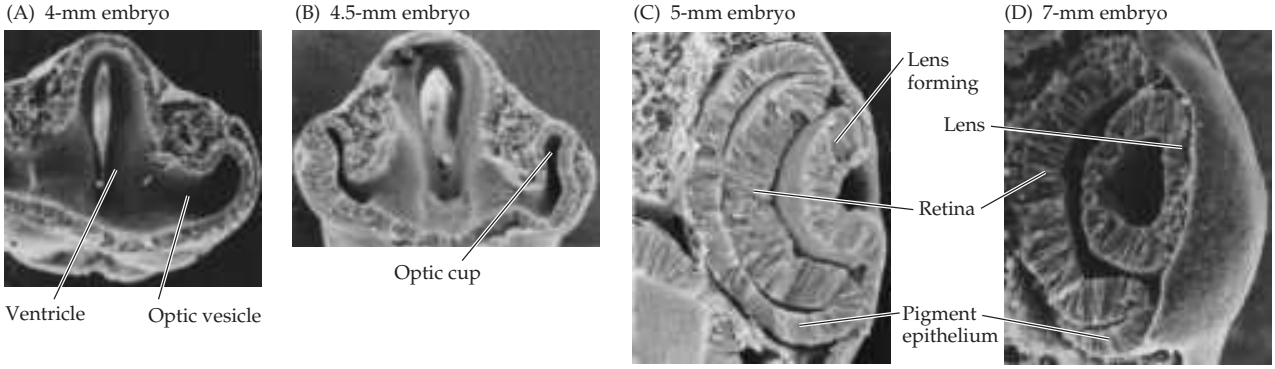


Figure 10.3 Development of the human eye. (A) The retina develops as an outpocketing from the neural tube, called the optic vesicle. (B) The optic vesicle invaginates to form the optic cup. (C, D) The inner wall of the optic cup becomes the neural retina, while the outer wall becomes the pigment epithelium. (A–C from Hilfer and Yang, 1980; D courtesy of K. Tosney.)

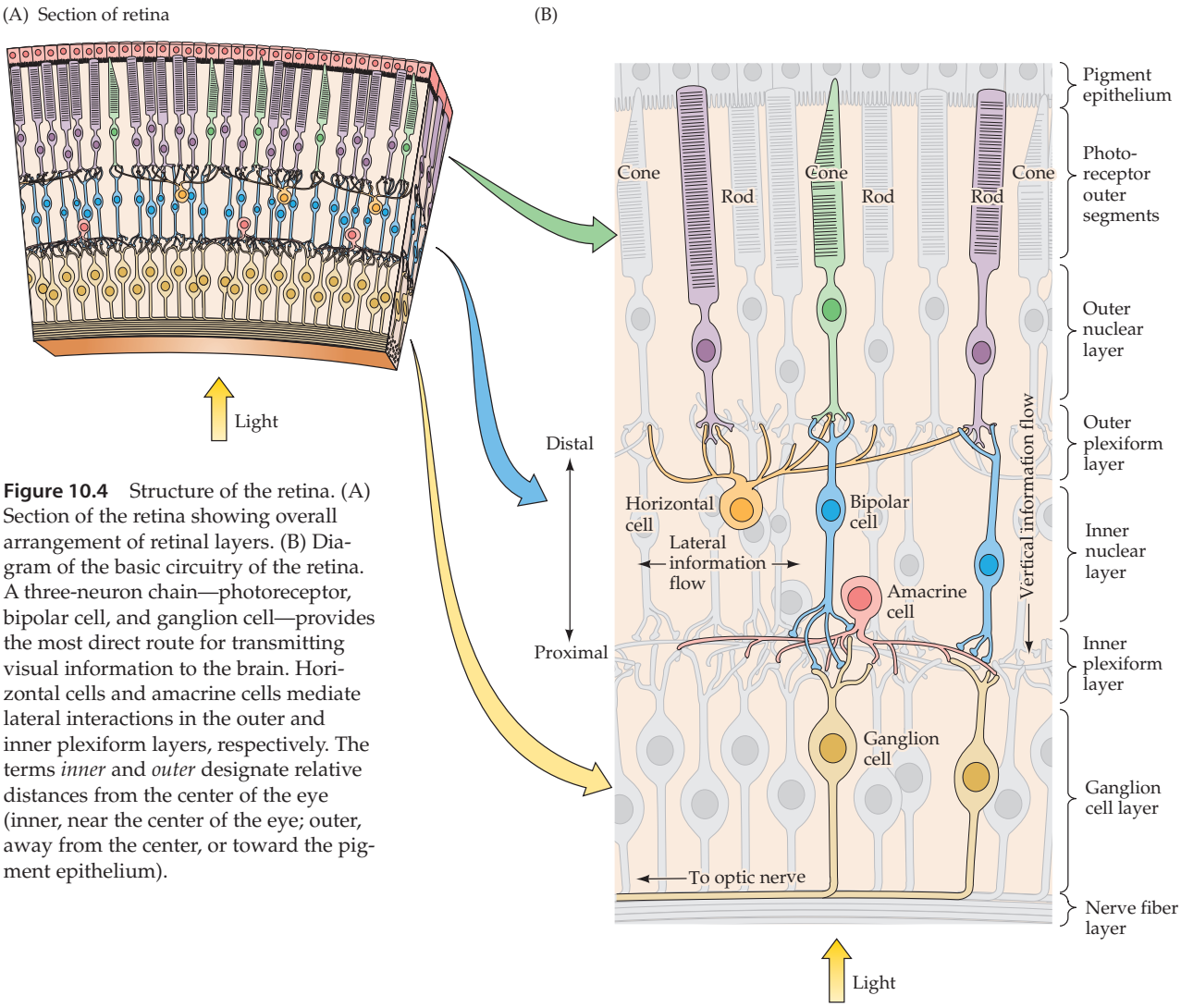


Figure 10.4 Structure of the retina. (A) Section of the retina showing overall arrangement of retinal layers. (B) Diagram of the basic circuitry of the retina. A three-neuron chain—photoreceptor, bipolar cell, and ganglion cell—provides the most direct route for transmitting visual information to the brain. Horizontal cells and amacrine cells mediate lateral interactions in the outer and inner plexiform layers, respectively. The terms *inner* and *outer* designate relative distances from the center of the eye (inner, near the center of the eye; outer, away from the center, or toward the pigment epithelium).

neuron chain—photoreceptor cell to bipolar cell to ganglion cell—is the major route of information flow from photoreceptors to the optic nerve.

There are two types of photoreceptors in the retina: **rods** and **cones**. Both types have an outer segment composed of membranous disks that contain light-sensitive photopigment and lies adjacent to the pigment epithelium, and an inner segment that contains the cell nucleus and gives rise to synaptic terminals that contact bipolar or horizontal cells (see also Figure 10.8). Absorption of light by the photopigment in the outer segment of the photoreceptors initiates a cascade of events that changes the membrane potential of the receptor, and therefore the amount of neurotransmitter released by the photoreceptor synapses onto the cells they contact. The synapses between photoreceptor terminals and bipolar cells (and horizontal cells) occur in the outer plexiform layer; more specifically, the cell bodies of photoreceptors make up the outer nuclear layer, whereas the cell bodies of bipolar cells lie in the inner nuclear layer. The short axonal processes of bipolar cells make synaptic contacts in turn on the dendritic processes of ganglion cells in the inner plexiform layer. The much larger axons of the ganglion cells form the **optic**

nerve and carry information about retinal stimulation to the rest of the central nervous system.

The two other types of neurons in the retina, **horizontal cells** and **amacrine cells**, have their cell bodies in the inner nuclear layer and have processes that are limited to the outer and inner plexiform layers respectively (see Figure 10.4). The processes of horizontal cells enable lateral interactions between photoreceptors and bipolar cells that maintain the visual system's sensitivity to luminance contrast over a wide range of light intensities. The processes of amacrine cells are postsynaptic to bipolar cell terminals and presynaptic to the dendrites of ganglion cells. Different subclasses of amacrine cells are thought to make distinct contributions to visual function. One class of amacrine cells, for example, plays an important role in transforming the sustained responses of bipolar cells to step changes in light intensity into transient onset or offset responses exhibited by some types of ganglion cells. Another type serves as an obligatory step in the pathway that transmits information from rod photoreceptors to retinal ganglion cells. The variety of amacrine cell subtypes illustrates the more general rule that although there are only five basic retinal cell types, there can be considerable diversity within a given cell type. This diversity is also a hallmark of retinal ganglion cells and the basis for pathways that convey different sorts of information to central targets in a parallel manner (see Chapter 11).

At first glance, the spatial arrangement of retinal layers seems counterintuitive, since light rays must pass through various non-light-sensitive elements of the retina as well as the retinal vasculature (which branches extensively on the inner surface of the retina—see Figure 11.1) before reaching the outer segments of the photoreceptors, where photons are absorbed (Figure 10.4). The reason for this curious feature of retinal organization lies in the special relationship that exists among the outer segments of the photoreceptors, the pigment epithelium, and the underlying choroid. Recall that the outer segments contain membranous disks that house the light-sensitive photopigment and other proteins involved in the transduction process. These disks are formed near the inner segment of the photoreceptor and move toward the tip of the outer segment, where they are shed. The pigment epithelium plays an essential role in removing the expended receptor disks; this is no small task, since all the disks in the outer segments are replaced every 12 days. In addition, the pigment epithelium contains the biochemical machinery that is required to regenerate photopigment molecules after they have been exposed to light. Finally, the capillaries in the choroid underlying the pigment epithelium are the primary source of nourishment for retinal photoreceptors. These functional considerations presumably explain why rods and cones are found in the outermost rather than the innermost layer of the retina. They also explain why disruptions in the normal relationships between the pigment epithelium and retinal photoreceptors such as those that occur in retinitis pigmentosa have severe consequences for vision (Box B).

Phototransduction

In most sensory systems, activation of a receptor by the appropriate stimulus causes the cell membrane to depolarize, ultimately stimulating an action potential and transmitter release onto the neurons it contacts. In the retina, however, photoreceptors do not exhibit action potentials; rather, light activation causes a graded change in membrane potential and a corresponding change in the rate of transmitter release onto postsynaptic neurons. Indeed,

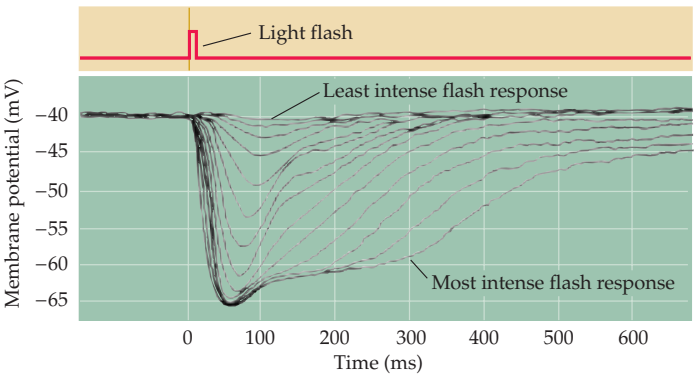


Figure 10.5 An intracellular recording from a single cone stimulated with different amounts of light (the cone has been taken from the turtle retina, which accounts for the relatively long time course of the response). Each trace represents the response to a brief flash that was varied in intensity. At the highest light levels, the response amplitude saturates (at about –65 mV). The hyperpolarizing response is characteristic of vertebrate photoreceptors; interestingly, some invertebrate photoreceptors depolarize in response to light. (After Schnapf and Baylor, 1987.)

much of the processing within the retina is mediated by graded potentials, largely because action potentials are not required to transmit information over the relatively short distances involved.

Perhaps even more surprising is that shining light on a photoreceptor, either a rod or a cone, leads to membrane *hyperpolarization* rather than depolarization (Figure 10.5). In the dark, the receptor is in a depolarized state, with a membrane potential of roughly –40 mV (including those portions of the cell that release transmitters). Progressive increases in the intensity of illumination cause the potential across the receptor membrane to become more negative, a response that saturates when the membrane potential reaches about –65 mV. Although the sign of the potential change may seem odd, the only logical requirement for subsequent visual processing is a consistent relationship between luminance changes and the rate of transmitter release from the photoreceptor terminals. As in other nerve cells, transmitter release from the synaptic terminals of the photoreceptor is dependent on voltage-sensitive Ca^{2+} channels in the terminal membrane. Thus, in the dark, when photoreceptors are relatively depolarized, the number of open Ca^{2+} channels in the synaptic terminal is high, and the rate of transmitter release is correspondingly great; in the light, when receptors are hyperpolarized, the number of open Ca^{2+} channels is reduced, and the rate of transmitter release is also reduced. The reason for this unusual arrangement compared to other sensory receptor cells is not known.

The relatively depolarized state of photoreceptors in the dark depends on the presence of ion channels in the outer segment membrane that permit Na^+ and Ca^{2+} ions to flow into the cell, thus reducing the degree of inside negativity (Figure 10.6). The probability of these channels in the outer segment being open or closed is regulated in turn by the levels of the nucleotide cyclic guanosine monophosphate (cGMP) (as in many other second messenger systems; see Chapter 7). In darkness, high levels of cGMP in the outer segment keep the channels open. In the light, however, cGMP levels drop and some of the channels close, leading to hyperpolarization of the outer segment membrane, and ultimately the reduction of transmitter release at the photoreceptor synapse.

The series of biochemical changes that ultimately leads to a reduction in cGMP levels begins when a photon is absorbed by the photopigment in the receptor disks. The photopigment contains a light-absorbing chromophore (**retinal**, an aldehyde of vitamin A) coupled to one of several possible proteins called **opsins** that tune the molecule's absorption of light to a particular region of the spectrum. Indeed, it is the different protein component of

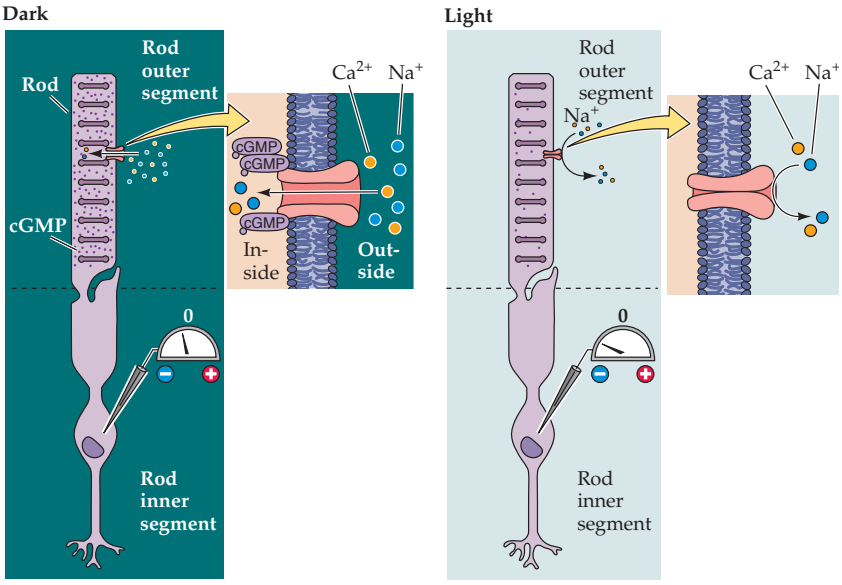


Figure 10.6 Cyclic GMP-gated channels in the outer segment membrane are responsible for the light-induced changes in the electrical activity of photoreceptors (a rod is shown here, but the same scheme applies to cones). In the dark, cGMP levels in the outer segment are high; this molecule binds to the Na^+-permeable channels in the membrane, keeping them open and allowing sodium (and other cations) to enter, thus depolarizing the cell. Exposure to light leads to a decrease in cGMP levels, a closing of the channels, and receptor hyperpolarization.

the photopigment in rods and cones that contributes to the functional specialization of these two receptor types. Most of what is known about the molecular events of phototransduction has been gleaned from experiments in rods, in which the photopigment is **rhodopsin** (Figure 10.7A). When the retinal moiety in the rhodopsin molecule absorbs a photon, its configuration changes from the 11-*cis* isomer to all-*trans* retinal; this change then triggers a series of alterations in the protein component of the molecule (Figure 10.7B). The changes lead, in turn, to the activation of an intracellular messenger called **transducin**, which activates a phosphodiesterase that hydrolyzes

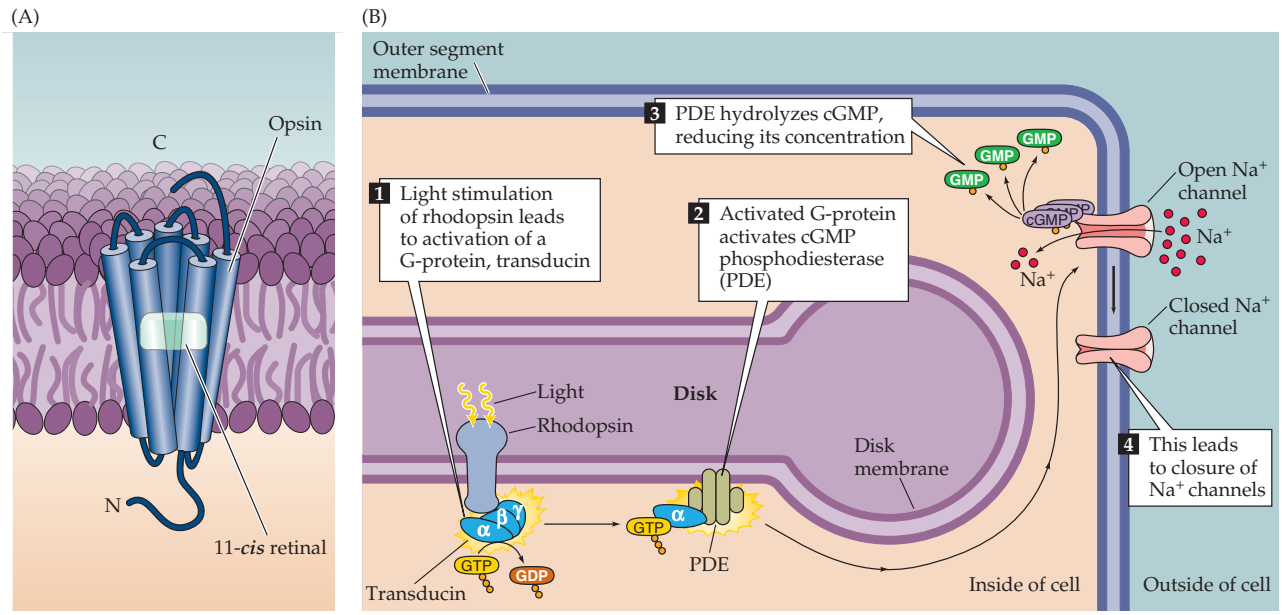


Figure 10.7 Details of phototransduction in rod photoreceptors. (A) The molecular structure of rhodopsin, the pigment in rods. (B) The second messenger cascade of phototransduction. Light stimulation of rhodopsin in the receptor disks leads to the activation of a G-protein (transducin), which in turn activates a phosphodiesterase (PDE). The phosphodiesterase hydrolyzes cGMP, reducing its concentration in the outer segment and leading to the closure of sodium channels in the outer segment membrane.

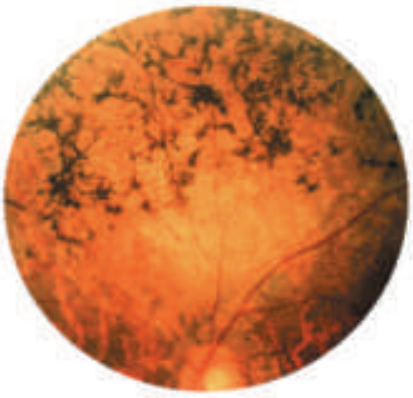
Box B

Retinitis Pigmentosa

Retinitis pigmentosa (**RP**) refers to a heterogeneous group of hereditary eye disorders characterized by progressive vision loss due to a gradual degeneration of photoreceptors. An estimated 100,000 people in the United States have RP. In spite of the name, inflammation is not a prominent part of the disease process; instead the photoreceptor cells appear to die by apoptosis (determined by the presence of DNA fragmentation).

Classification of this group of disorders under one rubric is based on the clinical features commonly observed in these patients. The hallmarks of RP are night blindness, a reduction of peripheral vision, narrowing of the retinal vessels, and the migration of pigment from disrupted retinal pigment epithelium into the retina, forming clumps of various sizes, often next to retinal blood vessels (see figure).

Typically, patients first notice difficulty seeing at night due to the loss of rod photoreceptors; the remaining cone photoreceptors then become the mainstay of visual function. Over many years, the cones also degenerate, leading to a progressive loss of vision. In most RP patients, visual field defects begin in the midperiphery, between 30° and 50° from the point of foveal fixation. The defective regions gradually enlarge, leaving islands of vision in the periphery and a constricted central field—a condition known as tunnel vision. When the visual field contracts to 20° or less and/or central vision is 20/200 or worse, the patient is categorized as legally blind.



Characteristic appearance of the retina in patients with retinitis pigmentosa. Note the dark clumps of pigment that are the hallmark of this disorder.

Inheritance patterns indicate that RP can be transmitted in an X-linked (XLRP), autosomal dominant (ADRP), or recessive (ARRP) manner. In the United States, the percentage of these genetic types is estimated to be 9%, 16%, and 41%, respectively. When only one member of a pedigree has RP, the case is classified as "simplex," which accounts for about a third of all cases.

Among the three genetic types of RP, ADRP is the mildest. These patients often retain good central vision until 60 years of age or older. In contrast, patients with the XLRP form of the disease are usually legally blind by 30 to 40 years of age. However, the severity and age of onset of the symptoms varies greatly among patients with the same type of RP, and even within the same family (when, presumably, all the affected members have the same genetic mutation).

To date, RP-inducing mutations of 30 genes have been identified. Many of these genes encode photoreceptor-specific proteins, several being associated with phototransduction in the rods. Among the latter are genes for rhodopsin, subunits of the cGMP phosphodiesterase, and the cGMP-gated channel. Multiple mutations have been found in each of these cloned genes. For example, in the case of the rhodopsin gene, 90 different mutations have been identified among ADRP patients.

The heterogeneity of RP at all levels, from genetic mutations to clinical symptoms, has important implications for understanding the pathogenesis of the disease and designing therapies. Given the complex molecular etiology of RP, it is unlikely that a single cellular mechanism will explain the disease in all cases. Regardless of the specific mutation or causal sequence, the vision loss that is most critical to RP patients is due to the gradual degeneration of cones. In many cases, the protein that the RP-causing mutation affects is not even expressed in the cones; the prime example is rhodopsin—the rod-specific visual pigment. Therefore, the loss of cones may be an indirect result of a rod-specific mutation. In consequence, understanding and treating this disease presents a particularly difficult challenge.

References

Weleber, R. G. and K. Gregory-Evans (2001) Retinitis pigmentosa and allied disorders. In *Retina*, 3rd Ed., Vol. 1: *Basic Science and Inherited Retinal Diseases.* S. J. Ryan (ed. in chief). St. Louis, MO: Mosby Year Book, pp. 362–460.

Rattner, A., A. Sun and J. Nathans (1999) Molecular genetics of human retinal disease. Annu. Rev. Genet. 33: 89–131.

The Foundation Fighting Blindness of Hunt Valley, MD, maintains a web site that provides updated information about many forms of retinal degeneration: www.blindness.org

RetNet provides updated information, including references to original articles, on genes and mutations associated with retinal diseases: www.sph.uth.tmc.edu/RetNet

cGMP. All of these events take place within the disk membrane. The hydrolysis by phosphodiesterase at the disk membrane lowers the concentration of cGMP throughout the outer segment, and thus reduces the number of cGMP molecules that are available for binding to the channels in the surface of the outer segment membrane, leading to channel closure.

One of the important features of this complex biochemical cascade initiated by photon capture is that it provides enormous signal amplification. It has been estimated that a single light-activated rhodopsin molecule can activate 800 transducin molecules, roughly eight percent of the transducin molecules on the disk surface. Although each transducin molecule activates only one phosphodiesterase molecule, each of these is in turn capable of catalyzing the breakdown of as many as six cGMP molecules. As a result, the absorption of a single photon by a rhodopsin molecule results in the closure of approximately 200 ion channels, or about 2% of the number of channels in each rod that are open in the dark. This number of channel closures causes a net change in the membrane potential of about 1 mV.

Equally important is the fact that the magnitude of this amplification varies with the prevailing levels of illumination, a phenomenon known as **light adaptation**. At low levels of illumination, photoreceptors are the most sensitive to light. As levels of illumination increase, sensitivity decreases, preventing the receptors from saturating and thereby greatly extending the range of light intensities over which they operate. The concentration of Ca^{2+} in the outer segment appears to play a key role in the light-induced modulation of photoreceptor sensitivity. The cGMP-gated channels in the outer segment are permeable to both Na^{+} and Ca^{2+}; thus, light-induced closure of these channels leads to a net decrease in the internal Ca^{2+} concentration. This decrease triggers a number of changes in the phototransduction cascade, all of which tend to reduce the sensitivity of the receptor to light. For example, the decrease in Ca^{2+} increases the activity of quanylate cyclase, the cGMP synthesizing enzyme, leading to an increase in cGMP levels. Likewise, the decrease in Ca^{2+} increases the affinity of the cGMP-gated channels for cGMP, reducing the impact of the light-induced reduction of cGMP levels. The regulatory effects of Ca^{2+} on the phototransduction cascade are only one part of the mechanism that adapts retinal sensitivity to background levels of illumination; another important contribution comes from neural interactions between horizontal cells and photoreceptor terminals (see below).

Once initiated, additional mechanisms limit the duration of this amplifying cascade and restore the various molecules to their inactivated states. The protein **arrestin**, for instance, blocks the ability of activated rhodopsin to activate transducin, and facilitates the breakdown of activated rhodopsin. The all-*trans* retinal then dissociates from the opsin, diffuses into the cytosol of the outer segment, is converted to all-*trans* retinol and is transported out of the outer segment and into the pigment epithelium, where appropriate enzymes ultimately convert it to 11-*cis* retinal. After it is transported back into the outer segment, the 11-*cis* retinal recombines with opsin in the receptor disks. The recycling of rhodopsin is critically important for maintaining the light sensitivity of photoreceptors. Even under intense levels of illumination, the rate of regeneration is sufficient to maintain a significant number of active photopigment molecules.

Functional Specialization of the Rod and Cone Systems

The two types of photoreceptors, rods and cones, are distinguished by shape (from which they derive their names), the type of photopigment they con-

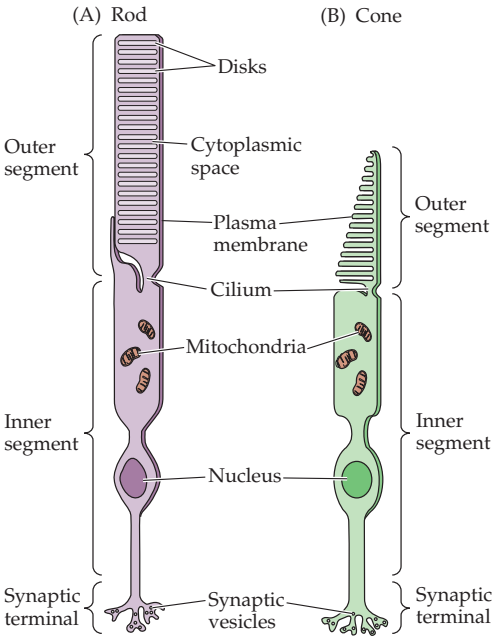


Figure 10.8 Structural differences between rods and cones. Although generally similar in structure, rods (A) and cones (B) differ in their size and shape, as well as in the arrangement of the membranous disks in their outer segments.

tain, distribution across the retina, and pattern of synaptic connections (Figure 10.8). These properties reflect the fact that the rod and cone systems (the receptors and their connections within the retina) are specialized for different aspects of vision. The rod system has very low spatial resolution but is extremely sensitive to light; it is therefore specialized for sensitivity at the expense of resolution. Conversely, the cone system has very high spatial resolution but is relatively insensitive to light; it is therefore specialized for acuity at the expense of sensitivity. The properties of the cone system also allow humans and many other animals to see color.

The range of illumination over which the rods and cones operate is shown in Figure 10.9. At the lowest levels of light, only the rods are activated. Such rod-mediated perception is called **scotopic vision**. The difficulty of making fine visual discriminations under very low light conditions where only the rod system is active is a common experience. The problem is primarily the poor resolution of the rod system (and, to a lesser degree, the fact that there is no perception of color in dim light because the cones are not involved to a significant degree). Although cones begin to contribute to visual perception at about the level of starlight, spatial discrimination at this light level is still very poor. As illumination increases, cones become more and more dominant in determining what is seen, and they are the major determinant of perception under relatively bright conditions such as normal indoor lighting or sunlight. The contributions of rods to vision drops out nearly entirely in so-called **photopic vision** because their response to light saturates—that is, the membrane potential of individual rods no longer varies as a function of illumination because all of the membrane channels are closed (see Figure 10.5). **Mesopic vision** occurs in levels of light at which both rods and cones contribute—at twilight, for example. From these considerations it should be clear that most of what we think of as normal "seeing" is mediated by the cone system, and that loss of cone function is devastating, as occurs in

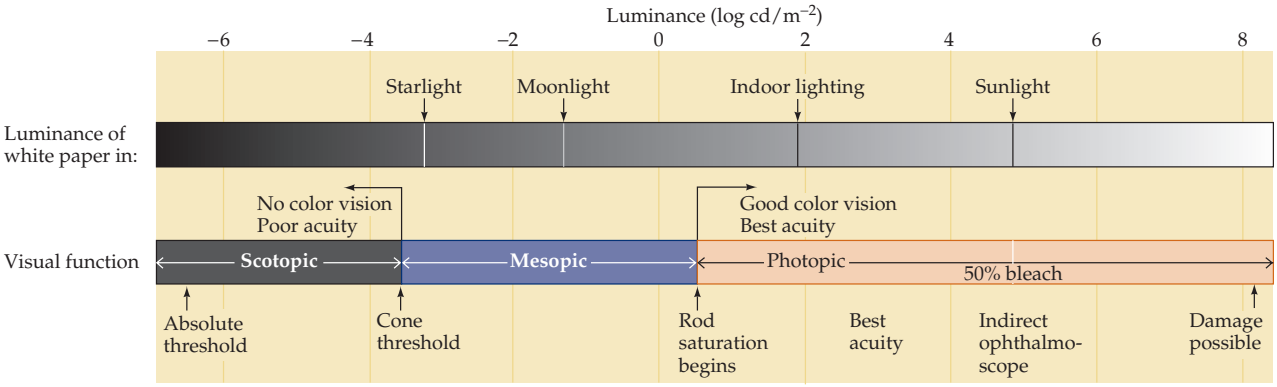


Figure 10.9 The range of luminance values over which the visual system operates. At the lowest levels of illumination, only rods are activated. Cones begin to contribute to perception at about the level of starlight and are the only receptors that function under relatively bright conditions.

elderly individuals suffering from macular degeneration (Box C). People who have lost cone function are legally blind, whereas those who have lost rod function only experience difficulty seeing at low levels of illumination (night blindness; see Box B).

Differences in the transduction mechanisms utilized by the two receptor types is a major factor in the ability of rods and cones to respond to different ranges of light intensity. For example, rods produce a reliable response to a single photon of light, whereas more than 100 photons are required to produce a comparable response in a cone. It is not, however, that cones fail to effectively capture photons. Rather, the change in current produced by single photon capture in cones is comparatively small and difficult to distinguish from noise. Another difference is that the response of an individual cone does not saturate at high levels of steady illumination, as does the rod response. Although both rods and cones adapt to operate over a range of luminance values, the adaptation mechanisms of the cones are more effective. This difference in adaptation is apparent in the time course of the response of rods and cones to light flashes. The response of a cone, even to a bright light flash that produces the maximum change in photoreceptor current, recovers in about 200 milliseconds, more than four times faster than rod recovery.

The arrangement of the circuits that transmit rod and cone information to retinal ganglion cells also contributes to the different characteristics of scotopic and photopic vision. In most parts of the retina, rod and cone signals converge on the same ganglion cells; i.e., individual ganglion cells respond to both rod and cone inputs, depending on the level of illumination. The early stages of the pathways that link rods and cones to ganglion cells, however, are largely independent. For example, the pathway from rods to ganglion cells involves a distinct class of bipolar cell (called rod bipolar) that, unlike cone bipolar cells, does not contact retinal ganglion cells. Instead, rod bipolar cells synapse with the dendritic processes of a specific class of amacrine cell that makes gap junctions and chemical synapses with the terminals of cone bipolars; these processes, in turn, make synaptic contacts on the dendrites of ganglion cells in the inner plexiform layer. As a consequence, the circuits linking the rods and cones to retinal ganglion cells differ dramatically in their degree of convergence. Each rod bipolar cell is contacted by a number of rods, and many rod bipolar cells contact a given amacrine cell. In contrast, the cone system is much less convergent. Thus, each retinal ganglion cell that dominates central vision (called midget gan-

Box C
Macular Degeneration

An estimated six million people in the United States suffer from a condition known as **age-related macular degeneration** (**AMD**), which causes a progressive loss of central vision. Since central vision is critical for sight, diseases that affect the macula (see Figure 11.1) severely limit the ability to perform visual tasks. Indeed, AMD is the most common cause of vision loss in people over age 55, and its incidence is rising with the increasing percentage of elderly individuals in the population.

The underlying problem, which remains poorly understood, is degeneration of the photoreceptors. Usually, patients first notice a blurring of central vision when performing tasks such as reading. Images may also appear distorted. A graph paper-like chart known as the Amsler grid is used as a simple test for early signs of AMD. By focusing on a marked spot in the middle of the grid, the patient can assess whether the parallel and perpendicular lines on the grid appear blurred or distorted. Blurred central vision often progresses to having blind spots within central vision, and in most cases both eyes are eventually involved.

Although the risk of developing AMD clearly increases with age, the causes of the disease are not known. Various studies have implicated hereditary factors, cardiovascular disease, environmental factors such as smoking and light exposure, and nutritional causes. Indeed, it may be that all these contribute to the risk of developing AMD.

In descriptive terms, macular degeneration is broadly divided into two types. In the *exudative-neovascular form*, or "wet" AMD, which accounts for 10% of all cases, abnormal blood vessel growth occurs under the macula. These blood vessels leak fluid and blood into the retina and cause damage to the photoreceptors. Wet AMD tends to progress rapidly and can cause severe damage; rapid loss of central vision may occur over just a few months. The treatment for this form of the disease is laser therapy. By transferring thermal energy, the laser beam destroys the leaky blood vessels under the macula, thus slowing the rate of vision loss. A disadvantage of this approach is that the high thermal energy delivered by the beam also destroys nearby healthy tissue. An improvement in the laser treatment of AMD involves a light-activated drug to target abnormal blood vessels. Once the drug is administered, relatively low energy laser pulses aimed at the abnormal blood vessels are delivered to stimulate the drug, which in turn destroys the abnormal blood vessels with minimal damage to the surrounding tissue.

The remaining 90% of AMD cases are the nonexudative, or "dry" form. In these patients there is a gradual disappearance of the retinal pigment epithelium, resulting in circumscribed areas of atrophy. Since photoreceptor loss follows the disappearance of the pigment epithelium, the affected retinal areas have little or no visual function. Vision loss from dry AMD occurs more gradually, typically over the course of many years. These patients usually retain some central vision, although the loss can be severe enough to compromise performance of tasks that require seeing details. Unfortunately, at the present time there is no treatment for dry AMD. A radical and quite fascinating new approach that offers some promise entails surgically repositioning the retina away from the abnormal area.

Occasionally, macular degeneration occurs in much younger individuals. Many of these cases are caused by various mutations, each with its own clinical manifestations and genetic cause. The most common form of juvenile macular degeneration is known as *Stargardt disease*, which is inherited as an autosomal recessive. Patients are usually diagnosed before they reach the age of 20. Although the progression of vision loss is variable, most of these patients are legally blind by age 50. Mutations that cause Stargardt disease have been identified in the *ABCR* gene, which codes for a protein that transports retinoids across the photoreceptor membrane. Thus, the visual cycle of photopigment regeneration may be disrupted in this form of macular degeneration, presumably by dysfunctional proteins encoded by the abnormal gene. Interestingly, the *ABCR* gene is expressed only in rods, suggesting that the cones may have their own visual cycle enzymes.

References

Fine, S. L., J. W. Berger, M. G. Maguire and A. C. Ho (2000) Drug therapy: Age-related macular degeneration. NEJM 342: 483–492.

Sarks, S. H. and J. P. Sarks (2001) Age-related macular degeneration—atrophic form. In *Retina,* 3rd Ed., Vol. 2: *Medical Retina*. S. J. Ryan (ed.-in-chief). St. Louis, MO: Mosby Year Book, pp. 1071–1102.

Elman, M. J. and S. L. Fine (2001) Exudative age-related macular degeneration. In *Retina,* 3rd Ed., Vol. 2: *Medical Retina*. S. J. Ryan (ed.-in-chief). St. Louis, MO: Mosby Year Book, pp. 1103–114.

Deutman, A. F. (2001) Macular dystrophies. In *Retina,* 3rd Ed., Vol. 2: *Medical Retina*. S. J. Ryan (ed.-in-chief). St. Louis, MO: Mosby Year Book, pp. 1186–1240.

The Foundation Fighting Blindness of Hunt Valley, MD, maintains a web site that provides updated information about many forms of retinal degeneration: www.blindness.org

RetNet provides updated information, including references to original articles, on genes and mutations associated with retinal diseases: www.sph.uth.tmc.edu/RetNet

glion cells) receives input from only one cone bipolar cell, which, in turn, is contacted by a single cone. Convergence makes the rod system a better detector of light, because small signals from many rods are pooled to generate a large response in the bipolar cell. At the same time, convergence reduces the spatial resolution of the rod system, since the source of a signal in a rod bipolar cell or retinal ganglion cell could have come from anywhere within a relatively large area of the retinal surface. The one-to-one relationship of cones to bipolar and ganglion cells is, of course, just what is required to maximize acuity.

Anatomical Distribution of Rods and Cones

The distribution of rods and cones across the surface of the retina also has important consequences for vision (Figure 10.10). Despite the fact that perception in typical daytime light levels is dominated by cone-mediated vision, the total number of rods in the human retina (about 90 million) far exceeds the number of cones (roughly 4.5 million). As a result, the density of rods is much greater than cones throughout most of the retina. However, this relationship changes dramatically in the **fovea**, a highly specialized region of the central retina that measures about 1.2 millimeters in diameter (Figure 10.11). In the fovea, cone density increases almost 200-fold, reaching, at its center, the highest receptor packing density anywhere in the retina. This high density is achieved by decreasing the diameter of the cone outer segments such that foveal cones resemble rods in their appearance. The increased density of cones in the fovea is accompanied by a sharp decline in the density of rods. In fact, the central 300 μm of the fovea, called the **foveola**, is totally rod-free.

The extremely high density of cone receptors in the fovea, and the one-to-one relationship with bipolar cells and retinal ganglion cells (see earlier), endows this component of the cone system with the capacity to mediate high visual acuity. As cone density declines with eccentricity and the degree of convergence onto retinal ganglion cells increases, acuity is markedly reduced. Just 6° eccentric to the line of sight, acuity is reduced by 75%, a fact

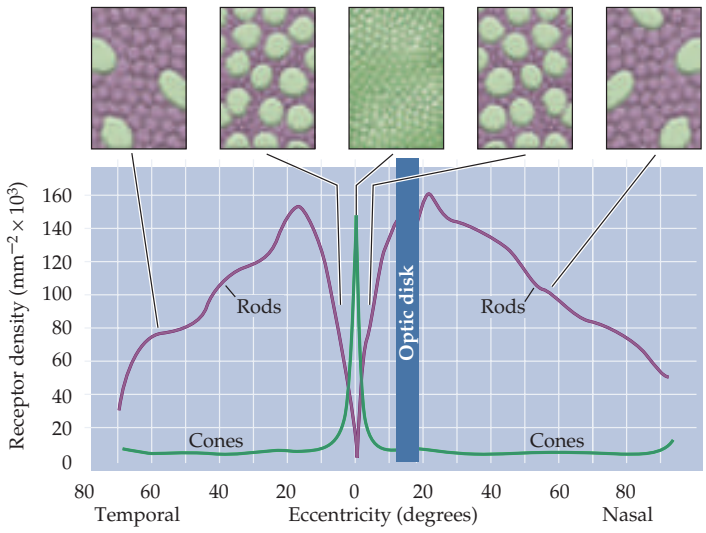


Figure 10.10 Distribution of rods and cones in the human retina. Graph illustrates that cones are present at a low density throughout the retina, with a sharp peak in the center of the fovea. Conversely, rods are present at high density throughout most of the retina, with a sharp decline in the fovea. Boxes at top illustrate the appearance of face on sections through the outer segments of the photoreceptors at different eccentricities. The increased density of cones in the fovea is accompanied by a striking reduction in the diameter of their outer segments.

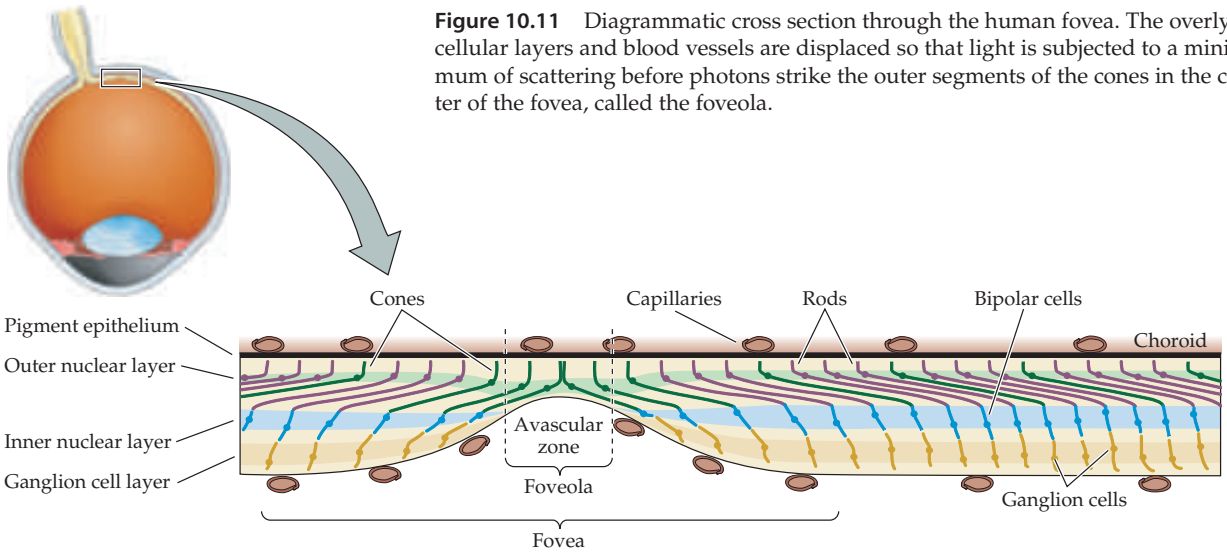


Figure 10.11 Diagrammatic cross section through the human fovea. The overlying cellular layers and blood vessels are displaced so that light is subjected to a minimum of scattering before photons strike the outer segments of the cones in the center of the fovea, called the foveola.

that can be readily appreciated by trying to read the words on any line of this page beyond the word being fixated on. The restriction of highest acuity vision to such a small region of the retina is the main reason humans spend so much time moving their eyes (and heads) around—in effect directing the foveas of the two eyes to objects of interest (see Chapter 19). It is also the reason why disorders that affect the functioning of the fovea have such devastating effects on sight (see Box C). Conversely, the exclusion of rods from the fovea, and their presence in high density away from the fovea, explain why the threshold for detecting a light stimulus is lower outside the region of central vision. It is easier to see a dim object (such as a faint star) by looking slightly away from it, so that the stimulus falls on the region of the retina that is richest in rods (see Figure 10.10).

Another anatomical feature of the fovea (which literally means "pit") that contributes to the superior acuity of the cone system is that the layers of cell bodies and processes that overlie the photoreceptors in other areas of the retina are displaced around the fovea, and especially the foveola (see Figure 10.11). As a result, photons are subjected to a minimum of scattering before they strike the photoreceptors. Finally, another potential source of optical distortion that lies in the light path to the receptors—the retinal blood vessels—are diverted away from the foveola. This central region of the fovea is therefore dependent on the underlying choroid and pigment epithelium for oxygenation and metabolic sustenance.

Cones and Color Vision

A special property of the cone system is color vision. Perceiving color allows humans (and many other animals) to discriminate objects on the basis of the distribution of the wavelengths of light that they reflect to the eye. While differences in luminance (i.e., overall light intensity) are often sufficient to distinguish objects, color adds another perceptual dimension that is especially useful when differences in luminance are subtle or nonexistent. Color obviously gives us a quite different way of perceiving and describing the world we live in.

Unlike rods, which contain a single photopigment, there are three types of cones that differ in the photopigment they contain. Each of these photopigments has a different sensitivity to light of different wavelengths, and for this reason are referred to as "blue," "green," and "red" or, more appropriately, short (S), medium (M), and long (L) wavelength cones—terms that more or less describe their spectral sensitivities (Figure 10.12). This nomenclature implies that individual cones provide color information for the wavelength of light that excites them best. In fact, individual cones, like rods, are entirely color blind in that their response is simply a reflection of the number of photons they capture, regardless of the wavelength of the photon (or, more properly, its vibrational energy). It is impossible, therefore, to determine whether the change in the membrane potential of a particular cone has arisen from exposure to many photons at wavelengths to which the receptor is relatively insensitive, or fewer photons at wavelengths to which it is most sensitive. This ambiguity can only be resolved by *comparing* the activity in different classes of cones. Based on the responses of individual ganglion cells, and cells at higher levels in the visual pathway (see Chapter 11), comparisons of this type are clearly involved in how the visual system extracts color information from spectral stimuli. Despite these insights, a full understanding of the neural mechanisms that underlie color perception has been elusive (Box D).

Much additional information about color vision has come from studies of individuals with abnormal color detecting abilities. Color vision deficiencies result either from the inherited failure to make one or more of the cone pigments or from an alteration in the absorption spectra of cone pigments (or, rarely, from lesions in the central stations that process color information; see Chapter 11). Under normal conditions, most people can match any color in a test stimulus by adjusting the intensity of three superimposed light sources generating long, medium, and short wavelengths. The fact that only three such sources are needed to match (nearly) all the perceived colors is strong

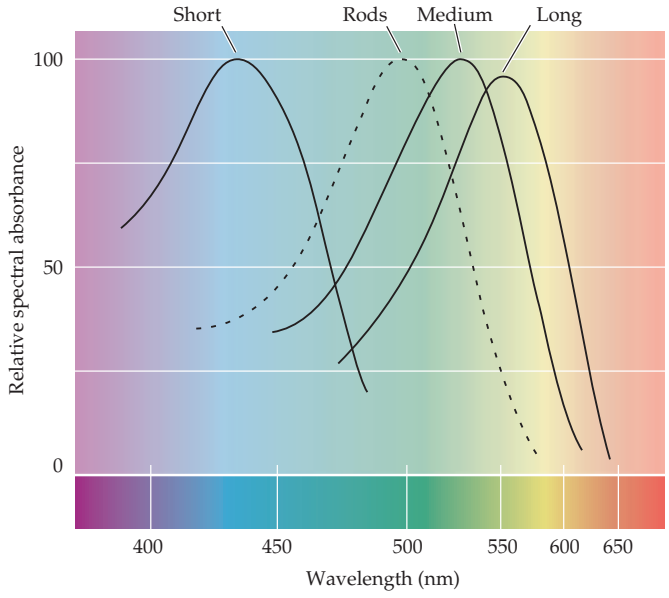


Figure 10.12 Color vision. The light absorption spectra of the four photopigments in the normal human retina. (Recall that light is defined as electromagnetic radiation having wavelengths between ~400 and 700 nm.) The solid curves indicate the three kinds of cone opsins; the dashed curve shows rod rhodopsin for comparison. Absorbance is defined as the log value of the intensity of incident light divided by intensity of transmitted light.

Box D
The Importance of Context in Color Perception

Seeing color logically demands that retinal responses to different wavelengths in some way be *compared*. The discovery of the three human cone types and their different absorption spectra is correctly regarded, therefore, as the basis for human color vision. Nevertheless, how these human cone types and the higher-order neurons they contact (see Chapter 11) produce the sensations of color is still unclear. Indeed, this issue has been debated by some of the greatest minds in science (Hering, Helmholtz, Maxwell, Schroedinger, and Mach, to name only a few) since Thomas Young first proposed that humans must have three different receptive "particles"—i.e., the three cone types.

A fundamental problem has been that, although the relative activities of three cone types can more or less explain the colors perceived in color-matching experiments performed in the laboratory, the perception of color is strongly influenced by context. For example, a patch returning the exact same spectrum of wavelengths to the eye can appear quite different depending on its surround, a phenomenon called *color contrast* (Figure A). Moreover, test patches returning different spectra to the eye can appear to be the same color, an effect called *color constancy* (Figure B). Although these phenomena were well known in the nineteenth century, they were not accorded a central place in color vision theory until Edwin Land's work in the 1950s. In his most famous demonstration, Land (who among other achievements founded the Polaroid company and became a billionaire) used a collage of colored papers that have been referred to as "the Land Mondrians" because of their similarity to the work of the Dutch artist Piet Mondrian.

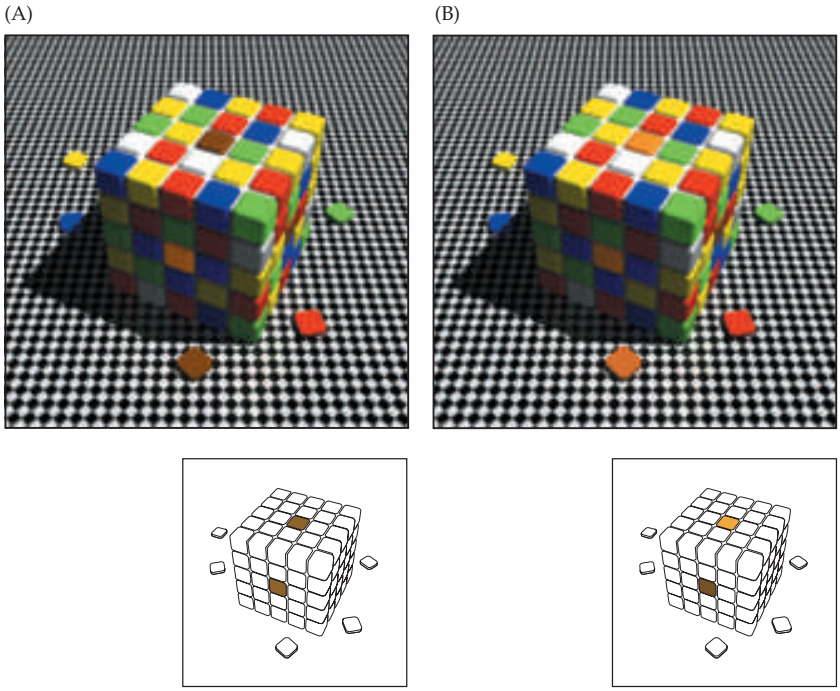
Using a telemetric photometer and three adjustable illuminators generating short, middle, and long wavelength light, Land showed that two patches that in white light appeared quite different in color (e. g., green and brown) continued to look their respective colors even when the three illuminators were adjusted so that the light being returned from the "green" surfaces produced exactly the same readings on the three telephotometers as had previously come from the "brown" surface—a striking demonstration of color constancy.

The phenomena of color contrast and color constancy have led to a heated modern debate about how color percepts are generated that now spans several decades. For Land, the answer lay in a series of ratiometric equations that could integrate the spectral returns of different regions over the entire scene. It was recognized even before Land's death in 1991, however, that his so-called retinex theory did not work in all circumstances and was in any event a description rather than an explanation. An alternative explanation of these contextual aspects of color vision is that color, like brightness, is generated empirically according to what spectral stimuli have typically signified in past experience (see Box E).

References

LAND, E. (1986) Recent advances in Retinex theory. Vis. Res. 26: 7–21.

PURVES, D. AND R. B. LOTTO (2003) *Why We See What We Do: An Empirical Theory of Vision*, Chapters 5 and 6. Sunderland MA: Sinauer Associates, pp. 89–138.



The genesis of contrast and constancy effects by exactly the same context. The two panels demonstrate the effects on apparent color when two *similarly* reflective target surfaces (A) or two *differently* reflective target surfaces (B) are presented in the *same* context in which all the information provided is consistent with illumination that differs only in intensity. The appearances of the relevant target surfaces in a neutral context are shown in the insets below. (From Purves and Lotto, 2003)

confirmation of the fact that color sensation is based on the relative levels of activity in three sets of cones with different absorption spectra. That color vision is **trichromatic** was first recognized by Thomas Young at the beginning of the nineteenth century (thus, people with normal color vision are called *trichromats*). For about 5–6% of the male population in the United States and a much smaller percentage of the female population, however, color vision is more limited. Only two bandwidths of light are needed to match all the colors that these individuals can perceive; the third color category is simply not seen. Such **dichromacy**, or "color blindness" as it is commonly called, is inherited as a recessive, sex-linked characteristic and exists in two forms: *protanopia*, in which all color matches can be achieved by using only green and blue light, and *deuteranopia*, in which all matches can be achieved by using only blue and red light. In another major class of color deficiencies, all three light sources (i.e., short, medium, and long wavelengths) are needed to make all possible color matches, but the matches are made using values that are significantly different from those used by most individuals. Some of these *anomalous trichromats* require more red than normal to match other colors (protanomalous trichromats); others require more green than normal (deuteranomalous trichromats).

Jeremy Nathans and his colleagues at Johns Hopkins University have provided a deeper understanding of these color vision deficiencies by identifying and sequencing the genes that encode the three human cone pigments (Figure 10.13). The genes that encode the red and green pigments show a high degree of sequence homology and lie adjacent to each other on the X chromosome, thus explaining the prevalence of color blindness in males. In contrast, the blue-sensitive pigment gene is found on chromosome 7 and is quite different in its amino acid sequence. These facts suggest that the red and green pigment genes evolved relatively recently, perhaps as a result of the duplication of a single ancestral gene; they also explain why most color vision abnormalities involve the red and green cone pigments.

Human dichromats lack one of the three cone pigments, either because the corresponding gene is missing or because it exists as a hybrid of the red and green pigment genes (see Figure 10.13). For example, some dichromats lack the green pigment gene altogether, while others have a hybrid gene that is thought to produce a red-like pigment in the "green" cones. Anomalous trichromats also possess hybrid genes, but these genes elaborate pigments

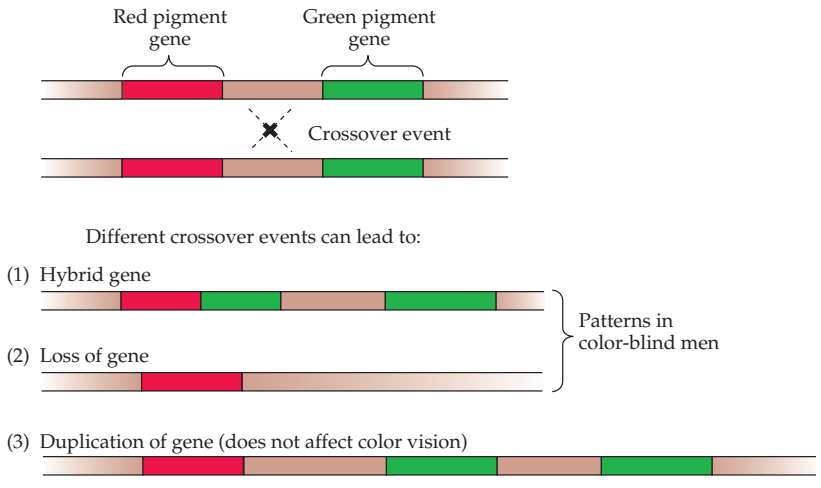


Figure 10.13 Many deficiencies of color vision are the result of genetic alterations in the red or green cone pigments due to the crossing over of chromosomes during meiosis. This recombination can lead to the loss of a gene, the duplication of a gene, or the formation of a hybrid with characteristics distinct from those of normal genes.

whose spectral properties lie between those of the normal red and green pigments. Thus, although most anomalous trichromats have distinct sets of medium and long-wavelength cones, there is more overlap in their absorption spectra than in normal trichromats, and thus less difference in how the two sets of cones respond to a given wavelength (with resulting anomalies in color perception).

Retinal Circuits for Detecting Luminance Change

Despite the esthetically pleasing nature of color vision, most of the information in visual scenes consists of spatial variations in light intensity (a black and white movie, for example, has most of the information a color version has, although it is deficient in some respects and usually is less fun to watch). How the spatial patterns of light and dark that fall on the photoreceptors are deciphered by central targets has been a vexing problem (Box E). To understand what is accomplished by the complex neural circuits within the retina during this process, it is useful to start by considering the responses of individual retinal ganglion cells to small spots of light. Stephen Kuffler, working at Johns Hopkins University in the 1950s, pioneered this approach by characterizing the responses of single ganglion cells in the cat retina. He found that each ganglion cell responds to stimulation of a small circular patch of the retina, which defines the cell's receptive field (see Chapter 8 for discussion of receptive fields). Based on these responses, Kuffler distinguished two classes of ganglion cells, "on"-center and "off"-center (Figure 10.14).

Turning on a spot of light in the receptive field center of an **on-center ganglion cell** produces a burst of action potentials. The same stimulus applied to the receptive field center of an **off-center ganglion cell** reduces the rate of

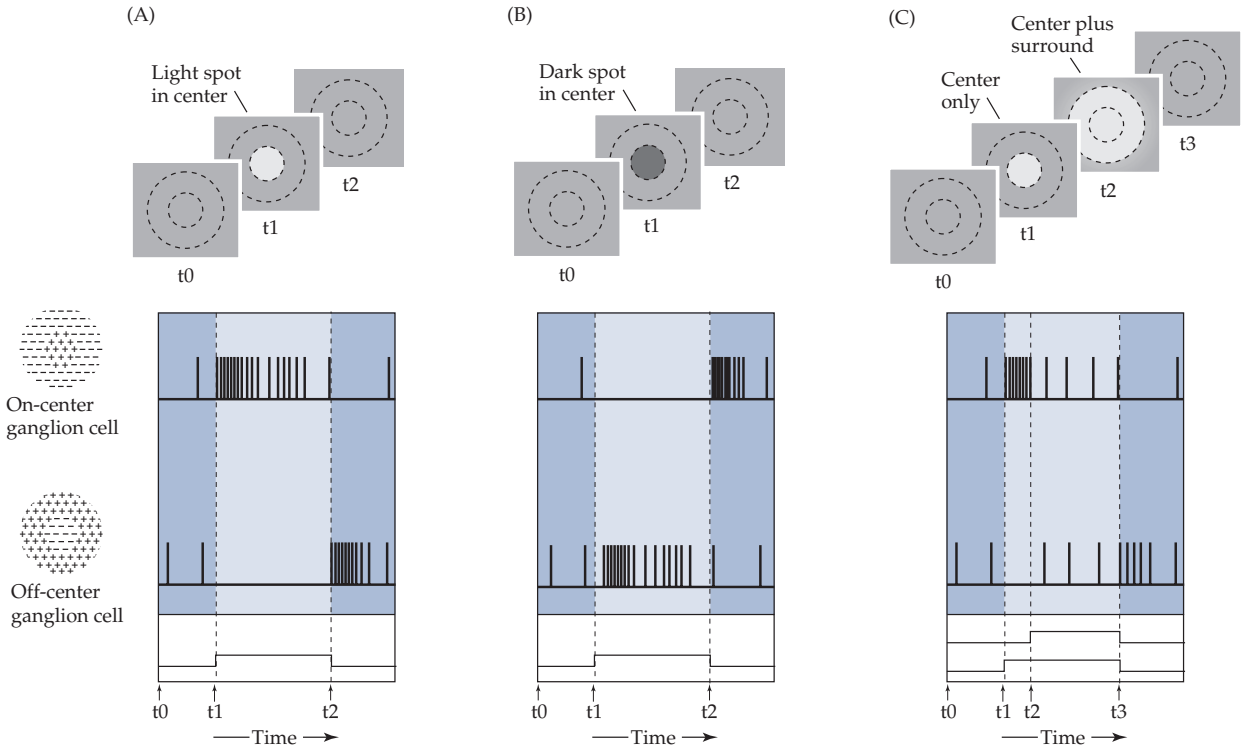


Figure 10.14 The responses of on-center and off-center retinal ganglion cells to stimulation of different regions of their receptive fields. Upper panels indicate the time sequence of stimulus changes. (A) Effects of light spot in the receptive field center. (B) Effects of dark spot in the receptive field center. (C) Effects of light spot in the center followed by the addition of light in the surround.

Box E

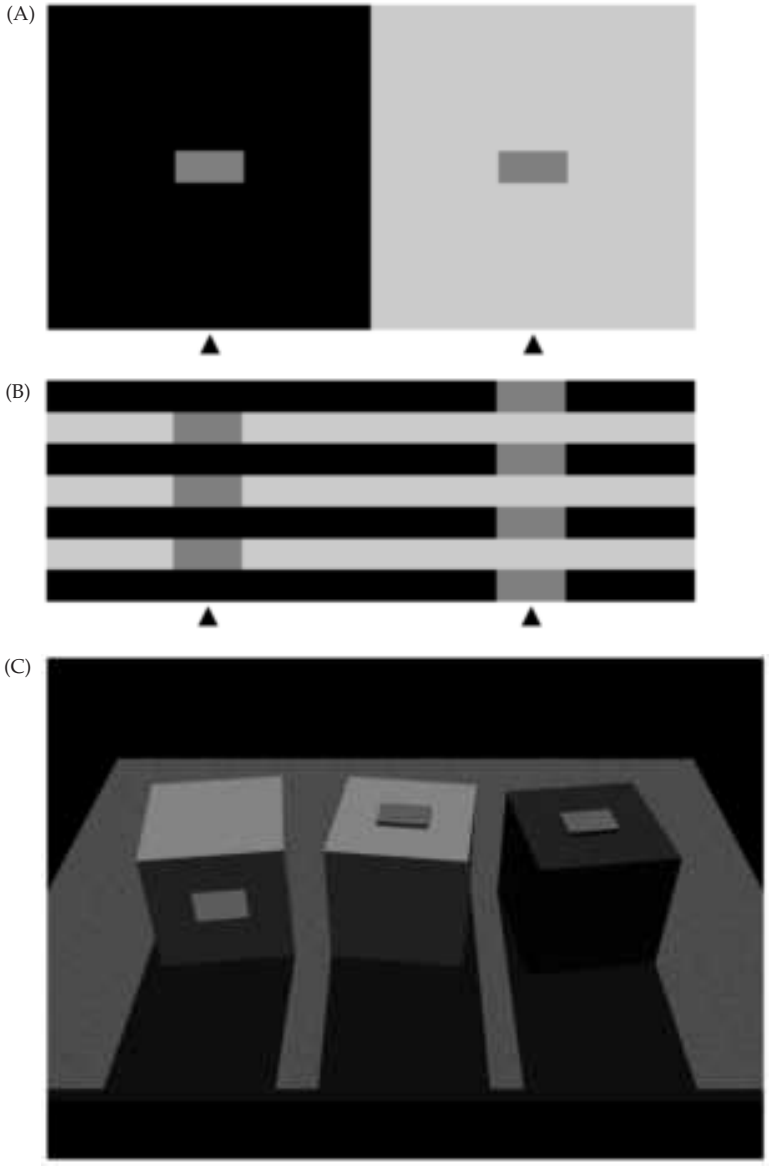
The Perception of Light Intensity

Understanding the link between retinal stimulation and what we see (perception) is arguably the central problem in vision, and the relation of luminance (a physical measurement of light intensity) and brightness (the sensation elicited by light intensity) is probably the simplest place to consider this challenge.

As indicated in the text, how we see the brightness differences (i.e., contrast) between adjacent territories with distinct luminances depends in the first instance on the relative firing rate of retinal ganglion cells, modified by lateral interactions. However, there is a problem with the assumption that the central nervous system simply "reads out" these relative rates of ganglion cell activity to sense brightness. The difficulty, as in perceiving color, is that the brightness of a given target is markedly affected by its context in ways that are difficult or impossible to explain in terms of the retinal output as such. The accompanying figures, which illustrate two simultaneous brightness contrast illusions, help make this point. In Figure A, two photometrically identical (equiluminant) gray squares appear differently bright as a function of the background in which they are presented.

A conventional interpretation of this phenomenon is that the receptive field properties illustrated in Figures 10.14 through 10.17 cause ganglion cells to fire differently depending on whether the surround of the equiluminant target is dark or light. The demonstration in Figure B, however, undermines this explanation, since in this case the target surrounded by more dark area actually looks *darker* than the same target surrounded by more light area.

An alternative interpretation of luminance perception that can account for these puzzling phenomena is that brightness percepts are generated on a statistical basis as a means of contending with the inherent ambiguity of luminance (i.e., the fact that a given value of luminance can be generated by many different combinations of illumination and surface reflectance properties). Since to be successful an observer has to respond to the real-world sources of luminance and not to light intensity as such, this ambiguity of the retinal stimulus presents a quandary. A plausible solution to



(A) Standard illusion of simultaneous brightness contrast. (B) Another illusion of simultaneous brightness contrast that is difficult to explain in conventional terms. (C) Cartoons of some possible sources of the standard simultaneous brightness contrast illusion in (A). (Courtesy of R. Beau Lotto and Dale Purves.)

the inherent uncertainty of the relationship between luminance values and their actual sources would be to generate the sensation of brightness elicited by a given luminance (e.g., in the brightness of the identical test patches in the figure) on the basis of what the luminance of the test patches had typically turned out to be in the past experience of human observers. To get the gist of this explanation consider Figure C, which illustrates the point that the two equiluminant target patches in Figure A could have been generated by two differently painted surfaces in different illuminants, as in a comparison of the target patches on the left and middle cubes, or two similarly reflecting surfaces in similar amounts of light, as in a comparison of the target patches on the middle and right cubes. An expedient—and perhaps the only—way the visual system can cope with this ambiguity is to generate the perception of the stimulus in Figure A (and in Figure B) empirically, i.e., based on what the target patches typically turned out to signify in the past. Since the equiluminant targets will have arisen from a variety of possible sources, it makes sense to have the brightness elicited by the patches determined statistically by the relative frequency of occurrence of that luminance in the particular context in which it is presented. The advantage of seeing luminance according to the relative probabilities of the possible sources of the stimulus is that percepts generated in this way give the observer the best chance of making appropriate behavioral responses to profoundly ambiguous stimuli.

References

ADELSON, E. H. (1999) Light perception and lightness illusions. In *The Cognitive Neurosciences*, 2nd Ed. M. Gazzaniga (ed.). Cambridge, MA: MIT Press, pp. 339–351.

PURVES, D. AND R. B. LOTTO (2003) *Why We See What We Do: An Empirical Theory of Vision*, Chapters 3 and 4. Sunderland MA: Sinauer Associates, pp. 41–87.

discharge, and when the spot of light is turned off, the cell responds with a burst of action potentials (Figure 10.14A). Complementary patterns of activity are found for each cell type when a dark spot is placed in the receptive field center (Figure 10.14B). Thus, on-center cells increase their discharge rate to luminance *increments* in the receptive field center, whereas off-center cells increase their discharge rate to luminance *decrements* in the receptive field center.

On- and off-center ganglion cells are present in roughly equal numbers. The receptive fields have overlapping distributions, so that every point on the retinal surface (that is, every part of visual space) is analyzed by several on-center and several off-center ganglion cells. A rationale for having these two distinct types of retinal ganglion cells was suggested by Peter Schiller and his colleagues at the Massachusetts Institute of Technology, who examined the effects of pharmacologically inactivating on-center ganglion cells on a monkey's ability to detect a variety of visual stimuli. After silencing on-center ganglion cells, the animals showed a deficit in their ability to detect stimuli that were brighter than the background; however, they could still see objects that were darker than the background.

These observations imply that information about increases or decreases in luminance is carried separately to the brain by the axons of these two different types of retinal ganglion cells. Having separate luminance "channels" means that changes in light intensity, whether increases or decreases, are always conveyed to the brain by an increased number of action potentials. Because ganglion cells rapidly adapt to changes in luminance, their "resting" discharge rate in constant illumination is relatively low. Although an increase in discharge rate above resting level serves as a reliable signal, a decrease in firing rate from an initially low rate of discharge might not. Thus, having luminance changes signaled by two classes of adaptable cells provides unambiguous information about both luminance increments and decrements.

The functional differences between these two ganglion cell types can be understood in terms of both their anatomy and their physiological proper-

ties and relationships (Figure 10.15). On- and off-center ganglion cells have dendrites that arborize in separate strata of the inner plexiform layer, forming synapses selectively with the terminals of on- and off-center bipolar cells that respond to luminance increases and decreases, respectively (Figure 10.15A). As mentioned previously, the principal difference between ganglion cells and bipolar cells lies in the nature of their electrical response. Like most other cells in the retina, bipolar cells have graded potentials rather than action potentials. Graded depolarization of bipolar cells leads to an increase in transmitter release (glutamate) at their synapses and consequent depolarization of the on-center ganglion cells that they contact via AMPA, kainate, and NMDA receptors.

The selective response of on- and off-center bipolar cells to light increments and decrements is explained by the fact that they express different types of glutamate receptors (Figure 10.15A). Off-center bipolar cells have ionotropic receptors (AMPA and kainate) that cause the cells to depolarize in response to glutamate released from photoreceptor terminals. In contrast, on-center bipolar cells express a G-protein-coupled metabotropic glutamate receptor (mGluR6). When bound to glutamate, these receptors activate an intracellular cascade that closes cGMP-gated Na^+ channels, reducing inward

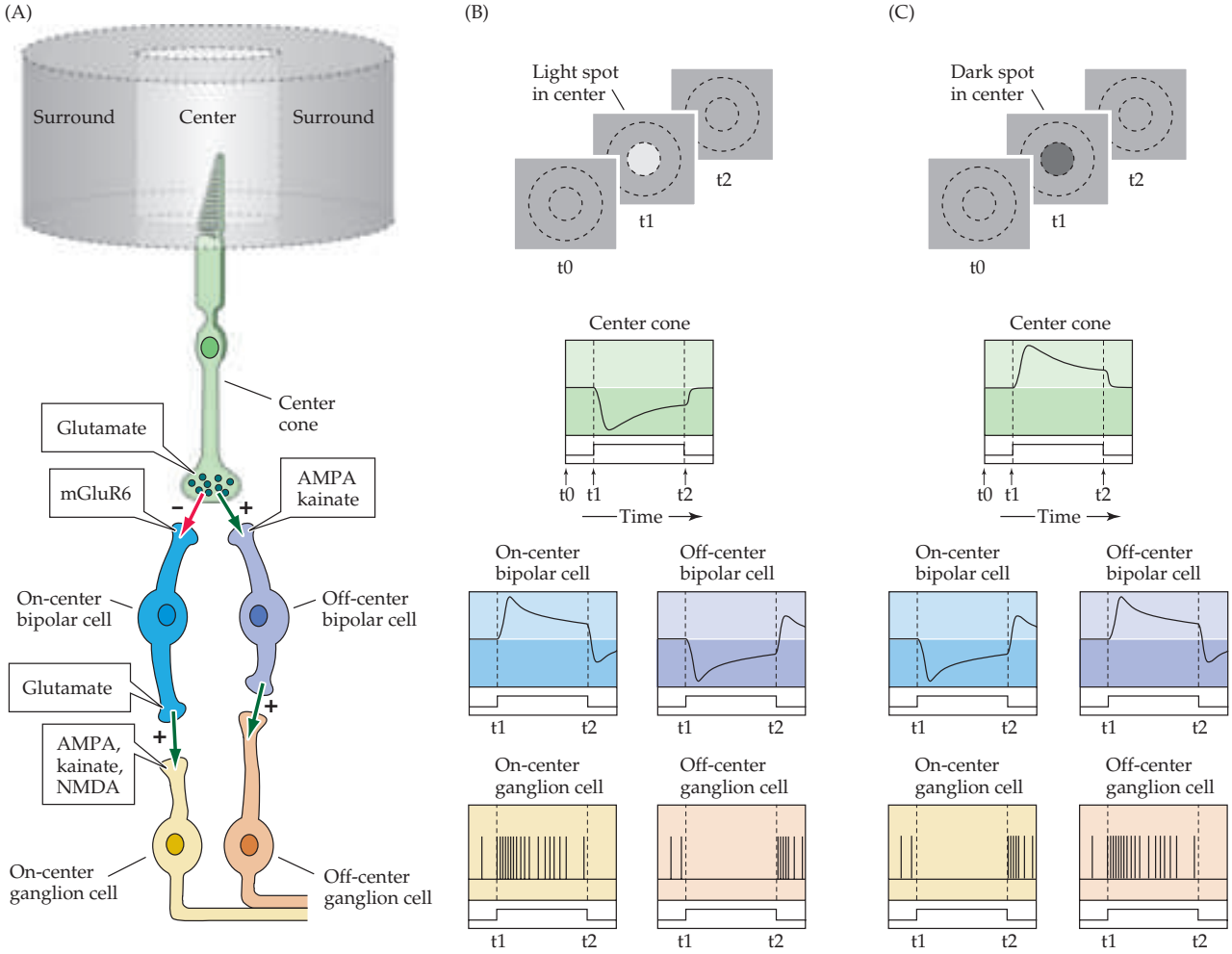


Figure 10.15 Circuitry responsible for generating receptive field center responses of retinal ganglion cells. (A) Functional anatomy of cone inputs to the center of a ganglion cell receptive field. A plus indicates a sign-conserving synapse; a minus represents a sign-inverting synapse. (B) Responses of various cell types to the presentation of a light spot in the center of the ganglion cell receptive field. (C) Responses of various cell types to the presentation of a dark spot in the center of the ganglion cell receptive field.

current and hyperpolarizing the cell. Thus, glutamate has opposite effects on these two classes of cells, depolarizing off-center bipolar cells and hyperpolarizing on-center cells. Photoreceptor synapses with off-center bipolar cells are called sign-conserving, since the sign of the change in membrane potential of the bipolar cell (depolarization or hyperpolarization) is the same as that in the photoreceptor (Figure 10.15B,C). Photoreceptor synapses with on-center bipolar cells are called sign-inverting because the change in the membrane potential of the bipolar cell is the opposite of that in the photoreceptor.

In order to understand the response of on- and off-center bipolar cells to changes in light intensity, recall that photoreceptors hyperpolarize in response to light increments, decreasing their release of neurotransmitter (Figure 10.15B). Under these conditions, on-center bipolar cells contacted by the photoreceptors are freed from the hyperpolarizing influence of the photoreceptor's transmitter, and they depolarize. In contrast, for off-center cells, the reduction in glutamate represents the withdrawal of a depolarizing influence, and these cells hyperpolarize. Decrements in light intensity naturally have the opposite effect on these two classes of bipolar cells, hyperpolarizing on-center cells and depolarizing off-center ones (Figure 10.15C).

Kuffler's work also called attention to the fact that retinal ganglion cells do not act as simple photodetectors. Indeed, most ganglion cells are relatively poor at signaling differences in the level of diffuse illumination. Instead, they are sensitive to *differences* between the level of illumination that falls on the receptive field center and the level of illumination that falls on the surround—that is, to **luminance contrast**. The center of a ganglion cell receptive field is surrounded by a concentric region that, when stimulated, antagonizes the response to stimulation of the receptive field center (see Figure 10.14C). For example, as a spot of light is moved from the center of the receptive field of an on-center cell toward its periphery, the response of the cell to the spot of light decreases (Figure 10.16). When the spot falls completely outside the center (that is, in the surround), the response of the cell falls below its resting level; the cell is effectively inhibited until the distance from the center is so great that the spot no longer falls on the receptive field at all, in which case the cell returns to its resting level of firing. Off-center

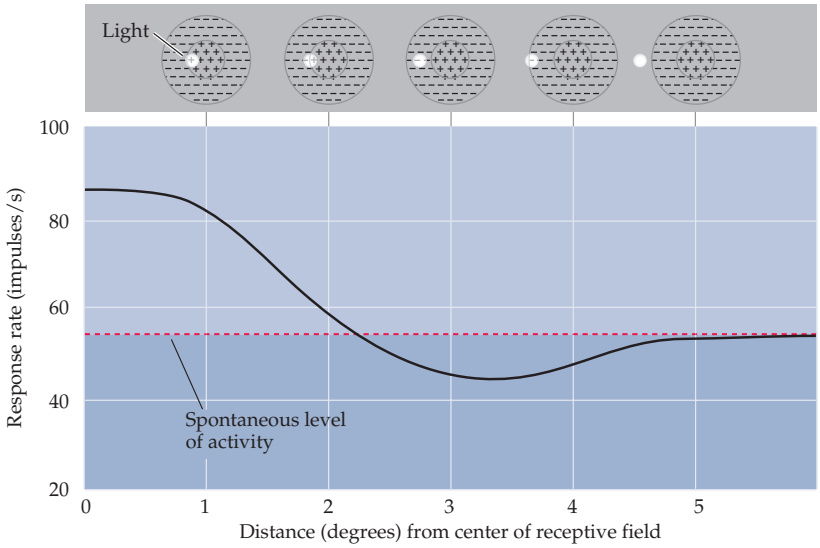


Figure 10.16 Rate of discharge of an on-center ganglion cell to a spot of light as a function of the distance of the spot from the receptive field center. Zero on the *x* axis corresponds to the center; at a distance of 5°, the spot falls outside the receptive field.

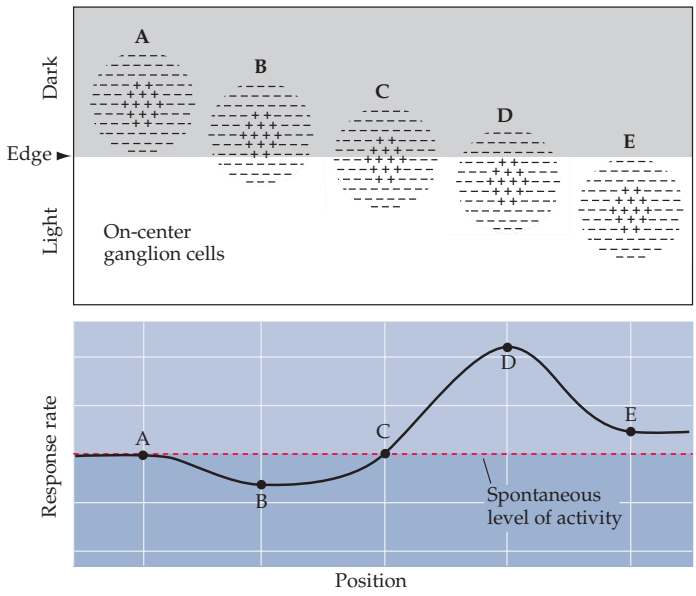


Figure 10.17 Responses of a hypothetical population of on-center ganglion cells whose receptive fields (A–E) are distributed across a light-dark edge. Those cells whose activity is most affected have receptive fields that lie along the light-dark edge.

cells exhibit a similar surround antagonism. Stimulation of the surround by light opposes the decrease in firing rate that occurs when the center is stimulated alone, and reduces the response to light decrements in the center (compare Figures 10.14A and 10.14C).

Because of their antagonistic surrounds, ganglion cells respond much more vigorously to small spots of light confined to their receptive field centers than to large spots, or to uniform illumination of the visual field (see Figure 10.14C).

To appreciate how center-surround antagonism makes the ganglion cell sensitive to luminance contrast, consider the activity levels in a hypothetical population of on-center ganglion cells whose receptive fields are distributed across a retinal image of a light-dark edge (Figure 10.17). The neurons whose firing rates are most affected by this stimulus—either increased (neuron D) or decreased (neuron B)—are those with receptive fields that lie along the light-dark border; those with receptive fields completely illuminated (or completely darkened) are less affected (neurons A and E). Thus, the information supplied by the retina to central visual stations for further processing does not give equal weight to all regions of the visual scene; rather, it emphasizes the regions where there are differences in luminance.

Contribution of Retinal Circuits to Light Adaptation

In addition to making ganglion cells especially sensitive to light-dark borders in the visual scene, center-surround mechanisms make a significant contribution to the process of **light adaptation**. As illustrated for an on-center cell in Figure 10.18, the response rate of a ganglion cell to a small spot of light turned on in its receptive field center varies as a function of the spot's intensity. In fact, response rate is proportional to the spot's intensity over a range of about one log unit. However, the intensity of spot illumination required to evoke a given discharge rate is dependent on the background level of illumination. Increases in background level of illumination are accompanied by adaptive shifts in the cell's operating range such that

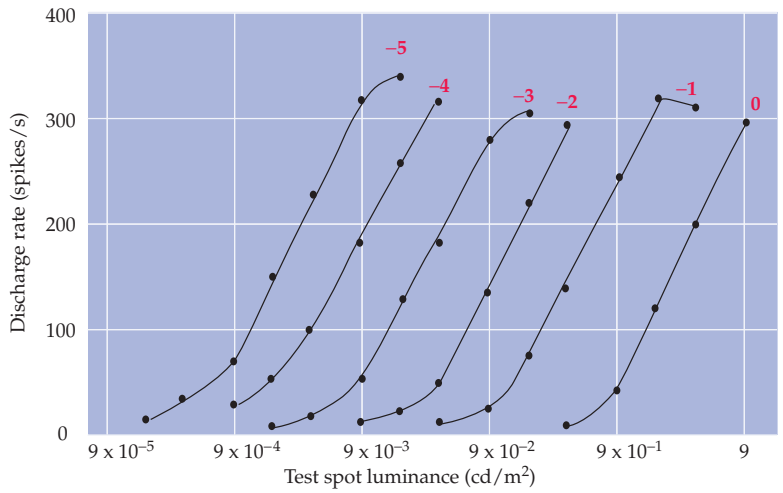


Figure 10.18 A series of curves illustrating the discharge rate of a single on-center ganglion cell to the onset of a small test spot of light in the center of its receptive field. Each curve represents the discharge rate evoked by spots of varying intensity at a constant background level of illumination, which is given by the red numbers at the top of each curve (the highest background level is 0, the lowest –5). The response rate is proportional to stimulus intensity over a range of 1 log unit, but the operating range shifts to the right as the background level of illumination increases.

greater stimulus intensities are required to achieve the same discharge rate. Thus, firing rate is not an absolute measure of light intensity, but rather signals the difference from background level of illumination.

Because the range of light intensities over which we can see is enormous compared to the narrow range of ganglion cell discharge rates (see Figure 10.9), adaptational mechanisms are essential. By scaling the ganglion cell's response to ambient levels of illumination, the entire dynamic range of a neuron's firing rate is used to encode information about intensity differences over the range of luminance values that are relevant for a given visual scene. Due to the antagonistic center-surround organization of retinal ganglion cells, the signal sent to the brain from the retina downplays the background level of illumination (see Figure 10.14). This arrangement presumably explains why the relative brightness of objects remains much the same over a wide range of lighting conditions. In bright sunlight, for example, the print on this page reflects considerably more light to the eye than it does in room light. In fact, the *print* reflects more light in sunlight than the *paper* reflects in room light; yet it continues to look black and the page white, indoors or out.

Like the mechanism responsible for generating the on- and off-center response, the antagonistic surround of ganglion cells is a product of interactions that occur at the early stages of retinal processing (Figure 10.19). Much of the antagonism is thought to arise via lateral connections established by horizontal cells and receptor terminals. Horizontal cells receive synaptic inputs from photoreceptor terminals and are linked via gap junctions with a vast network of other horizontal cells distributed over a wide area of the retinal surface. As a result, the activity in horizontal cells reflects levels of illumination over a broad area of the retina. Although the details of their actions are not entirely clear, horizontal cells are thought to exert their influence via the release of neurotransmitter directly onto photoreceptor terminals, regulating the amount of transmitter that the photoreceptors release onto bipolar cell dendrites.

Glutamate release from photoreceptor terminals has a depolarizing effect on horizontal cells (sign-conserving synapse), while the transmitter released from horizontal cells (GABA) has a hyperpolarizing influence on photoreceptor terminals (sign-inverting synapse) (Figure 10.19A). As a result, the net effect of inputs from the horizontal cell network is to oppose changes in the

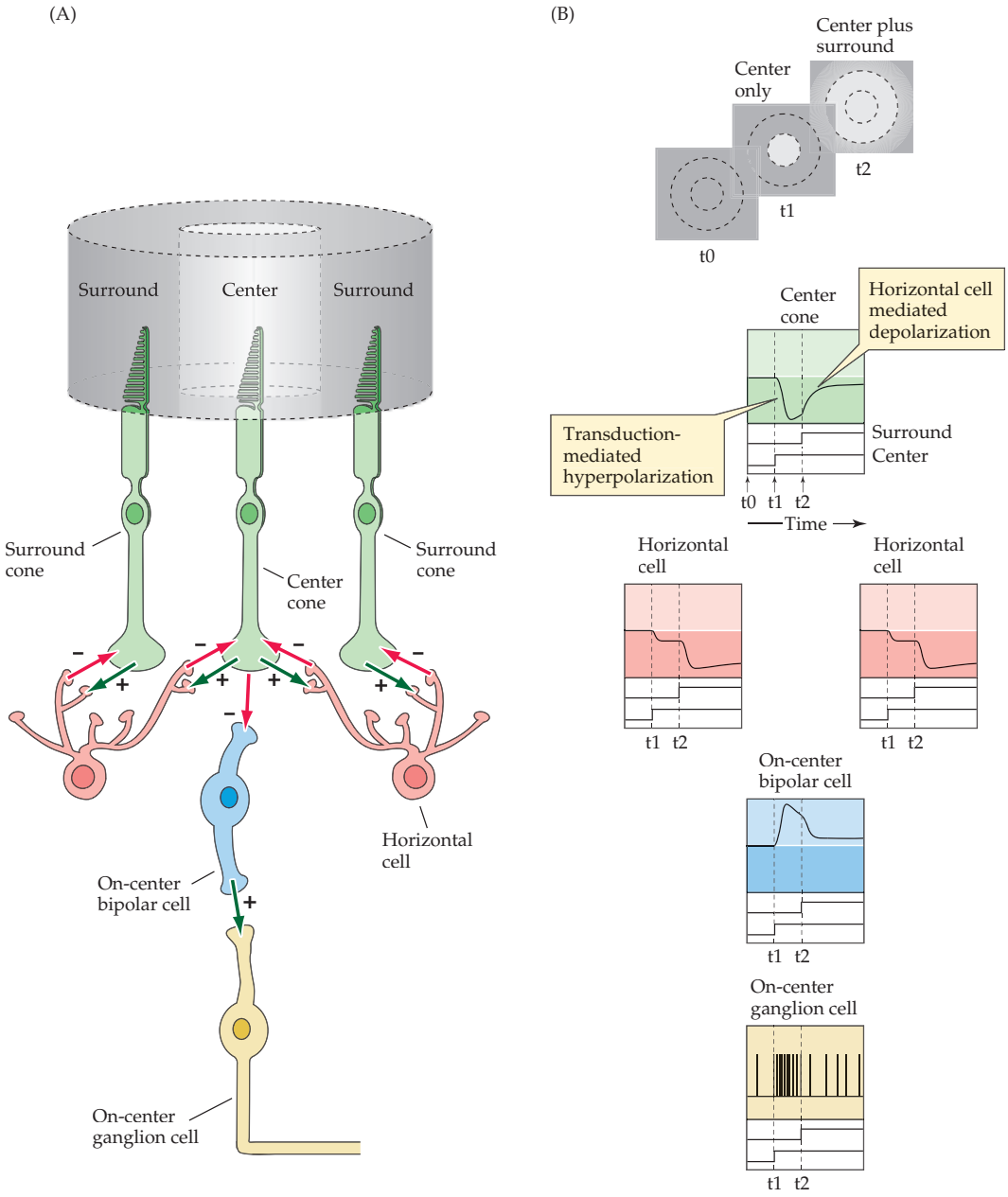


Figure 10.19 Circuitry responsible for generating the receptive field surround of an on-center retinal ganglion cell. (A) Functional anatomy of horizontal cell inputs responsible for surround antagonism. A plus indicates a sign-conserving synapse; a minus represents a sign-inverting synapse. (B) Responses of various cell types to the presentation of a light spot in the center of the receptive field (t1) followed by the addition of light stimulation in the surround (t2). Light stimulation of the surround leads to hyperpolarization of the horizontal cells and a decrease in the release of inhibitory transmitter (GABA) onto the photoreceptor terminals. The net effect is to depolarize the center cone terminal, offsetting much of the hyperpolarization induced by the transduction cascade in the center cone's outer segment.

membrane potential of the photoreceptor that are induced by phototransduction events in the outer segment. How these events lead to surround suppression in an on-center ganglion cell is illustrated in Figure 10.19. A small spot of light centered on a photoreceptor supplying input to the center of the ganglion cell's receptive field produces a strong hyperpolarizing response in the photoreceptor. Under these conditions, changes in the membrane potential of the horizontal cells that synapse with the photoreceptor terminal are relatively small, and the response of the photoreceptor to light is largely determined by its phototransduction cascade (Figure10.19B). With the addition of light to the surround, however, the impact of the horizontal network becomes significantly greater; the light-induced reduction in the release of glutamate from the photoreceptors in the surround leads to a strong hyperpolarization of the horizontal cells whose processes converge on the terminal of the photoreceptor in the receptive field center. The reduction in GABA release from the horizontal cells has a depolarizing effect on the membrane potential of the central photoreceptor, reducing the light-evoked response and ultimately reducing the firing rate of the on-center ganglion cell.

Thus, even at the earliest stages in visual processing, neural signals do not represent the absolute numbers of photons that are captured by receptors, but rather the relative intensity of stimulation—how much the current level of stimulation differs from ambient levels. While it may seem that the actions of horizontal cells decrease the sensitivity of the retina, they play a critical role in allowing the full range of the photoreceptor's electrical response (about 30 mV) to be applied to the limited range of stimulus intensities that are present at any given moment. The network mechanisms of adaptation described here function in conjunction with cellular mechanisms in the receptor outer segments that regulate the sensitivity of the phototransduction cascade at different light levels. Together, they allow retinal circuits to convey the most salient aspects of luminance changes to the central stages of the visual system described in the following chapter.

Summary

The light that falls on photoreceptors is transformed by retinal circuitry into a pattern of action potentials that ganglion cell axons convey to the visual centers in the brain. This process begins with phototransduction, a biochemical cascade that ultimately regulates the opening and closing of ion channels in the membrane of the photoreceptor's outer segment, and thereby the amount of neurotransmitter the photoreceptor releases. Two systems of photoreceptors—rods and cones—allow the visual system to meet the conflicting demands of sensitivity and acuity, respectively. Retinal ganglion cells operate quite differently from the photoreceptor cells. The center-surround arrangement of ganglion cell receptive fields makes these neurons particularly sensitive to luminance contrast and relatively insensitive to the overall level of illumination. It also allows the retina to adapt, such that it can respond effectively over the enormous range of illuminant intensities in the world. The underlying organization is generated by the synaptic interactions between photoreceptors, horizontal cells, and bipolar cells in the outer plexiform layer. As a result, the signal sent to the visual centers in the brain is already highly processed when it leaves the retina, emphasizing those aspects of the visual scene that convey the most information.

Additional Reading

Reviews

ARSHAVSKY, V. Y., T. D. LAMB AND E. N. PUGH JR. (2002) G proteins and phototransduction. Annu. Rev. Physiol. 64: 153–187.

BURNS, M. E. AND D. A. BAYLOR (2001) Activation, deactivation, and adaptation in vertebrate photoreceptor cells. Annu. Rev. Neurosci. 24: 779–805.

NATHANS, J. (1987) Molecular biology of visual pigments. Annu. Rev. Neurosci. 10: 163–194.

SCHNAPF, J. L. AND D. A. BAYLOR (1987) How photoreceptor cells respond to light. Sci. Amer. 256 (April): 40–47.

STERLING, P. (1990) Retina. In *The Synaptic Organization of the Brain*, G. M. Shepherd (ed.). New York: Oxford University Press, pp. 170–213.

STRYER, L. (1986) Cyclic GMP cascade of vision. Annu. Rev. Neurosci. 9: 87–119.

Important Original Papers

BAYLOR, D. A., M. G. F. FUORTES AND P. M. O'BRYAN (1971) Receptive fields of cones in the retina of the turtle. J. Physiol. (Lond.) 214: 265–294.

DOWLING, J. E. AND F. S. WERBLIN (1969) Organization of the retina of the mud puppy, *Necturus maculosus*. I. Synaptic structure. J. Neurophysiol. 32: 315–338.

ENROTH-CUGELL, C. AND R. M. SHAPLEY (1973) Adaptation and dynamics of cat retinal ganglion cells. J. Physiol. 233: 271–309.

FASENKO, E. E., S. S. KOLESNIKOV AND A. L. LYUBARSKY (1985) Induction by cyclic GMP of cationic conductance in plasma membrane of retinal rod outer segment. Nature 313: 310–313.

KUFFLER, S. W. (1953) Discharge patterns and functional organization of mammalian retina. J. Neurophysiol. 16: 37–68.

NATHANS, J., D. THOMAS AND D. S. HOGNESS (1986) Molecular genetics of human color vision: The genes encoding blue, green and red pigments. Science 232: 193–202.

NATHANS, J., T. P. PIANTANIDA, R. EDDY, T. B. SHOWS AND D. S. HOGNESS (1986) Molecular genetics of inherited variation in human color vision. Science 232: 203–210.

SCHILLER, P. H., J. H. SANDELL AND J. H. R. MAUNSELL (1986) Functions of the "on" and "off" channels of the visual system. Nature 322: 824–825.

WERBLIN, F. S. AND J. E. DOWLING (1969) Organization of the retina of the mud puppy, *Necturus maculosus*. II. Intracellular recording. J. Neurophysiol. 32: 339–354.

Books

BARLOW, H. B. AND J. D. MOLLON (1982) *The Senses*. London: Cambridge University Press.

DOWLING, J. E. (1987) *The Retina: An Approachable Part of the Brain*. Cambridge, MA: Belknap Press.

FAIN, G. L. (2003) *Sensory Transduction*. Sunderland, MA: Sinauer Associates.

HART, W. M. J. (ed.) (1992) *Adler's Physiology of the Eye: Clinical Application*, 9th Ed. St. Louis, MO: Mosby Year Book.

HELMHOLTZ, H. L. F. VON (1924) *Helmholtz's Treatise on Physiological Optics*, Vol. I–III. Transl. from the Third German Edition by J. P. C. Southall. Menasha, WI: George Banta Publishing Company.

HOGAN, M. J., J. A. ALVARADO AND J. E. WEDDELL (1971) *Histology of the Human Eye: An Atlas and Textbook*. Philadelphia: Saunders.

HUBEL, D. H. (1988) *Eye, Brain, and Vision*, Scientific American Library Series. New York: W. H. Freeman.

HURVICH, L. (1981) *Color Vision*. Sunderland, MA: Sinauer Associates, pp. 180–194.

OGLE, K. N. (1964) *Researches in Binocular Vision*. Hafner: New York.

OYSTER, C. (1999) *The Human Eye: Structure and Function*. Sunderland, MA: Sinauer Associates.

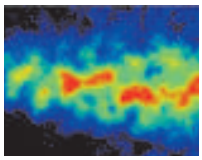
POLYAK, S. (1957) *The Vertebrate Visual System*. Chicago: The University of Chicago Press.

RODIECK, R. W. (1973) *The Vertebrate Retina*. San Francisco: W. H. Freeman.

RODIECK, R. W. (1998) *First Steps in Seeing*. Sunderland, MA: Sinauer Associates.

WANDELL, B. A. (1995) *Foundations of Vision*. Sunderland, MA: Sinauer Associates.

Chapter 11



Central Visual Pathways

Overview

Information supplied by the retina initiates interactions between multiple subdivisions of the brain that eventually lead to conscious perception of the visual scene, at the same time stimulating more conventional reflexes such as adjusting the size of the pupil, directing the eyes to targets of interest, and regulating homeostatic behaviors that are tied to the day/night cycle. The pathways and structures that mediate this broad range of functions are necessarily diverse. Of these, the primary visual pathway from the retina to the dorsal lateral geniculate nucleus in the thalamus and on to the primary visual cortex is the most important and certainly the most thoroughly studied component of the visual system. Different classes of neurons within this pathway encode the varieties of visual information—luminance, spectral differences, orientation, and motion—that we ultimately see. The parallel processing of different categories of visual information continues in cortical pathways that extend beyond primary visual cortex, supplying a variety of visual areas in the occipital, parietal, and temporal lobes. Visual areas in the temporal lobe are primarily involved in object recognition, whereas those in the parietal lobe are concerned with motion. Normal vision depends on the integration of information in all these cortical areas. The processes underlying visual perception are not understood and remain one of the central challenges of modern neuroscience.

Central Projections of Retinal Ganglion Cells

Ganglion cell axons exit the retina through a circular region in its nasal part called the **optic disk** (or optic papilla), where they bundle together to form the **optic nerve**. This region of the retina contains no photoreceptors and, because it is insensitive to light, produces the perceptual phenomenon known as the **blind spot** (Box A). The optic disk is easily identified as a whitish circular area when the retina is examined with an ophthalmoscope; it also is recognized as the site from which the ophthalmic artery and veins enter (or leave) the eye (Figure 11.1). In addition to being a conspicuous retinal landmark, the appearance of the optic disk is a useful gauge of intracranial pressure. The subarachnoid space surrounding the optic nerve is continuous with that of the brain; as a result, increases in intracranial pressure—a sign of serious neurological problems such as a space-occupying lesion—can be detected as *papilledema*, a swelling of the optic disk.

Axons in the optic nerve run a straight course to the **optic chiasm** at the base of the diencephalon. In humans, about 60% of these fibers cross in the chiasm, while the other 40% continue toward the thalamus and midbrain targets on the same side. Once past the chiasm, the ganglion cell axons on each

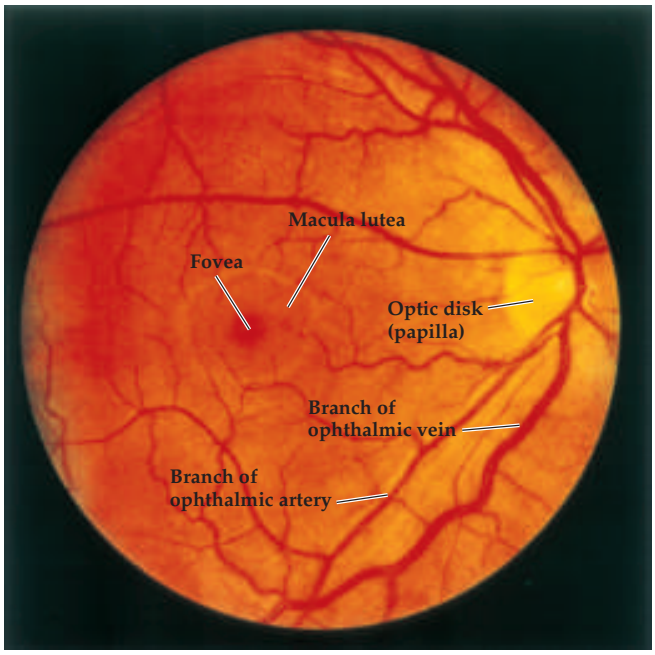


Figure 11.1 The retinal surface of the left eye, viewed with an ophthalmoscope. The optic disk is the region where the ganglion cell axons leave the retina to form the optic nerve; it is also characterized by the entrance and exit, respectively, of the ophthalmic arteries and veins that supply the retina. The macula lutea can be seen as a distinct area at the center of the optical axis (the optic disk lies nasally); the macula is the region of the retina that has the highest visual acuity. The fovea is a depression or pit about 1.5 mm in diameter that lies at the center of the macula (see Chapter 10).

side form the **optic tract**. Thus, the optic tract, unlike the optic nerve, contains fibers from *both* eyes. The partial crossing (or decussation) of ganglion cell axons at the optic chiasm allows information from corresponding points on the two retinas to be processed by approximately the same cortical site in each hemisphere, an important issue that is considered in the next section.

The ganglion cell axons in the optic tract reach a number of structures in the diencephalon and midbrain (Figure 11.2). The major target in the diencephalon is the **dorsal lateral geniculate nucleus** of the thalamus. Neurons in the lateral geniculate nucleus, like their counterparts in the thalamic relays of other sensory systems, send their axons to the cerebral cortex via the internal capsule. These axons pass through a portion of the internal capsule called the **optic radiation** and terminate in the **primary visual cortex**, or **striate cortex** (also referred to as **Brodmann's area 17** or **V1**), which lies largely along and within the calcarine fissure in the occipital lobe. The **retinogeniculostriate pathway**, or **primary visual pathway**, conveys information that is essential for most of what is thought of as seeing. Thus, damage anywhere along this route results in serious visual impairment.

A second major target of the ganglion cell axons is a collection of neurons that lies between the thalamus and the midbrain in a region known as the **pretectum**. Although small in size compared to the lateral geniculate nucleus, the pretectum is particularly important as the coordinating center for the **pupillary light reflex** (i.e., the reduction in the diameter of the pupil that occurs when sufficient light falls on the retina) (Figure 11.3). The initial component of the pupillary light reflex pathway is a bilateral projection from the retina to the pretectum. Pretectal neurons, in turn, project to the **Edinger-Westphal nucleus**, a small group of nerve cells that lies close to the nucleus of the oculomotor nerve (cranial nerve III) in the midbrain. The Edinger-Westphal nucleus contains the preganglionic parasympathetic neurons that send their axons via the oculomotor nerve to terminate on neurons in the ciliary

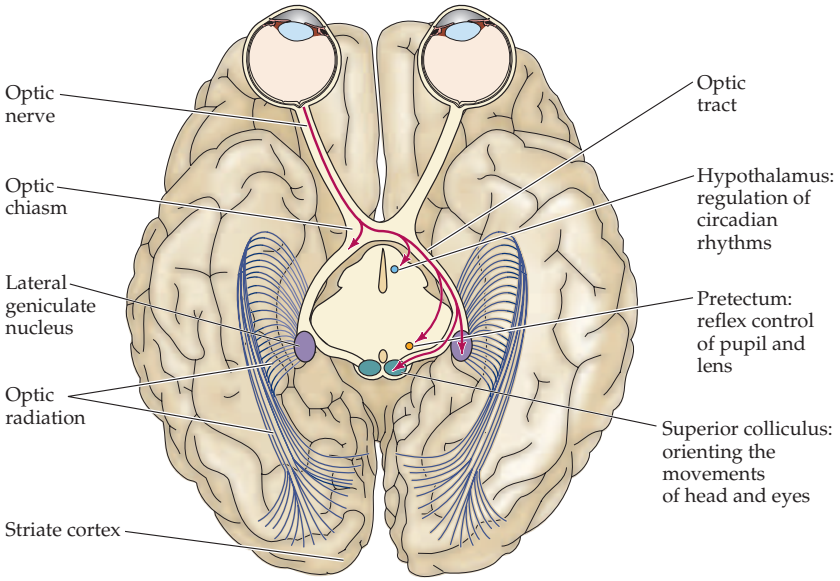
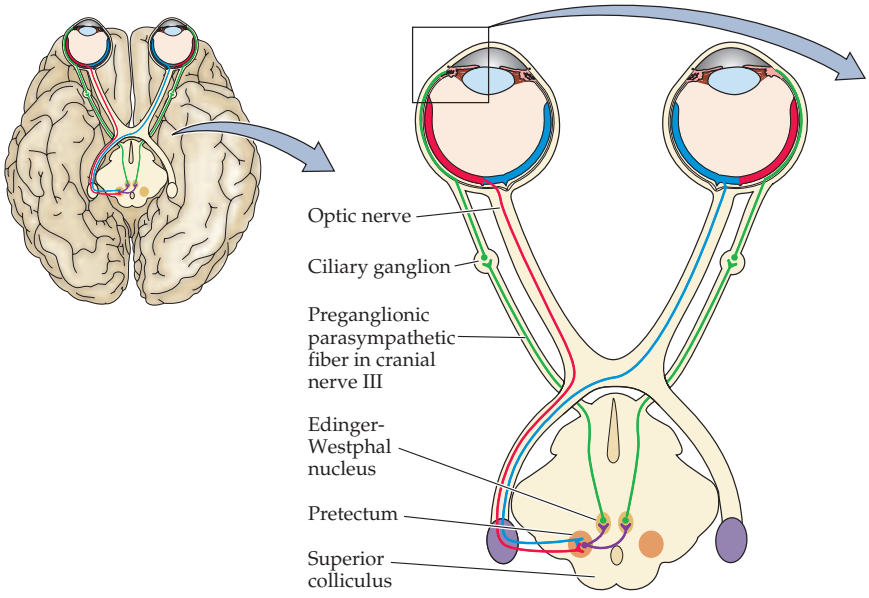


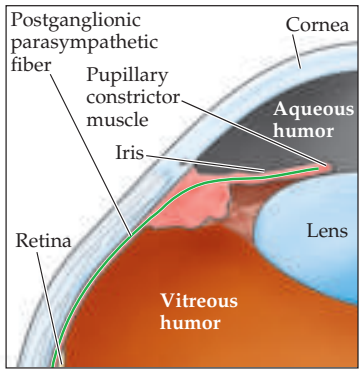
Figure 11.2 Central projections of retinal ganglion cells. Ganglion cell axons terminate in the lateral geniculate nucleus of the thalamus, the superior colliculus, the pretectum, and the hypothalamus. For clarity, only the crossing axons of the right eye are shown (view is looking up at the inferior surface of the brain).

ganglion (see Chapter 19). Neurons in the ciliary ganglion innervate the constrictor muscle in the iris, which decreases the diameter of the pupil when activated. Shining light in the eye thus leads to an increase in the activity of pretectal neurons, which stimulates the Edinger-Westphal neurons and the ciliary ganglion neurons they innervate, thus constricting the pupil.

In addition to its normal role in regulating the amount of light that enters the eye, the pupillary reflex provides an important diagnostic tool that allows the physician to test the integrity of the visual sensory apparatus, the motor outflow to the pupillary muscles, and the central pathways that medi-

Figure 11.3 The circuitry responsible for the pupillary light reflex. This pathway includes bilateral projections from the retina to the pretectum and projections from the pretectum to the Edinger-Westphal nucleus. Neurons in the Edinger-Westphal nucleus terminate in the ciliary ganglion, and neurons in the ciliary ganglion innervate the pupillary constrictor muscles. Notice that the afferent axons activate both Edinger-Westphal nuclei via the neurons in the pretectum.





Box A
The Blind Spot

It is logical to suppose that a visual field defect (called a *scotoma*) arising from damage to the retina or central visual pathways would be obvious to the individual suffering from such pathology. When the deficit involves a peripheral region of the visual field, however, a scotoma often goes unnoticed until a car accident or some other mishap all too dramatically reveals the sensory loss. In fact, all of us have a physiological scotoma of which we are quite unaware, the so-called "blind spot." The blind spot is the substantial gap in each monocular visual field that corresponds to the location of the optic disk, the receptor-free region of the retina where the optic nerve leaves the eye (see Figure 11.1).

To find the "blind spot" of the right eye, close the left eye and fixate on the X shown in the figure here, holding the book about 30–40 centimeters away. Now take a pencil in your right hand and, without breaking fixation, move the tip slowly toward the X from the right side of the page. At some point, the tip of the pencil (indeed the whole end of the pencil) will disappear; mark this point and continue to move the pencil to the left until it reappears; then make another mark. The borders of the blind spot along the vertical axis can be determined in the same way by moving the pencil up and down so that its path falls between the two horizontal marks. To prove that information from the region of visual space bounded by the marks is really not perceived, put a penny inside the demarcated area. When you fixate the X with both eyes and then close the left eye, the penny will disappear, a seemingly magical event that amazed the French royal court when it was first reported by the natural philosopher Edmé Mariotte in 1668.

How can we be unaware of such a large defect in the visual field (typically about 5°–8°)? The optic disk is located in the nasal retina of each eye. With both eyes open, information about the corresponding region of visual space is, of course, available from the temporal retina of the other eye. But this fact does not explain why the blind spot remains undetected with one eye closed. When the world is viewed monocularly, the visual system appears to "fill-in" the missing part of the scene based on the information supplied by the regions surrounding the optic disk. To observe this phenomenon, notice what happens when a pencil or some other object lies *across* the optic disk representation. Remarkably, the pencil looks complete! Although electrophysiological recordings have shown that neurons in the visual cortex whose receptive fields lie in the optic disk representation can be activated by stimulating the regions that surround the optic disk of the contralateral eye, suggesting that "filling-in" the blind spot is based on cortical mechanisms that integrate information from different points in the visual field, the mechanism of this striking phenomenon is not clear. Herman von Helmholtz pointed out in the nineteenth century that it may just be that this part of the visual world is ignored, the pencil being completed across the blind spot because the rest of the scene simply "collapses" around it.

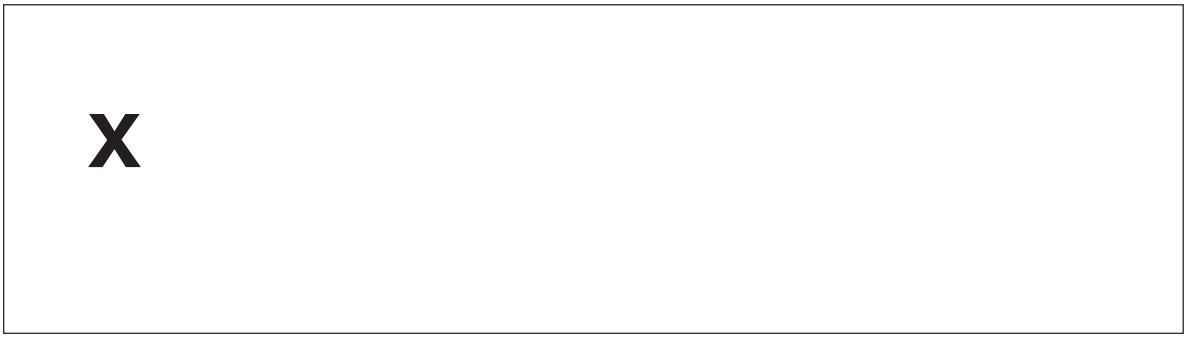
References

Fiorani, M., M. G. P. Rosa, R. Gattass and C. E. Rocha-Miranda (1992) Dynamic surrounds of receptive fields in striate cortex: A physiological basis for perceptual completion? Proc. Natl. Acad. Sci. USA 89: 8547–8551.

Gilbert, C. D. (1992) Horizontal integration and cortical dynamics. Neuron 9: 1–13.

Ramachandran, V. S. and T. L. Gregory (1991) Perceptual filling in of artificially induced scotomas in human vision. Nature 350: 699–702.

Von Helmholtz, H. (1968). *Helmholtz's Treatise on Physiological Optics*, Vols. I–III (Translated from the Third German Ed. published in 1910). J. P. C. Southall (ed.). New York: Dover Publications. See pp. 204ff in Vol. III.



ate the reflex. Under normal conditions, the pupils of both eyes respond identically, regardless of which eye is stimulated; that is, light in one eye produces constriction of both the stimulated eye (the direct response) and the unstimulated eye (the consensual response; see Figure 11.3). Comparing the response in the two eyes is often helpful in localizing a lesion. For example, a direct response in the left eye without a consensual response in the right eye suggests a problem with the visceral motor outflow to the right eye, possibly as a result of damage to the oculomotor nerve or Edinger-Westphal nucleus in the brainstem. Failure to elicit a response (either direct or indirect) to stimulation of the left eye if both eyes respond normally to stimulation of the right eye suggests damage to the sensory input from the left eye, possibly to the left retina or optic nerve.

There are several other important targets of retinal ganglion cell axons. One is the **suprachiasmatic nucleus** of the hypothalamus, a small group of neurons at the base of the diencephalon (see Box A in Chapter 20). The **retinohypothalamic pathway** is the route by which variation in light levels influences the broad spectrum of visceral functions that are entrained to the day/night cycle (see Chapters 20 and 27). Another target is the **superior colliculus**, a prominent structure visible on the dorsal surface of the midbrain (see Figure 1.14). The superior colliculus coordinates head and eye movements to visual (as well as other) targets; its functions are considered in Chapter 19.

The type of visual information required to perform the functions of these different retinal targets is quite different. Reading the text on this page, for example, requires a high-resolution sampling of the retinal image, whereas regulating circadian rhythms and adjusting the pupil accordingly require only a measure of overall changes in light levels, and little or no information about the features of the image. It should come as no surprise, then, that there is a diversity of ganglion cell types that provide information appropriate to the functions of these different targets.

Projections to the lateral geniculate nucleus (which are described in more detail later) arise from at least three broad classes of ganglion cells, whose tuning properties are appropriate for mediating the richness of visual perception (high acuity, color, motion). In contrast, projections to the hypothalamus and the pretectum arise from ganglion cells that lack these properties and are highly suited for detecting luminance flux. The retinal specializations responsible for constructing these distinct classes of retinal ganglion cells are only beginning to be identified; they include not only differences in ganglion cell synaptic connections, but in the locus of the phototransduction event itself. Unlike the majority of ganglion cells, which depend on rods and cones for their sensitivity to light, the ganglion cells that project to the hypothalamus and pretectum express their own light-sensitive photopigment (*melanopsin*) and are capable of modulating their response to changes in light levels in the absence of signals from rods and cones. The presence of light sensitivity within this class of ganglion cells presumably explains why normal circadian rhythms are maintained in animals that have completely lost form vision due to degeneration of rod and cone photoreceptors.

The Retinotopic Representation of the Visual Field

The spatial relationships among the ganglion cells in the retina are maintained in most of their central targets as orderly representations or "maps" of visual space. Most of these structures receive information from both eyes, requiring that these inputs be integrated to form a coherent map of individ-

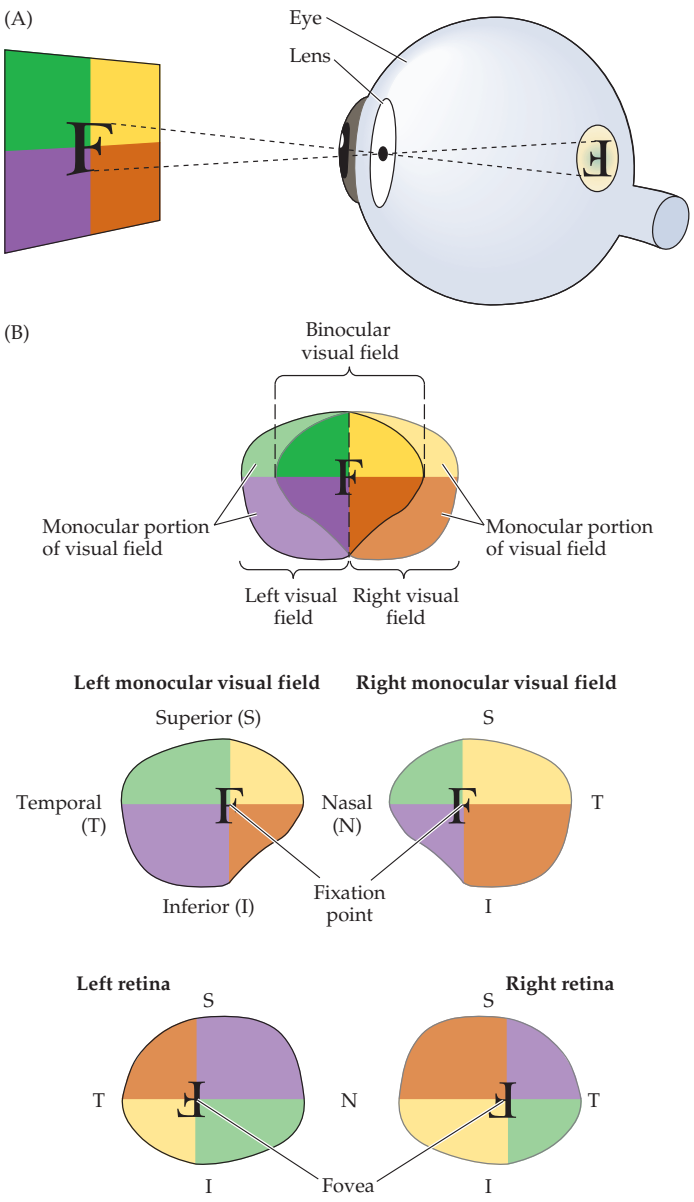


Figure 11.4 Projection of the visual fields onto the left and right retinas. (A) Projection of an image onto the surface of the retina. The passage of light rays through the pupil of the eye results in images that are inverted and left–right reversed on the retinal surface. (B) Retinal quadrants and their relation to the organization of monocular and binocular visual fields, as viewed from the back surface of the eyes. Vertical and horizontal lines drawn through the center of the fovea define retinal quadrants (bottom). Comparable lines drawn through the point of fixation define visual field quadrants (center). Color coding illustrates corresponding retinal and visual field quadrants. The overlap of the two monocular visual fields is shown at the top.

ual points in space. As a general rule, information from the left half of the visual world, whether it originates from the left or right eye, is represented in the right half of the brain, and vice versa.

Understanding the neural basis for the appropriate arrangement of inputs from the two eyes requires considering how images are projected onto the two retinas, and the central destination of the ganglion cells located in different parts of the retina. Each eye sees a part of visual space that defines its **visual field** (Figure 11.4A). For descriptive purposes, each retina and its corresponding visual field are divided into quadrants. In this scheme, the surface of the retina is subdivided by vertical and horizontal lines that intersect at the center of the fovea (Figure 11.4B). The vertical line divides the retina into **nasal** and **temporal divisions** and the horizontal line divides the retina

into **superior** and **inferior divisions**. Corresponding vertical and horizontal lines in visual space (also called meridians) intersect at the **point of fixation** (the point in visual space that falls on the fovea) and define the quadrants of the visual field. The crossing of light rays diverging from different points on an object at the pupil causes the images of objects in the visual field to be inverted and left-right reversed on the retinal surface. As a result, objects in the temporal part of the visual field are seen by the nasal part of the retina, and objects in the superior part of the visual field are seen by the inferior part of the retina. (It may help in understanding Figure 11.4B to imagine that you are looking at the back surfaces of the retinas, with the corresponding visual fields projected onto them.)

With both eyes open, the two foveas are normally aligned on a single target in visual space, causing the visual fields of both eyes to overlap extensively (see Figure 11.4B and Figure 11.5). This **binocular field** of view consists of two symmetrical visual hemifields (left and right). The left binocular hemifield includes the nasal visual field of the right eye and the temporal visual field of the left eye; the right hemifield includes the temporal visual field of

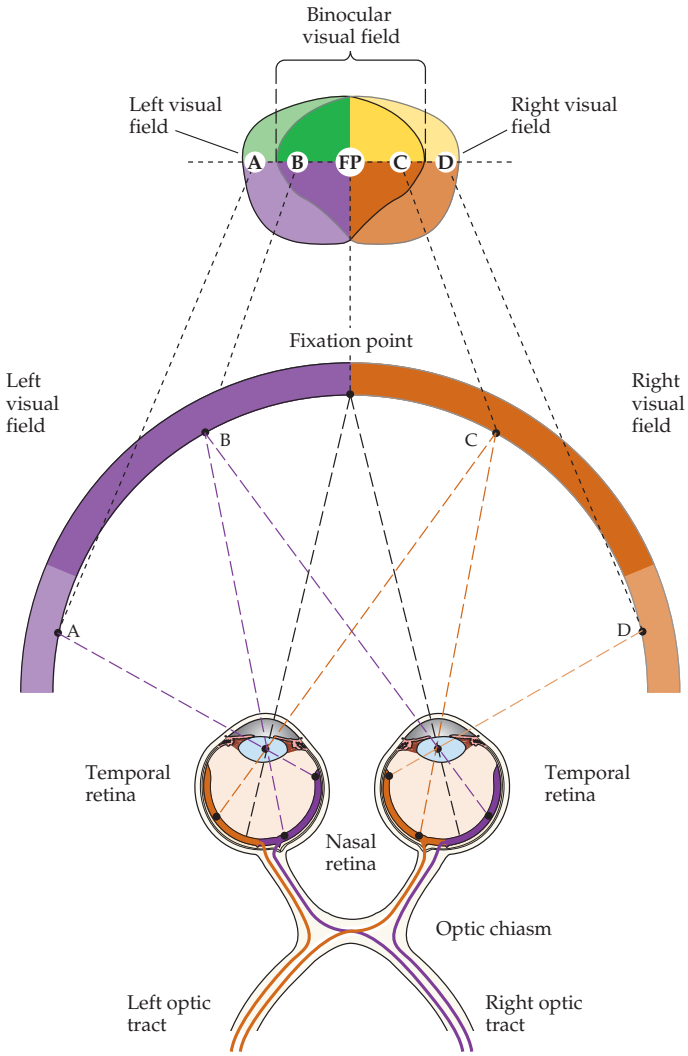


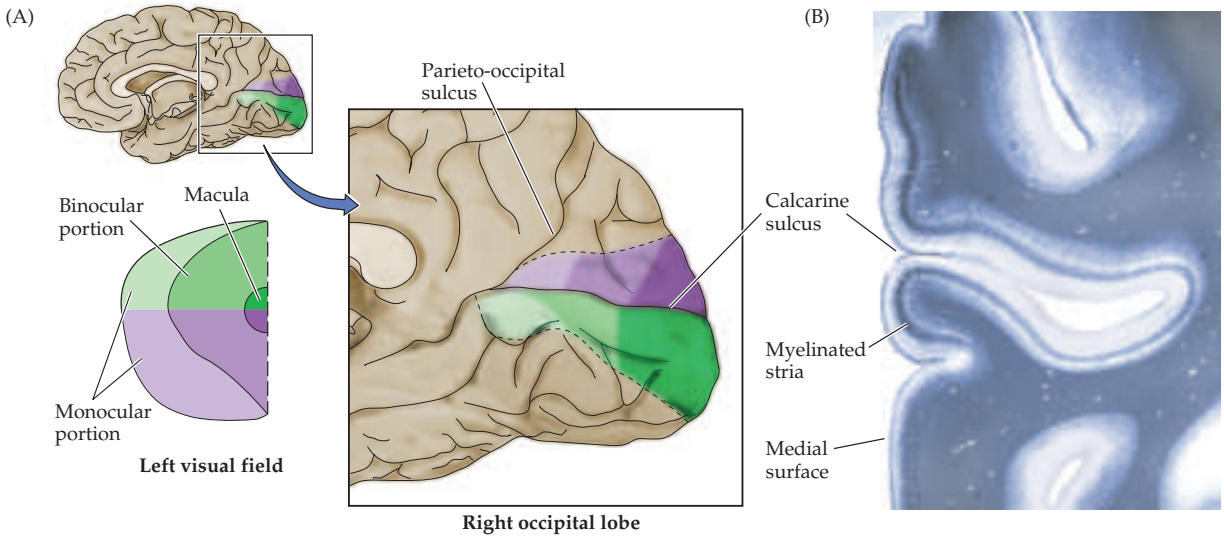
Figure 11.5 Projection of the binocular field of view onto the two retinas and its relation to the crossing of fibers in the optic chiasm. Points in the binocular portion of the left visual field (B) fall on the nasal retina of the left eye and the temporal retina of the right eye. Points in the binocular portion of the right visual field (C) fall on the nasal retina of the right eye and the temporal retina of the left eye. Points that lie in the monocular portions of the left and right visual fields (A and D) fall on the left and right nasal retinas, respectively. The axons of ganglion cells in the nasal retina cross in the optic chiasm, whereas those from the temporal retina do not. As a result, information from the left visual field is carried in the right optic tract, and information from the right visual field is carried in the left optic tract.

the right eye and the nasal visual field of the left eye. The temporal visual fields are more extensive than the nasal visual fields, reflecting the size of the nasal and temporal retinas respectively. As a result, vision in the periphery of the field of view is strictly monocular, mediated by the most medial portion of the nasal retina. Most of the rest of the field of view can be seen by both eyes; i.e., individual points in visual space lie in the nasal visual field of one eye and the temporal visual field of the other. It is worth noting, however, that the shape of the face and nose impact the extent of this region of binocular vision. In particular, the inferior nasal visual fields are less extensive than the superior nasal fields, and consequently the binocular field of view is smaller in the lower visual field than in the upper (see Figure 11.4B).

Ganglion cells that lie in the nasal division of each retina give rise to axons that cross in the chiasm, while those that lie in the temporal retina give rise to axons that remain on the same side (see Figure 11.5). The boundary (or line of decussation) between contralaterally and ipsilaterally projecting ganglion cells runs through the center of the fovea and defines the border between the nasal and temporal hemiretinas. Images of objects in the left visual hemifield (such as point B in Figure 11.5) fall on the nasal retina of the left eye and the temporal retina of the right eye, and the axons from ganglion cells in these regions of the two retinas project through the right optic tract. Objects in the right visual hemifield (such as point C in Figure 11.5) fall on the nasal retina of the right eye and the temporal retina of the left eye; the axons from ganglion cells in these regions project through the left optic tract. As mentioned previously, objects in the monocular portions of the visual hemifields (points A and D in Figure 11.5) are seen only by the most peripheral nasal retina of each eye; the axons of ganglion cells in these regions (like the rest of the nasal retina) run in the contralateral optic tract. Thus, unlike the optic nerve, the optic tract contains the axons of ganglion cells that originate in both eyes and represent the contralateral field of view.

Optic tract axons terminate in an orderly fashion within their target structures thus generating well ordered maps of the contralateral hemifield. For the primary visual pathway, the map of the contralateral hemifield that is established in the lateral geniculate nucleus is maintained in the projections of the lateral geniculate nucleus to the striate cortex (Figure 11.6). Thus the

Figure 11.6 Visuotopic organization of the striate cortex in the right occipital lobe, as seen in mid-sagittal view. (A) The primary visual cortex occupies a large part of the occipital lobe. The area of central vision (the fovea) is represented over a disproportionately large part of the caudal portion of the lobe, whereas peripheral vision is represented more anteriorly. The upper visual field is represented below the calcarine sulcus, the lower field above the calcarine sulcus. (B) Photomicrograph of a coronal section of the human striate cortex, showing the characteristic myelinated band, or stria, that gives this region of the cortex its name. The calcarine sulcus on the medial surface of the occipital lobe is indicated. (B courtesy of T. Andrews and D. Purves.)



fovea is represented in the posterior part of the striate cortex, whereas the more peripheral regions of the retina are represented in progressively more anterior parts of the striate cortex. The upper visual field is mapped below the calcarine sulcus, and the lower visual field above it. As in the somatic sensory system, the amount of cortical area devoted to each unit area of the sensory surface is not uniform, but reflects the density of receptors and sensory axons that supply the peripheral region. Like the representation of the hand region in the somatic sensory cortex, the representation of the macula is therefore disproportionately large, occupying most of the caudal pole of the occipital lobe.

Visual Field Deficits

A variety of retinal or more central pathologies that involve the primary visual pathway can cause visual field deficits that are limited to particular regions of visual space. Because the spatial relationships in the retinas are maintained in central visual structures, a careful analysis of the visual fields can often indicate the site of neurological damage. Relatively large visual field deficits are called **anopsias** and smaller ones are called **scotomas** (see Box A). The former term is combined with various prefixes to indicate the specific region of the visual field from which sight has been lost (Figures 11.7 and 11.8).

Damage to the retina or one of the optic nerves before it reaches the chiasm results in a loss of vision that is limited to the eye of origin. In contrast, damage in the region of the optic chiasm—or more centrally—results in specific types of deficits that involve the visual fields of both eyes (Figure 11.8). Damage to structures that are central to the optic chiasm, including the optic tract, lateral geniculate nucleus, optic radiation, and visual cortex, results in deficits that are limited to the contralateral visual hemifield. For example, interruption of the optic tract on the right results in a loss of sight in the left visual field (that is, blindness in the temporal visual field of the left eye and the nasal visual field of the right eye). Because such damage affects corresponding parts of the visual field in each eye, there is a complete loss of vision in the affected region of the binocular visual field, and the deficit is referred to as a **homonymous hemianopsia** (in this case, a left homonymous hemianopsia).

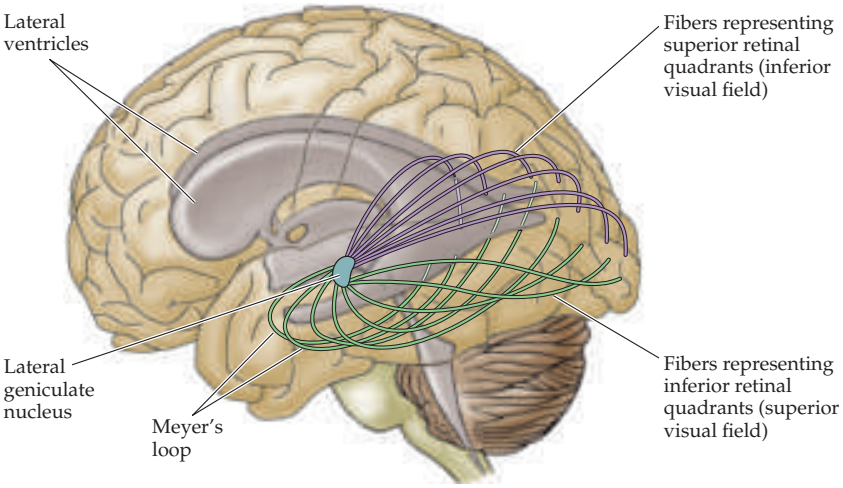


Figure 11.7 Course of the optic radiation to the striate cortex. Axons carrying information about the superior portion of the visual field sweep around the lateral horn of the ventricle in the temporal lobe (Meyer's loop) before reaching the occipital lobe. Those carrying information about the inferior portion of the visual field travel in the parietal lobe.

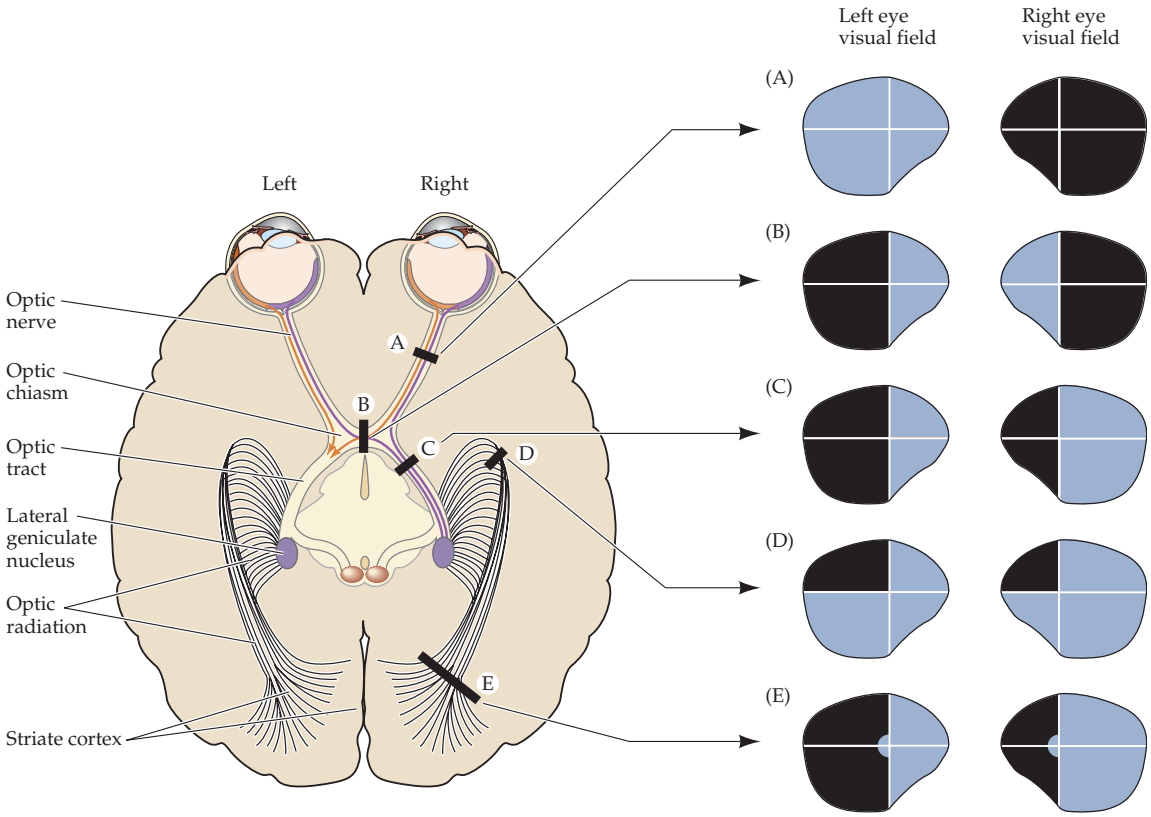


Figure 11.8 Visual field deficits resulting from damage at different points along the primary visual pathway. The diagram on the left illustrates the basic organization of the primary visual pathway and indicates the location of various lesions. The right panels illustrate the visual field deficits associated with each lesion. (A) Loss of vision in right eye. (B) Bitemporal (heteronomous) hemianopsia. (C) Left homonymous hemianopsia. (D) Left superior quadrantanopsia. (E) Left homonymous hemianopsia with macular sparing.

In contrast, damage to the optic chiasm results in visual field deficits that involve noncorresponding parts of the visual field of each eye. For example, damage to the middle portion of the optic chiasm (which is often the result of pituitary tumors) can affect the fibers that are crossing from the nasal retina of each eye, leaving the uncrossed fibers from the temporal retinas intact. The resulting loss of vision is confined to the temporal visual field of each eye and is known as **bitemporal hemianopsia**. It is also called **heteronomous hemianopsia** to emphasize that the parts of the visual field that are lost in each eye do not overlap. Individuals with this condition are able to see in both left and right visual fields, provided both eyes are open. However, all information from the most peripheral parts of visual fields (which are seen only by the nasal retinas) is lost.

Damage to central visual structures is rarely complete. As a result, the deficits associated with damage to the chiasm, optic tract, optic radiation, or visual cortex are typically more limited than those shown in Figure 11.8. This is especially true for damage along the optic radiation, which fans out under the temporal and parietal lobes in its course from the lateral geniculate nucleus to the striate cortex. Some of the optic radiation axons run out into the temporal lobe on their route to the striate cortex, a branch called **Meyer's loop** (see Figure 11.7). Meyer's loop carries information from the superior portion of the contralateral visual field. More medial parts of the optic radiation, which pass under the cortex of the parietal lobe, carry information from the inferior portion of the contralateral visual field. Damage to parts of the temporal lobe with involvement of Meyer's loop can thus result in a superior

homonymous quadrantanopsia; damage to the optic radiation underlying the parietal cortex results in an inferior homonymous quadrantanopsia.

Injury to central visual structures can also lead to a phenomenon called *macular sparing*, i.e., the loss of vision throughout wide areas of the visual field, with the exception of foveal vision. Macular sparing is commonly found with damage to the cortex, but can be a feature of damage anywhere along the length of the visual pathway. Although several explanations for macular sparing have been offered, including overlap in the pattern of crossed and uncrossed ganglion cells supplying central vision, the basis for this selective preservation is not clear.

The Functional Organization of the Striate Cortex

Much in the same way that Stephen Kuffler explored the response properties of individual retinal ganglion cells (see Chapter 10), David Hubel and Torsten Wiesel used microelectrode recordings to examine the properties of neurons in more central visual structures.

The responses of neurons in the lateral geniculate nucleus were found to be remarkably similar to those in the retina, with a center-surround receptive field organization and selectivity for luminance increases or decreases. However, the small spots of light that were so effective at stimulating neurons in the retina and lateral geniculate nucleus were largely ineffective in visual cortex. Instead, most cortical neurons in cats and monkeys responded vigorously to light–dark bars or edges, and only if the bars were presented at a particular range of orientations within the cell's receptive field (Figure 11.9). The responses of cortical neurons are thus tuned to the orientation of edges, much like cone receptors are tuned to the wavelength of light; the peak in the tuning curve (the orientation to which a cell is most responsive) is referred to as the neuron's preferred orientation. By sampling the responses of a large number of single cells, Hubel and Weisel demonstrated that all edge orientations were roughly equally represented in visual cortex. As a

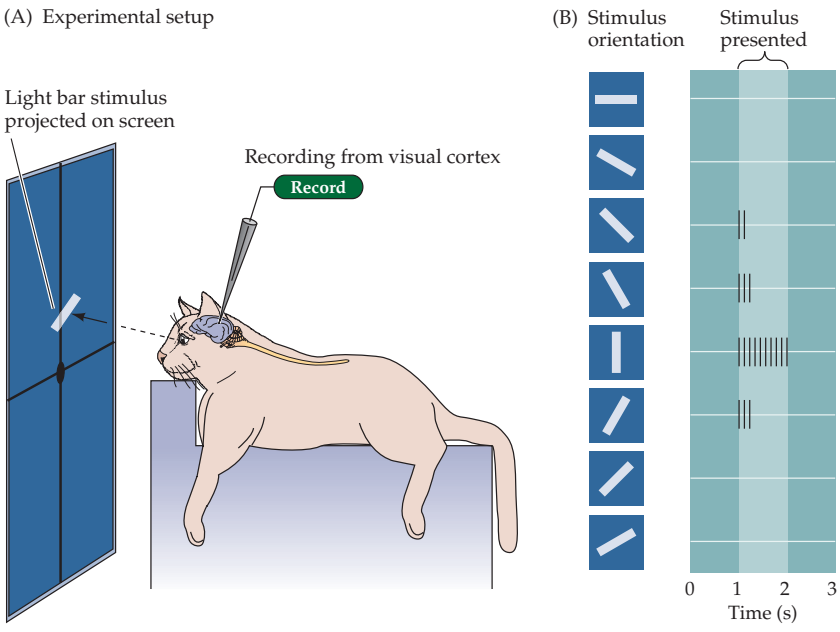


Figure 11.9 Neurons in the primary visual cortex respond selectively to oriented edges. (A) An anesthetized animal is fitted with contact lenses to focus the eyes on a screen, where images can be projected; an extracellular electrode records the neuronal responses. (B) Neurons in the primary visual cortex typically respond vigorously to a bar of light oriented at a particular angle and weakly—or not at all—to other orientations.

result, a given orientation in a visual scene appears to be "encoded" in the activity of a distinct population of **orientation-selective neurons**.

Hubel and Wiesel also found that there are subtly different subtypes within a class of neurons that preferred the same orientation. For example, the receptive fields of some cortical cells, which they called **simple cells**, were composed of spatially separate "on" and "off" response zones, as if the "on" and "off" centers of lateral geniculate cells that supplied these neurons were arrayed in separate parallel bands. Other neurons, referred to as **complex cells**, exhibited mixed "on" and "off" responses throughout their receptive field, as if they received their inputs from a number of simple cells. Further analysis uncovered cortical neurons sensitive to the *length* of the bar of light that was moved across their receptive field, decreasing their rate of response when the bar exceeded a certain length. Still other cells responded selectively to the *direction* in which an edge moved across their receptive field. Although the mechanisms responsible for generating these selective responses are still not well understood, there is little doubt that the specificity of the receptive field properties of neurons in the striate cortex (and beyond) plays an important role in determining the basic attributes of visual scenes.

Another feature that distinguishes the responses of neurons in the striate cortex from those at earlier stages in the primary visual pathway is **binocularity**. Although the lateral geniculate nucleus receives inputs from both eyes, the axons terminate in separate layers, so that individual geniculate

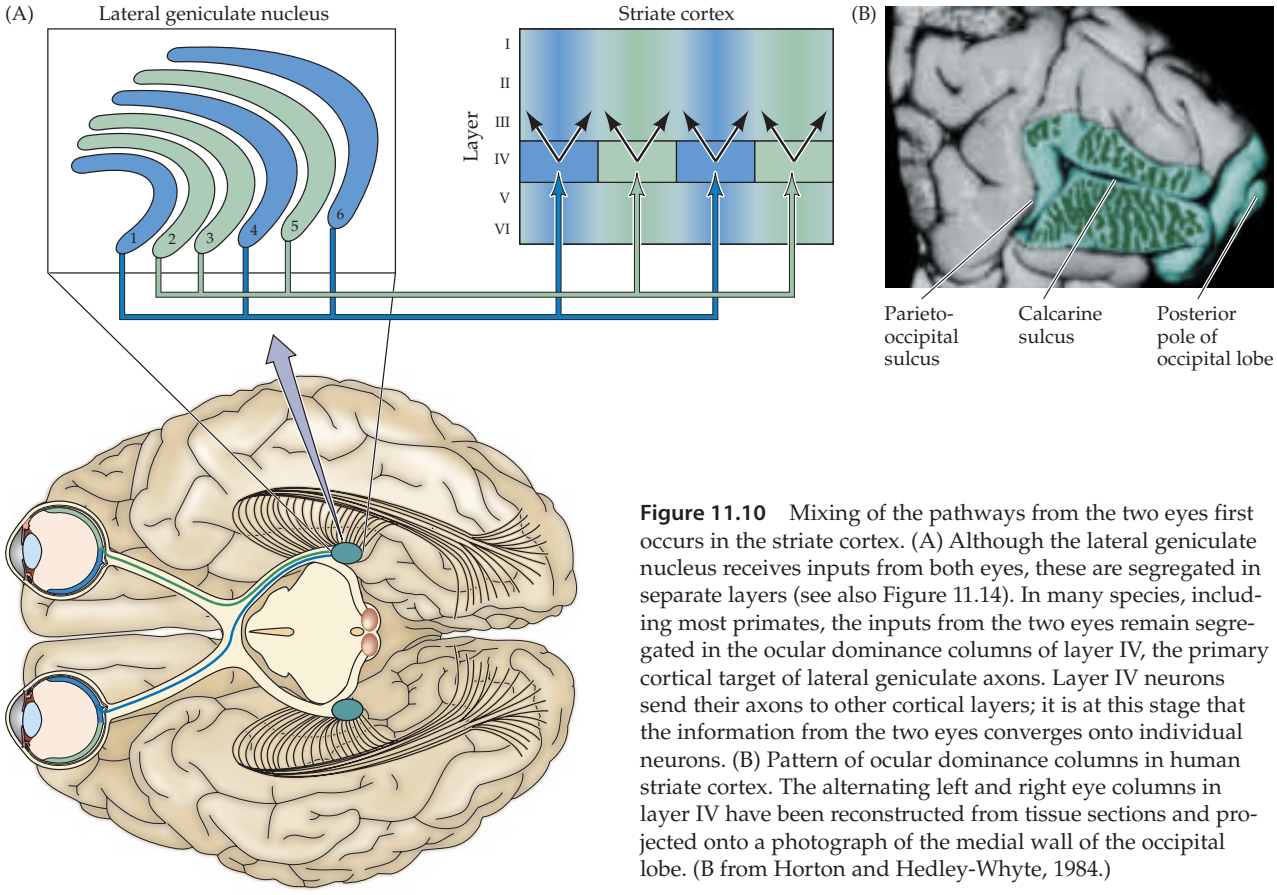


Figure 11.10 Mixing of the pathways from the two eyes first occurs in the striate cortex. (A) Although the lateral geniculate nucleus receives inputs from both eyes, these are segregated in separate layers (see also Figure 11.14). In many species, including most primates, the inputs from the two eyes remain segregated in the ocular dominance columns of layer IV, the primary cortical target of lateral geniculate axons. Layer IV neurons send their axons to other cortical layers; it is at this stage that the information from the two eyes converges onto individual neurons. (B) Pattern of ocular dominance columns in human striate cortex. The alternating left and right eye columns in layer IV have been reconstructed from tissue sections and projected onto a photograph of the medial wall of the occipital lobe. (B from Horton and Hedley-Whyte, 1984.)

neurons are monocular, driven by either the left or right eye but not by both (Figure 11.10; see also Figure 11.14). In some species, including most (but not all) primates, inputs from the left and right eyes remain segregated to some degree even beyond the geniculate because the axons of geniculate neurons terminate in alternating eye-specific columns within cortical layer IV—the so-called **ocular dominance columns** (see the next section). Beyond this point, the signals from the two eyes are combined at the cellular level. Thus, most cortical neurons have binocular receptive fields, and these fields are almost identical, having the same size, shape, preferred orientation, and roughly the same position in the visual field of each eye.

Bringing together the inputs from the two eyes at the level of the striate cortex provides a basis for **stereopsis**, the special sensation of depth that arises from viewing nearby objects with two eyes instead of one. Because the two eyes look at the world from slightly different angles, objects that lie in front of or behind the plane of fixation project to noncorresponding points on the two retinas. To convince yourself of this fact, hold your hand at arm's length and fixate on the tip of one finger. Maintain fixation on the finger as you hold a pencil in your other hand about half as far away. At this distance, the image of the pencil falls on noncorresponding points on the two retinas and will therefore be perceived as two separate pencils (a phenomenon called double vision, or *diplopia*). If the pencil is now moved toward the finger (the point of fixation), the two images of the pencil fuse and a single pencil is seen in front of the finger. Thus, for a small distance on either side of the plane of fixation, where the disparity between the two views of the world remains modest, a single image is perceived; the disparity between the two eye views of objects nearer or farther than the point of fixation is interpreted as depth (Figure 11.11).

Although the neurophysiological basis of stereopsis is not understood, some neurons in the striate cortex and in other visual cortical areas have receptive field properties that make them good candidates for extracting information about binocular disparity. Unlike many binocular cells whose monocular receptive fields sample the same region of visual space, these neurons have monocular fields that are slightly displaced (or perhaps differ in their internal organization) so that the cell is maximally activated by stimuli that fall on noncorresponding parts of the retinas. Some of these neurons (so-called **far cells**) discharge to disparities beyond the plane of fixation, while others (**near cells**) respond to disparities in front of the plane of fixation. The pattern of activity in these different classes of neurons seems likely to contribute to sensations of stereoscopic depth (Box B).

Interestingly, the preservation of the binocular responses of cortical neurons is contingent on the normal activity from the two eyes during early postnatal life. Anything that creates an imbalance in the activity of the two eyes—for example, the clouding of one lens or the abnormal alignment of the eyes during infancy (strabismus)—can permanently reduce the effectiveness of one eye in driving cortical neurons, and thus impair the ability to use binocular information as a cue for depth. Early detection and correction of visual problems is therefore essential for normal visual function in maturity (see Chapter 23).

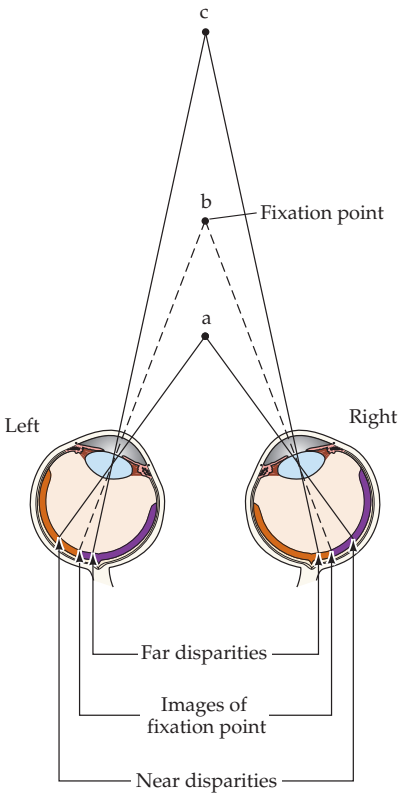


Figure 11.11 Binocular disparities are generally thought to be the basis of stereopsis. When the eyes are fixated on point b, points that lie beyond the plane of fixation (point c) or in front of the point of fixation (point a) project to noncorresponding points on the two retinas. When these disparities are small, the images are fused and the disparity is interpreted by the brain as small differences in depth. When the disparities are greater, double vision occurs (although this normal phenomenon is generally unnoticed).

The Columnar Organization of the Striate Cortex

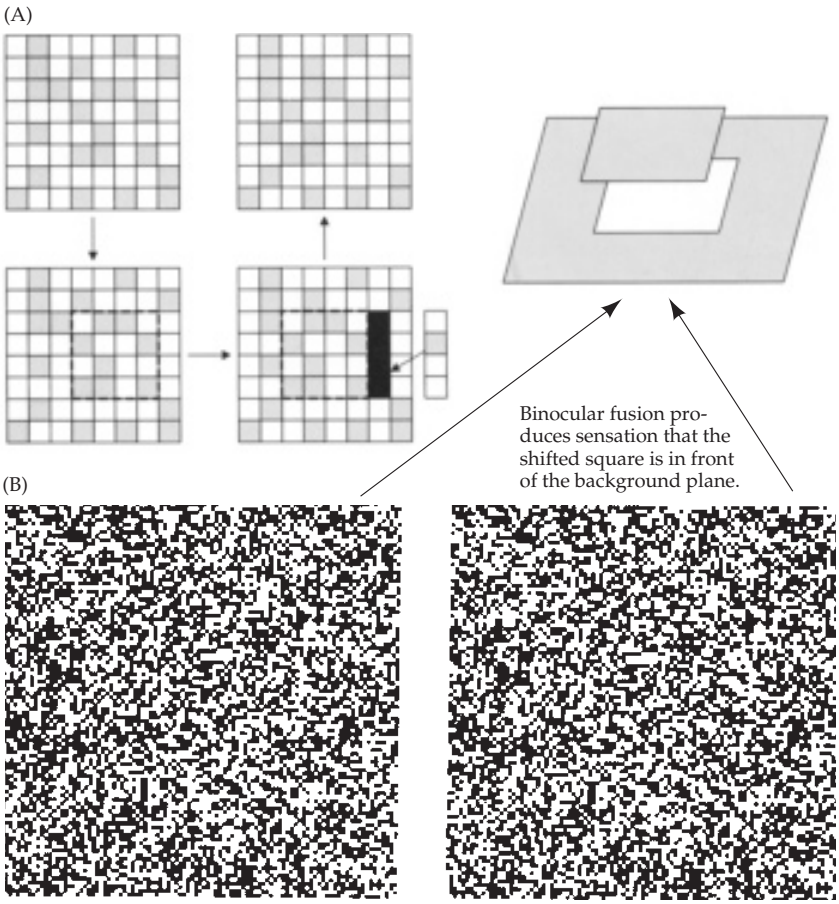
The variety of response properties exhibited by cortical neurons raises the question of how neurons with different receptive fields are arranged within striate cortex. For the most part, the responses of neurons are qualitatively

Box B

Random Dot Stereograms and Related Amusements

An important advance in studies of stereopsis was made in 1959 when Bela Julesz, then working at the Bell Laboratories in Murray Hill, New Jersey, discovered an ingenious way of showing that stereoscopy depends on matching information seen by the two eyes without any prior recognition of what object(s) such matching might generate. Julesz, a Hungarian whose background was in engineering and physics, was working on the problem of how to "break" camouflage. He surmised that the brain's ability to fuse the slightly different views of the two eyes to bring out new information would be an aid in overcoming military camouflage. Julesz also realized that, if his hypothesis was correct, a hidden figure in a random pattern presented to the two eyes should emerge when a portion of the otherwise identical pattern was shifted horizontally in the view of one eye or the other. A horizontal shift in one direction would cause the hidden object to appear in front of the plane of the background, whereas a shift in the other direction would cause the hidden object to appear in back of the plane. Such a figure, called a random dot stereogram, and the method of its creation are shown in Figures A and B. The two images can be easily fused in a stereoscope (like the familiar Viewmaster® toy) but can also be fused simply by allowing the eyes to diverge. Most people find it easiest to do this by imagining that they are looking "through" the figure; after some seconds, during which the brain tries to make sense of what it is presented with, the two images merge and the hidden figure appears (in this case, a square that occupies the middle portion of the figure). The random dot stereogram has been widely used in stereoscopic research for about 40 years, although how such stimuli elicit depth remains very much a matter of dispute.

An impressive—and extraordinarily popular—derivative of the random dot stereogram is the autostereogram (Figure C). The possibility of autostereograms was first discerned by the nineteenth-century British physicist David Brewster. While staring at a Victorian wallpaper with an iterated but offset pattern, he noticed that when the patterns were fused, he perceived two different planes. The plethora of autostereograms that can be seen today in posters, books, and newspapers are close cousins of the random dot stereogram in that computers are used to shift patterns of iterated



Random dot stereograms and autostereograms. (A) to construct a random dot stereogram, a random dot pattern is created to be observed by one eye. The stimulus for the other eye is created by copying the first image, displacing a particular region horizontally, and then filling in the gap with a random sample of dots. (B) When the right and left images are viewed simultaneously but independently by the two eyes (by using a stereoscope or fusing the images by converging or diverging the eyes), the shifted region (a square) appears to be in a different plane from the other dots. (A after Wandell, 1995.)

information with respect to each other. The result is that different planes emerge from what appears to be a meaningless array of visual information (or, depending on the taste of the creator, an apparently "normal" scene in which the iterated and displaced information is hidden). Some autostereograms are designed to reveal the hidden figure when the eyes diverge, and others when they converge. (Looking at a plane more distant than the plane of the surface causes divergence; looking at a plane in front of the picture causes the eyes to converge; see Figure 11.11.)

The elevation of the autostereogram to a popular art form should probably be attributed to Chris W. Tyler, a student of Julesz's and a visual psychophysicist, who was among the first to create commercial autostereograms. Numerous graphic artists—preeminently in Japan, where the popularity of the autostereogram has been enormous—have generated many of such images. As with the random dot stereogram, the task in viewing the autostereogram is not clear to the observer. Nonetheless, the hidden figure emerges, often after minutes of effort in which the brain automatically tries to make sense of the occult information.

References

Julesz, B. (1971) *Foundations of Cyclopean Perception*. Chicago: The University of Chicago Press.

Julesz, B. (1995) *Dialogues on Perception*. Cambridge, MA: MIT Press.

N. E. Thing Enterprises (1993) *Magic Eye: A New Way of Looking at the World*. Kansas City: Andrews and McMeel.

(C)

(C) An autostereogram. The hidden figure (three geometrical forms) emerges by diverging the eyes in this case. (C courtesy of Jun Oi.)

similar at any one point in primary visual cortex, but tend to shift smoothly across its surface. With respect to orientation, for example, all the neurons encountered in an electrode penetration perpendicular to the surface at a particular point will very likely have the same orientation preference, forming a "column" of cells with similar response properties. Adjacent columns, however, usually have slightly different orientation preferences; the sequence of orientation preferences encountered along a tangential electrode penetration gradually shifts as the electrode advances (Figure 11.12). Thus, orientation preference is mapped in the cortex, much like receptive field

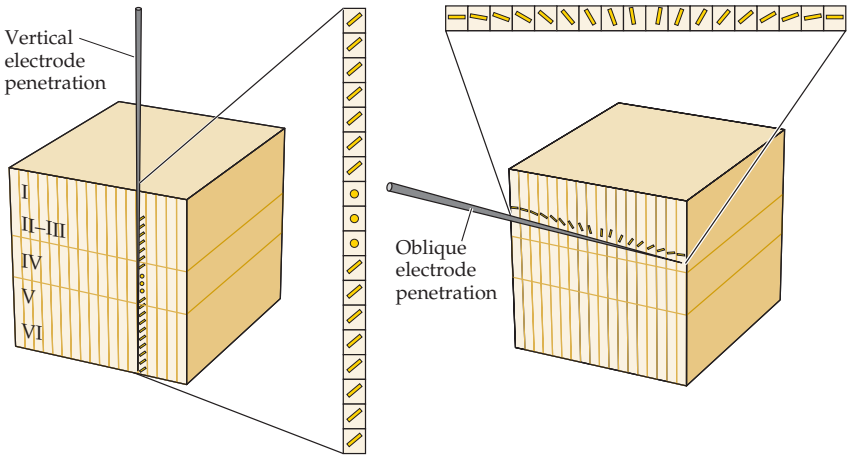


Figure 11.12 Columnar organization of orientation selectivity in the monkey striate cortex. Vertical electrode penetrations encounter neurons with the same preferred orientations, whereas oblique penetrations show a systematic change in orientation across the cortical surface. The circles denote the lack of orientation-selective cells in layer IV.

location (Box C). Unlike the map of visual space, however, the map of orientation preference is iterated many times, such that the same orientation preference is repeated at approximately 1-mm intervals across the striate cortex. This iteration presumably ensures that there are neurons for each region of visual space that represent the full range of orientation values. The orderly progression of orientation preference (as well as other properties that are mapped in this systematic way) is accommodated within the orderly map of visual space by the fact that the mapping is relatively coarse. Each small region of visual space is represented by a set of neurons whose receptive fields cover the full range of orientation preferences, the set being distributed over several millimeters of the cortical surface

The columnar organization of the striate cortex is equally apparent in the binocular responses of cortical neurons. Although most neurons in the striate cortex respond to stimulation of both eyes, the relative strength of the inputs from the two eyes varies from neuron to neuron. At the extremes of this continuum are neurons that respond almost exclusively to the left or right eye; in the middle are those that respond equally well to both eyes. As in the case of orientation preference, vertical electrode penetrations tend to encounter neurons with similar ocular preference (or **ocular dominance**, as it is usually called), whereas tangential penetrations show gradual shifts in ocular dominance. And, like the arrangement of orientation preference, a movement of about a millimeter across the surface is required to sample the full complement of ocular dominance values (Figure 11.13). These shifts in ocular dominance result from the ocular segregation of the inputs from lateral geniculate nucleus within cortical layer IV (see Figure 11.10).

Although the modular arrangement of the visual cortex was first recognized on the basis of these orientation and ocular dominance columns, further work has shown that other stimulus features such as color, direction of motion, and spatial frequency also tend to be distributed in iterated patterns that are systematically related to each other (for example, orientation columns tend to intersect ocular dominance columns at right angles). In short, the striate cortex is composed of repeating units, or modules, that contain all the neuronal machinery necessary to analyze a small region of visual space for a variety of different stimulus attributes. As described in Box D in Chapter 8, a number of other cortical regions show a similar columnar arrangement of their processing circuitry.

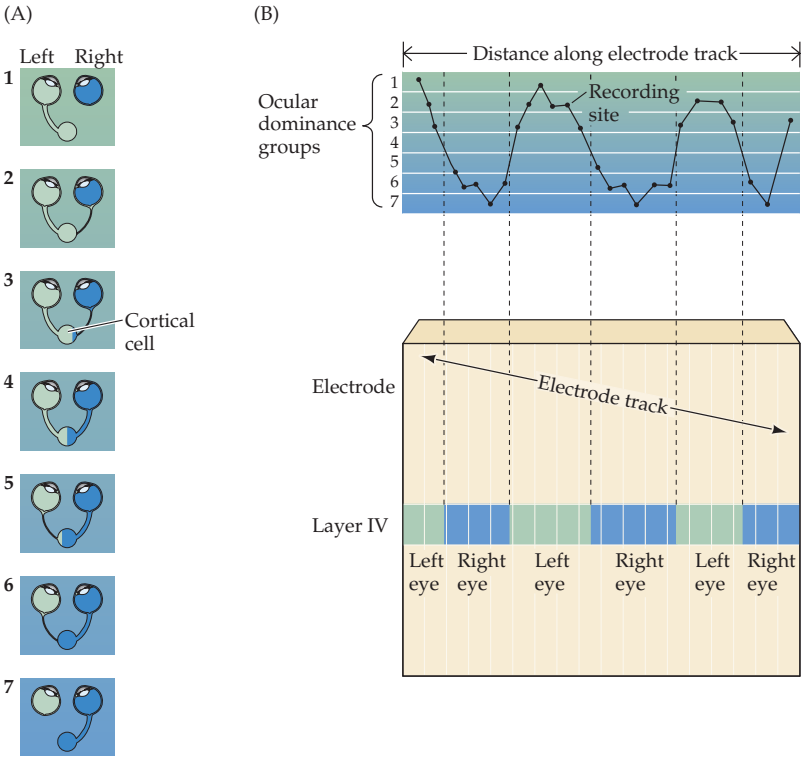


Figure 11.13 Columnar organization of ocular dominance. (A) Cortical neurons in all layers vary in the strength of their response to the inputs from the two eyes, from complete domination by one eye to equal influence of the two eyes. (B) Tangential electrode penetration across the superficial cortical layers reveals a gradual shift in the ocular dominance of the recorded neurons from one eye to the other. In contrast, all neurons encountered in a vertical electrode penetration (other than those neurons that lie in layer IV) tend to have the same ocular dominance.

Division of Labor within the Primary Visual Pathway

In addition to being specific for input from one eye or the other, the layers in the lateral geniculate are also distinguished on the basis of cell size: Two ventral layers are composed of large neurons and are referred to as the **magnocellular layers**, while more dorsal layers are composed of small neurons and are referred to as the **parvocellular layers**. The magno- and parvocellular layers receive inputs from distinct populations of ganglion cells that exhibit corresponding differences in cell size. M ganglion cells that terminate in the magnocellular layers have larger cell bodies, more extensive dendritic fields, and larger-diameter axons than the P ganglion cells that terminate in the parvocellular layers (Figure 11.14A). Moreover, the axons of relay cells in the magno- and parvocellular layers of the lateral geniculate nucleus terminate on distinct populations of neurons located in separate strata within layer 4 of striate cortex. Thus the retinogeniculate pathway is composed of parallel **magnocellular and parvocellular streams** that convey distinct types of information to the initial stages of cortical processing.

The response properties of the M and P ganglion cells provide important clues about the contributions of the magno- and parvocellular streams to visual perception. M ganglion cells have larger receptive fields than P cells, and their axons have faster conduction velocities. M and P ganglion cells also differ in ways that are not so obviously related to their morphology. M cells respond transiently to the presentation of visual stimuli, while P cells respond in a sustained fashion. Moreover, P ganglion cells can transmit information about color, whereas M cells cannot. P cells convey color information because their receptive field centers and surrounds are driven by different classes of cones (i.e., cones responding with greatest sensitivity to

Box C

Optical Imaging of Functional Domains in the Visual Cortex

The recent availability of optical imaging techniques has made it possible to visualize how response properties, such as the selectivity for edge orientation or ocular dominance, are mapped across the cortical surface. These methods generally rely on intrinsic signals (changes in the amount of light reflected from the cortical surface) that correlate with levels of neural activity. Such signals are thought to arise at least in part from local changes in the ratio of oxyhemoglobin and deoxyhemoglobin that accompany such activity, more active areas having a higher deoxyhemoglobin/oxyhemoglobin ratio (see also Box A in Chapter 1). This change can be detected when the cortical surface is illuminated with red light (605–700 nm). Under these conditions, active cortical regions absorb more light than less active ones. With the use of a sensitive video camera, and averaging over a number of trials (the changes are small, 1 or 2 parts per thousand), it is possible to visualize these differences and use them to map cortical patterns of activity (Figure A).

This approach has now been successfully applied to both striate and extrastriate areas in both experimental animals and human patients undergoing neurosurgery. The results emphasize that maps of stimulus features are a general principle of cortical organization. For example, orientation preference is mapped in a continuous fashion such that adjacent positions on the cortical surface tend to have only slightly shifted orientation preferences. However, there are points where continuity breaks down. Around these points, orientation preference is represented in a radial pattern resembling a pinwheel, covering the whole 180° of possible orientation values (Figure B).

This powerful technique can also be used to determine how maps for different stimulus properties are arranged relative to one another, and to detect additional maps such as that for direction of motion. A comparison of ocular dominance bands and orientation preference maps, for example, shows that pinwheel centers are generally located in the center of ocular dominance bands, and that the iso-orientation contours that emanate from the pinwheel centers run orthogonal to the borders of ocular dominance bands (Figure C). An orderly relationship between maps of orientation selectivity and direction selectivity has also been demonstrated. These systematic relationships between the functional maps that coexist within primary visual cortex are thought to ensure that all combinations of stimulus features (orientation, direction, ocular dominance, and spatial frequency) are analyzed for all regions of visual space.

References

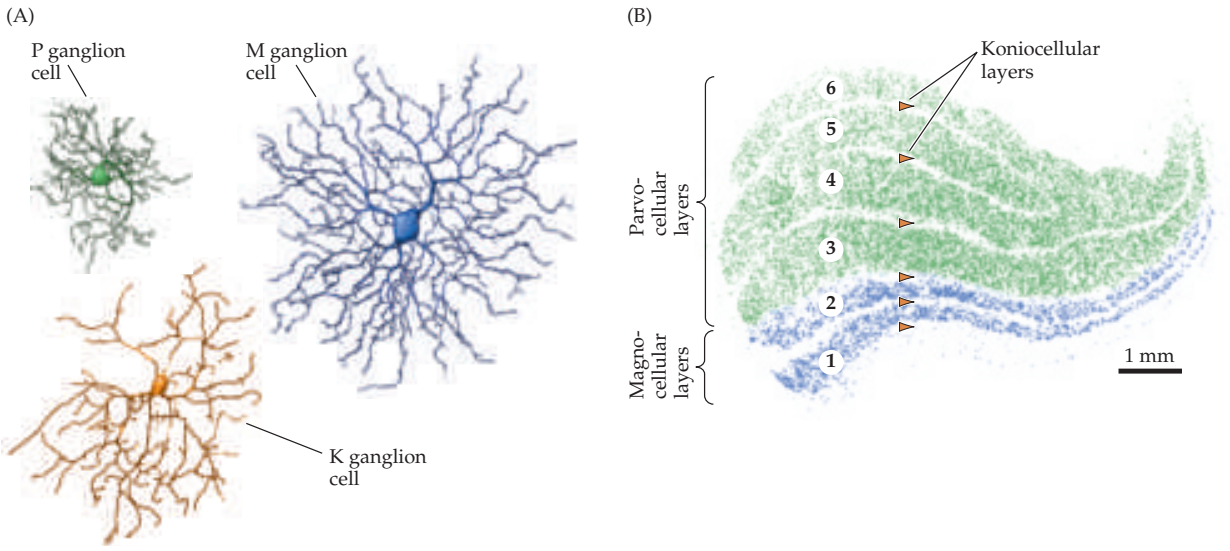
Blasdel, G. G. and G. Salama (1986) Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. Nature 321: 579–585.

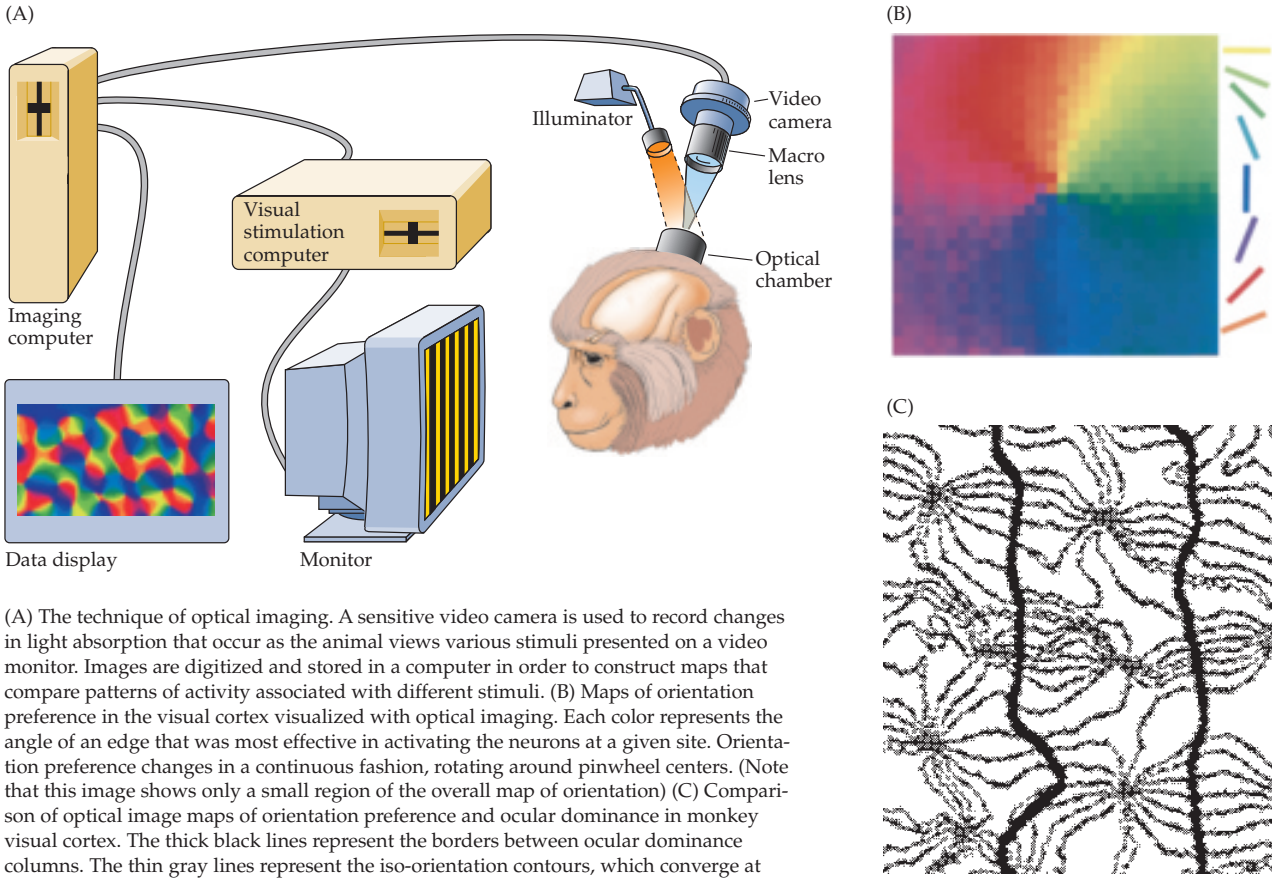
Bonhoeffer, T. and A. Grinvald (1993) The layout of iso-orientation domains in area 18 of the cat visual cortex: Optical imaging reveals a pinwheel-like organization. J. Neurosci 13: 4157–4180.

Bonhoeffer, T. and A. Grinvald (1996) Optical imaging based on intrinsic signals: The methodology. In *Brain Mapping: The Methods*, A. Toge (ed.). New York: Academic Press.

Obermayer, K. and G. G. Blasdel (1993) Geometry of orientation and ocular dominance columns in monkey striate cortex. J. Neurosci. 13: 4114–4129.

Weliky, M., W. H. Bosking and D. Fitzpatrick (1996) A systematic map of direction preference in primary visual cortex. Nature 379: 725–728.





(A) The technique of optical imaging. A sensitive video camera is used to record changes in light absorption that occur as the animal views various stimuli presented on a video monitor. Images are digitized and stored in a computer in order to construct maps that compare patterns of activity associated with different stimuli. (B) Maps of orientation preference in the visual cortex visualized with optical imaging. Each color represents the angle of an edge that was most effective in activating the neurons at a given site. Orientation preference changes in a continuous fashion, rotating around pinwheel centers. (Note that this image shows only a small region of the overall map of orientation) (C) Comparison of optical image maps of orientation preference and ocular dominance in monkey visual cortex. The thick black lines represent the borders between ocular dominance columns. The thin gray lines represent the iso-orientation contours, which converge at orientation pinwheel centers (arrow). Iso-orientation contour lines generally intersect the borders of ocular dominance bands at right angles. (B from Bonhoeffer and Grinvald, 1993; C from Obermeyer and Blasdel, 1993.)

short-, medium-, or long-wavelength light). For example, some P ganglion cells have centers that receive inputs from long-wavelength ("red") sensitive cones and surrounds that receive inputs from medium-wavelength ("green") cones. Others have centers that receive inputs from "green cones" and surrounds from "red cones" (see Chapter 10). As a result, P cells are sensitive to differences in the wavelengths of light striking their receptive field center

◀ **Figure 11.14** Magno- and parvocellular streams. (A) Tracings of M and P ganglion cells as seen in flat mounts of the retina after staining by the Golgi method. M cells have large-diameter cell bodies and large dendritic fields. They supply the magnocellular layers of the lateral geniculate nucleus. P cells have smaller cell bodies and dendritic fields. They supply the parvocellular layers of the lateral geniculate nucleus. (B) Photomicrograph of the human lateral geniculate nucleus showing the magnocellular and parvocellular layers. (A after Watanabe and Rodieck, 1989; B courtesy of T. Andrews and D. Purves.)

and surround. Although M ganglion cells also receive inputs from cones, there is no difference in the type of cone input to the receptive field center and surround; the center and surround of each M cell receptive field is driven by all cone types. The absence of cone specificity to center-surround antagonism makes M cells largely insensitive to differences in the wavelengths of light that strike their receptive field centers and surrounds, and they are thus unable to transmit color information to their central targets.

The contribution of the magno- and parvocellular streams to visual perception has been tested experimentally by examining the visual capabilities of monkeys after selectively damaging either the magno- or parvocellular layers of the lateral geniculate nucleus. Damage to the magnocellular layers has little effect on visual acuity or color vision, but sharply reduces the ability to perceive rapidly changing stimuli. In contrast, damage to the parvocellular layers has no effect on motion perception but severely impairs visual acuity and color perception. These observations suggest that the visual information conveyed by the parvocellular stream is particularly important for high spatial resolution vision—the detailed analysis of the shape, size, and color of objects. The magnocellular stream, on the other hand, appears critical for tasks that require high temporal resolution, such as evaluating the location, speed and direction of a rapidly moving object.

In addition to the magno- and parvocellular streams, a third distinct anatomical pathway—the **koniocellular**, or **K-cell pathway**—has been identified within the lateral geniculate nucleus. Neurons contributing to the K-cell pathway reside in the interlaminar zones that separate lateral geniculate layers; these neurons receive inputs from fine-caliber retinal axons and project in a patchy fashion to the superficial layers (layers II and III) of striate cortex. Although the contribution of the K-cell pathway to perception is not understood, it appears that some aspects of color vision, especially information derived from short-wavelength-sensitive cones, may be transmitted via the K-cell rather than the P-cell pathway. Why short-wavelength-sensitive cone signals should be processed differently from middle- and long-wavelength information is not clear, but the distinction may reflect the earlier evolutionary origin of the K-cell pathway (see Chapter 10).

The Functional Organization of Extrastriate Visual Areas

Anatomical and electrophysiological studies in monkeys have led to the discovery of a multitude of areas in the occipital, parietal, and temporal lobes that are involved in processing visual information (Figure 11.15). Each of these areas contains a map of visual space, and each is largely dependent on the primary visual cortex for its activation. The response properties of the neurons in some of these regions suggest that they are specialized for different aspects of the visual scene. For example, the **middle temporal area** (**MT**) contains neurons that respond selectively to the direction of a moving edge without regard to its color. In contrast, neurons in another cortical area called **V4** respond selectively to the color of a visual stimulus without regard to its direction of movement. These physiological findings are supported by behavioral evidence; thus, damage to area MT leads to a specific impairment in a monkey's ability to perceive the direction of motion in a stimulus pattern, while other aspects of visual perception remain intact.

Recent functional imaging studies have indicated a similar arrangement of visual areas within human extrastriate cortex. Using retinotopically restricted stimuli, it has been possible to localize at least 10 separate repre-

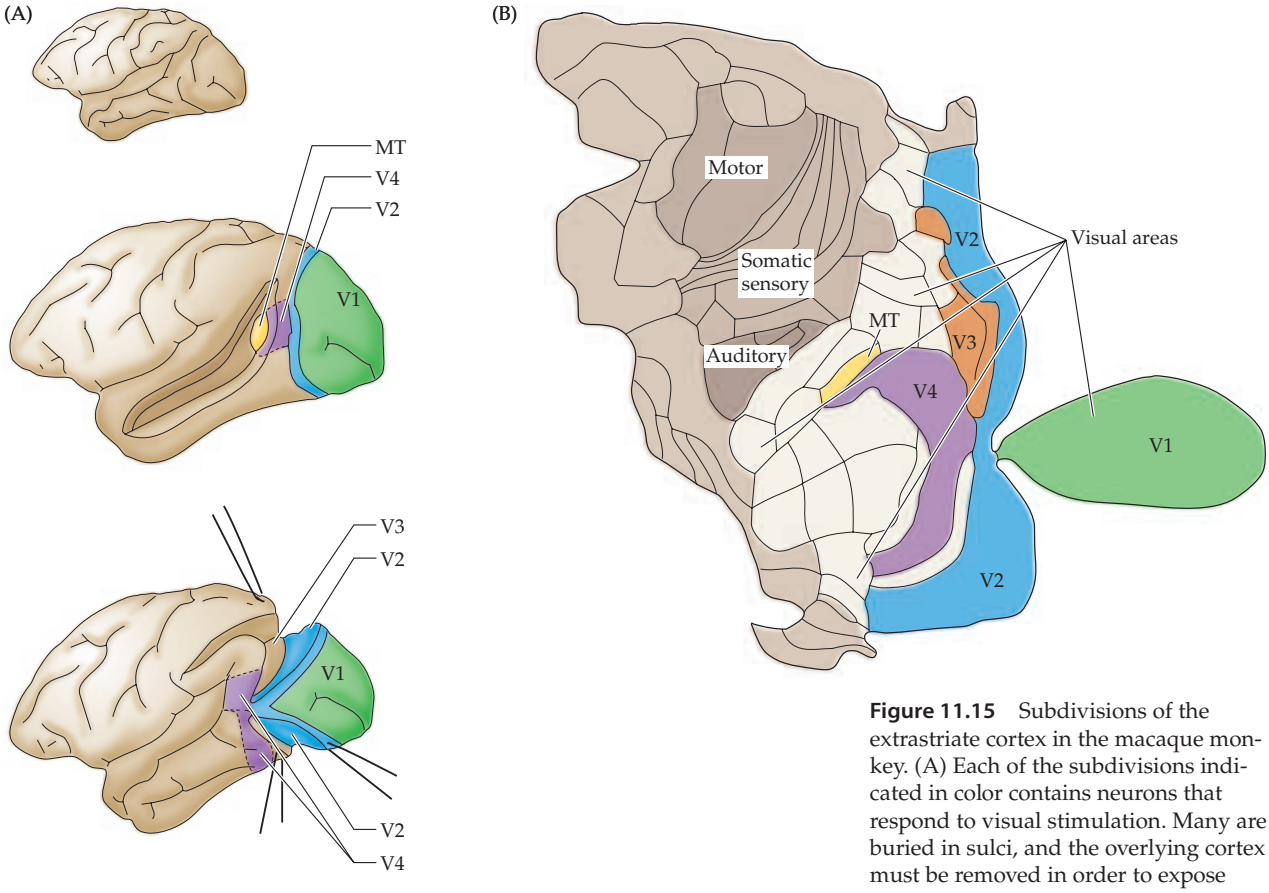


Figure 11.15 Subdivisions of the extrastriate cortex in the macaque monkey. (A) Each of the subdivisions indicated in color contains neurons that respond to visual stimulation. Many are buried in sulci, and the overlying cortex must be removed in order to expose them. Some of the more extensively studied extrastriate areas are specifically identified (V2, V3, V4, and MT). V1 is the primary visual cortex; MT is the middle temporal area. (B) The arrangement of extrastriate and other areas of neocortex in a flattened view of the monkey neocortex. There are at least 25 areas that are predominantly or exclusively visual in function, plus 7 other areas suspected to play a role in visual processing. (A after Maunsell and Newsome, 1987; B after Felleman and Van Essen, 1991.)

sentations of the visual field (Figure 11.16). One of these areas exhibits a large motion-selective signal, suggesting that it is the homologue of the motion-selective middle temporal area described in monkeys. Another area exhibits color-selective responses, suggesting that it may be similar to V4 in non-human primates. A role for these areas in the perception of motion and color, respectively, is further supported by evidence for increases in activity not only during the presentation of the relevant stimulus, but also during periods when subjects experience motion or color afterimages.

The clinical description of selective visual deficits after localized damage to various regions of extrastriate cortex also supports functional specialization of extrastriate visual areas in humans. For example, a well-studied patient who suffered a stroke that damaged the extrastriate region thought to be comparable to area MT in the monkey was unable to appreciate the motion of objects. The neurologist who treated her noted that she had difficulty in pouring tea into a cup because the fluid seemed to be "frozen." In addition, she could not stop pouring at the right time because she was unable to perceive when the fluid level had risen to the brim. The patient also had trouble following a dialogue because she could not follow the movements of the speaker's mouth. Crossing the street was potentially terrifying because she couldn't judge the movement of approaching cars. As the patient related, "When I'm looking at the car first, it seems far away. But

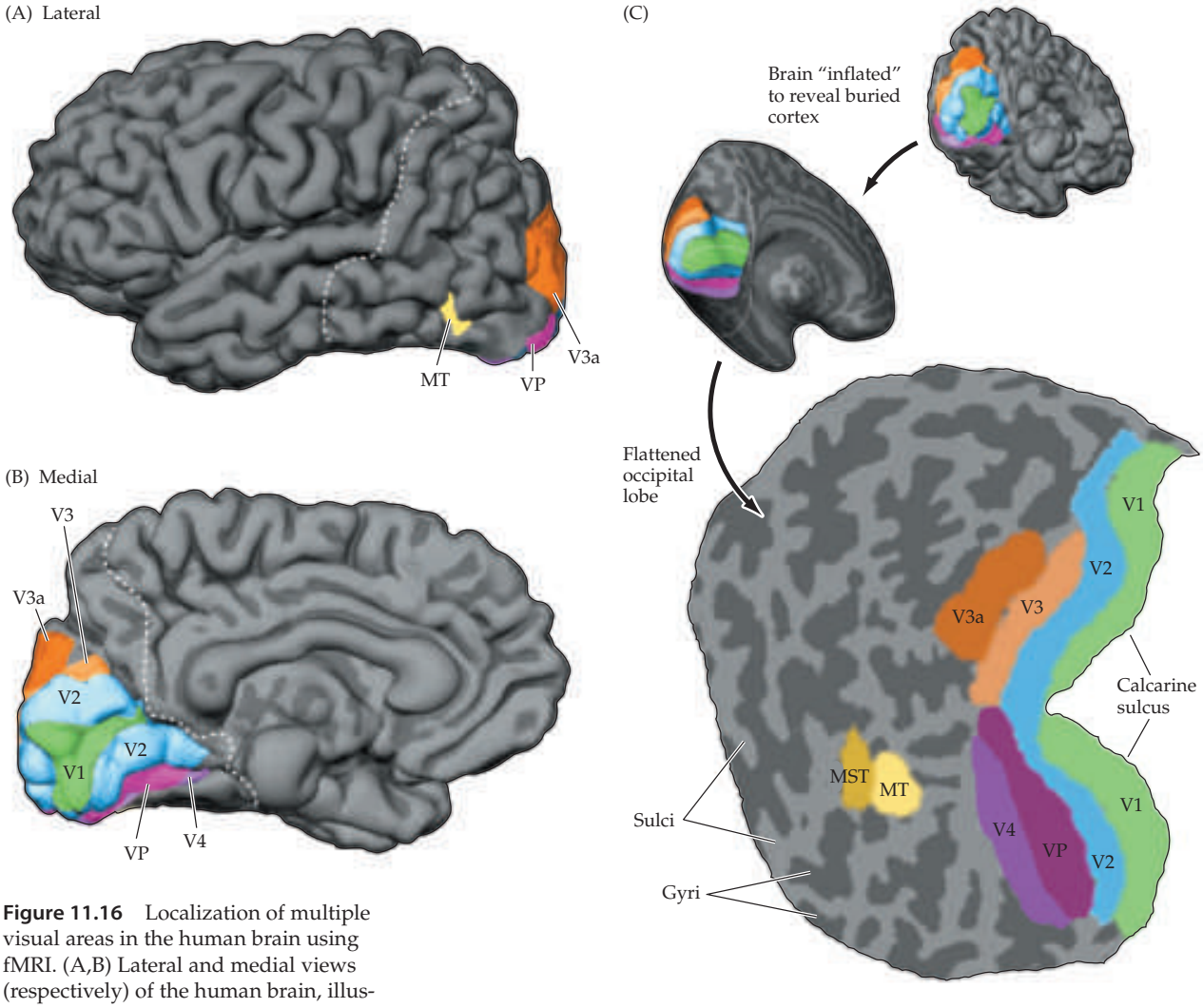


Figure 11.16 Localization of multiple visual areas in the human brain using fMRI. (A,B) Lateral and medial views (respectively) of the human brain, illustrating the location of primary visual cortex (V1) and additional visual areas V2, V3, VP (ventral posterior area), V4, MT (middle temporal area), and MST (medial superior temporal area). (C) Unfolded and flattened view of retinotopically defined visual areas in the occipital lobe. Dark grey areas correspond to cortical regions that were buried in sulci; light regions correspond to regions that were located on the surface of gyri. Visual areas in humans show a close resemblance to visual areas originally defined in monkeys (compare with Figure 11.15). (After Sereno et al., 1995.)

then, when I want to cross the road, suddenly the car is very near." Her ability to perceive other features of the visual scene, such as color and form, was intact.

Another example of a specific visual deficit as a result of damage to extrastriate cortex is **cerebral achromatopsia**. These patients lose the ability to see the world in color, although other aspects of vision remain in good working order. The normal colors of a visual scene are described as being replaced by "dirty" shades of gray, much like looking at a poor quality black-and-white movie. Achromatopsic individuals know the normal colors of objects—that a school bus is yellow, an apple red—but can no longer see them. Thus, when asked to draw objects from memory, they have no difficulty with shapes but are unable to appropriately color the objects they have represented. It is important to distinguish this condition from the color blindness that arises from the congenital absence of one or more cone pigments in the retina (see Chapter 10). In achromatopsia, the three types of cones are functioning normally; it is damage to specific extrastriate cortical areas that renders the patient unable to use the information supplied by the retina.

Based on the anatomical connections between visual areas, differences in electrophysiological response properties, and the effects of cortical lesions, a consensus has emerged that extrastriate cortical areas are organized into two largely separate systems that eventually feed information into cortical association areas in the temporal and parietal lobes (see Chapter 25). One system, called the ventral stream, includes area V4 and leads from the striate cortex into the inferior part of the temporal lobe. This system is thought to be responsible for high-resolution form vision and object recognition. The dorsal stream, which includes the middle temporal area, leads from striate cortex into the parietal lobe. This system is thought to be responsible for spatial aspects of vision, such as the analysis of motion, and positional relationships between objects in the visual scene (Figure 11.17).

The functional dichotomy between these two streams is supported by observations on the response properties of neurons and the effects of selective cortical lesions. Neurons in the ventral stream exhibit properties that are important for object recognition, such as selectivity for shape, color, and texture. At the highest levels in this pathway, neurons exhibit even greater selectivity, responding preferentially to faces and objects (see Chapter 25). In contrast, those in the dorsal stream are not tuned to these properties, but show selectivity for direction and speed of movement. Consistent with this interpretation, lesions of the parietal cortex severely impair an animal's ability to distinguish objects on the basis of their position, while having little effect on its ability to perform object recognition tasks. In contrast, lesions of the inferotemporal cortex produce profound impairments in the ability to perform recognition tasks but no impairment in spatial tasks. These effects are remarkably similar to the syndromes associated with damage to the parietal and temporal lobe in humans (see Chapters 25 and 26).

What, then, is the relationship between these "higher-order" extrastriate visual pathways and the magno- and parvocellular pathways that supply the primary visual cortex? Not long ago, it seemed that these intracortical pathways were simply a continuation of the geniculostriate pathways—that is, the magnocellular pathway provided input to the dorsal stream and the parvocellular pathway provided input to the ventral stream. However, more recent work has indicated that the situation is more complicated. The temporal pathway clearly has access to the information conveyed by both the magno- and parvocellular streams; and the parietal pathway, while dominated by inputs from the magnocellular stream, also receives inputs from the parvocellular stream. Thus, interaction and cooperation between the magno- and parvocellular streams appear to be the rule in complex visual perceptions.

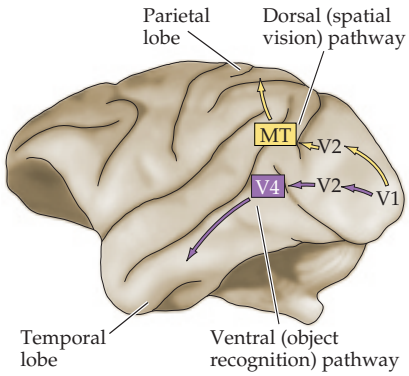


Figure 11.17 The visual areas beyond the striate cortex are broadly organized into two pathways: a ventral pathway that leads to the temporal lobe, and a dorsal pathway that leads to the parietal lobe. The ventral pathway plays an important role in object recognition, the dorsal pathway in spatial vision.

Summary

Distinct populations of retinal ganglion cells send their axons to a number of central visual structures that serve different functions. The most important projections are to the pretectum for mediating the pupillary light reflex, to the hypothalamus for the regulation of circadian rhythms, to the superior colliculus for the regulation of eye and head movements, and—most important of all—to the lateral geniculate nucleus for mediating vision and visual perception. The retinogeniculostriate projection (the primary visual pathway) is arranged topographically such that central visual structures contain an organized map of the contralateral visual field. Damage anywhere along the primary visual pathway, which includes the optic nerve, optic tract, lateral geniculate nucleus, optic radiation, and striate cortex, results in a loss of vision confined to a predictable region of visual space. Compared to retinal

ganglion cells, neurons at higher levels of the visual pathway become increasingly selective in their stimulus requirements. Thus, most neurons in the striate cortex respond to light–dark edges only if they are presented at a certain orientation; some are selective for the length of the edge, and others to movement of the edge in a specific direction. Indeed, a point in visual space is related to a set of cortical neurons, each of which is specialized for processing a limited set of the attributes in the visual stimulus. The neural circuitry in the striate cortex also brings together information from the two eyes; most cortical neurons (other than those in layer IV, which are segregated into eye-specific columns) have binocular responses. Binocular convergence is presumably essential for the detection of binocular disparity, an important component of depth perception. The primary visual pathway is composed of separate functional streams that convey information from different types of retinal ganglion cells to the initial stages of cortical processing. The magnocellular stream conveys information that is critical for the detection of rapidly changing stimuli, the parvocellular stream mediates high acuity vision and appears to share responsibility for color vision with the koniocellular stream. Finally, beyond striate cortex, parcellation of function continues in the ventral and dorsal streams that lead to the extrastriate and association areas in the temporal and parietal lobes, respectively. Areas in the inferotemporal cortex are especially important in object recognition, whereas areas in the parietal lobe are critical for understanding the spatial relations between objects in the visual field.

Additional Reading

Reviews

Berson, D. M. (2003) Strange vision: Ganglion cells as circadian photoreceptors. Trends Neurosci. 26: 314–320.

Courtney, S. M. and L. G. Ungerleider (1997) What fMRI has taught us about human vision. Curr. Op. Neurobiol. 7: 554–561.

Felleman, D. J. and D. C. Van Essen (1991) Distributed hierarchical processing in primate cerebral cortex. Cerebral Cortex 1: 1–47.

Horton, J. C. (1992) The central visual pathways. In *Alder's Physiology of the Eye*. W. M. Hart (ed.). St. Louis: Mosby Yearbook.

Hendry, S. H. and R. C. Reid (2000) The koniocellular pathway in primate vision. Annu. Rev. Neurosci. 23: 127–153.

Hubel, D. H. and T. N. Wiesel (1977) Functional architecture of macaque monkey visual cortex. Proc. R. Soc. (Lond.) 198: 1–59.

Maunsell, J. H. R. (1992) Functional visual streams. Curr. Opin. Neurobiol. 2: 506–510.

Schiller, P. H. and N. K. Logothetis (1990) The color-opponent and broad-band channels of the primate visual system. Trends Neurosci. 13: 392–398.

Tootell, R.B., A. M. Dale, M. I. Sereno and R. Malach (1996) New images from human visual cortex. Trends Neurosci. 19: 481–489.

Ungerleider, J. G. and M. Mishkin (1982) Two cortical visual systems. In *Analysis of Visual Behavior*. D. J. Ingle, M. A. Goodale and R. J. W. Mansfield (eds.). Cambridge, MA: MIT Press, pp. 549–586.

Important Original Papers

Hattar, S., H. W. Liao, M. Takao, D. M. Berson and K. W. Yau (2002) Melanopsin-containing retinal ganglion cells: Architecture, projections, and intrinsic photosensitivity. Science 295: 1065–1070.

Hubel, D. H. and T. N. Wiesel (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. (Lond.) 160: 106–154.

Hubel, D. H. and T. N. Wiesel (1968) Receptive fields and functional architecture of monkey striate cortex. J. Physiol. (Lond.) 195: 215–243.

Sereno, M. I. and 7 Others (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268: 889–893.

Zihl, J., D. von Cramon and N. Mai (1983) Selective disturbance of movement vision after bilateral brain damage. Brain 106: 313–340.

Books

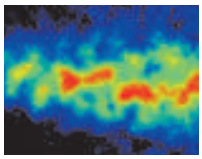
Chalupa, L. M. and J. S. Werner (eds.) (2004) *The Visual Neurosciences*. Cambridge, MA: MIT Press.

Hubel, D. H. (1988) *Eye, Brain, and Vision*. New York: Scientific American Library.

Rodieck, R. W. (1998) *The First Steps in Seeing*. Sunderland, MA: Sinauer Associates.

Zeki, S. (1993) *A Vision of the Brain*. Oxford: Blackwell Scientific Publications.

Chapter 12



The Auditory System

Overview

The auditory system is one of the engineering masterpieces of the human body. At the heart of the system is an array of miniature acoustical detectors packed into a space no larger than a pea. These detectors can faithfully transduce vibrations as small as the diameter of an atom, and they can respond a thousand times faster than visual photoreceptors. Such rapid auditory responses to acoustical cues facilitate the initial orientation of the head and body to novel stimuli, especially those that are not initially within the field of view. Although humans are highly visual creatures, much human communication is mediated by the auditory system; indeed, loss of hearing can be more socially debilitating than blindness. From a cultural perspective, the auditory system is essential not only to understanding speech, but also to music, one of the most aesthetically sophisticated forms of human expression. For these and other reasons, audition represents a fascinating and especially important mode of sensation.

Sound

In physical terms, *sound* refers to pressure waves generated by vibrating air molecules (somewhat confusingly, sound is used more casually to refer to an auditory percept). Sound waves are much like the ripples that radiate outward when a rock is thrown in a pool of water. However, instead of occurring across a two-dimensional surface, sound waves propagate in three dimensions, creating spherical shells of alternating compression and rarefaction. Like all wave phenomena, sound waves have four major features: **waveform**, **phase**, **amplitude** (usually expressed in log units known as decibels, abbreviated dB), and **frequency** (expressed in cycles per second or Hertz, abbreviated Hz). For human listeners, the amplitude and frequency of a sound pressure change at the ear roughly correspond to **loudness** and **pitch**, respectively.

The waveform of a sound stimulus is its amplitude plotted against time. It helps to begin by visualizing an acoustical waveform as a sine wave. At the same time, it must be kept in mind that sounds composed of single sine waves (i.e., pure tones) are extremely rare in nature; most sounds in speech, for example, consist of acoustically complex waveforms. Interestingly, such complex waveforms can often be modeled as the sum of sinusoidal waves of varying amplitudes, frequencies, and phases. In engineering applications, an algorithm called the Fourier transform decomposes a complex signal into its sinusoidal components. In the auditory system, as will be apparent later in the chapter, the inner ear acts as a sort of acoustical prism, decomposing complex sounds into a myriad of constituent tones.

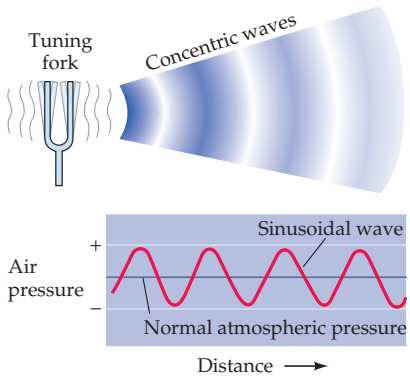


Figure 12.1 Diagram of the periodic condensation and rarefaction of air molecules produced by the vibrating tines of a tuning fork. The molecular disturbance of the air is pictured as if frozen at the instant the constituent molecules responded to the resultant pressure wave. Shown below is a plot of the air pressure versus distance from the fork. Note its sinusoidal quality.

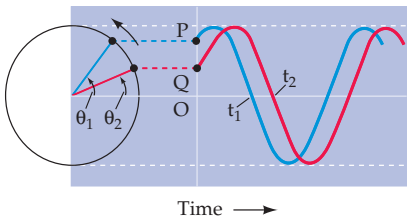


Figure 12.2 A sine wave and its projection as circular motion. The two sinusoids shown are at different phases, such that point P corresponds to phase angle θ_1 and point Q corresponds to phase angle θ_2.

Figure 12.1 diagrams the behavior of air molecules near a tuning fork that vibrates sinusoidally when struck. The vibrating tines of the tuning fork produce local displacements of the surrounding molecules, such that when the tine moves in one direction, there is molecular condensation; when it moves in the other direction, there is rarefaction. These changes in density of the air molecules are equivalent to local changes in air pressure.

Such regular, sinusoidal cycles of compression and rarefaction can be thought of as a form of circular motion, with one complete cycle equivalent to one full revolution (360°). This point can be illustrated with two sinusoids of the same frequency projected onto a circle, a strategy that also makes it easier to understand the concept of phase (Figure 12.2). Imagine that two tuning forks, both of which resonate at the same frequency, are struck at slightly different times. At a given time $t = 0$, one wave is at position P and the other at position Q. By projecting P and Q onto the circle, their respective phase angles, θ_1 and θ_2, are apparent. The sine wave that starts at P reaches a particular point on the circle, say 180°, at time t_1, whereas the wave that starts at Q reaches 180° at time t_2. Thus, phase differences have corresponding time differences, a concept that is important in appreciating how the auditory system locates sounds in space.

The human ear is extraordinarily sensitive to sound pressure. At the threshold of hearing, air molecules are displaced an average of only 10 picometers (10^{-11} m), and the intensity of such a sound is about one-trillionth of a watt per square meter! This means a listener on an otherwise noiseless planet could hear a 1-watt, 3-kHz sound source located over 450 km away (consider that even a very dim light bulb consumes more than 1 watt of power). Even dangerously high sound pressure levels (>100 dB) have power at the eardrum that is only in the milliwatt range (Box A).

The Audible Spectrum

Humans can detect sounds in a frequency range from about 20 Hz to 20 kHz. Human infants can actually hear frequencies slightly higher than 20 kHz, but lose some high-frequency sensitivity as they mature; the upper limit in average adults is closer to 15–17 kHz. Not all mammalian species are sensitive to the same range of frequencies. Most small mammals are sensitive to very high frequencies, but not to low frequencies. For instance, some species of bats are sensitive to tones as high as 200 kHz, but their lower limit is around 20 kHz—the upper limit for young people with normal hearing.

One reason for these differences is that small objects, including the auditory structures of these small mammals, resonate at high frequencies, whereas large objects tend to resonate at low frequencies—which explains why the violin has a higher pitch than the cello. Different animal species tend to emphasize frequency bandwidths in both their vocalizations and their range of hearing. In general, vocalizations by virtue of their periodicity can be distinguished from the noise "barrier" created by environmental sounds, such as wind and rustling leaves. Animals that echolocate, such as bats and dolphins, rely on very high-frequency vocal sounds to maximally resolve spatial features of the target, while animals intent on avoiding predation have auditory systems "tuned" to the low frequency vibrations that approaching predators transmit through the substrate. These behavioral differences are mirrored by a wealth of anatomical and functional specializations throughout the auditory system.

Box A

Four Causes of Acquired Hearing Loss

Acquired hearing loss is an increasingly common sensory deficit that can often lead to impaired oral communication and social isolation. Four major causes of acquired hearing loss are acoustical trauma, infection of the inner ear, ototoxic drugs, and presbyacusis (literally, the hearing of the old).

The exquisite sensitivity of the auditory periphery, combined with the direct mechanical linkage between the acoustical stimulus and the receptor cells, make the ear especially susceptible to acute or chronic acoustical trauma. Extremely loud, percussive sounds, such as those generated by explosives or gunfire, can rupture the eardrum and so severely distort the inner ear that the organ of Corti is torn. The resultant loss of hearing is abrupt and often quite severe. Less well appreciated is the fact that repeated exposure to less dramatic but nonetheless loud sounds, including those produced by industrial or household machinery or by amplified musical instruments, can also damage the inner ear. Although these sounds leave the eardrum intact, specific damage is done to the hair bundle itself; the stereocilia of cochlear hair cells of animals exposed to loud sounds shear off at their pivot points with the hair cell body, or fuse together in a platelike fashion that impedes movement. In humans, the mechanical resonance of the ear to stimulus frequencies centered about 3 kHz means that exposure to loud, broadband noises (such as those generated by jet engines) results in especially pronounced deficits near this resonant frequency.

Ototoxic drugs include aminoglycoside antibiotics (such as gentamycin and kanamycin), which directly affect hair cells, and ethacrynic acid, which poisons the potassium-extruding cells of the stria vascularis that generate the endocochlear potential. In the absence of these ion pumping cells, the endocochlear potential, which supplies the energy to drive the transduction process, is lost. Although still a matter of some debate, the relatively nonselective transduction channel apparently affords a means of entry for aminoglycoside antibiotics, which then poison hair cells by disrupting phosphoinositide metabolism. In particular, outer hair cells and those inner hair cells that transduce high-frequency stimuli are more affected, simply because of their greater energy requirements.

Finally, *presbyacusis*, the hearing loss associated with aging, may in part stem from atherosclerotic damage to the especially fine microvasculature of the inner ear, as well as from genetic predispositions to hair cell damage. Recent advances in understanding the genetic transmission of acquired hearing loss in both humans and mice point to mutations in myosin isoforms unique to hair cells as a likely culprit.

References

Holt, J. R. and D. P. Corey (1999) Ion channel defects in hereditary hearing loss. Neuron 22: 217–219.

Keats, B. J. and D. P. Corey (1999) The usher syndromes. Amer. J. Med. Gen. 89: 158–166.

Priuska, E. M. and J. Schact (1997) Mechanism and prevention of aminoglycoside ototoxicity: Outer hair cells as targets and tools. Ear, Nose, Throat J. 76: 164–171.

A Synopsis of Auditory Function

The auditory system transforms sound waves into distinct patterns of neural activity, which are then integrated with information from other sensory systems to guide behavior, including orienting movements to acoustical stimuli and intraspecies communication. The first stage of this transformation occurs at the external and middle ears, which collect sound waves and amplify their pressure, so that the sound energy in the air can be successfully transmitted to the fluid-filled cochlea of the inner ear. In the inner ear, a series of biomechanical processes occur that break up the signal into simpler, sinusoidal components, with the result that the frequency, amplitude, and phase of the original signal are all faithfully transduced by the sensory **hair cells** and encoded by the electrical activity of the **auditory nerve fibers**. One product of this process of acoustical decomposition is the systematic representation of sound frequency along the length of the cochlea, referred to as **tonotopy**, which is an important organizational feature preserved

Box B
Music

Even though we all recognize it when we hear it, the concept of music is vague. The *Oxford English Dictionary* defines it as "The art or science of combining vocal or instrumental sounds with a view toward beauty or coherence of form and expression of emotion." In terms of the present chapter, music chiefly concerns the aspect of human audition that is experienced as tones. The stimuli that give rise to tonal percepts are periodic, meaning that they repeat systematically over time, as in the sine wave in Figure 12.1. Periodic stimuli, which do not occur naturally as sine waves but rather as complex repetitions involving a number of different frequencies, give rise to a sense of harmony when sounded together in appropriate combinations, and a sense of melody when they occur sequentially.

Although we usually take the way tone-evoking stimuli are heard for granted, this aspect of audition presents some profoundly puzzling qualities. The most obvious of these is that humans perceive periodic stimuli whose fundamental frequencies have a 2:1 ratio as highly similar, and, for the most part, musically interchangeable. Thus in Western musical terminology, any two tones related by an interval of one or more octaves are given the same name (i.e., A, B, C…G), and are distinguished only by a qualifier that denotes relative ordinal position (e.g., C_1, C_2, C_3, etc.). As a result, music is framed in repeating intervals (called octaves) defined by these more or less interchangeable tones. A key question, then, is why periodic sound stimuli whose fundamentals have a 2:1 ratio are perceived as similar when there is no obvious physical or physiological basis for this phenomenon.

A second puzzling feature is that most if not all musical traditions subdivide octaves into a relatively small set of intervals for composition and performance, each interval being defined by its relationship to the lowest tone of the set. Such sets are called musical scales. The scales predominantly employed in all cultures over the centuries have used some (or occasionally all) of the 12 tonal intervals that in Western musical terminology are referred to as the chromatic scale (see figure). Moreover, some intervals of the chromatic scale, such as the fifth, the fourth, the major third, and the major sixth, are more often used in composition and performance than others. These form the majority of the intervals employed in the pentatonic and diatonic major scales, the two most frequently used scales in music world-wide. Again,

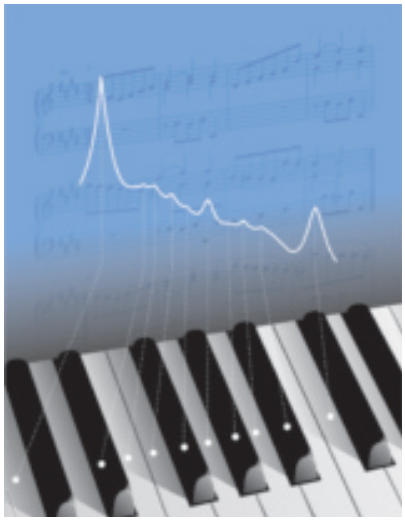


Illustration of 10 of the 12 tones in the chromatic scale, related to a piano keyboard. The function above the keyboard indicates that these tones correspond statistically to peaks of power in normalized human speech. (After Schwartz et al., 2003.)

throughout the central auditory pathways. The earliest stage of central processing occurs at the cochlear nucleus, where the peripheral auditory information diverges into a number of parallel central pathways. Accordingly, the output of the cochlear nucleus has several targets. One of these is the superior olivary complex, the first place that information from the two ears interacts and the site of the initial processing of the cues that allow listeners to localize sound in space. The cochlear nucleus also projects to the inferior colliculus of the midbrain, a major integrative center and the first place where auditory information can interact with the motor system. The inferior colliculus is an obligatory relay for information traveling to the thalamus and cortex, where additional integrative aspects (such as harmonic and temporal combinations) of sound especially germane to speech and music are processed (Box B). The large number of stations between the auditory periphery and the cortex far exceeds those in other sensory systems, providing a hint that the perception of communication and environmental sounds

there is no principled explanation of these preferences among all the possible intervals within the octave.

Perhaps the most fundamental question in music—and arguably the common denominator of all musical tonality—is why certain combinations of tones are perceived as relatively consonant or 'harmonious' and others relatively dissonant or 'inharmonious'. These perceived differences among the possible combinations of tones making up the chromatic scale are the basis for polytonal music, in which the perception of relative harmoniousness guides the composition of chords and melodic lines. The more compatible of these combinations are typically used to convey 'resolution' at the end of a musical phrase or piece, whereas less compatible combinations are used to indicate a transition, a lack of resolution, or to introduce a sense of tension in a chord or melodic sequence. Like octaves and scales, the reason for this phenomenology remains a mystery.

The classical approaches to rationalizing octaves, scales and consonance have been based on the fact that the musical intervals corresponding to octaves, fifths, and fourths (in modern musical terminology) are produced by physical sources whose relative proportions (e.g., the relative lengths of two plucked strings or their fundamental frequencies) have ratios of 2:1, 3:2, or 4:3, respectively (these relationships were first described by Pythagoras). This coincidence of numerical simplicity and perceptual effect has been so impressive over the centuries that attempts to rationalize phenomena such as consonance and scale structure in terms of mathematical relationships have tended to dominate the thinking about these issues. This conceptual framework, however, fails to account for many of the perceptual observations that have been made over the last century.

Another way to consider the problem is in terms of the biological rationale for evolving a sense of tonality in the first place. A pertinent fact in this regard is that only a small minority of naturally occurring sound stimuli are periodic. Since the auditory system evolved in the world of natural sounds, this point is presumably critical for thinking about the biological purposes of tonality and music. Indeed, the majority of periodic sounds that humans would have been exposed to during evolution are those made by the human vocal tract in the process of communication, initially prelinguistic but more recently speech sounds (see Chapter 26). Thus developing a sense of tonality would enable listeners to respond not only the distinctions among the different speech sounds that are important for understanding spoken language, but to information about the probable sex, age, and emotional state of the speaker. It may thus be that music reflects the advantage of facilitating a listener's ability to glean the linguistic intent and biological state of fellow humans through vocal utterances.

References

Burns, E. M. (1999) Intervals, scales, and tuning. In *The Psychology of Music*, D. Deutsch (ed.). New York: Academic Press, pp. 215–264.

Carterette, E. C. and R. A. Kendall (1999) Comparative music perception and cognition. In *The Psychology of Music*, D. Deutsch (ed.). New York: Academic Press.

Lewicki, M. S. (2002) Efficient coding of natural sounds. Nature Neurosci. 5: 356–363.

Pierce, J. R. (1983, 1992) *The Science of Musical Sound*. New York: W.H. Freeman and Co., Chapters 4–6.

Plomp, R. and W. J. Levelt (1965) Tonal consonance and critical bandwidth. *J. Acoust. Soc. Amer.* 28: 548–560.

Rasch, R. and R. Plomp (1999) The perception of musical tones. In *The Psychology of Music*, D. Deutsch (ed.). New York: Academic Press, pp. 89–112.

Schwartz, D. A., C. Q. Howe and D. Purves (2003) The statistical structure of human speech sounds predicts musical universals. J. Neurosci. 23: 7160–7168.

Terhardt, E. (1974) Pitch, consonance, and harmony. *J. Acoust. Soc. Amer.* 55: 1061–1069.

is an especially intensive neural process. Furthermore, both the peripheral and central auditory system are "tuned" to conspecific communication vocalizations, pointing to the interdependent evolution of neural systems used for generating and perceiving these signals.

The External Ear

The external ear, which consists of the **pinna**, **concha**, and **auditory meatus**, gathers sound energy and focuses it on the eardrum, or **tympanic membrane** (Figure 12.3). One consequence of the configuration of the human auditory meatus is that it selectively boosts the sound pressure 30- to 100-fold for frequencies around 3 kHz via passive resonance effects. This amplification makes humans especially sensitive to frequencies in the range of 2–5 kHz—and also explains why they are particularly prone to hearing loss near this frequency following exposure to loud broadband noises, such as those

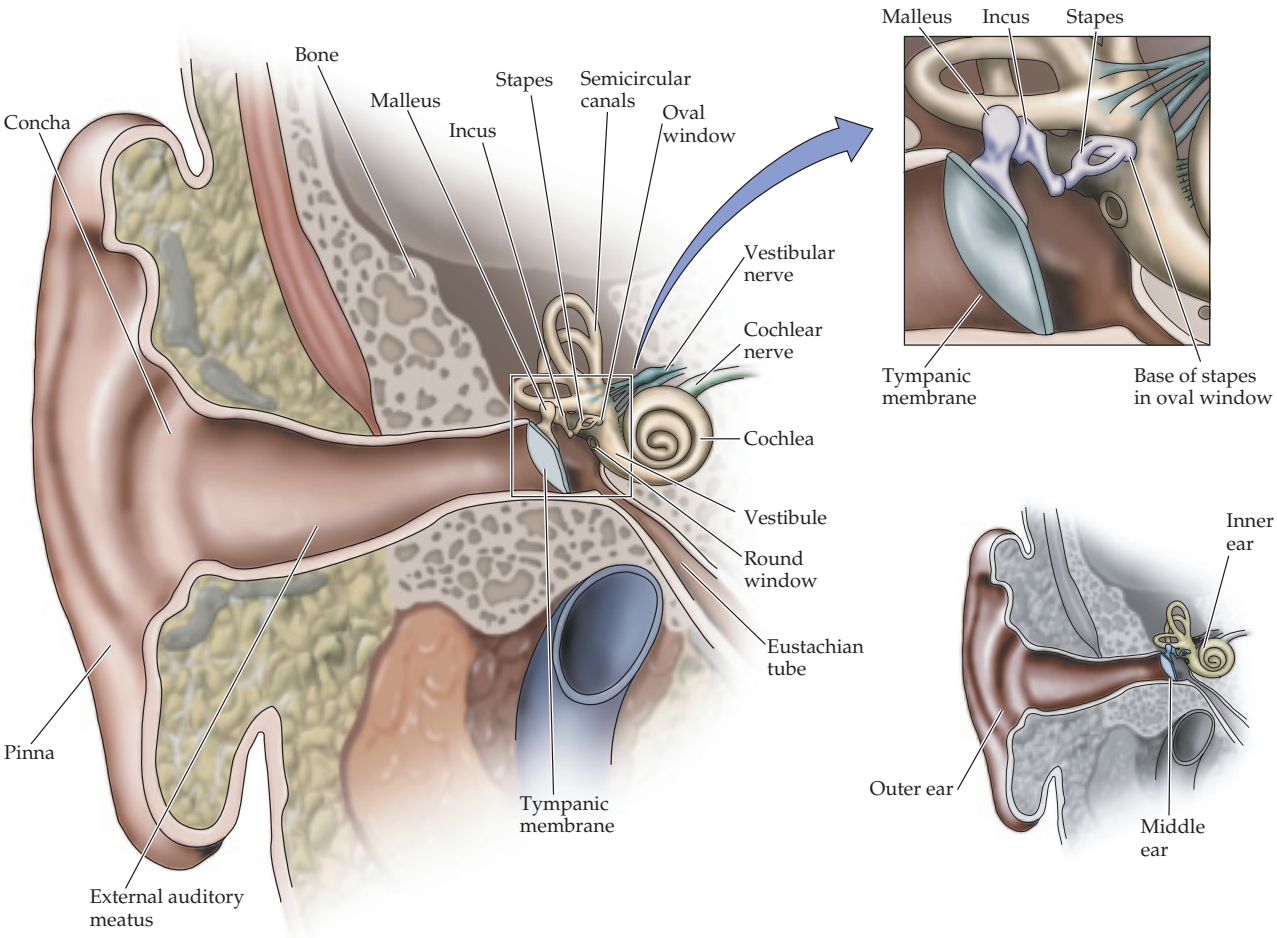


Figure 12.3 The human ear. Note the large surface area of the tympanic membrane (eardrum) relative to the oval window, a feature that facilitates transmission of airborne sounds to the fluid-filled cochlea.

generated by heavy machinery or high explosives (see Box A). The sensitivity to this frequency range in the human auditory system appears to be directly related to speech perception: although human speech is a broadband signal, the energy of the plosive consonants (e.g., *ba* and *pa*) that distinguish different phonemes (the elementary human speech sounds) is concentrated around 3 kHz (see Box A in Chapter 26). Therefore, selective hearing loss in the 2–5 kHz range disproportionately degrades speech recognition. Most vocal communication occurs in the low-kHz range to overcome environmental noise; as already noted, generation of higher frequencies is difficult for animals the size of humans.

A second important function of the pinna and concha is to selectively filter different sound frequencies in order to provide cues about the elevation of the sound source. The vertically asymmetrical convolutions of the pinna are shaped so that the external ear transmits more high-frequency components from an elevated source than from the same source at ear level. This effect can be demonstrated by recording sounds from different elevations after they have passed through an "artificial" external ear; when the recorded sounds are played back via earphones, so that the whole series is at the same elevation relative to the listener, the recordings from higher elevations are perceived as coming from positions higher in space than the recordings from lower elevations.

The Middle Ear

Sounds impinging on the external ear are airborne; however, the environment within the inner ear, where the sound-induced vibrations are converted to neural impulses, is aqueous. The major function of the middle ear is to match relatively low-impedance airborne sounds to the higher-impedance fluid of the inner ear. The term "impedance" in this context describes a medium's resistance to movement. Normally, when sound waves travel from a low-impedance medium like air to a much higher-impedance medium like water, almost all (more than 99.9%) of the acoustical energy is reflected. The middle ear (see Figure 12.3) overcomes this problem and ensures transmission of the sound energy across the air–fluid boundary by boosting the pressure measured at the tympanic membrane almost 200-fold by the time it reaches the inner ear.

Two mechanical processes occur within the middle ear to achieve this large pressure gain. The first and major boost is achieved by focusing the force impinging on the relatively large-diameter tympanic membrane on to the much smaller-diameter **oval window**, the site where the bones of the middle ear contact the inner ear. A second and related process relies on the mechanical advantage gained by the lever action of the three small interconnected middle ear bones, or **ossicles** (i.e., the malleus, incus, and stapes; see Figure 12.3), which connect the tympanic membrane to the oval window. **Conductive hearing losses**, which involve damage to the external or middle ear, lower the efficiency at which sound energy is transferred to the inner ear and can be partially overcome by artificially boosting sound pressure levels with an external hearing aid (Box C). In normal hearing, the efficiency of sound transmission to the inner ear also is regulated by two small muscles in the middle ear, the tensor tympani, innervated by cranial nerve V, and the stapedius, innervated by cranial nerve VII (see Appendix A). Flexion of these muscles, which is triggered automatically by loud noises or during self-generated vocalization, stiffens the ossicles and reduces the amount of sound energy transmitted to the cochlea, serving to protect the inner ear. Conversely, conditions that lead to flaccid paralysis of either of these muscles, such as Bell's palsy (nerve VII), can trigger a painful sensitivity to moderate or even low intensity sounds known as **hyperacusis**.

Bony and soft tissues, including those surrounding the inner ear, have impedances close to that of water. Therefore, even without an intact tympanic membrane or middle ear ossicles, acoustical vibrations can still be transferred directly through the bones and tissues of the head to the inner ear. In the clinic, bone conduction can be exploited using a simple test involving a tuning fork to determine whether hearing loss is due to conductive problems or is due to damage to the hair cells of the inner ear or to the auditory nerve itself (**sensorineural hearing loss**; see Boxes A and C)

The Inner Ear

The **cochlea** of the inner ear is arguably the most critical structure in the auditory pathway, for it is there that the energy from sonically generated pressure waves is transformed into neural impulses. The cochlea not only amplifies sound waves and converts them into neural signals, but it also acts as a mechanical frequency analyzer, decomposing complex acoustical waveforms into simpler elements. Many features of auditory perception derive from aspects of the physical properties of the cochlea; hence, it is important to consider this structure in some detail.

Box C

Sensorineural Hearing Loss and Cochlear Implants

The same features that make the auditory periphery exquisitely sensitive to detecting airborne sounds also make it highly vulnerable to damage. By far the most common forms of hearing loss involve the peripheral auditory system, namely to those structures that transmit and transduce sounds into neural impulses. Monaural hearing deficits are the defining symptom of a peripheral hearing loss, because unilateral damage at or above the auditory brainstem results in a binaural deficit (due to the extensive bilateral organization of the central auditory system). Peripheral hearing insults can be further divided into conductive hearing losses, which involve damage to the outer or middle ear, and sensorineural hearing losses, which stem from damage to the inner ear, most typically the cochlear hair cells or the VIIIth nerve itself. Although both forms of peripheral hearing loss manifest themselves as a raised threshold for hearing on the affected side, their diagnoses and treatments differ.

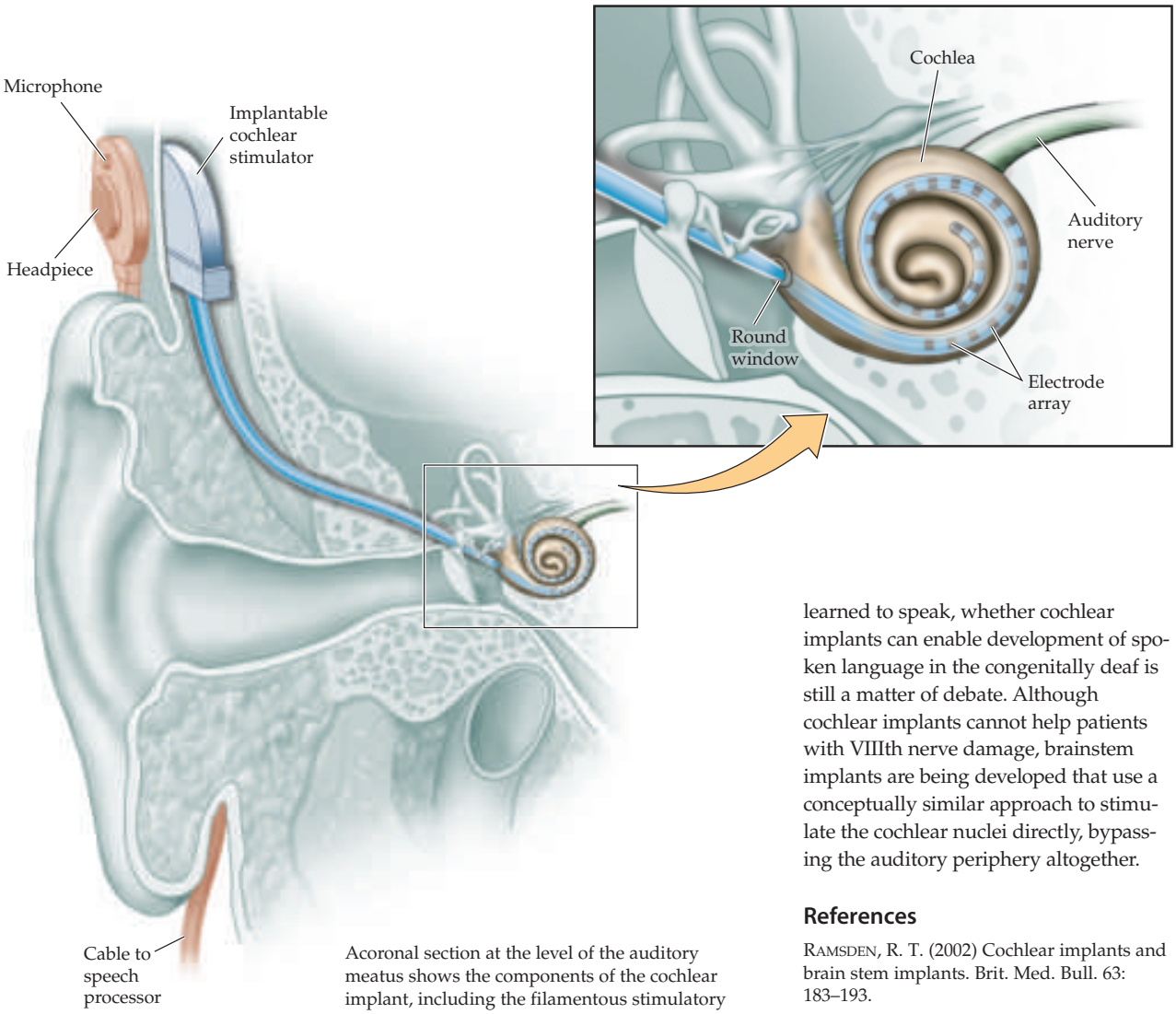
Conductive hearing loss can be due to occlusion of the ear canal by wax or foreign objects, rupture of the tympanic membrane itself, or arthritic ossification of the middle ear bones. In contrast, sensorineural hearing loss usually is due to congenital or environmental insults that lead to hair cell death (see Box A) or damage to the eighth nerve. As hair cells are relatively few in number and do not regenerate in humans, their depletion leads to a diminished ability to detect sounds. The Weber test, a simple test involving a tuning fork, can be used to distinguish between these two forms of hearing loss. If a resonating tuning fork (~256 Hz) is placed on the vertex, a patient with conductive hearing loss will report that the sound is louder in the affected ear. In the "plugged" state, sounds propagating through the skull do not dissipate so freely back out through the auditory meatus, and thus a greater amount of sound energy is transmitted to the cochlea on the blocked side. In contrast, a patient with a monaural sensorineural hearing loss will report that a Weber test sounds louder on the intact side, because even though the inner ear may vibrate equally on the two sides, the damaged side cannot transduce this vibration into a neural signal.

Treatment also differs for these two types of deafness. An external hearing aid is used to boost sounds to compensate for the reduced efficiency of the conductive apparatus in conductive hearing losses. These miniature devices are inserted in the ear canal, and contain a microphone and speaker, as well as an amplifier. One limitation of hearing aids is that they often provide rather flat amplification curves, which can interfere with listening in noisy environments; moreover, they do not achieve a high degree of directionality. The use of digital signal processing strategies partly overcomes these problems, and hearing aids obviously provide significant benefits to many people.

The treatment of sensorineural hearing loss is more complicated and invasive; conventional hearing aids are useless, because no amount of mechanical amplification can compensate for the inability to generate or convey a neural impulse from the cochlea. However, if the VIIIth nerve is intact, cochlear implants can be used to partially restore hearing. The cochlear implant consists of a peripherally mounted microphone and digital signal processor that transforms a sound into its spectral components, and additional electronics that use this information to activate different combinations of contacts on a threadlike multi-site stimulating electrode array. The electrode is inserted into the cochlea through the round window (see figure) and positioned along the length of the tonotopically organized basilar membrane and VIIIth nerve endings. This placement enables electrical stimulation of the nerve in a manner that mimics some aspects of the spectral decomposition naturally performed by the cochlea.

Cochlear implants can be remarkably effective in restoring hearing to people with hair cell damage, permitting them to engage in spoken communication. Despite such success in treating those who have lost their hearing *after* having

The cochlea (from the Latin for "snail") is a small (about 10 mm wide) coiled structure, which, were it uncoiled, would form a tube about 35 mm long (Figures 12.4 and 12.5). Both the oval window and, the **round window,** another region where the bone is absent surrounding the cochlea, are at the basal end of this tube. The cochlea is bisected from its basal almost to its apical end by the cochlear partition, which is a flexible structure that supports the **basilar membrane** and the **tectorial membrane**. There are fluid-filled chambers on each side of the cochlear partition, named the **scala vestibuli** and the **scala tympani**; a distinct channel, the **scala media**, runs within the



Acoronal section at the level of the auditory meatus shows the components of the cochlear implant, including the filamentous stimulatory electrode inserted into the cochlea through the round window.

learned to speak, whether cochlear implants can enable development of spoken language in the congenitally deaf is still a matter of debate. Although cochlear implants cannot help patients with VIIIth nerve damage, brainstem implants are being developed that use a conceptually similar approach to stimulate the cochlear nuclei directly, bypassing the auditory periphery altogether.

References

RAMSDEN, R. T. (2002) Cochlear implants and brain stem implants. Brit. Med. Bull. 63: 183–193.

RAUSCHECKER, J. P. AND R. V. SHANNON (2002) Sending sound to the brain. Science. 295: 1025–1029.

cochlear partition. The cochlear partition does not extend all the way to the apical end of the cochlea; instead there is an opening, known as the **helicotrema**, that joins the scala vestibuli to the scala tympani, allowing their fluid, known as **perilymph**, to mix. One consequence of this structural arrangement is that inward movement of the oval window displaces the fluid of the inner ear, causing the round window to bulge out slightly and deforming the cochlear partition.

The manner in which the basilar membrane vibrates in response to sound is the key to understanding cochlear function. Measurements of the vibra-

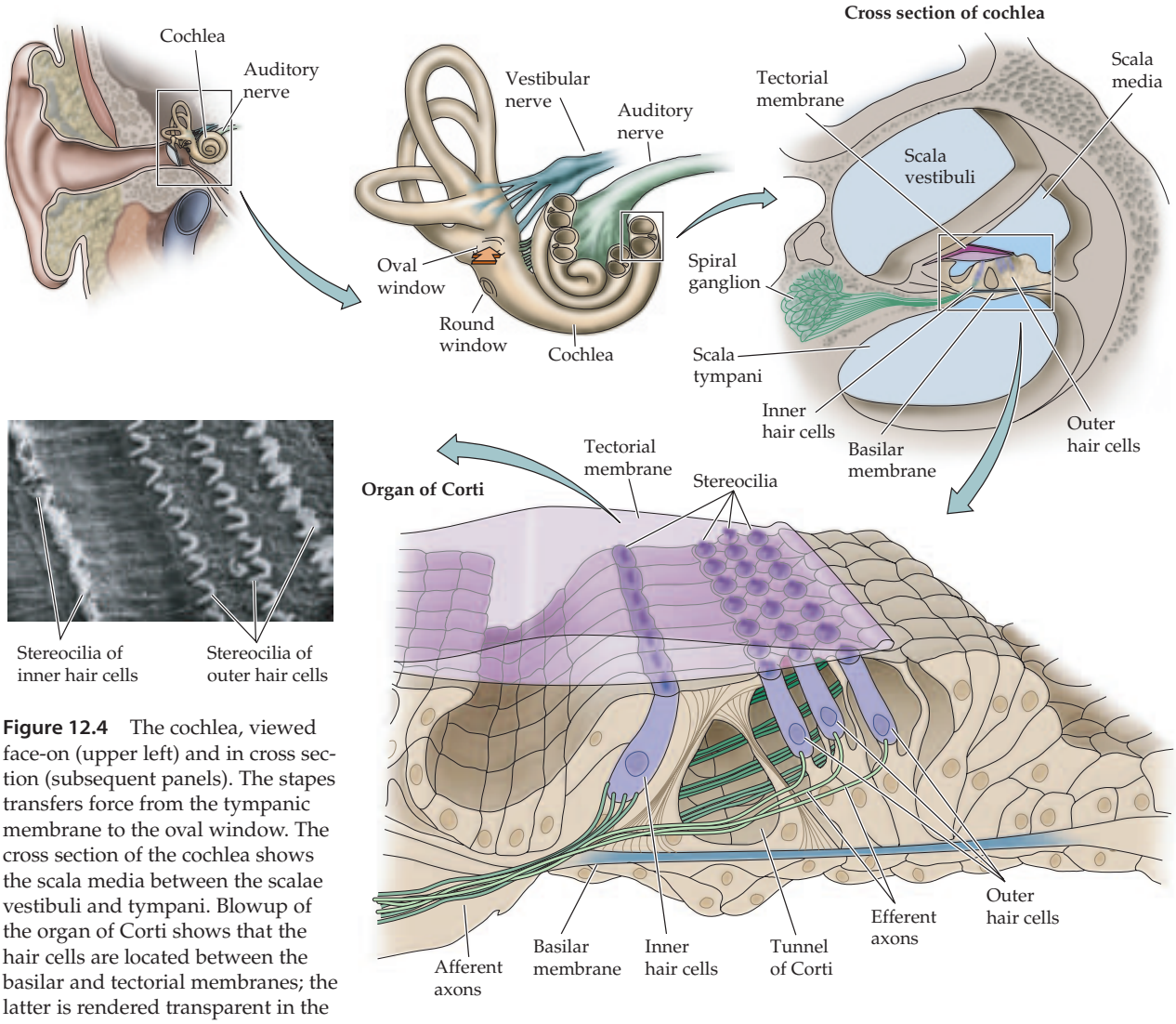


Figure 12.4 The cochlea, viewed face-on (upper left) and in cross section (subsequent panels). The stapes transfers force from the tympanic membrane to the oval window. The cross section of the cochlea shows the scala media between the scalae vestibuli and tympani. Blowup of the organ of Corti shows that the hair cells are located between the basilar and tectorial membranes; the latter is rendered transparent in the line drawing and removed in the scanning electron micrograph. The hair cells are named for their tufts of stereocilia; inner hair cells receive afferent inputs from cranial nerve VIII, whereas outer hair cells receive mostly efferent input. (Micrograph from Kessel and Kardon, 1979.)

tion of different parts of the basilar membrane, as well as the discharge rates of individual auditory nerve fibers that terminate along its length, show that both these features are highly tuned; that is, they respond most intensely to a sound of a specific frequency. Frequency tuning within the inner ear is attributable in part to the geometry of the basilar membrane, which is wider and more flexible at the apical end and narrower and stiffer at the basal end. One feature of such a system is that regardless of where energy is supplied to it, movement always begins at the stiff end (i.e., the base), and then propagates to the more flexible end (i.e., the apex). Georg von Békésy, working at Harvard University, showed that a membrane that varies systematically in its width and flexibility vibrates maximally at different positions as a function of the stimulus frequency (Figure 12.5). Using tubular models and human cochleas taken from cadavers, he found that an acoustical stimulus initiates a **traveling wave** of the same frequency in the cochlea, which propagates from the base toward the apex of the basilar membrane, growing in

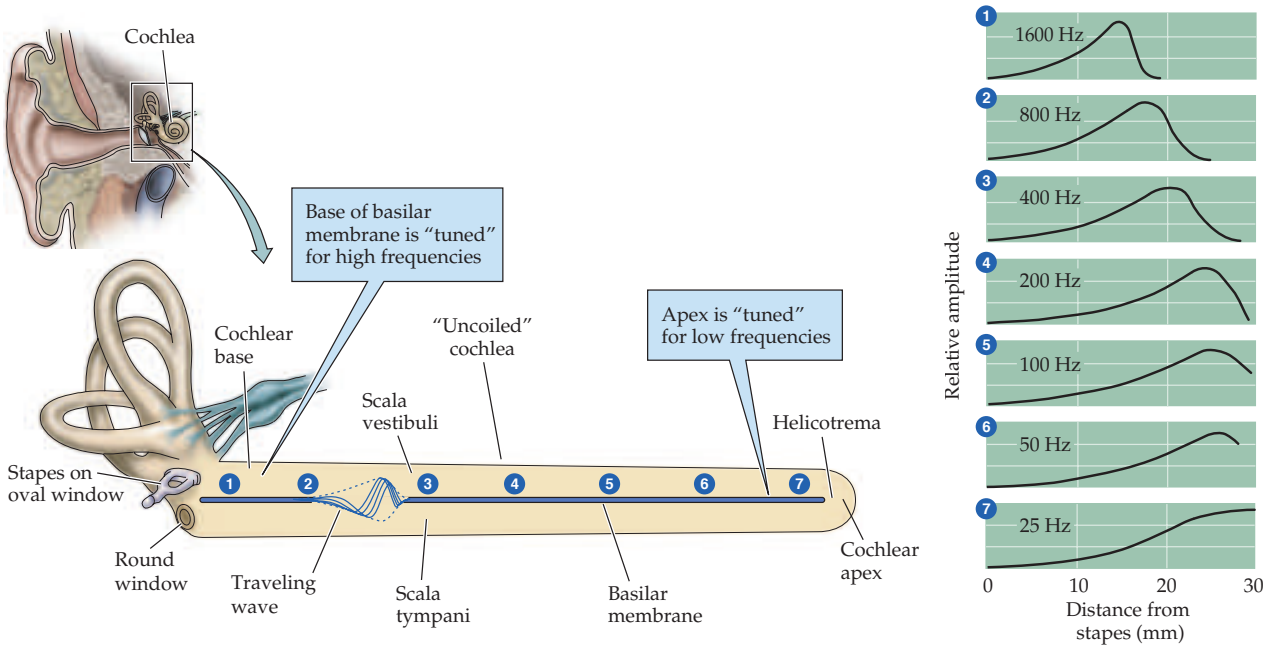


Figure 12.5 Traveling waves along the cochlea. A traveling wave is shown at a given instant along the cochlea, which has been uncoiled for clarity. The graphs on the right profile the amplitude of the traveling wave along the basilar membrane for different frequencies and show that the position (i.e., 1–7) where the traveling wave reaches its maximum amplitude varies directly with the frequency of stimulation. (Drawing after Dallos, 1992; graphs after von Békésy, 1960.)

amplitude and slowing in velocity until a point of maximum displacement is reached. This point of maximal displacement is determined by the sound frequency. The points responding to high frequencies are at the base of the basilar membrane where it is stiffer, and the points responding to low frequencies are at the apex, giving rise to a topographical mapping of frequency (that is, to **tonotopy**). An important feature is that complex sounds cause a pattern of vibration equivalent to the superposition of the vibrations generated by the individual tones making up that complex sound, thus accounting for the decompositional aspects of cochlear function mentioned earlier. This process of spectral decomposition appears to be an important strategy for detecting the various harmonic combinations that distinguish different natural sounds. Indeed, tonotopy is conserved throughout much of the auditory system, including the auditory cortex, suggesting that it is important to speech processing.

Von Békésy's model of cochlear mechanics was a passive one, resting on the premise that the basilar membrane acts like a series of linked resonators, much as a concatenated set of tuning forks. Each point on the basilar membrane was postulated to have a characteristic frequency at which it vibrated most efficiently; because it was physically linked to adjacent areas of the membrane, each point also vibrated (if somewhat less readily) at other frequencies, thus permitting propagation of the traveling wave. It is now clear, however, that the tuning of the auditory periphery, whether measured at the basilar membrane or recorded as the electrical activity of auditory nerve fibers, is too sharp to be explained by passive mechanics alone. At very low sound intensities, the basilar membrane vibrates one hundred-fold more than would be predicted by linear extrapolation from the motion measured at high intensities. Therefore, the ear's sensitivity arises from an active biomechanical process, as well as from its passive resonant properties (Box D). The outer hair cells, which together with the inner hair cells comprise the

Box D
The Sweet Sound of Distortion

As early as the first half of the eighteenth century, musical composers such as G. Tartini and W. A. Sorge discovered that upon playing pairs of tones, other tones not present in the original stimulus are also heard. These combination tones, *fc*, are mathematically related to the played tones f_1 and f_2 ($f_2 > f_1$) by the formula

$$fc = mf_1 \pm nf_2$$

where m and n are positive integers. Combination tones have been used for a variety of compositional effects, as they can strengthen the harmonic texture of a chord. Furthermore, organ builders sometimes use the difference tone ($f_2 - f_1$) created by two smaller organ pipes to produce the extremely low tones that would otherwise require building one especially large pipe.

Modern experiments suggest that this distortion product is due at least in part to the nonlinear properties of the inner ear. M. Ruggero and his colleagues placed small glass beads (10–30 mm in diameter) on the basilar membrane of an anesthetized animal and then determined the velocity of the basilar membrane in response to different combinations of tones by measuring the Doppler shift of laser light reflected from the beads. When two tones were played into the ear, the basilar membrane vibrated not only at those two frequencies, but also at other frequencies predicted by the above formula.

Related experiments on hair cells studied in vitro suggest that these nonlinearities result from the properties of the mechanical linkage of the transduction apparatus. By moving the hair bundle sinusoidally with a metal-coated glass fiber, A. J. Hudspeth and his coworkers found that the hair bundle exerts a force at the same frequency. However, when two sinusoids were applied simultaneously, the forces exerted by the hair bundle occurred not only at the primary frequencies, but at several combination frequencies as well. These distortion products are due to the transduction apparatus, since blocking the transduction channels causes the forces exerted at the combination frequencies to disappear, even though the forces at the primary frequencies remain unaffected. It seems that the tip links add a certain extra springiness to the hair bundle in the small range of motions over which the transduction channels are changing between closed and open states. If nonlinear distortions of basilar membrane vibrations arise from the properties of the hair bundle, then it is likely that hair cells can indeed influence basilar membrane motion, thereby accounting for the cochlea's extreme sensitivity. When we hear difference tones, we may be paying the price in distortion for an exquisitely fast and sensitive transduction mechanism.

References

Jaramillo, F., V. S. Markin and A. J. Hudspeth (1993) Auditory illusions and the single hair cell. Nature 364: 527–529.

Planchart, A. E. (1960) A study of the theories of Giuseppe Tartini. J. Music Theory 4: 32–61.

Robles, L., M. A. Ruggero and N. C. Rich (1991) Two-tone distortion in the basilar membrane of the cochlea. Nature 439: 413–414.

sensory cells of the inner ear, are the most likely candidates for driving this active process.

The motion of the traveling wave initiates sensory transduction by displacing the hair cells that sit atop the basilar membrane. Because these structures are anchored at different positions, the vertical component of the traveling wave is translated into a shearing motion between the basilar membrane and the overlying tectorial membrane (Figure 12.6). This motion bends the tiny processes, called **stereocilia**, that protrude from the apical ends of the hair cells, leading to voltage changes across the hair cell membrane. How the bending of stereocilia leads to receptor potentials in hair cells is considered in the following section.

Hair Cells and the Mechanoelectrical Transduction of Sound Waves

The hair cell is an evolutionary triumph that solves the problem of transforming vibrational energy into an electrical signal. The scale at which the

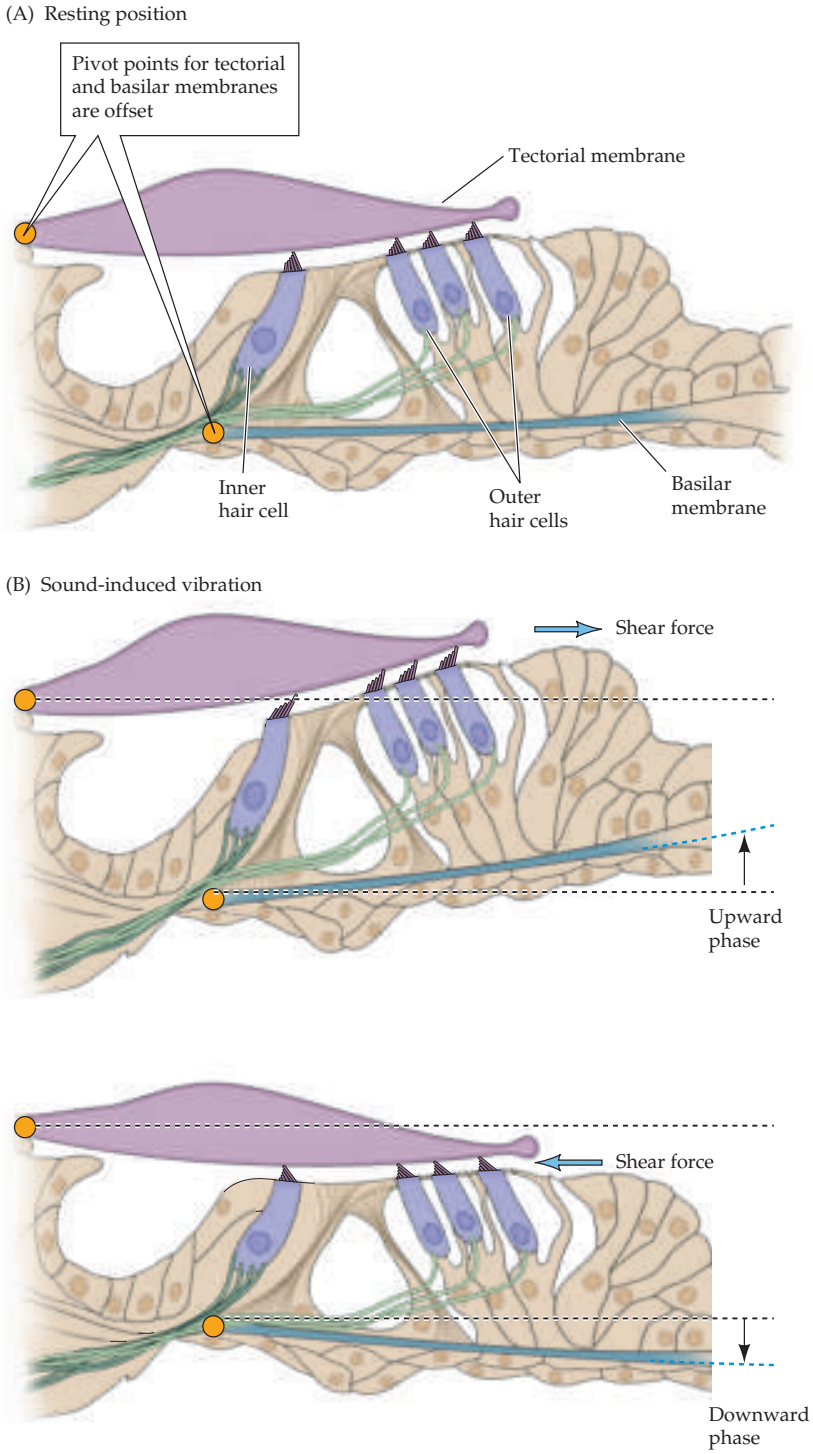


Figure 12.6 Movement of the basilar membrane creates a shearing force that bends the stereocilia of the hair cells. The pivot point of the basilar membrane is offset from the pivot point of the tectorial membrane, so that when the basilar membrane is displaced, the tectorial membrane moves across the tops of the hair cells, bending the stereocilia.

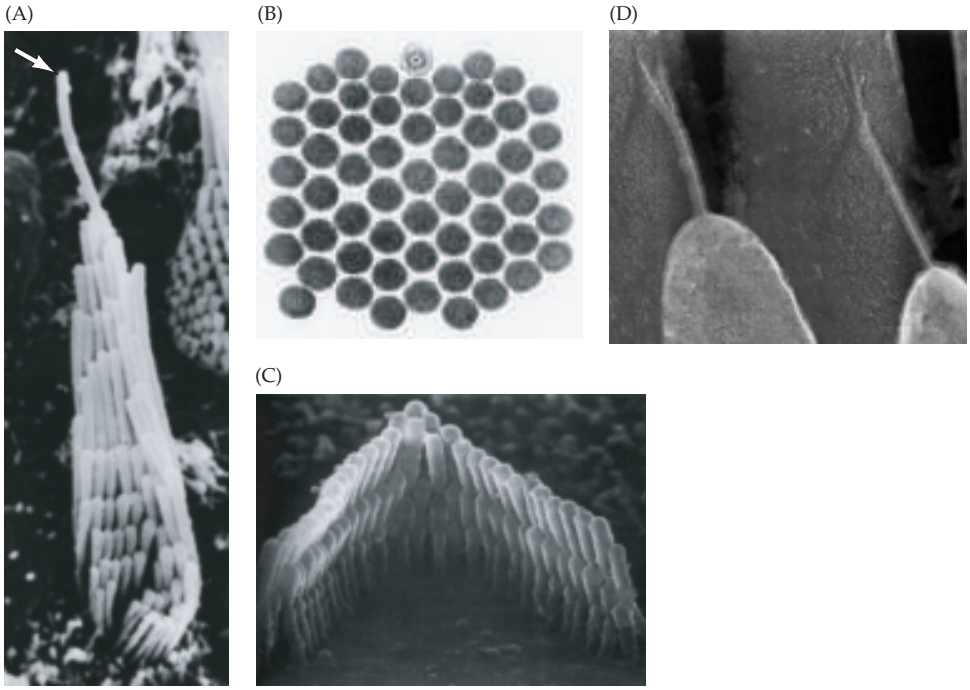


Figure 12.7 The structure and function of the hair bundle in vestibular and cochlear hair cells. The vestibular hair bundles shown here resemble those of cochlear hair cells, except for the presence of the kinocilium, which disappears in the mammalian cochlea shortly after birth. (A) The hair bundle of a guinea pig vestibular hair cell. This view shows the increasing height leading to the kinocilium (arrow). (B) Cross section through the vestibular hair bundle shows the 9 + 2 array of microtubules in the kinocilium (top), which contrasts with the simpler actin filament structure of the stereocilia. (C) Scanning electron micrograph of a guinea pig cochlear outer hair cell bundle viewed along the plane of mirror symmetry. Note the graded lengths of the stereocilia, and the absence of a kinocilium. (D) Tip links that connect adjacent stereocilia are believed to be the mechanical linkage that opens and closes the transduction channel. (A from Lindeman, 1973; B from Hudspeth, 1983; C from Pickles, 1988; D from Fain, 2003.)

hair cell operates is truly amazing: At the limits of human hearing, hair cells can faithfully detect movements of atomic dimensions and respond in the tens of microseconds! Furthermore, hair cells can adapt rapidly to constant stimuli, thus allowing the listener to extract signals from a noisy background.

The hair cell is a flask-shaped epithelial cell named for the bundle of hairlike processes that protrude from its apical end into the scala media. Each hair bundle contains anywhere from 30 to a few hundred hexagonally arranged stereocilia, with one taller **kinocilium** (Figure 12.7A). Despite their names, only the kinocilium is a true ciliary structure, with the characteristic two central tubules surrounded by nine doublet tubules that define cilia (Figure 12.7B). The function of the kinocilium is unclear, and in the cochlea of humans and other mammals it actually disappears shortly after birth (Figure 12.7C). The stereocilia are simpler, containing only an actin cytoskeleton. Each stereocilium tapers where it inserts into the apical membrane, forming a hinge about which each stereocilium pivots (Figure 12.7D). The stereocilia are graded in height and are arranged in a bilaterally symmetric fashion (in vestibular hair cells, this plane runs through the kinocilium). Displacement of the hair bundle parallel to this plane toward the tallest stereocilia depolarizes the hair cell, while movements parallel to this plane toward the shortest stereocilia cause hyperpolarization. In contrast, displacements perpendicular to the plane of symmetry do not alter the hair cell's membrane potential. The hair bundle movements at the threshold of hearing are approximately 0.3 nm, about the diameter of an atom of gold. Hair cells can convert the displacement of the stereociliary bundle into an electrical potential in as little as 10 microseconds; as described below, such speed is required for the accurate localization of the source of the sound. The need for microsecond resolution places certain constraints on the transduction mechanism, ruling out the rela-

tively slow second messenger pathways used in visual and olfactory transduction (see Chapters 7, 10, and 14); a direct, mechanically gated transduction channel is needed to operate this quickly. Evidently the filamentous structures that connect the tips of adjacent stereocilia, known as **tip links**, directly open cation-selective transduction channels when stretched, allowing K^+ ions to flow into the cell (see Figure 12.7D). As the linked stereocilia pivot from side to side, the tension on the tip link varies, modulating the ionic flow and resulting in a graded receptor potential that follows the movements of the stereocilia (Figures 12.8 and 12.9). The tip link model also explains why only deflections along the axis of the hair bundle activate transduction channels, since tip links join adjacent stereocilia along the axis directed toward the tallest stereocilia (see also Box B in Chapter 13). The exquisite mechanical sensitivity of the stereocilia also presents substantial risks: high intensity sounds can shear off the hair bundle, resulting in profound hearing deficits. Because human stereocilia, unlike those in fishes and birds, do not regenerate such damage is irreversible. The small number of hair cells (a total of about 30,000 in a human, or 15,000 per ear) further compounds the sensitivity of the inner

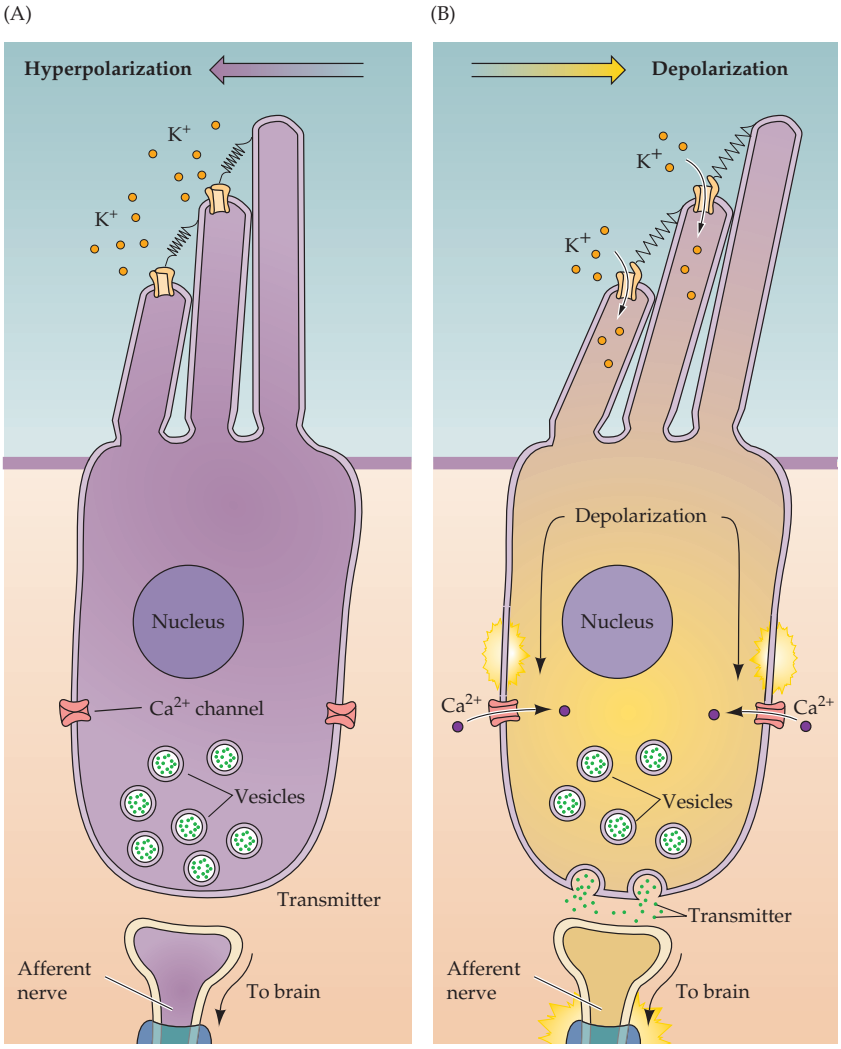


Figure 12.8 Mechanoelectrical transduction mediated by hair cells. (A,B) When the hair bundle is deflected toward the tallest stereocilium, cation-selective channels open near the tips of the stereocilia, allowing K^+ ions to flow into the hair cell down their electrochemical gradient (see text on next page for the explanation of this peculiar situation). The resulting depolarization of the hair cell opens voltage-gated Ca^{2+} channels in the cell soma, allowing calcium entry and release of neurotransmitter onto the nerve endings of the auditory nerve. (After Lewis and Hudspeth, 1983)

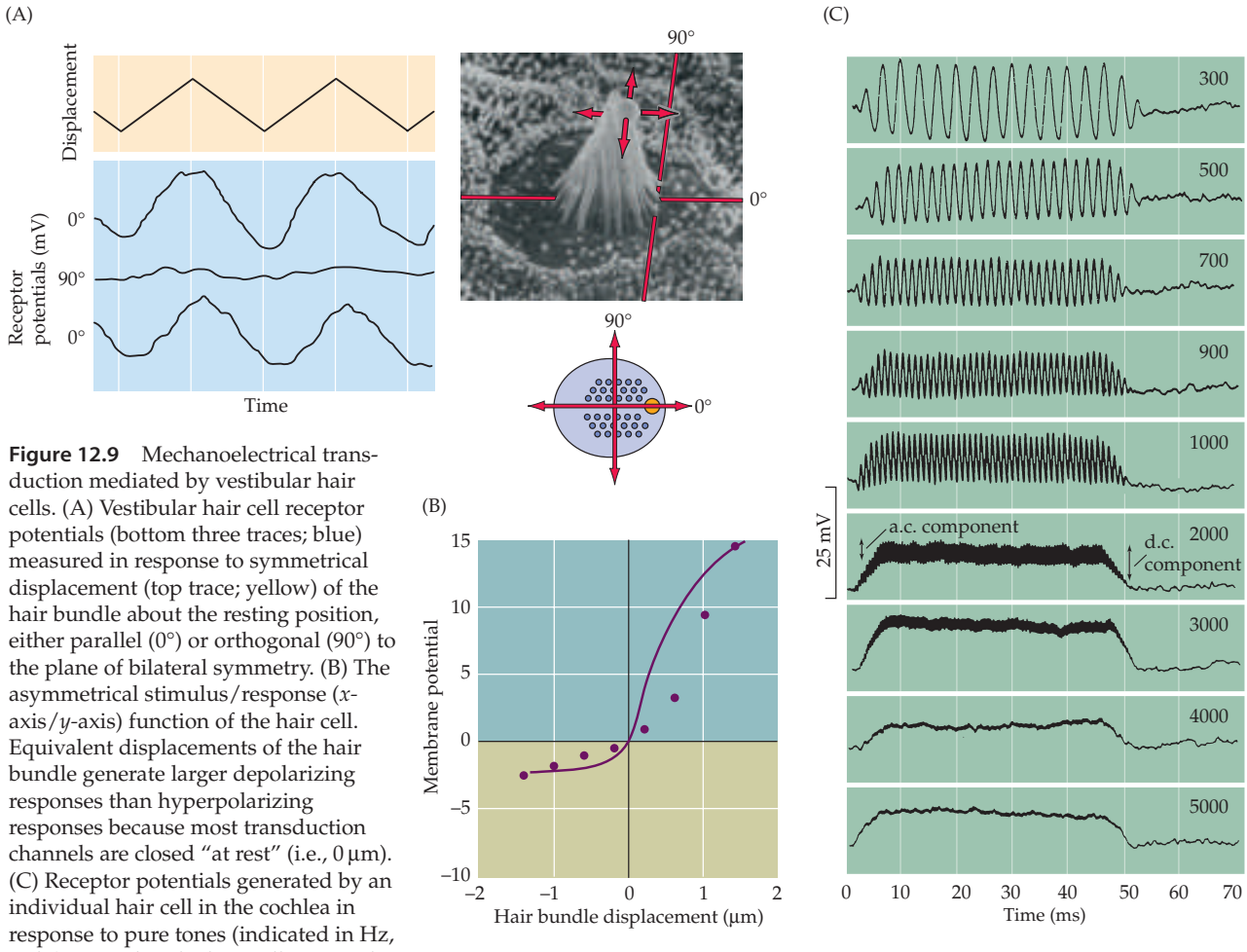


Figure 12.9 Mechanoelectrical transduction mediated by vestibular hair cells. (A) Vestibular hair cell receptor potentials (bottom three traces; blue) measured in response to symmetrical displacement (top trace; yellow) of the hair bundle about the resting position, either parallel (0°) or orthogonal (90°) to the plane of bilateral symmetry. (B) The asymmetrical stimulus/response (*x*-axis/*y*-axis) function of the hair cell. Equivalent displacements of the hair bundle generate larger depolarizing responses than hyperpolarizing responses because most transduction channels are closed "at rest" (i.e., 0 μm). (C) Receptor potentials generated by an individual hair cell in the cochlea in response to pure tones (indicated in Hz, right). Note that the hair cell potential faithfully follows the waveform of the stimulating sinusoids for low frequencies (< 3kHz), but still responds with a DC offset to higher frequencies. (A after Shotwell et al., 1981; B after Hudspeth and Corey, 1977; C after Palmer and Russell, 1986.)

ear to environmental and genetic insults. An important goal of current research is to identify the stem cells and factors that could contribute to the regeneration of human hair cells, thus affording a possible therapy for some forms of sensorineural hearing loss.

Understanding the ionic basis of hair cell transduction has been greatly advanced by intracellular recordings made from these tiny structures. The hair cell has a resting potential between –45 and –60 mV relative to the fluid that bathes the basal end of the cell. At the resting potential, only a small fraction of the transduction channels are open. When the hair bundle is displaced in the direction of the tallest stereocilium, more transduction channels open, causing depolarization as K^+ enters the cell. Depolarization in turn opens voltage-gated calcium channels in the hair cell membrane, and the resultant Ca^{2+} influx causes transmitter release from the basal end of the cell onto the auditory nerve endings (Figure 12.8A,B). Such calcium-dependent exocytosis is similar to chemical neurotransmission elsewhere in the central and peripheral nervous system (see Chapters 5 and 6); thus the hair cell has become a useful model for studying calcium-dependent transmitter release. Because some transduction channels are open at rest, the receptor potential is biphasic: Movement toward the tallest stereocilia depolarizes the cell, while move-

ment in the opposite direction leads to hyperpolarization. This situation allows the hair cell to generate a sinusoidal receptor potential in response to a sinusoidal stimulus, thus preserving the temporal information present in the original signal up to frequencies of around 3 kHz (Figure 12.9). Hair cells still can signal at frequencies above 3 kHz, although without preserving the exact temporal structure of the stimulus: the asymmetric displacement-receptor current function of the hair cell bundle is filtered by the cell's membrane time constant to produce a tonic depolarization of the soma, augmenting transmitter release and thus exciting VIIIth nerve terminals.

The high-speed demands of mechanoelectrical transduction have resulted in some impressive ionic specializations within the inner ear. An unusual adaptation of the hair cell in this regard is that K^+ serves both to depolarize *and* repolarize the cell, enabling the hair cell's K^+ gradient to be largely maintained by passive ion movement alone. As with other epithelial cells, the basal and apical surfaces of the hair cell are separated by tight junctions, allowing separate extracellular ionic environments at these two surfaces. The apical end (including the stereocilia) protrudes into the scala media and is exposed to the K^+-rich, Na^+-poor **endolymph**, which is produced by dedicated ion pumping cells in the **stria vascularis** (Figure 12.10). The basal end of the hair cell body is bathed in the same fluid that fills the scala tympani, the perilymph, which resembles other extracellular fluids in that it is K^+-poor and Na^+-rich. In addition, the compartment containing endolymph is about 80 mV more positive than the perilymph compartment (this difference is known as the endocochlear potential), while the inside of the hair cell is about 45 mV more negative than the perilymph (and 125 mV more negative than the endolymph). The resulting electrical gradient across the membrane of the stereocilia (about 125 mV; the difference between the hair cell resting potential and the endocochlear potential) drives K^+ through open transduc-

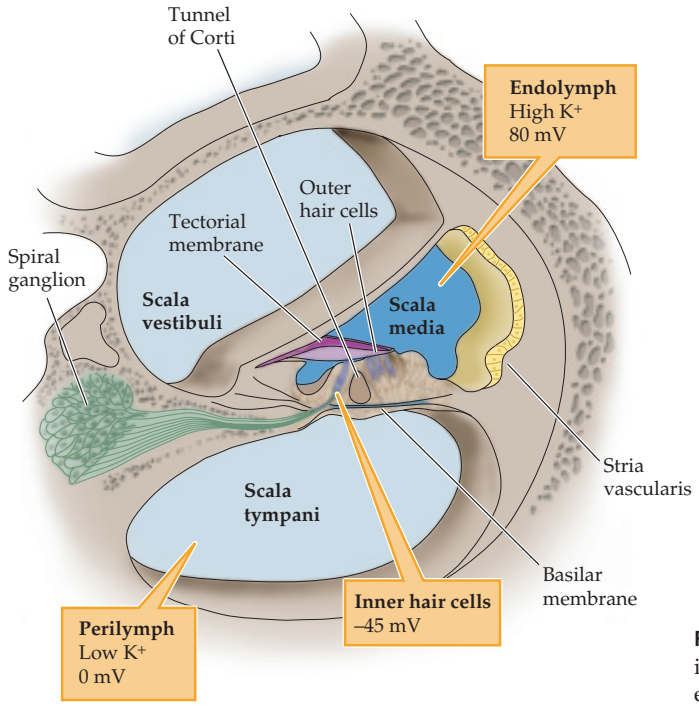


Figure 12.10 The stereocilia of the hair cells protrude into the endolymph, which is high in K^+ and has an electrical potential of +80 mV relative to the perilymph.

tion channels into the hair cell, even though these cells already have a high internal K^+ concentration. K^+ entry via the transduction channels electrotonically depolarizes the hair cell, opening voltage-gated Ca^{2+} and K^+ channels located in the membrane of the hair cell soma (see Box B in Chapter 13). The opening of *somatic* K^+ channels favors K^+ efflux, and thus repolarization; the efflux occurs because the perilymph surrounding the basal end is low in K^+ relative to the cytosol, and because the equilibrium potential for K^+ is more negative than the hair cell's resting potential ($E_{KBasal} \approx -85$ mV). Repolarization of the hair cell via K^+ efflux is also facilitated by Ca^{2+} entry. In addition to modulating the release of neurotransmitter, Ca^{2+} entry opens Ca^{2+}-dependent K^+ channels, which provide another avenue for K^+ to enter the perilymph. Indeed, the interaction of Ca^{2+} influx and Ca^{2+}-dependent K^+ efflux can lead to electrical resonances that enhance the tuning of response properties within the inner ear (also explained in Box B in Chapter 13). In essence, the hair cell operates as two distinct compartments, each dominated by its own Nernst equilibrium potential for K^+; this arrangement ensures that the hair cell's ionic gradient does not run down, even during prolonged stimulation. At the same time, rupture of Reissner's membrane, which normally separates the scalae media and vestibuli, or compounds such as ethacrynic acid (see Box A), which selectively poison the ion-pumping cells of the stria vascularis, can cause the endocochlear potential to dissipate, resulting in a sensorineural hearing deficit. In short, the hair cell exploits the different ionic milieus of its apical and basal surfaces to provide extremely fast and energy-efficient repolarization.

Two Kinds of Hair Cells in the Cochlea

The cochlear hair cells in humans consist of one row of **inner hair cells** and three rows of **outer hair cells** (see Figure 12.4). The inner hair cells are the actual sensory receptors, and 95% of the fibers of the auditory nerve that project to the brain arise from this subpopulation. The terminations on the outer hair cells are almost all from efferent axons that arise from cells in the superior olivary complex.

A clue to the significance of this efferent pathway was provided by the discovery that an active process within the cochlea, as mentioned, influences basilar membrane motion. First, it was found that the cochlea actually emits sound under certain conditions. These otoacoustical emissions can be detected by placing a sensitive microphone at the eardrum and monitoring the response after briefly presenting a tone or click, and provide a useful means to assess cochlear function in the newborn (this test is now done routinely to rule out congenital deafness). Such emissions can also occur spontaneously, especially in certain pathological states, and are thus one potential source of **tinnitus** (ringing in the ears). These observations clearly indicate that a process within the cochlea is capable of producing sound. Second, stimulation of the crossed olivocochlear bundle, which supplies efferent input to the outer hair cells, can broaden VIIIth nerve tuning curves. Third, the high sensitivity notch of VIIIth nerve tuning curves is lost when the outer hair cells are selectively inactivated. Finally, isolated outer hair cells contract and expand in response to small electrical currents, thus providing a potential source of energy to drive an active process within the cochlea. Thus, it seems likely that the outer hair cells sharpen the frequency-resolving power of the cochlea by actively contracting and relaxing, thus changing the stiffness of the tectorial membrane at particular locations. This active

process explains the nonlinear vibration of the basilar membrane at low sound intensities (see Box D).

Tuning and Timing in the Auditory Nerve

The rapid response time of the transduction apparatus allows the membrane potential of the hair cell to follow deflections of the hair bundle up to moderately high frequencies of oscillation. In humans, the receptor potentials of certain hair cells and the action potentials of their associated auditory nerve fiber can follow stimuli of up to about 3 kHz in a one-to-one fashion. Such real-time encoding of stimulus frequency by the pattern of action potentials in the auditory nerve is known as the "volley theory" of auditory information transfer. Even these extraordinarily rapid processes, however, fail to follow frequencies above 3 kHz (see Figure 12.9). Accordingly, some other mechanism must be used to transmit auditory information at higher frequencies. The tonotopically organized basilar membrane provides an alternative to temporal coding, namely a "labeled-line" coding mechanism. In this case, frequency information is specified by preserving the tonotopy of the cochlea at higher levels in the auditory pathway. Because the auditory nerve fibers associate with the inner hair cells in approximately a one-to-one ratio (although several or more VIIIth nerve fibers synapse on a single hair cell), each auditory nerve fiber transmits information about only a small part of the audible frequency spectrum. As a result, auditory nerve fibers related to the apical end of the cochlea respond to low frequencies, and fibers that are related to the basal end respond to high frequencies (see Figure 12.5). The limitations of specific fibers can be seen in electrophysiological recordings of responses to sound (Figure 12.11). These threshold functions are called **tuning curves**, and the lowest threshold of the tuning curve is called the **characteristic frequency**. Since the topographical order of the characteristic frequency of neurons is retained throughout the system, information about frequency is also preserved. Cochlear implants (see Box C) exploit the tonotopic organization of the cochlea, and particularly its eighth nerve afferents, to roughly recreate the patterns of VIIIth nerve activity elicited by sounds. In patients with damaged hair cells, such implants can effectively bypass the impaired transduction apparatus, and thus restore some degree of auditory function.

The other prominent feature of hair cells—their ability to follow the waveform of low-frequency sounds—is also important in other more subtle aspects of auditory coding. As mentioned earlier, hair cells have biphasic response properties. Because hair cells release transmitter only when depolarized, auditory nerve fibers fire only during the positive phases of low-frequency sounds (see Figure 12.11). The resultant "phase locking" that results provides temporal information from the two ears to neural centers that compare interaural time differences. The evaluation of interaural time differences provides a critical cue for sound localization and the perception of auditory "space." That auditory space can be perceived is remarkable, given that the cochlea, unlike the retina, cannot represent space directly. A final point is that VIIIth nerve activity patterns are not simply faithful neural replicas of the auditory stimulus itself. Indeed, William Bialek and his colleagues at Princeton University have shown that the VIIIth nerve in the bullfrog encodes conspecific mating calls more efficiently than artificial sounds with similar frequency and amplitude characteristics. Thus both animal and human studies support the idea that the auditory periphery is optimized to

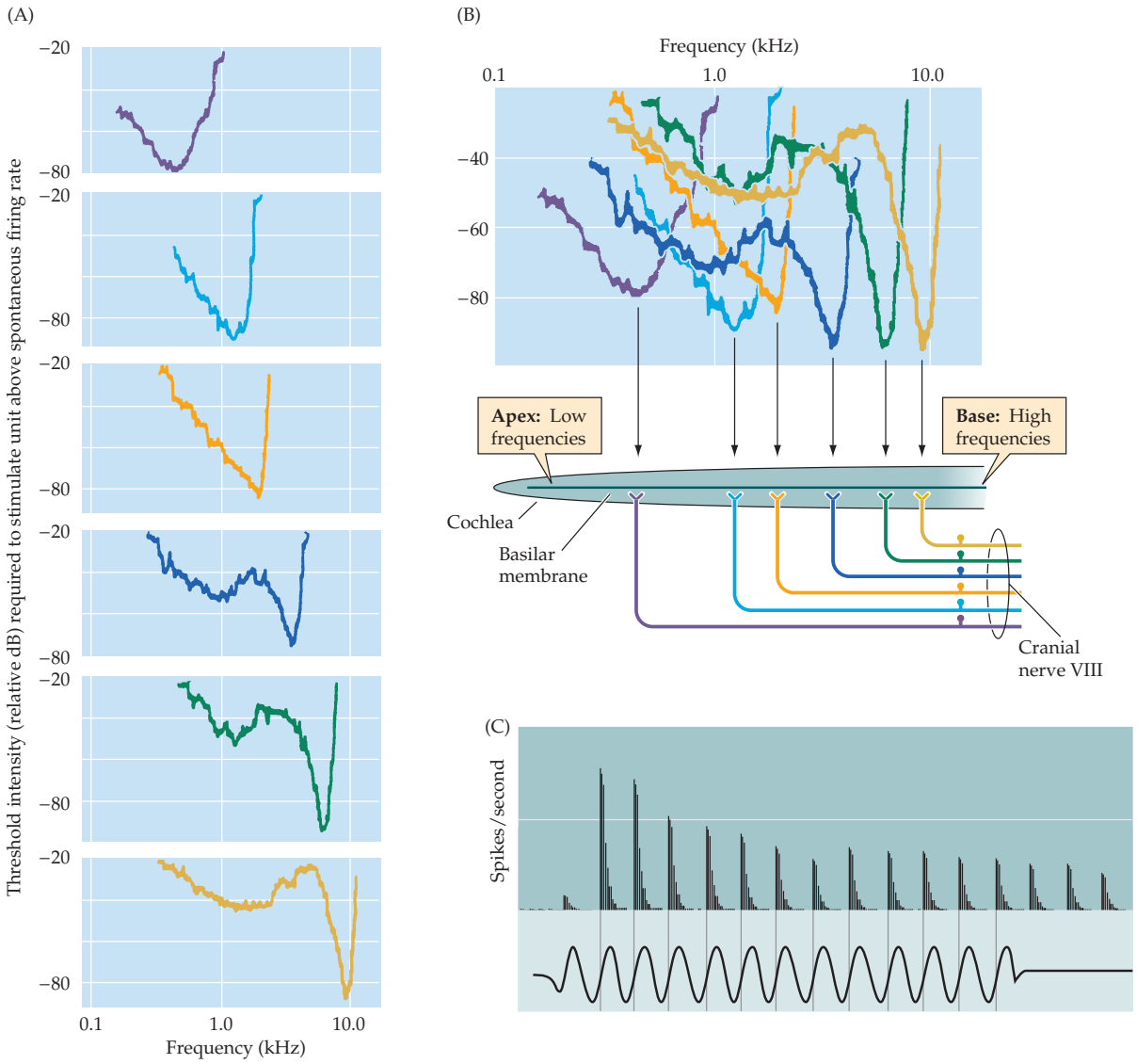


Figure 12.11 Response properties of auditory nerve fibers. (A) Frequency tuning curves of six different fibers in the auditory nerve. Each graph plots, across all frequencies to which the fiber responds, the minimum sound level required to increase the fiber's firing rate above its spontaneous firing level. The lowest point in the plot is the weakest sound intensity to which the neuron will respond. The frequency at this point is called the neuron's characteristic frequency. (B) The frequency tuning curves of auditory nerve fibers superimposed and aligned with their approximate relative points of innervation along the basilar membrane. (In the side view schematic, the basilar membrane is represented as a black line within the unrolled cochlea.) (C) Temporal response patterns of a low-frequency axon in the auditory nerve. The stimulus waveform is indicated beneath the histograms, which show the phase-locked responses to a 50-ms tone pulse of 260 Hz. Note that the spikes are all timed to the same phase of the sinusoidal stimulus. (A after Kiang and Moxon, 1972; C after Kiang, 1984.)

transmit species-typical vocal sounds, rather than simply transmitting all sounds equally well to central auditory areas.

How Information from the Cochlea Reaches Targets in the Brainstem

A hallmark of the ascending auditory system is its parallel organization. This arrangement becomes evident as soon as the auditory nerve enters the brainstem, where it branches to innervate the three divisions of the cochlear nucleus. The auditory nerve (the major component of cranial nerve VIII) comprises the central processes of the bipolar spiral ganglion cells in the cochlea (see Figure 12.4); each of these cells sends a peripheral process to contact one inner hair cell and a central process to innervate the cochlear nucleus. Within the cochlear nucleus, each auditory nerve fiber branches, sending an ascending branch to the anteroventral cochlear nucleus and a descending branch to the posteroventral cochlear nucleus and the dorsal cochlear nucleus (Figure 12.12). The tonotopic organization of the cochlea is maintained in the three parts of the cochlear nucleus, each of which contains different populations of cells with quite different properties. In addition, the patterns of termination of the auditory nerve axons differ in density and type; thus, there are several opportunities at this level for transformation of the information from the hair cells.

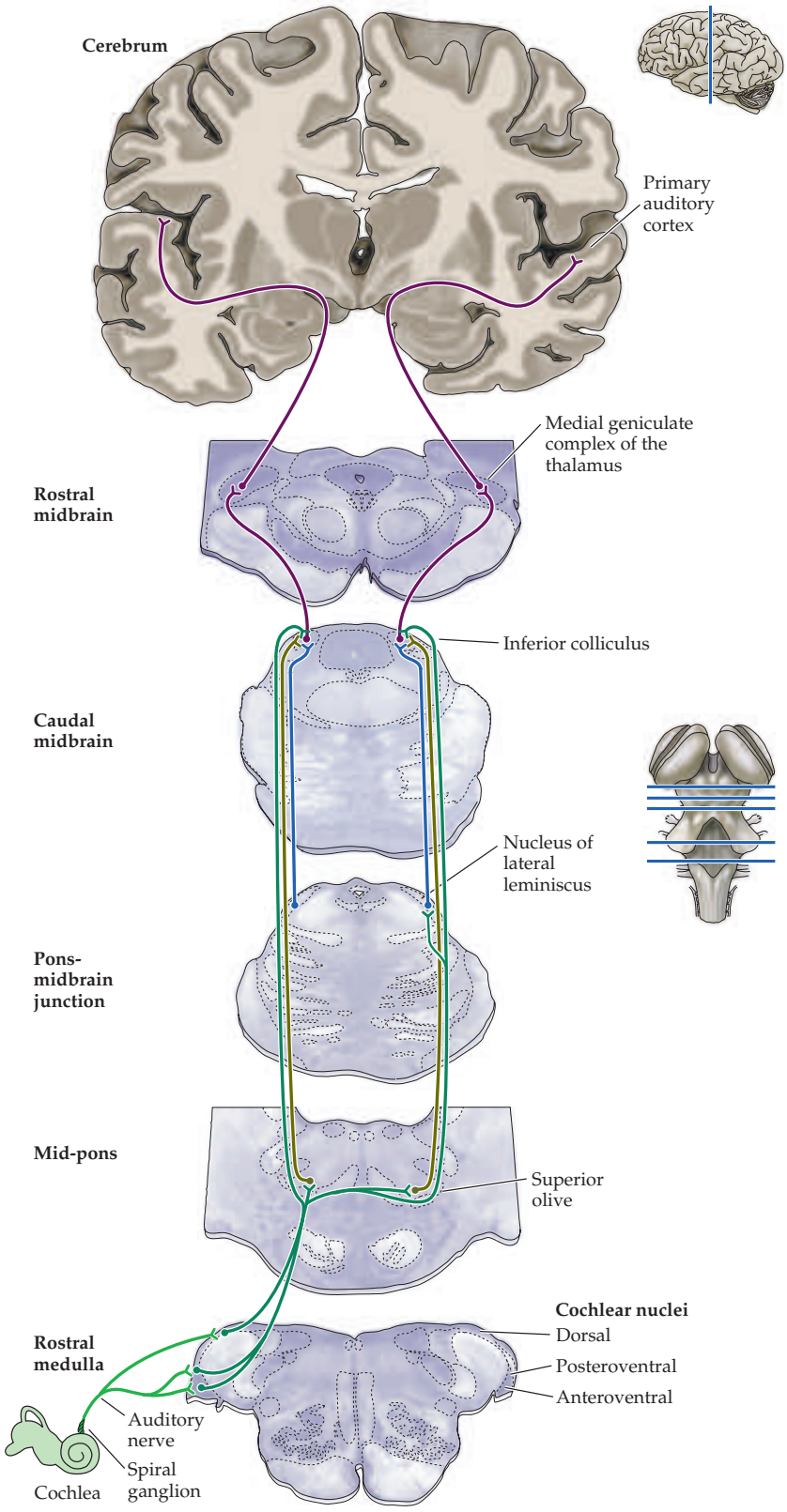
Integrating Information from the Two Ears

Just as the auditory nerve branches to innervate several different targets in the cochlear nuclei, the neurons in these nuclei give rise to several different pathways (see Figure 12.12). One clinically relevant feature of the ascending projections of the auditory brainstem is a high degree of bilateral connectivity, which means that damage to central auditory structures is almost never manifested as a monaural hearing loss. Indeed, monaural hearing loss strongly implicates unilateral peripheral damage, either to the middle or inner ear, or to the VIIIth nerve itself (see Box C). Given the relatively byzantine organization already present at the level of the auditory brainstem, it is useful to consider these pathways in the context of their functions.

The best-understood function mediated by the auditory brainstem nuclei, and certainly the one most intensively studied, is sound localization. Humans use at least two different strategies to localize the horizontal position of sound sources, depending on the frequencies in the stimulus. For frequencies below 3 kHz (which can be followed in a phase-locked manner), interaural *time* differences are used to localize the source; above these frequencies, interaural *intensity* differences are used as cues. Parallel pathways originating from the cochlear nucleus serve each of these strategies for sound localization.

The human ability to detect interaural time differences is remarkable. The longest interaural time differences, which are produced by sounds arising directly lateral to one ear, are on the order of only 700 microseconds (a value given by the width of the head divided by the speed of sound in air, about 340 m/s). Psychophysical experiments show that humans can actually detect interaural time differences as small as 10 microseconds; two sounds presented through earphones separated by such small interaural time differences are perceived as arising from the side of the leading ear. This sensitivity translates into accuracy for sound localization of about 1 degree.

Figure 12.12 Diagram of the major auditory pathways. Although many details are missing from this diagram, two important points are evident: (1) the auditory system entails several parallel pathways, and (2) information from each ear reaches both sides of the system, even at the level of the brainstem.



How is timing in the 10 microseconds range accomplished by neural components that operate in the millisecond range? The neural circuitry that computes such tiny interaural time differences consists of binaural inputs to the **medial superior olive** (**MSO**) that arise from the right and left anteroventral cochlear nuclei (Figure 12.13; see also Figure 12.12). The medial superior olive contains cells with bipolar dendrites that extend both medially and laterally. The lateral dendrites receive input from the ipsilateral anteroventral cochlear nucleus, and the medial dendrites receive input from the contralateral anteroventral cochlear nucleus (both inputs are excitatory). As might be expected, the MSO cells work as **coincidence detectors**, responding when both excitatory signals arrive at the same time. For a coincidence mechanism to be useful in localizing sound, different neurons must be maximally sensitive to different interaural time delays. The axons that project from the anteroventral cochlear nucleus evidently vary systematically in length to create delay lines. (Remember that the length of an axon divided by its conduction velocity equals the conduction time.) These anatomical differences compensate for sounds arriving at slightly different times at the two ears, so that the resultant neural impulses arrive at a particular MSO neuron simultaneously, making each cell especially sensitive to sound sources in a particular place. The mechanisms enabling MSO neurons to function as coincidence detectors at the microsecond level are still poorly understood, but certainly reflect one of the more impressive biophysical specializations in the nervous system.

Sound localization perceived on the basis of interaural time differences requires phase-locked information from the periphery, which, as already

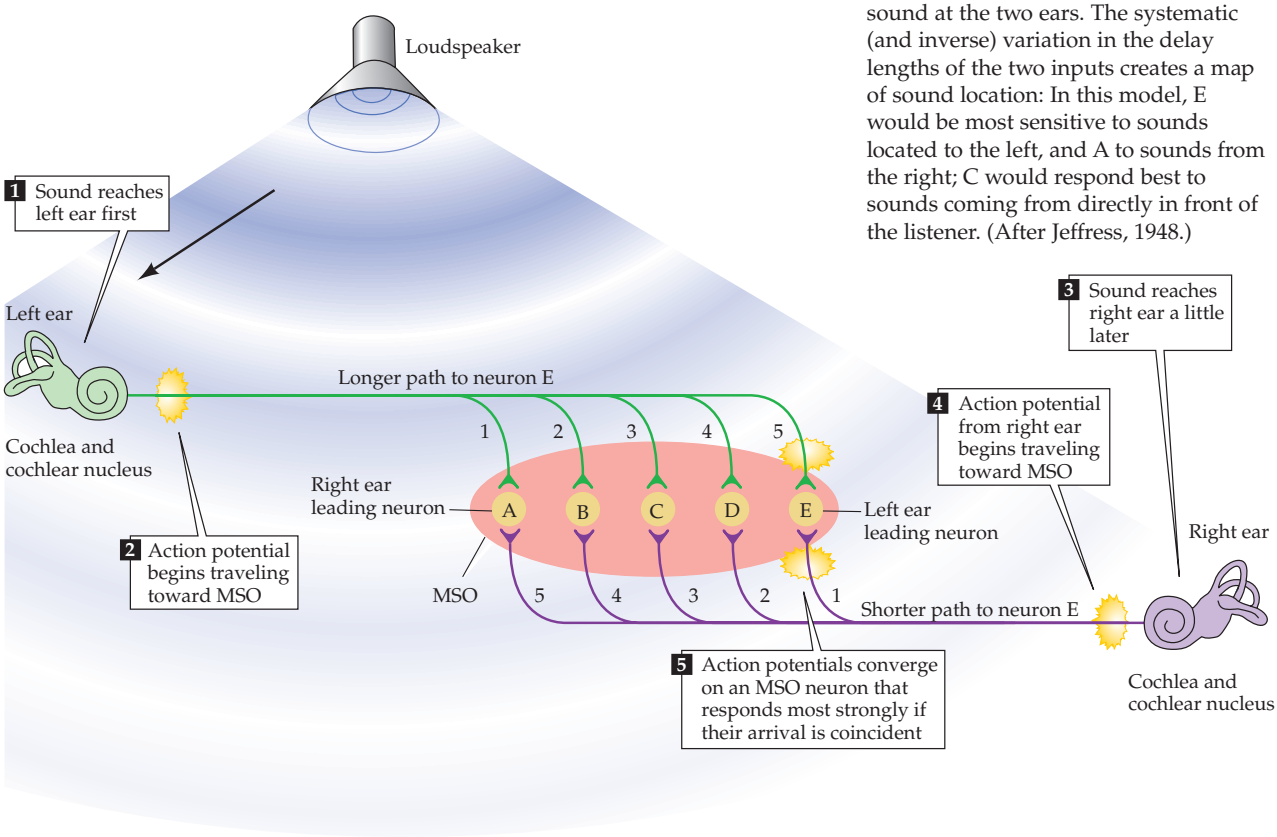


Figure 12.13 Diagram illustrating how the MSO computes the location of a sound by interaural time differences. A given MSO neuron responds most strongly when the two inputs arrive simultaneously, as occurs when the contralateral and ipsilateral inputs precisely compensate (via their different lengths) for differences in the time of arrival of a sound at the two ears. The systematic (and inverse) variation in the delay lengths of the two inputs creates a map of sound location: In this model, E would be most sensitive to sounds located to the left, and A to sounds from the right; C would respond best to sounds coming from directly in front of the listener. (After Jeffress, 1948.)

emphasized, is available to humans only for frequencies below 3 kHz. (In barn owls, the reigning champions of sound localization, phase locking occurs at up to 9 kHz.) Therefore, a second mechanism must come into play at higher frequencies. At frequencies higher than about 2 kHz, the human head begins to act as an acoustical obstacle because the wavelengths of the sounds are too short to bend around it. As a result, when high-frequency sounds are directed toward one side of the head, an acoustical "shadow" of lower intensity is created at the far ear. These intensity differences provide a second cue about the location of a sound. The circuits that compute the position of a sound source on this basis are found in the **lateral superior olive** (**LSO**) and the **medial nucleus of the trapezoid body** (**MNTB**) (Figure 12.14). Excitatory axons project directly from the ipsilateral anteroventral cochlear nucleus to the LSO (as well as to the MSO; see Figure 12.13). Note that the LSO also receives inhibitory input from the contralateral ear, via an inhibitory neuron in the MNTB. This excitatory/inhibitory interaction

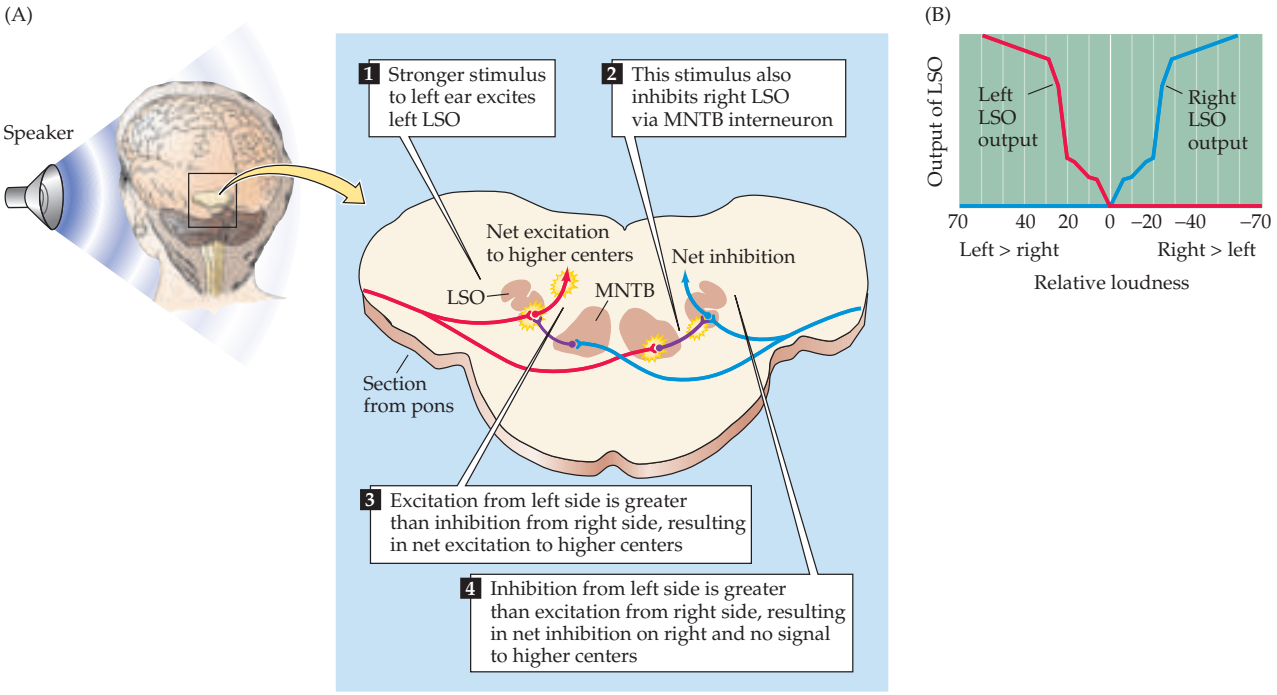


Figure 12.14 Lateral superior olive neurons encode sound location through interaural intensity differences. (A) LSO neurons receive direct excitation from the ipsilateral cochlear nucleus; input from the contralateral cochlear nucleus is relayed via inhibitory interneurons in the MNTB. (B) This arrangement of excitation–inhibition makes LSO neurons fire most strongly in response to sounds arising directly lateral to the listener on the same side as the LSO, because excitation from the ipsilateral input will be great and inhibition from the contralateral input will be small. In contrast, sounds arising from in front of the listener, or from the opposite side, will silence the LSO output, because excitation from the ipsilateral input will be minimal, but inhibition driven by the contralateral input will be great. Note that LSOs are paired and bilaterally symmetrical; each LSO only encodes the location of sounds arising on the same side of the body as its location.

results in a net excitation of the LSO on the same side of the body as the sound source. For sounds arising directly lateral to the listener, firing rates will be highest in the LSO on that side; in this circumstance, the excitation via the ipsilateral anteroventral cochlear nucleus will be maximal, and inhibition from the contralateral MNTB minimal. In contrast, sounds arising closer to the listener's midline will elicit lower firing rates in the ipsilateral LSO because of increased inhibition arising from the contralateral MNTB. For sounds arising at the midline, or from the other side, the increased inhibition arising from the MNTB is powerful enough to completely silence LSO activity. Note that each LSO only encodes sounds arising in the ipsilateral hemifield; it therefore takes both LSOs to represent the full range of horizontal positions.

In summary, there are two separate pathways—and two separate mechanisms—for localizing sound. Interaural time differences are processed in the medial superior olive, and interaural intensity differences are processed in the lateral superior olive. These two pathways are eventually merged in the midbrain auditory centers.

Monaural Pathways from the Cochlear Nucleus to the Lateral Lemniscus

The binaural pathways for sound localization are only part of the output of the cochlear nucleus. This fact is hardly surprising, given that auditory perception involves much more than locating the position of the sound source. A second major set of pathways from the cochlear nucleus bypasses the superior olive and terminates in the **nuclei of the lateral lemniscus** on the contralateral side of the brainstem (see Figure 12.12). These particular pathways respond to sound arriving at one ear only and are thus referred to as monaural. Some cells in the lateral lemniscus nuclei signal the onset of sound, regardless of its intensity or frequency. Other cells in the lateral lemniscus nuclei process other temporal aspects of sound, such as duration. The precise role of these pathways in processing temporal features of sound is not yet known. As with the outputs of the superior olivary nuclei, the pathways from the nuclei of the lateral lemniscus converge at the midbrain.

Integration in the Inferior Colliculus

Auditory pathways ascending via the olivary and lemniscal complexes, as well as other projections that arise directly from the cochlear nucleus, project to the midbrain auditory center, the **inferior colliculus**. In examining how integration occurs in the inferior colliculus, it is again instructive to turn to the most completely analyzed auditory mechanism, the binaural system for localizing sound. As already noted, space is not mapped on the auditory receptor surface; thus the perception of auditory space must somehow be synthesized by circuitry in the lower brainstem and midbrain. Experiments in the barn owl, an extraordinarily proficient animal at localizing sounds, show that the convergence of binaural inputs in the midbrain produces something entirely new relative to the periphery—namely, a computed topographical representation of auditory space. Neurons within this **auditory space map** in the colliculus respond best to sounds originating in a specific region of space and thus have both a preferred elevation and a preferred horizontal location, or azimuth. Although comparable maps of auditory space have not yet been found in mammals, humans have a clear perception of

both the elevational and azimuthal components of a sound's location, suggesting that we have a similar auditory space map.

Another important property of the inferior colliculus is its ability to process sounds with complex temporal patterns. Many neurons in the inferior colliculus respond only to frequency-modulated sounds, while others respond only to sounds of specific durations. Such sounds are typical components of biologically relevant sounds, such as those made by predators, or intraspecific communication sounds, which in humans include speech.

The Auditory Thalamus

Despite the parallel pathways in the auditory stations of the brainstem and midbrain, the **medial geniculate complex** (**MGC**) in the thalamus is an obligatory relay for all ascending auditory information destined for the cortex (see Figure 12.12). Most input to the MGC arises from the inferior colliculus, although a few auditory axons from the lower brainstem bypass the inferior colliculus to reach the auditory thalamus directly. The MGC has several divisions, including the ventral division, which functions as the major thalamocortical relay, and the dorsal and medial divisions, which are organized like a belt around the ventral division.

In some mammals, the strictly maintained tonotopy of the lower brainstem areas is exploited by convergence onto MGC neurons, generating specific responses to certain spectral combinations. The original evidence for this statement came from research on the response properties of cells in the MGC of echolocating bats. Some cells in the so-called belt areas of the bat MGC respond only to combinations of widely spaced frequencies that are specific components of the bat's echolocation signal and of the echoes that are reflected from objects in the bat's environment. In the mustached bat, where this phenomenon has been most thoroughly studied, the echolocation pulse has a changing frequency (frequency-modulated, or FM) component that includes a fundamental frequency and one or more harmonics. The fundamental frequency (FM_1) has low intensity and sweeps from 30 kHz to 20 kHz. The second harmonic (FM_2) is the most intense component and sweeps from 60 kHz to 40 kHz. Note that these frequencies do not overlap. Most of the echoes are from the intense FM_2 sound, and virtually none arise from the weak FM_1, even though the emitted FM_1 is loud enough for the bat to hear. Apparently, the bat assesses the distance to an object by measuring the delay between the FM_1 emission and the FM_2 echo. Certain MGC neurons respond when FM_2 follows FM_1 by a specific delay, providing a mechanism for sensing such frequency combinations. Because each neuron responds best to a particular delay, the population of MGC neurons encodes a range of distances.

Bat sonar illustrates two important points about the function of the auditory thalamus. First, the MGC is the first station in the auditory pathway where selectivity for combinations of frequencies is found. The mechanism responsible for this selectivity is presumably the ultimate convergence of inputs from cochlear areas with different spectral sensitivities. Second, cells in the MGC are selective not only for frequency combinations, but also for specific time intervals between the two frequencies. The principle is the same as that described for binaural neurons in the medial superior olive, but in this instance, two monaural signals with different frequency sensitivity coincide, and the time difference is in the millisecond rather than the microsecond range.

In summary, neurons in the medial geniculate complex receive convergent inputs from spectrally and temporally separate pathways. This complex, by

virtue of its convergent inputs, mediates the detection of specific spectral and temporal combinations of sounds. In many species, including humans, varying spectral and temporal cues are especially important features of communication sounds. It is not known whether cells in the human medial geniculate are selective to combinations of sounds, but the processing of speech certainly requires both spectral and temporal combination sensitivity.

The Auditory Cortex

The ultimate target of afferent auditory information is the auditory cortex. Although the auditory cortex has a number of subdivisions, a broad distinction can be made between a primary area and peripheral, or belt, areas. The **primary auditory cortex** (**A1**) is located on the superior temporal gyrus in the temporal lobe and receives point-to-point input from the ventral division of the medial geniculate complex; thus, it contains a precise tonotopic map. The **belt areas** of the auditory cortex receive more diffuse input from the belt areas of the medial geniculate complex and therefore are less precise in their tonotopic organization.

The primary auditory cortex (A1) has a topographical map of the cochlea (Figure 12.15), just as the primary visual cortex (V1) and the primary somatic sensory cortex (S1) have topographical maps of their respective sensory

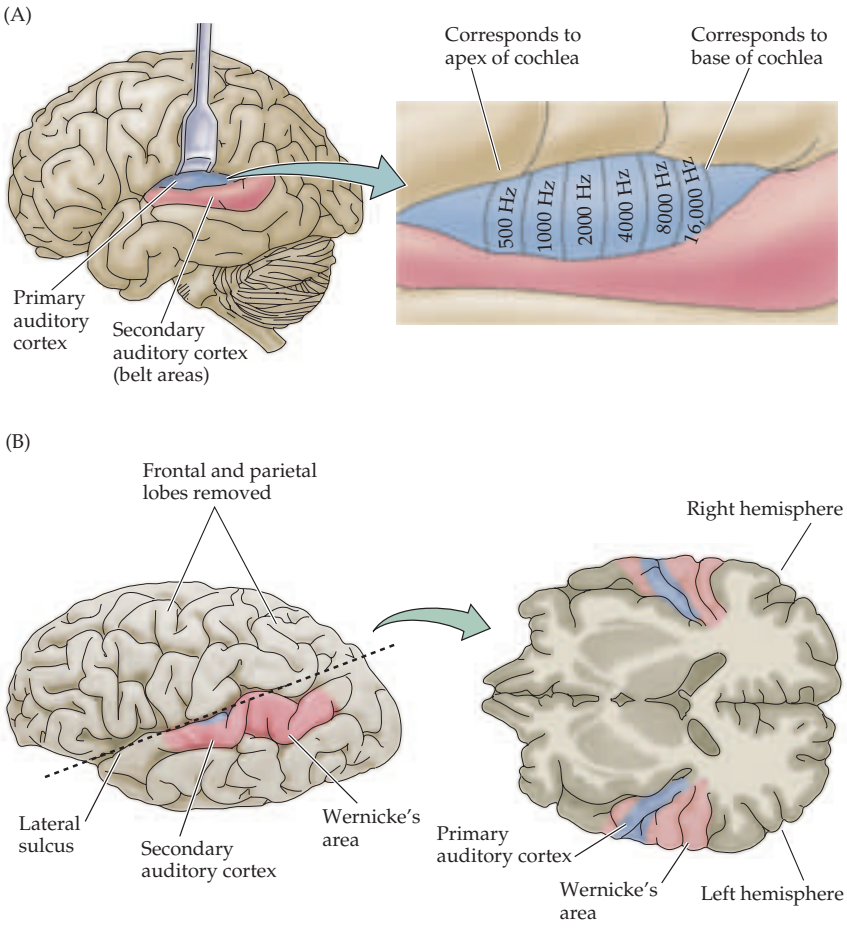
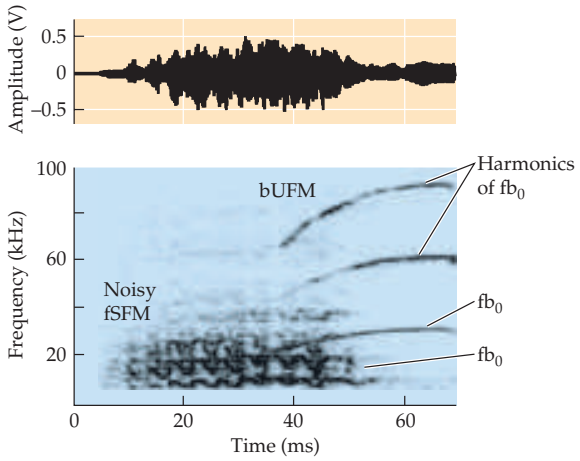


Figure 12.15 The human auditory cortex. (A) Diagram showing the brain in left lateral view, including the depths of the lateral sulcus, where part of the auditory cortex occupying the superior temporal gyrus normally lies hidden. The primary auditory cortex (A1) is shown in blue; the surrounding belt areas of the auditory cortex are in red. The primary auditory cortex has a tonotopic organization, as shown in this blowup diagram of a segment of A1 (right). (B) Diagram of the brain in left lateral view, showing locations of human auditory cortical areas related to processing speech sounds in the intact hemisphere. *Right:* An oblique section (plane of dashed line) shows the cortical areas on the superior surface of the temporal lobe. Note that Wernicke's area, a region important in comprehending speech, is just posterior to the primary auditory cortex.

Box E

Representing Complex Sounds in the Brains of Bats and Humans

Most natural sounds are complex, meaning that they differ from the pure tones or clicks that are frequently used in neurophysiological studies of the auditory system. Rather, natural sounds are tonal: they have a fundamental frequency that largely determines the "pitch" of the sound, and one or more harmonics of different intensities that contribute to the quality or "timbre" of a sound. The frequency of a harmonic is, by definition, a multiple of the fundamental frequency, and both may be modulated over time. Such *frequency-modulated* (FM) sweeps can rise or fall in frequency, or change in a sinusoidal or some other fashion. Occasionally, multiple nonharmonic frequencies may be simultaneously present in some communication or musical sounds. In some sounds, a level of spectral splatter or "broadband noise" is embedded within tonal or frequency modulated sounds. The variations in the sound spectrum are typically accompanied by a modulation of the amplitude envelop of the complex sound as well. All of these features can be visualized by performing a spectrographic analysis.

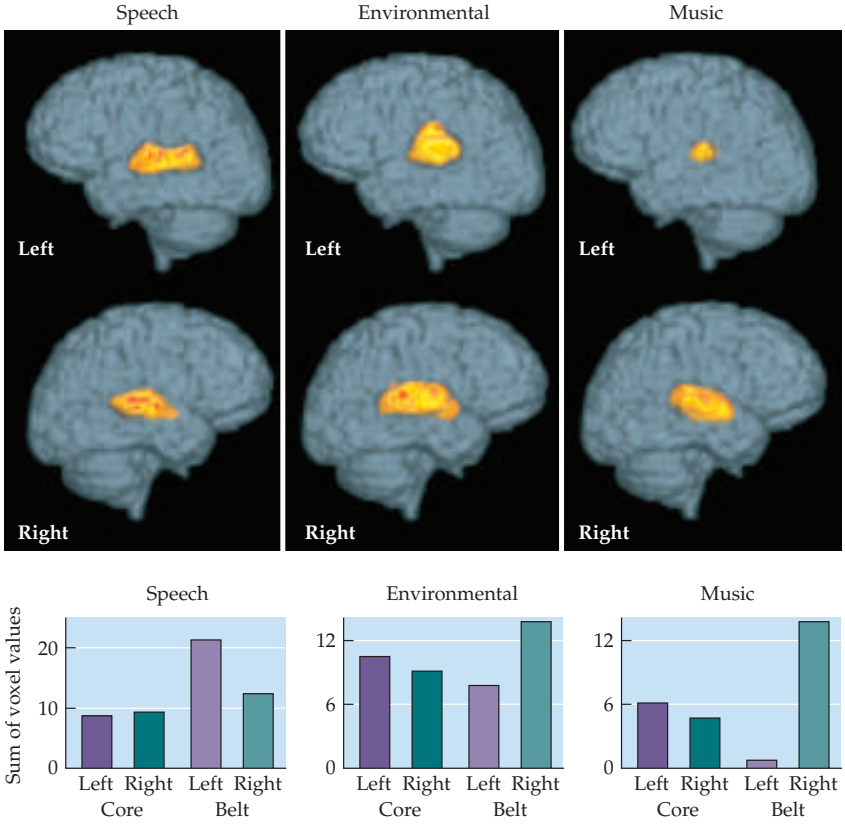


(A) Amplitude envelope (above) and spectrogram (below) of a composite syllable emitted by mustached bats for social communication. This composite consists of two simple syllables, a fixed Sinusoidal FM (fSFM) and a bent Upward FM (bUFM) that emerges from the fSFM after some overlap. Each syllable has its own fundamental (fa_0 and fb_0) and multiple harmonics. (Courtesy of Jagmeet Kanwal.)

How does the brain represent such complex natural sounds? Cognitive studies of complex sound perception provide some understanding of how a large but limited number of neurons in the brain can dynamically represent an infinite variety of natural stimuli in the sensory environment of humans and other animals. In bats, specializations for processing complex sounds are apparent. Studies in echolocating bats show that both communication and echolocation sounds (Figure A) are processed not only within some of the same areas, but also within the same neurons in the auditory cortex. In humans, multiple modes of processing are also likely, given the large overlap within the superior and middle temporal gyri in the temporal lobe for the representation of different types of complex sounds.

Asymmetrical representation is another common principle of complex sound processing that results in lateralized (though largely overlapping) representations of natural stimuli. Thus, speech sounds that are important for communication are lateralized to the left in the belt regions of the auditory cortex, whereas environmental sounds that are important for reacting to and recogniz-

epithelia. Unlike the visual and somatic sensory systems, however, the cochlea has already decomposed the acoustical stimulus so that it is arrayed tonotopically along the length of the basilar membrane. Thus, A1 is said to comprise a tonotopic map, as do most of the ascending auditory structures between the cochlea and the cortex. Orthogonal to the frequency axis of the tonotopic map is a striped arrangement of binaural properties. The neurons in one stripe are excited by both ears (and are therefore called EE cells), while the neurons in the next stripe are excited by one ear and inhibited by the other ear (EI cells). The EE and EI stripes alternate, an arrangement that is reminiscent of the ocular dominance columns in V1 (see Chapter 11).



(B) *Top:* Reconstructed functional magnetic resonance images of BOLD contrast signal change (average for 8 subjects) showing significant ($p < 0.001$) activation elicited by speech, environmental, and musical sounds on surface views of the left versus the right side of the human brain. *Bottom:* Bar graphs showing the total significant activation to each category of complex sounds in the core and belt areas of the auditory cortex for the left versus the right side. (Courtesy of Jagmeet Kanwal.)

ing aspects of the auditory environment are represented in each hemisphere (Figure B). Musical sounds that can either motivate us to march in war or to relax and meditate when coping with physical and emotional stress are highly lateralized to the right in the belt regions of the auditory cortex. The extent of lateralization for speech and possibly music may vary with sex, age, and training. In some species of bats, mice, and primates, processing of natural communication sounds appears to be lateralized to the left hemisphere. In summary, natural sounds are complex and their representation within the sensory cortex tends to be asymmetric across the two hemispheres.

References

Ehret, G. (1987) Left hemisphere advantage in the mouse brain for recognizing ultrasonic communication calls. Nature 325: 249–251.

Esser, K.-H., C. J. Condon, N. Suga and J. S. Kanwal (1997) Syntax processing by auditory cortical neurons in the FM-FM area of the mustached bat, *Pteronotus parnellii*. Proc. Natl. Acad. Sci. USA 94: 14019–14024.

Hauser, M. D. and K. Andersson (1994) Left hemisphere dominance for processing vocalizations in adult, but not infant, rhesus monkeys: Field experiments. Proc. Natl. Acad. Sci. USA 91: 3946-3948.

Kanwal, J. S., J. Kim and K. Kamada (2000) Separate, distributed processing of environmental, speech and musical sounds in the cerebral hemispheres. J. Cog. Neurosci. (Supp.): p. 32.

Kanwal, J. S., J. S. Matsumura, K. Ohlemiller and N. Suga (1994) Acoustic elements and syntax in communication sounds emitted by mustached bats. J. Acous. Soc. Am. 96: 1229–1254.

Kanwal, J. S. and N. Suga (1995) Hemispheric asymmetry in the processing of calls in the auditory cortex of the mustached bat. Assoc. Res. Otolaryngol. 18: 104.

The auditory cortex obviously does much more than provide a tonotopic map and respond differentially to ipsi- and contralateral stimulation. Although the sorts of sensory processing that occur in the auditory cortex are not well understood, they are likely to be important to higher-order processing of natural sounds, especially those used for communication (Box E; see also Chapter 26). One clue about such processing comes from work in marmosets, a small neotropical primate with a complex vocal repertoire. The primary auditory cortex and belt areas of these animals are indeed organized tonotopically, but also contain neurons that are strongly responsive to spectral combinations that characterize certain vocalizations. The responses

of these neurons to the tonal stimuli do not accurately predict their responses to the spectral combinations, suggesting that, in accord with peripheral optimization, cortical processing is in part dedicated to detecting particular intraspecific vocalizations.

Another clue about the role of the primary auditory cortex in the processing of intraspecific communication sounds comes from work in echolocating bats. Consistent with the essential role that echolocation plays in the survival of these crepuscular animals, certain regions of the bat primary auditory cortex, like those described in the MGC, are tuned in a systematic manner to the delays between frequency modulated pulses and their echoes, thus providing information about target distance and velocity. These delay-tuned neurons can exhibit highly specific responses to intraspecific communication calls, suggesting that the same cortical neurons can serve these two distinct auditory functions (see Box E). Evidently the general ability of the mammalian auditory cortex to detect certain spectral and temporal combinations of natural sounds has been exploited in bats to serve sonar-mediated navigation, yielding these dual function neurons.

Many of the dually specialized neurons are categorized as "combination-sensitive" neurons, i.e., neurons that show a nonlinear increase in their response magnitude when presented with a combination of tones and/or noise bands in comparison to the total magnitude of the response elicited by presenting each sound element separately. Combination-sensitive neurons are tuned to more than one frequency and are specialized to recognize complex species-specific sounds and extract information that is critical for survival. This sensitivity to combinations of simple sound elements appears to be a universal property of neurons for the perception of complex sounds by many animal species, such as frogs, birds bats and nonhuman primates. Therefore, combination-sensitive neurons most likely partake in the recognition of complex sounds in the human auditory cortex as well.

Sounds that are especially important for intraspecific communication often have a highly ordered temporal structure. In humans, the best example of such time-varying signals is speech, where different phonetic sequences are perceived as distinct syllables and words (see Box A in Chapter 26). Behavioral studies in cats and monkeys show that the auditory cortex is especially important for processing temporal sequences of sound. If the auditory cortex is ablated in these animals, they lose the ability to discriminate between two complex sounds that have the same frequency components but which differ in temporal sequence. Thus, without the auditory cortex, monkeys cannot discriminate one conspecific communication sound from another. The physiological basis of such temporal sensitivity likely requires neurons that are sensitive to time-varying cues in communication sounds. Indeed, electrophysiological recordings from the primary auditory cortex of both marmosets and bats show that some neurons that respond to intraspecific communication sounds do not respond as strongly when the sounds are played in reverse, indicating sensitivity to the sounds' temporal features. Studies of human patients with bilateral damage to the auditory cortex also reveal severe problems in processing the temporal order of sounds. It seems likely, therefore, that specific regions of the human auditory cortex are specialized for processing elementary speech sounds, as well as other temporally complex acoustical signals, such as music (Box B). Thus, Wernicke's area, which is critical to the comprehension of human language, lies within the secondary auditory area (Figure 12.15; see also Chapter 26).

Summary

Sound waves are transmitted via the external and middle ear to the cochlea of the inner ear, which exhibits a traveling wave when stimulated. For high-frequency sounds, the amplitude of the traveling wave reaches a maximum at the base of the cochlea; for low-frequency sounds, the traveling wave reaches a maximum at the apical end. The associated motions of the basilar membrane are transduced primarily by the inner hair cells, while the basilar membrane motion is itself actively modulated by the outer hair cells. Damage to the outer or middle ear results in conductive hearing loss, while hair cell damage results in a sensorineural hearing deficit. The tonotopic organization of the cochlea is retained at all levels of the central auditory system. Projections from the cochlea travel via the eighth nerve to the three main divisions of the cochlear nucleus. The targets of the cochlear nucleus neurons include the superior olivary complex and nuclei of the lateral lemniscus, where the binaural cues for sound localization are processed. The inferior colliculus is the target of nearly all of the auditory pathways in the lower brainstem and carries out important integrative functions, such as processing sound frequencies and integrating the cues for localizing sound in space. The primary auditory cortex, which is also organized tonotopically, is essential for basic auditory functions, such as frequency discrimination and sound localization, and also plays an important role in processing of intraspecific communication sounds. The belt areas of the auditory cortex have a less strict tonotopic organization and also process complex sounds, such as those that mediate communication. In the human brain, the major speech comprehension areas are located in the zone immediately adjacent to the auditory cortex.

Additional Reading

Reviews

Corey, D. P. and A. J. Hudspeth (1979) Ionic basis of the receptor potential in a vertebrate hair cell. Nature 281: 675–677.

Corey, D.P. (1999) Ion channel defects in hereditary hearing loss. Neuron. 22(2):217-9.

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

Garcia-Anoveros, J. and D. P. Corey (1997) The molecules of mechanosensation. Ann. Rev. Neurosci. 20: 567–597.

Heffner, H. E. and R. S. Heffner (1990) Role of primate auditory cortex in hearing. In *Comparative Perception, Volume II: Complex Signals*. W. C. Stebbins and M. A. Berkley (eds.). New York: John Wiley.

Hudspeth, A. J. (1997) How hearing happens. Neuron 19: 947–950.

Hudspeth, A. J. (2000) Hearing and deafness. Neurobiol. Dis. 7: 511–514.

Hudspeth, A. J. and M. Konishi (2000) Auditory neuroscience: Development, transduction, and integration. Proc. Natl. Acad. Sci. USA 97: 11690–11691.

Hudspeth, A. J., Y. Choe, A. D. Mehta and P. Martin (2000) Putting ion channels to work: Mechanoelectrical transduction, adaptation, and amplification by hair cells. Proc. Natl. Acad. Sci. USA 97: 11765-11772.

Kiang, N. Y. S. (1984) Peripheral neural processing of auditory information. In *Handbook of Physiology*, Section 1: *The Nervous System*, Volume III. *Sensory Processes*, Part 2. J. M. Brookhart, V. B. Mountcastle, I. Darian-Smith and S. R. Geiger (eds.). Bethesda, MD: American Physiological Society.

Neff, W. D., I. T. Diamond and J. H. Casseday (1975) Behavioral studies of auditory discrimination. In *Handbook of Sensory Physiology*, Volumes V–II. W. D. Keidel and W. D. Neff (eds.). Berlin: Springer-Verlag.

Nelken, I. (2002) Feature detection by the auditory cortex. In *Integrative Functions in the Mammalian Auditory Pathway, Springer Handbook of Auditory Research, Volume 15.* D. Oertel, R. Fay and A. N. Popper (eds.). New York: Springer-Verlag, pp. 358–416.

Suga, N. (1990) Biosonar and neural computation in bats. Sci. Am. 262 (June): 60–68.

Important Original Papers

Barbour, D. L. and X. Wang (2003) Contrast tuning in auditory cortex. Science. 299: 1073–1075.

Crawford, A. C. and R. Fettiplace (1981) An electrical tuning mechanism in turtle cochlear hair cells. J. Physiol. 312: 377–412.

Fitzpatrick, D. C., J. S. Kanwal, J. A. Butman and N. Suga (1993) Combination-sensitive neurons in the primary auditory cortex of the mustached bat. J. Neurosci. 13: 931–940.

Corey, D. P. and A. J. Hudspeth (1979) Ionic basis of the receptor potential in a vertebrate hair cell. Nature 281: 675–677.

Middlebrooks, J. C., A. E. Clock, L. Xu and D. M. Green (1994) A panoramic code for sound location by cortical neurons. Science 264: 842–844.

Knudsen, E. I. and M. Konishi (1978) A neural map of auditory space in the owl. Science 200: 795–797.

Jeffress, L. A. (1948) A place theory of sound localization. J. Comp. Physiol. Psychol. 41: 35–39.

NELKEN, I., Y. ROTMAN AND O. BAR YOSEF (1999) Responses of auditory-cortex neurons to structural features of natural sounds. Nature 397: 154–157.

SUGA, N., W. E. O'NEILL AND T. MANABE (1978) Cortical neurons sensitive to combinations of information-bearing elements of biosonar signals in the mustache bat. Science 200: 778–781.

VON BÉKÉSY, G. (1960) *Experiments in Hearing*. New York: McGraw-Hill. (A collection of von Békésy's original papers.)

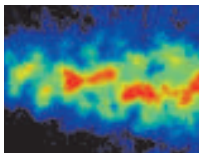
Books

PICKLES, J. O. (1988) *An Introduction to the Physiology of Hearing*. London: Academic Press.

YOST, W. A. AND G. GOUREVITCH (EDS.) (1987) *Directional Hearing*. Berlin: Springer Verlag.

YOST, W. A. AND D. W. NIELSEN (1985) *Fundamentals of Hearing*. Fort Worth: Holt, Rinehart and Winston.

Chapter 13



The Vestibular System

Overview

The vestibular system has important sensory functions, contributing to the perception of self-motion, head position, and spatial orientation relative to gravity. It also serves important motor functions, helping to stabilize gaze, head, and posture. The peripheral portion of the vestibular system includes inner ear structures that function as miniaturized accelerometers and inertial guidance devices, continually reporting information about the motions and position of the head and body to integrative centers in the brainstem, cerebellum, and somatic sensory cortices. The central portion of the system includes the vestibular nuclei, which make extensive connections with brainstem and cerebellar structures. The vestibular nuclei also directly innervate motor neurons controlling extraocular, cervical, and postural muscles. This motor output is especially important to stabilization of gaze, head orientation, and posture during movement. Although we are normally unaware of its functioning, the vestibular system is a key component in postural reflexes and eye movements. Balance, gaze stabilization during head movement, and sense of orientation in space are all adversely affected if the system is damaged. These manifestations of vestibular damage are especially important in the evaluation of brainstem injury. Because the circuitry of the vestibular system extends through a large part of the brainstem, simple clinical tests of vestibular function can be performed to determine brainstem involvement, even on comatose patients.

The Vestibular Labyrinth

The main peripheral component of the vestibular system is an elaborate set of interconnected chambers—the **labyrinth**—that has much in common, and is in fact continuous with, the cochlea (see Chapter 12). Like the cochlea, the labyrinth is derived from the otic placode of the embryo, and it uses the same specialized set of sensory cells—hair cells—to transduce physical motion into neural impulses. In the cochlea, the motion is due to airborne sounds; in the labyrinth, the motions transduced arise from head movements, inertial effects due to gravity, and ground-borne vibrations (Box A).

The labyrinth is buried deep in the temporal bone and consists of the two **otolith organs** (the **utricle** and **saccule**) and three **semicircular canals** (Figure 13.1). The elaborate and tortuous architecture of these components explains why this part of the vestibular system is called the labyrinth. The utricle and saccule are specialized primarily to respond to *linear accelerations* of the head and *static head position relative to the graviational axis*, whereas the semicircular canals, as their shapes suggest, are specialized for responding to *rotational accelerations* of the head.

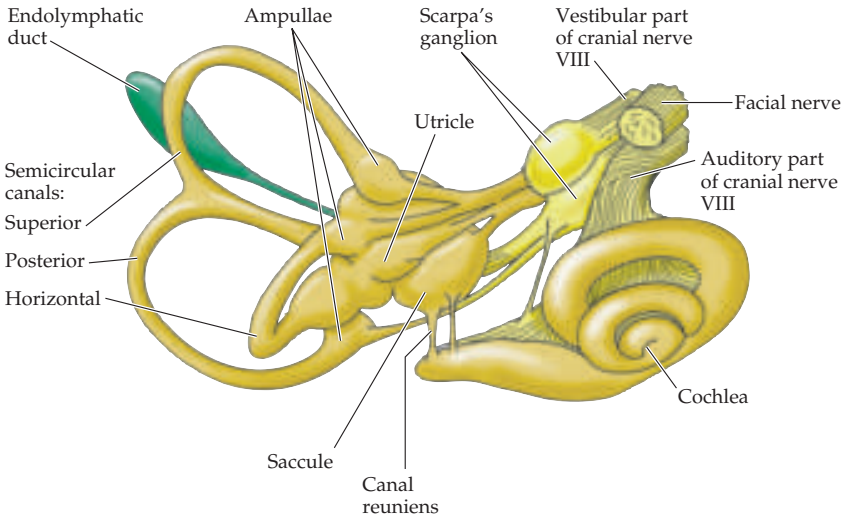


Figure 13.1 The labyrinth and its innervation. The vestibular and auditory portions of the eighth nerve are shown; the small connection from the vestibular nerve to the cochlea contains auditory efferent fibers. General orientation in head is shown in Figure 12.3; see also Figure 13.8.

The intimate relationship between the cochlea and the labyrinth goes beyond their common embryonic origin. Indeed, the cochlear and vestibular spaces are actually joined (see Figure 13.1), and the specialized ionic environments of the vestibular end organ parallel those of the cochlea. The membranous sacs within the bone are filled with fluid (endolymph) and are collectively called the membranous labyrinth. The endolymph (like the cochlear endolymph) is similar to intracellular solutions in that it is high in K^+ and low in Na^+. Between the bony walls (the osseous labyrinth) and the membranous labyrinth is another fluid, the perilymph, which is similar in composition to cerebrospinal fluid (i.e., low in K^+ and high in Na^+; see Chapter 12).

The vestibular hair cells are located in the utricle and saccule and in three juglike swellings called **ampullae**, located at the base of the semicircular canals next to the utricle. Within each ampulla, the vestibular hair cells extend their hair bundles into the endolymph of the membranous labyrinth. As in the cochlea, tight junctions seal the apical surfaces of the vestibular hair cells, ensuring that endolymph selectively bathes the hair cell bundle while remaining separate from the perilymph surrounding the basal portion of the hair cell.

Vestibular Hair Cells

The vestibular hair cells, which like cochlear hair cells transduce minute displacements into behaviorally relevant receptor potentials, provide the basis for vestibular function. Vestibular and auditory hair cells are quite similar; a detailed description of hair cell structure and function has already been given in Chapter 12. As in the case of auditory hair cells, movement of the stereocilia toward the kinocilium in the vestibular end organs opens mechanically gated transduction channels located at the tips of the stereocilia, depolarizing the hair cell and causing neurotransmitter release onto (and excitation of) the vestibular nerve fibers. Movement of the stereocilia in the direction away from the kinocilium closes the channels, hyperpolarizing the hair cell and thus reducing vestibular nerve activity. The biphasic nature of the receptor potential means that some transduction channels are open in the absence of stimulation, with the result that hair cells tonically release

transmitter, thereby generating considerable spontaneous activity in vestibular nerve fibers (see Figure 13.6). One consequence of these spontaneous action potentials is that the firing rates of vestibular fibers can increase or decrease in a manner that faithfully mimics the receptor potentials produced by the hair cells (Box B).

Importantly, the hair cell bundles in each vestibular organ have specific orientations (Figure 13.2). As a result, the organ as a whole is responsive to displacements in all directions. In a given semicircular canal, the hair cells in the ampulla are all polarized in the same direction. In the utricle and saccule, a specialized area called the **striola** divides the hair cells into two populations with opposing polarities (Figure 13.2C; see also Figure 13.4C). The directional polarization of the receptor surfaces is a basic principle of organization in the vestibular system, as will become apparent in the following descriptions of the individual vestibular organs.

The Otolith Organs: The Utricle and Saccule

Displacements and linear accelerations of the head, such as those induced by tilting or translational movements (see Box A), are detected by the two otolith organs: the saccule and the utricle. Both of these organs contain a

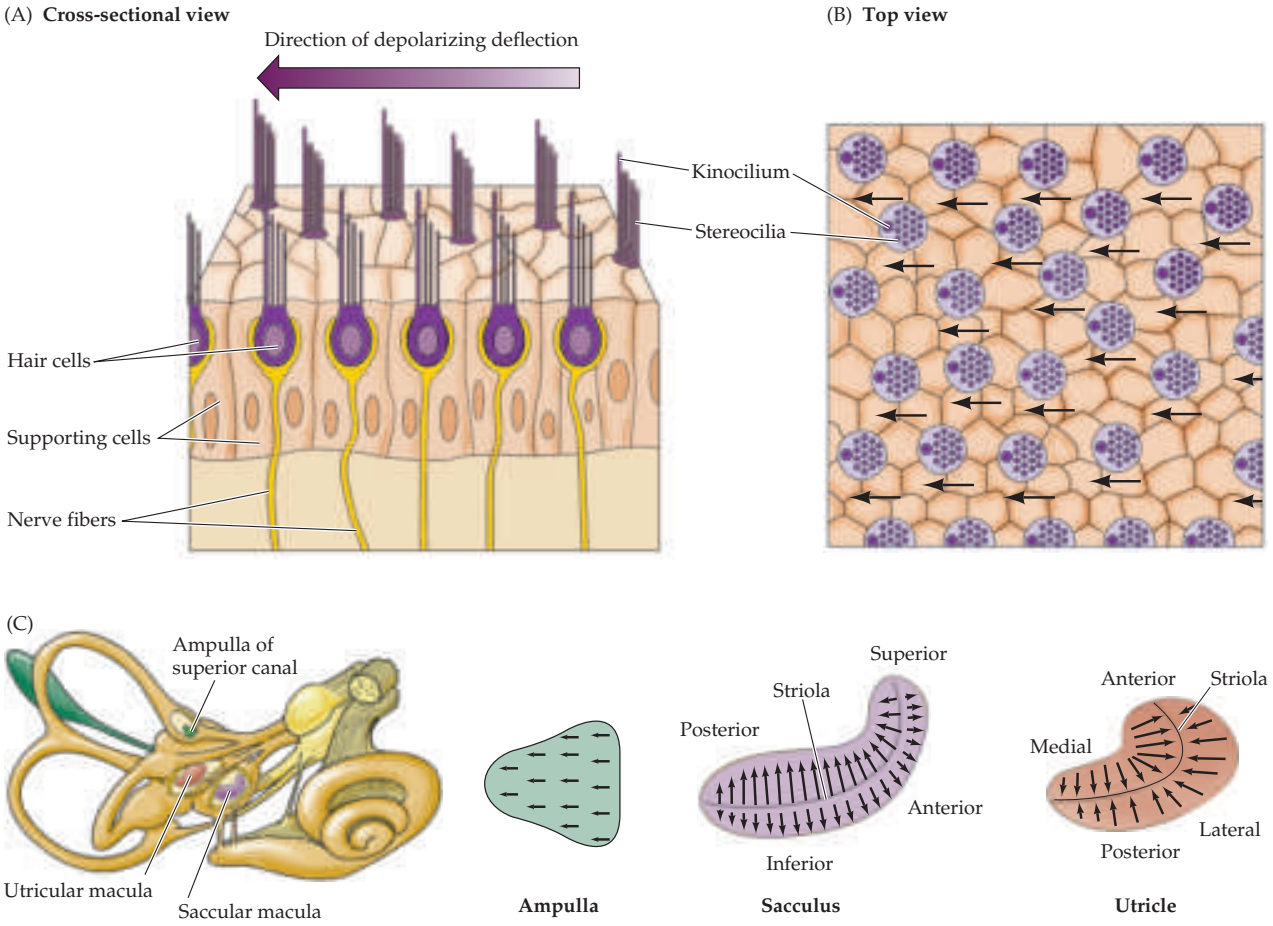
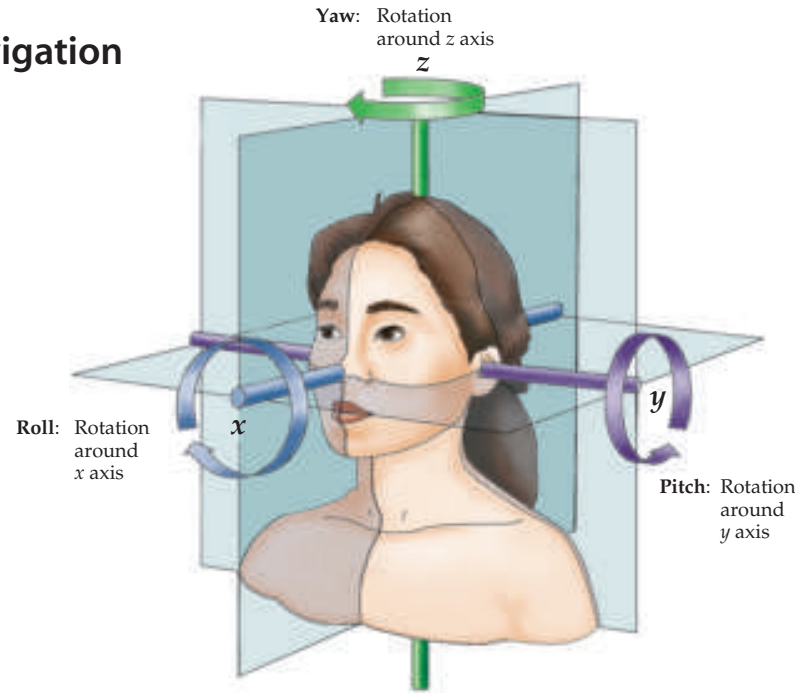


Figure 13.2 The morphological polarization of vestibular hair cells and the polarization maps of the vestibular organs. (A) A cross section of hair cells shows that the kinocilia of a group of hair cells are all located on the same side of the hair cell. The arrow indicates the direction of deflection that depolarizes the hair cell. (B) View looking down on the hair bundles. (C) In the ampulla located at the base of each semicircular canal, the hair bundles are oriented in the same direction. In the sacculus and utricle, the striola divides the hair cells into populations with opposing hair bundle polarities.

Box A

A Primer on Vestibular Navigation

The function of the vestibular system can be simplified by remembering some basic terminology of classical mechanics. All bodies moving in a three-dimensional framework have six degrees of freedom: three of these are translational and three are rotational. The translational elements refer to linear movements in the *x*, *y*, and *z* axes (the horizontal and vertical planes). Translational motion in these planes (linear acceleration and static displacement of the head) is the primary concern of the otolith organs. The three degrees of rotational freedom refer to a body's rotation relative to the *x*, *y*, and *z* axes and are commonly referred to as *roll*, *pitch*, and *yaw*. The semicircular canals are primarily responsible for sensing rotational accelerations around these three axes.



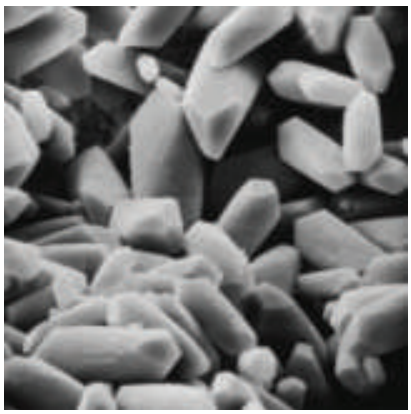


Figure 13.3 Scanning electron micrograph of calcium carbonate crystals (otoconia) in the utricular macula of the cat. Each crystal is about 50 mm long. (From Lindeman, 1973.)

sensory epithelium, the **macula**, which consists of hair cells and associated supporting cells. Overlying the hair cells and their hair bundles is a gelatinous layer; above this layer is a fibrous structure, the **otolithic membrane**, in which are embedded crystals of calcium carbonate called **otoconia** (Figures 13.3 and 13.4A). The crystals give the otolith organs their name (*otolith* is Greek for "ear stones"). The otoconia make the otolithic membrane considerably heavier than the structures and fluids surrounding it; thus, when the head tilts, gravity causes the membrane to shift relative to the sensory epithelium (Figure 13.4B). The resulting shearing motion between the otolithic membrane and the macula displaces the hair bundles, which are embedded in the lower, gelatinous surface of the membrane. This displacement of the hair bundles generates a receptor potential in the hair cells. A shearing motion between the macula and the otolithic membrane also occurs when the head undergoes linear accelerations (see Figure 13.5); the greater relative mass of the otolithic membrane causes it to lag behind the macula temporarily, leading to transient displacement of the hair bundle.

The similar effects exerted on otolithic hair cells by certain head tilts and linear accelerations would be expected to render these different stimuli perceptually equivalent when visual feedback is absent, as occurs in the dark or when the eyes are closed. Nevertheless, evidence suggests that subjects can discriminate between these two stimulus categories, apparently through combined activity of the otolith organs and the semicircular canals.

As already mentioned, the orientation of the hair cell bundles is organized relative to the striola, which demarcates the overlying layer of otoco-

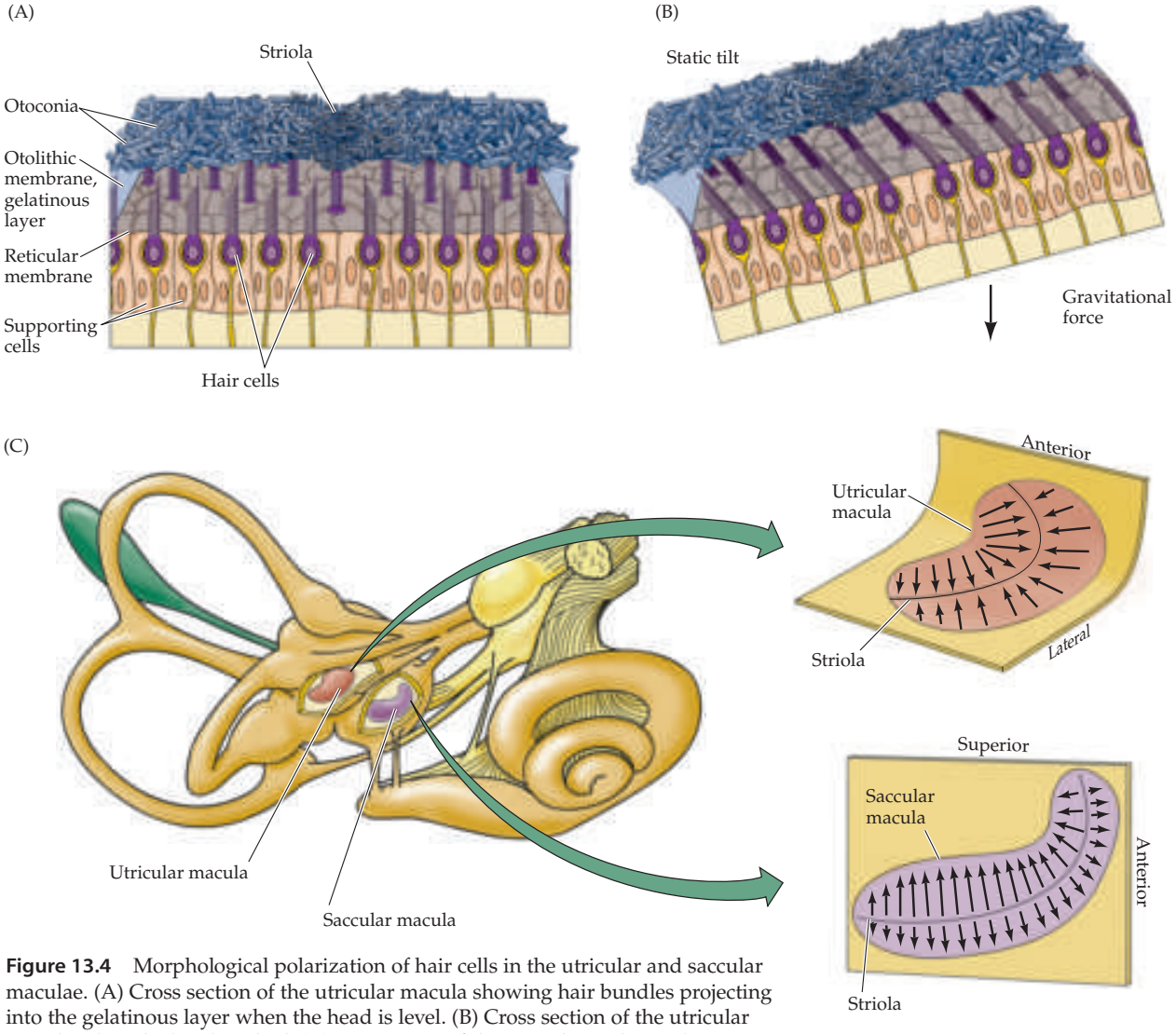


Figure 13.4 Morphological polarization of hair cells in the utricular and saccular maculae. (A) Cross section of the utricular macula showing hair bundles projecting into the gelatinous layer when the head is level. (B) Cross section of the utricular macula when the head is tilted. (C) Orientation of the utricular and saccular maculae in the head; arrows show orientation of the kinocilia, as in Figure 13.2. The *saccules* on either side are oriented more or less vertically, and the *utricles* more or less horizontally. The striola is a structural landmark consisting of small otoconia arranged in a narrow trench that divides each otolith organ. In the utricular macula, the kinocilia are directed toward the striola. In the saccular macula, the kinocilia point away from the striola. Note that, given the utricle and sacculus on both sides of the body, there is a continuous representation of all directions of body movement.

nia (see Figure 13.4A). The striola forms an axis of mirror symmetry such that hair cells on opposite sides of the striola have opposing morphological polarizations. Thus, a tilt along the axis of the striola will excite the hair cells on one side while inhibiting the hair cells on the other side. The saccular macula is oriented vertically and the utricular macula horizontally, with a continuous variation in the morphological polarization of the hair cells

Box B

Adaptation and Tuning of Vestibular Hair Cells

Hair Cell Adaptation

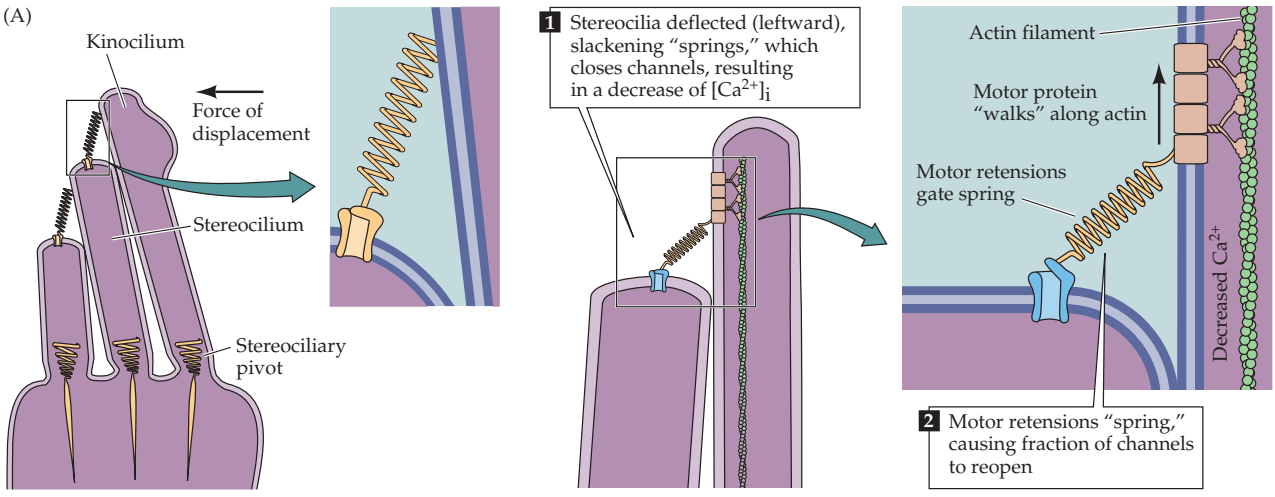
The minuscule movement of the hair bundle at sensory threshold has been compared to the displacement of the top of the Eiffel Tower by a thumb's breadth! Despite its great sensitivity, the hair cell can adapt quickly and continuously to static displacements of the hair bundle caused by large movements. Such adjustments are especially useful in the otolith organs, where adaptation permits hair cells to maintain sensitivity to small linear and angular accelerations of the head despite the constant input from gravitational forces that are over a million times greater. In other receptor cells, such as photoreceptors, adaptation is accomplished by regulating the second messenger cascade induced by the initial transduction event. The hair cell has to depend on a different strategy, however, because there is no second messenger system between the initial transduction event and the subsequent receptor potential (as might be expected for receptors that respond so rapidly).

Adaptation occurs in both directions in which the hair bundle displacement generates a receptor potential, albeit at different rates for each direction. When the hair bundle is pushed toward the kinocilium, tension is initially increased in the gating spring. During adaptation, tension decreases back to the resting level, perhaps because one end of the gating spring repositions itself along the shank of the stereocilium. When the hair bundle is displaced in the opposite direction, away from the kinocilium, tension in the spring initially decreases; adaptation then involves an increase in spring tension. One theory is that a calcium-regulated motor such as a myosin ATPase climbs along actin filaments in the stereocilium and actively resets the tension in the transduction spring. During sustained depolarization, some Ca^{2+} enters through the transduction channel, along with K^+. Ca^{2+} then causes the motor to spend a greater fraction of its time unbound from the actin, resulting in slippage of the spring down the side of the stereocilium. During sustained hyperpolarization (Figure A), Ca^{2+} levels drop below normal resting levels and the motor spends more of its time bound to the actin, thus climbing up the actin filaments and increasing the spring tension. As tension increases, some of the previously closed transduction channels open, admitting Ca^{2+} and thus slowing the motor's progress until a balance is struck between the climbing and slipping of the motor. In support of this model, when internal Ca^{2+} is reduced artificially, spring tension increases. This model of hair cell adaptation presents an elegant molecular solution to the regulation of a mechanical process.

Electrical Tuning

Although mechanical tuning plays an important role in generating frequency selectivity in the cochlea, there are other mechanisms that contribute to this process in vestibular and auditory nerve cells. These other tuning mechanisms are especially important in the otolith organs, where, unlike the cochlea, there are no

(A) Adaptation is explained in the gating spring model by adjustment of the insertion point of tips links. Movement of the insertion point up or down the shank of the stereocilium, perhaps driven by a Ca^{2+}-dependent protein motor, can continually adjust the resting tension of the tip link. (After Hudspeth and Gillespie, 1994.)



obvious macromechanical resonances to selectively filter and/or enhance biologically relevant movements. One such mechanism is an electrical resonance displayed by hair cells in response to depolarization: The membrane potential of a hair cell undergoes damped sinusoidal oscillations at a specific frequency in response to the injection of depolarizing current pulses (Figure B).

The ionic mechanism of this process involves two major types of ion channels located in the membrane of the hair cell soma. The first of these is a voltage-activated Ca^{2+} conductance, which lets Ca^{2+} into the cell soma in response to depolarization, such as that generated by the transduction current. The second is a Ca^{2+}-activated K^+ conductance, which is triggered by the rise in internal Ca^{2+} concentration. These two currents produce an interplay of depolarization and repolarization that results in electrical resonance (Figure C). Activation of the hair cell's calcium-activated K^+ conductance occurs 10 to 100 times faster than that of similar currents in other cells. Such rapid kinetics allow this conductance to generate an electrical response that usually requires the fast properties of a voltage-gated channel.

Although a hair cell responds to hair bundle movement over a wide range of frequencies, the resultant receptor potential is largest at the frequency of electrical resonance. The resonance frequency represents the characteristic frequency of the hair cell, and transduction at that frequency will be most efficient. This electrical resonance has important implications for structures like the utricle and sacculus, which may encode a range of characteristic frequencies based on the different resonance frequencies of their constituent hair cells. Thus, electrical tuning in the otolith organs can generate enhanced tuning to biologically relevant frequencies of stimulation, even in the absence of macromechanical resonances within these structures.

References

Assad, J. A. and D. P. Corey (1992) An active motor model for adaptation by vertebrate hair cells. J. Neurosci. 12: 3291–3309.

Crawford, A. C. and R. Fettiplace (1981) An electrical tuning mechanism in turtle cochlear hair cells. J. Physiol. 312: 377–412.

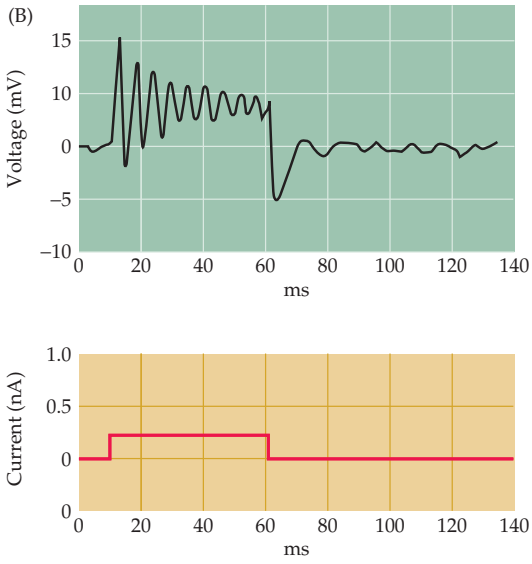
Hudspeth, A. J. (1985) The cellular basis of hearing: The biophysics of hair cells. Science 230: 745–752.

Hudspeth, A. J. and P. G. Gillespie (1994) Pulling strings to tune transduction: Adaptation by hair cells. Neuron 12: 1–9.

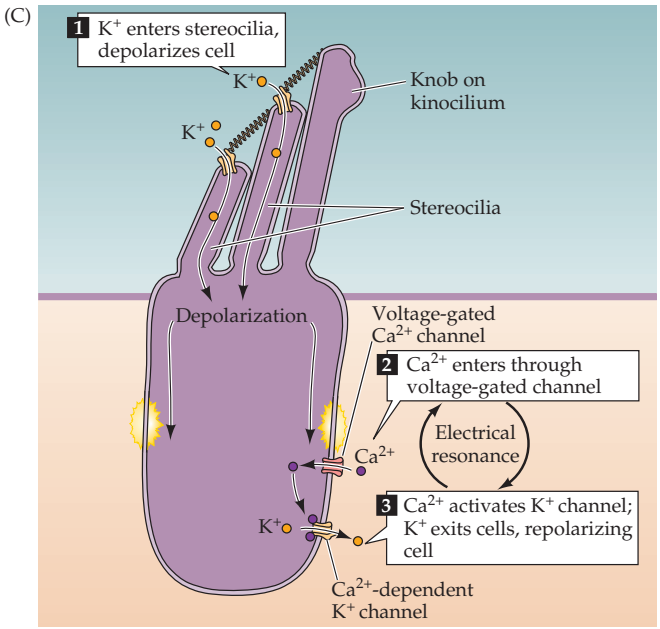
Lewis, R. S. and A. J. Hudspeth (1988) A model for electrical resonance and frequency tuning in saccular hair cells of the bull-frog, *Rana catesbeiana*. J. Physiol. 400: 275–297.

Lewis, R. S. and A. J. Hudspeth (1983) Voltage- and ion-dependent conductances in solitary vertebrate hair cells. Nature 304: 538–541.

Shepherd, G. M. G. and D. P. Corey (1994) The extent of adaptation in bullfrog saccular hair cells. J. Neurosci. 14: 6217–6229.



(B) Voltage oscillations (upper trace) in an isolated hair cell in response to a depolarizing current injection (lower trace). (After Lewis and Hudspeth, 1983.)



(C) Proposed ionic basis for electrical resonance in hair cells. (After Hudspeth, 1985.)

located in each macula (as shown in Figure 13.4C, where the arrows indicate the direction of movement that produces excitation). Inspection of the excitatory orientations in the maculae indicates that the utricle responds to movements of the head in the horizontal plane, such as sideways head tilts and rapid lateral displacements, whereas the saccule responds to movements in the vertical plane (up–down and forward–backward movements in the sagittal plane).

Note that the saccular and utricular maculae on one side of the head are mirror images of those on the other side. Thus, a tilt of the head to one side has opposite effects on corresponding hair cells of the two utricular maculae. This concept is important in understanding how the central connections of the vestibular periphery mediate the interaction of inputs from the two sides of the head (see the next section).

How Otolith Neurons Sense Linear Forces

The structure of the otolith organs enables them to sense both static displacements, as would be caused by tilting the head relative to the gravitational axis, and transient displacements caused by translational movements of the head. The mass of the otolithic membrane relative to the surrounding endolymph, as well as the otolithic membrane's physical uncoupling from the underlying macula, means that hair bundle displacement will occur transiently in response to linear accelerations and tonically in response to tilting of the head. Therefore, both tonic and transient information can be conveyed by these sense organs. Figure 13.5 illustrates some of the forces produced by head tilt and linear accelerations on the utricular macula.

These properties of hair cells are reflected in the responses of the vestibular nerve fibers that innervate the otolith organs. The nerve fibers have a

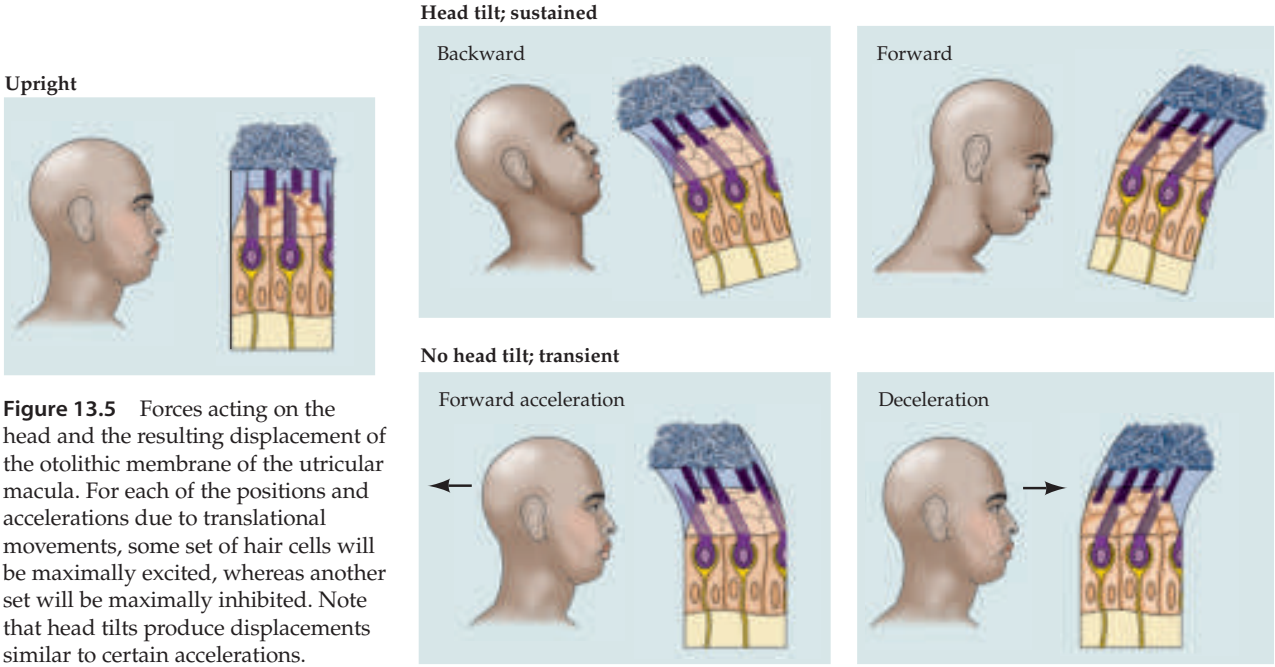


Figure 13.5 Forces acting on the head and the resulting displacement of the otolithic membrane of the utricular macula. For each of the positions and accelerations due to translational movements, some set of hair cells will be maximally excited, whereas another set will be maximally inhibited. Note that head tilts produce displacements similar to certain accelerations.

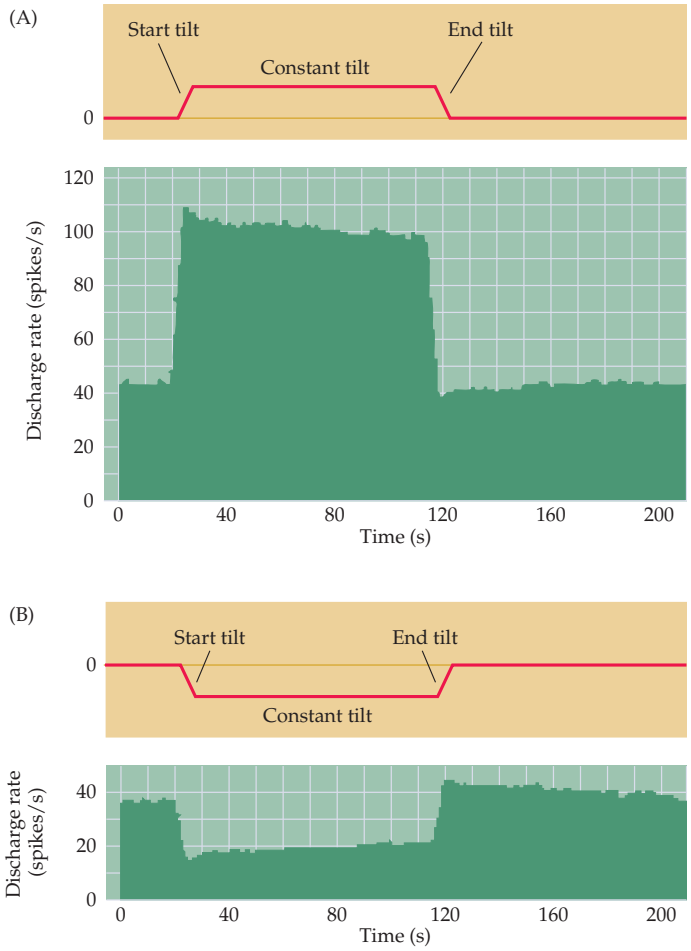


Figure 13.6 Response of a vestibular nerve axon from an otolith organ (the utricle in this example). (A) The stimulus (top) is a change in head tilt. The spike histogram shows the neuron's response to tilting in a particular direction. (B) A response of the same fiber to tilting in the opposite direction. (After Goldberg and Fernandez, 1976.)

steady and relatively high firing rate when the head is upright. The change in firing rate in response to a given movement can be either sustained or transient, thereby signaling either absolute head position or linear acceleration. An example of the sustained response of a vestibular nerve fiber innervating the utricle is shown in Figure 13.6. The responses were recorded from axons in a monkey seated in a chair that could be tilted for several seconds to produce a steady force. Prior to the tilt, the axon has a high firing rate, which increases or decreases depending on the direction of the tilt. Notice also that the response remains at a high level as long as the tilting force remains constant; thus, such neurons faithfully encode the static force being applied to the head (Figure 13.6A). When the head is returned to the original position, the firing level of the neurons returns to baseline value. Conversely, when the tilt is in the opposite direction, the neurons respond by decreasing their firing rate below the resting level (Figure 13.6B) and remain depressed as long as the static force continues. In a similar fashion, transient increases or decreases in firing rate from spontaneous levels signal the direction of linear accelerations of the head.

The range of orientations of hair bundles within the otolith organs enables them to transmit information about linear forces in every direction

the body moves (see Figure 13.4C). The utricle, which is primarily concerned with motion in the horizontal plane, and the saccule, which is concerned with vertical motion, combine to effectively gauge the linear forces acting on the head at any instant in three dimensions. Tilts of the head off the horizontal plane and translational movements of the head in any direction stimulate a distinct subset of hair cells in the saccular and utricular maculae, while simultaneously suppressing the responses of other hair cells in these organs. Ultimately, variations in hair cell polarity within the otolith organs produce patterns of vestibular nerve fiber activity that, at a population level, can unambiguously encode head position and the forces that influence it.

The Semicircular Canals

Whereas the otolith organs are primarily concerned with head translations and orientation with respect to gravity, the semicircular canals sense head *rotations*, arising either from self-induced movements or from angular accelerations of the head imparted by external forces. Each of the three semicircular canals has at its base a bulbous expansion called the **ampulla** (Figure 13.7), which houses the sensory epithelium, or **crista**, that contains the hair cells. The structure of the canals suggests how they detect the angular accelerations that arise through rotation of the head. The hair bundles extend out of the crista into a gelatinous mass, the **cupula**, that bridges the width of the ampulla, forming a fluid barrier through which endolymph cannot circulate. As a result, the relatively compliant cupula is distorted by movements of the endolymphatic fluid. When the head turns in the plane of one of the semicircular canals, the inertia of the endolymph produces a force across the cupula, distending it away from the direction of head movement and causing a displacement of the hair bundles within the crista (Figure 13.8A,B). In contrast, linear accelerations of the head produce equal forces on the two sides of the cupula, so the hair bundles are not displaced.

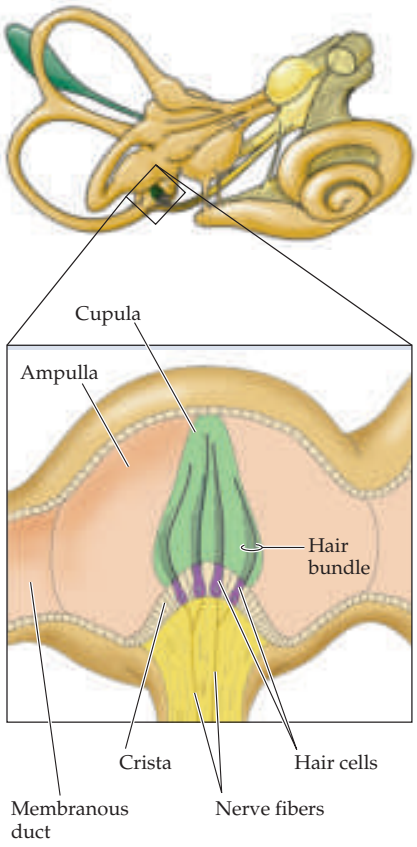


Figure 13.7 The ampulla of the posterior semicircular canal showing the crista, hair bundles, and cupula. The cupula is distorted by the fluid in the membranous canal when the head rotates.

Unlike the saccular and utricular maculae, all of the hair cells in the crista within each semicircular canal are organized with their kinocilia pointing in the same direction (see Figure 13.2C). Thus, when the cupula moves in the appropriate direction, the entire population of hair cells is depolarized and activity in all of the innervating axons increases. When the cupula moves in the opposite direction, the population is hyperpolarized and neuronal activity decreases. Deflections orthogonal to the excitatory–inhibitory direction produce little or no response.

Each semicircular canal works in concert with the partner located on the other side of the head that has its hair cells aligned oppositely. There are three such pairs: the two pairs of horizontal canals, and the superior canal on each side working with the posterior canal on the other side (Figure 13.8C). Head rotation deforms the cupula in opposing directions for the two partners, resulting in opposite changes in their firing rates (Box C). Thus, the orientation of the horizontal canals makes them selectively sensitive to rotation in the horizontal plane. More specifically, the hair cells in the canal towards which the head is turning are depolarized, while those on the other side are hyperpolarized.

For example, when the head accelerates to the left, the cupula is pushed toward the kinocilium in the left horizontal canal, and the firing rate of the relevant axons in the left vestibular nerve increases. In contrast, the cupula in the right horizontal canal is pushed away from the kinocilium, with a concomitant decrease in the firing rate of the related neurons. If the head movement is

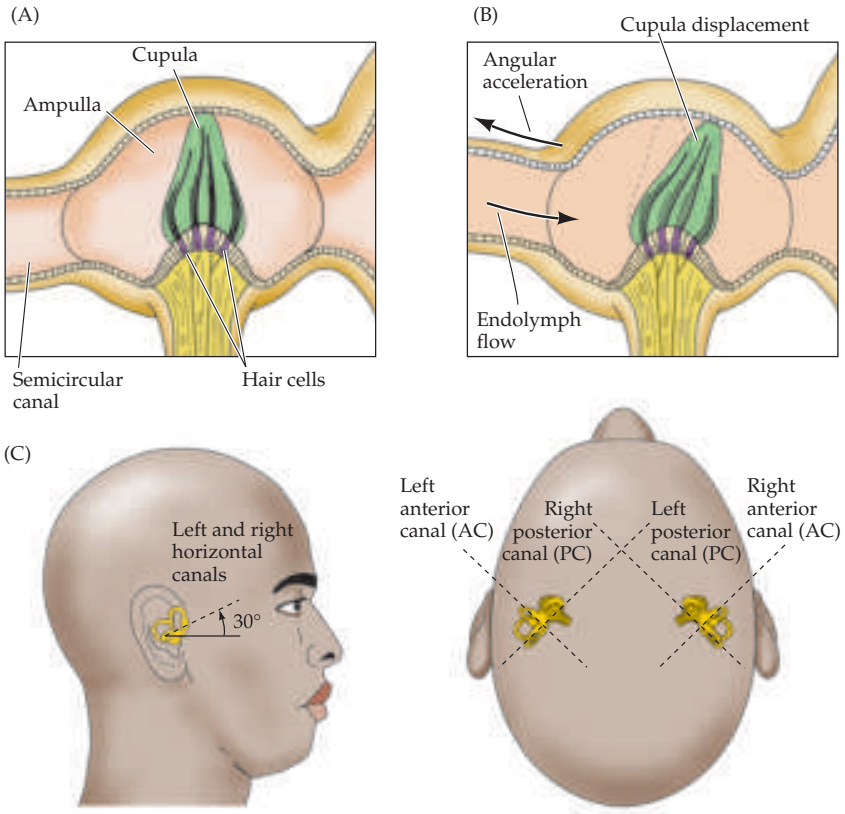


Figure 13.8 Functional organization of the semicircular canals. (A) The position of the cupula without angular acceleration. (B) Distortion of the cupula during angular acceleration. When the head is rotated in the plane of the canal (arrow outside canal), the inertia of the endolymph creates a force (arrow inside the canal) that displaces the cupula. (C) Arrangement of the canals in pairs. The two horizontal canals form a pair; the right anterior canal (AC) and the left posterior canal (PC) form a pair; and the left AC and the right PC form a pair.

to the right, the result is just the opposite. This push–pull arrangement operates for all three pairs of canals; the pair whose activity is modulated is in the plane of the rotation, and the member of the pair whose activity is increased is on the side toward which the head is turning. The net result is a system that provides information about the rotation of the head in any direction.

How Semicircular Canal Neurons Sense Angular Accelerations

Like axons that innervate the otolith organs, the vestibular fibers that innervate the semicircular canals exhibit a high level of spontaneous activity. As a result, they can transmit information by either increasing or decreasing their firing rate, thus more effectively encoding head movements (see above). The bidirectional responses of fibers innervating the hair cells of the semicircular canal have been studied by recording the axonal firing rates in a monkey's

Box C

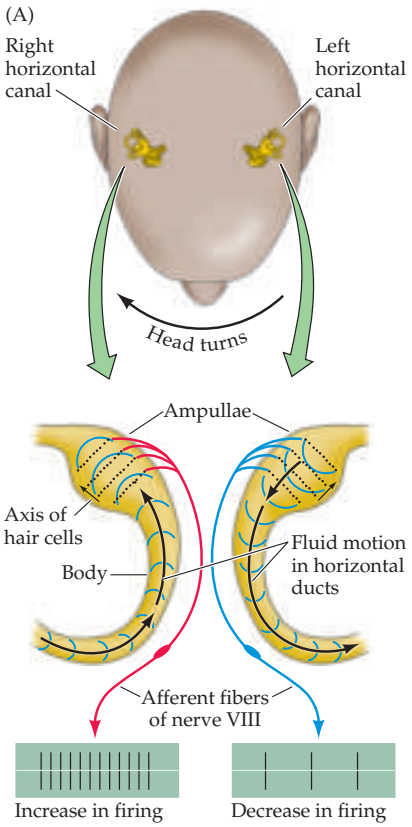
Throwing Cold Water on the Vestibular System

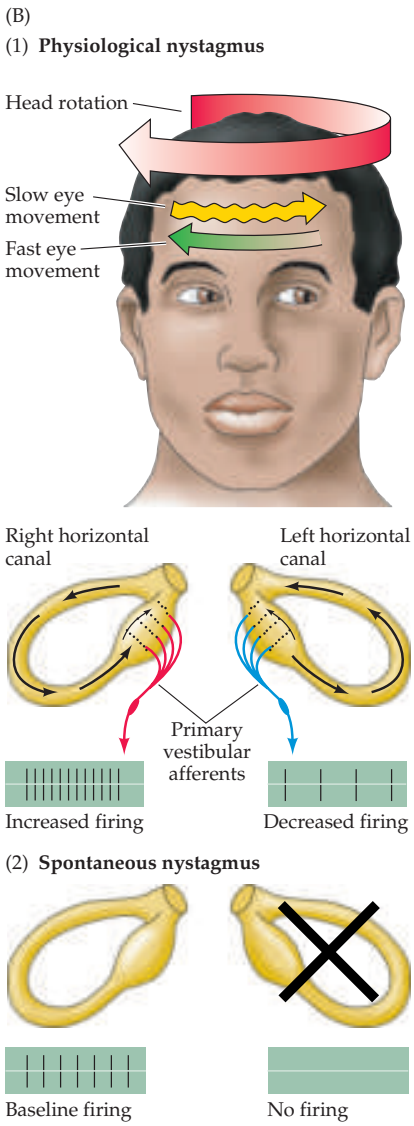
Testing the integrity of the vestibular system can indicate much about the condition of the brainstem, particularly in comatose patients.

Normally, when the head is not being rotated, the output of the nerves from the right and left sides are equal; thus, no eye movements occur. When the head is rotated in the horizontal plane, the vestibular afferent fibers on the side toward the turning motion increase their firing rate, while the afferents on the opposite side decrease their firing rate (Figures A and B). The net difference in firing rates then leads to slow movements of the eyes counter to the turning motion. This reflex response generates the slow component of a normal eye movement pattern called physiological nystagmus, which means "nodding" or oscillatory movements of the eyes (Figure B1). (The fast component is a saccade that resets the eye position; see Chapter 19.)

Pathological nystagmus can occur if there is unilateral damage to the vestibular system. In this case, the silencing of the spontaneous output from the damaged side results in an unphysiological difference in firing rate because the spontaneous discharge from the intact side remains (Figure B2). The difference in firing rates causes nystagmus, even though no head movements are being made.

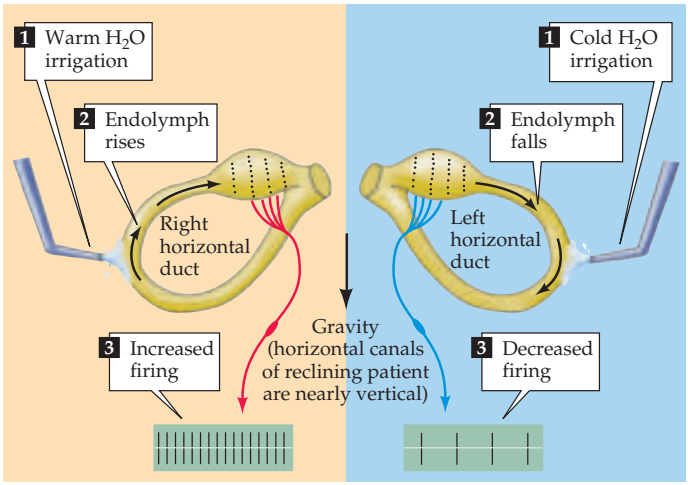
Responses to vestibular stimulation are thus useful in assessing the integrity of the brainstem in unconscious patients. If the individual is placed on his or her back and the head is elevated to about 30° above horizontal, the horizontal semicircular canals lie in an almost vertical orientation. Irrigating one ear with cold water will then lead to spontaneous eye movements because convection currents in the canal mimic rotatory head movements away from the irrigated ear (Figure C). In normal individuals, these eye movements consist of a slow movement toward the irrigated ear and a fast movement away from it. The fast movement is most readily detected by the observer, and the significance of its direction can be kept in mind by using the





(A) View looking down on the top of a person's head illustrates the fluid motion generated in the left and right horizontal canals, and the changes in vestibular nerve firing rates when the head turns to the right. (B) In normal individuals, rotating the head elicits physiological nystagmus (1), which consists of a slow eye movement counter to the direction of head turning. The slow component of the eye movements is due to the net differences in left and right vestibular nerve firing rates acting via the central circuit diagrammed in Figure 13.10. Spontaneous nystagmus (2), where the eyes move rhythmically from side to side in the absence of any head movements, occurs when one of the canals is damaged. In this situation, net differences in vestibular nerve firing rates exist even when the head is stationary because the vestibular nerve innervating the intact canal fires steadily when at rest, in contrast to a lack of activity on the damaged side.

(C)

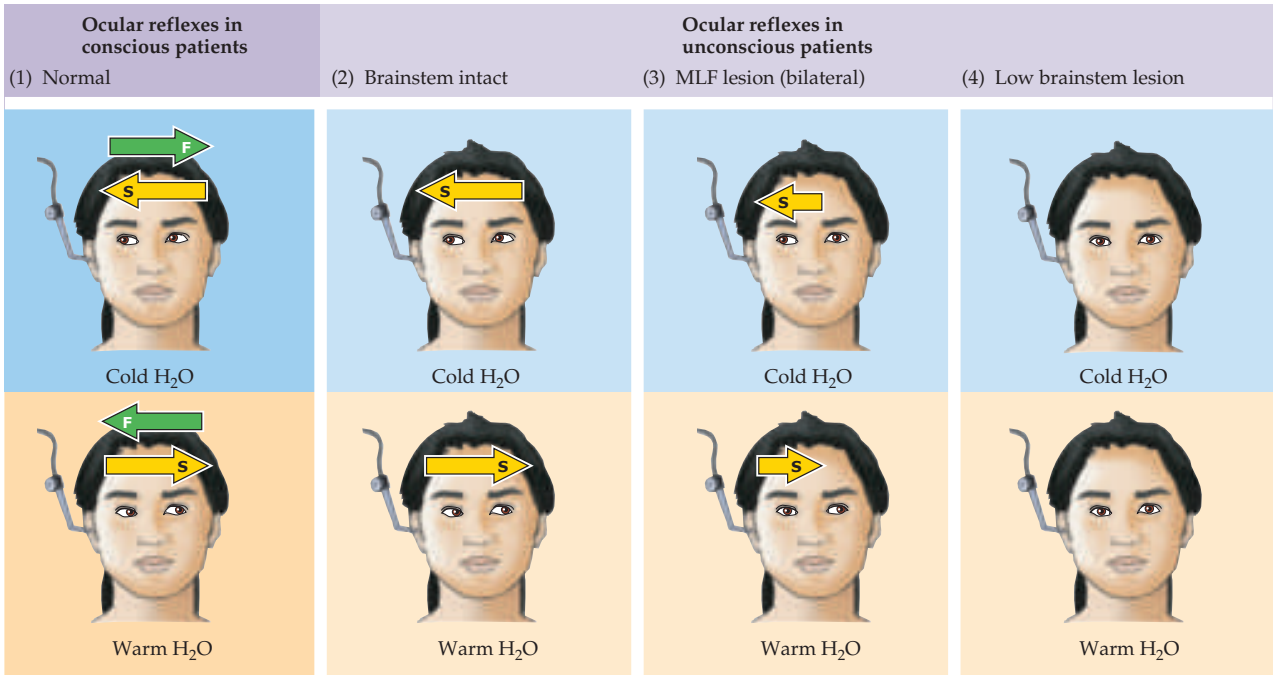


(C) Caloric testing of vestibular function is possible because irrigating an ear with water slightly warmer than body temperature generates convection currents in the canal that mimic the endolymph movement induced by turning the head to the irrigated side. Irrigation with cold water induces the opposite effect. These currents result in changes in the firing rate of the associated vestibular nerve, with an increased rate on the warmed side and a decreased rate on the chilled side. As in head rotation and spontaneous nystagmus, net differences in firing rates generate eye movements.

mnemonic COWS ("Cold Opposite, Warm Same"). This same test can also be used in unconscious patients. In patients who are comatose due to dysfunction of both cerebral hemispheres but whose brainstem is intact, saccadic movements are no longer made and the response to cold water consists of only the slow movement component of the eyes to side of the irrigated ear (Figure D). In the presence of brainstem lesions involving either the vestibular nuclei themselves, the connections from the vestibular nuclei to oculomotor nuclei (the third, fourth, or sixth cranial nerves), or the peripheral nerves exiting these nuclei, vestibular responses are abolished (or altered, depending on the severity of the lesion).

(D) Caloric testing can be used to test the function of the brainstem in an unconscious patient. The figures show eye movements resulting from cold or warm water irrigation in one ear for (1) a normal subject, and in three different conditions in an unconscious patient: (2) with the brainstem intact; (3) with a lesion of the medial longitudinal fasciculus (MLF; note that irrigation in this case results in lateral movement of the eye only on the less active side); and (4) with a low brainstem lesion (see Figure 13.10).

(D)



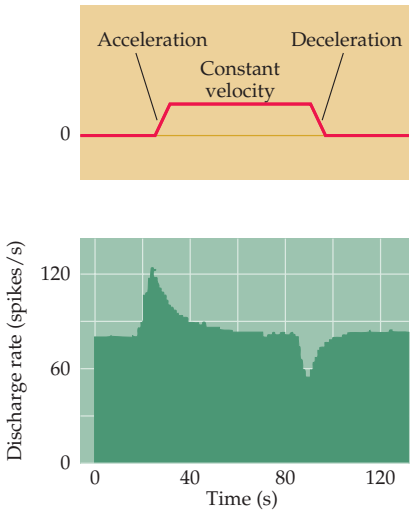


Figure 13.9 Response of a vestibular nerve axon from the semicircular canal to angular acceleration. The stimulus (top) is a rotation that first accelerates, then maintains constant velocity, and then decelerates the head. The axon increases its firing above resting level in response to the acceleration, returns to resting level during constant velocity, then decreases its firing rate below resting level during deceleration; these changes in firing rate reflect inertial effects on the displacement of the cupula. (After Goldberg and Fernandez, 1971.)

vestibular nerve. Seated in a chair, the monkey was rotated continuously in one direction during three phases: an initial period of acceleration, then a periord of several seconds at constant velocity, and finally a period of sudden deceleration to a stop (Figure 13.9). The maximum firing rates observed correspond to the period of acceleration; the maximum inhibition corresponds to the period of deceleration. During the constant-velocity phase, the response adapts so that the firing rate subsides to resting level; after the movement stops, the neuronal activity decreases transiently before returning to the resting level.

Neurons innervating paired semicircular canals have a complementary response pattern. Note that the rate of adaptation (on the order of tens of seconds) corresponds to the time it takes the cupula to return to its undistorted state (and for the hair bundles to return to their undeflected position); adaptation therefore can occur while the head is still turning, as long as a constant angular velocity is maintained. Such constant forces are rare in nature, although they are encountered on ships, airplanes, and space vehicles, where prolonged acceleratory arcs are sometimes described.

Central Pathways for Stabilizing Gaze, Head, and Posture

The vestibular end organs communicate via the vestibular branch of cranial nerve VIII with targets in the brainstem and the cerebellum that process much of the information necessary to compute head position and motion. As with the cochlear nerve, the vestibular nerves arise from a population of bipolar neurons, the cell bodies of which in this instance reside in the **vestibular nerve ganglion** (also called **Scarpa's ganglion**; see Figure 13.1). The distal processes of these cells innervate the semicircular canals and the otolith organs, while the central processes project via the vestibular portion of cranial nerve VIII to the **vestibular nuclei** (and also directly to the cerebellum; Figure 13.10). The vestibular nuclei are important centers of integration, receiving input from the vestibular nuclei of the opposite side, as well as from the cerebellum and the visual and somatic sensory systems. Because vestibular and auditory fibers run together in the eighth nerve, damage to this structure often results in both auditory and vestibular disturbances.

The central projections of the vestibular system participate in three major classes of reflexes: (1) helping to maintain equilibrium and gaze during movement, (2) maintaining posture, and (3) maintaining muscle tone. The first of these reflexes helps coordinate head and eye movements to keep gaze fixated on objects of interest during movements (other functions include protective or escape reactions; see Box D). The **vestibulo-ocular reflex** (**VOR**) in particular is a mechanism for producing eye movements that counter head movements, thus permitting the gaze to remain fixed on a particular point (Box C; see also Chapter 19). For example, activity in the left horizontal canal induced by leftward rotary acceleration of the head excites neurons in the left vestibular nucleus and results in compensatory eye movements to the right. This effect is due to excitatory projections from the vestibular nucleus to the contralateral nucleus abducens that, along with the oculomotor nucleus, help execute conjugate eye movements.

For instance, horizontal movement of the two eyes toward the right requires contraction of the left medial and right lateral rectus muscles. Vestibular nerve fibers originating in the left horizontal semicircular canal project to the medial and lateral vestibular nuclei (see Figure 13.10). Excitatory fibers from the medial vestibular nucleus cross to the contralateral abducens nucleus, which has two outputs. One of these is a motor pathway

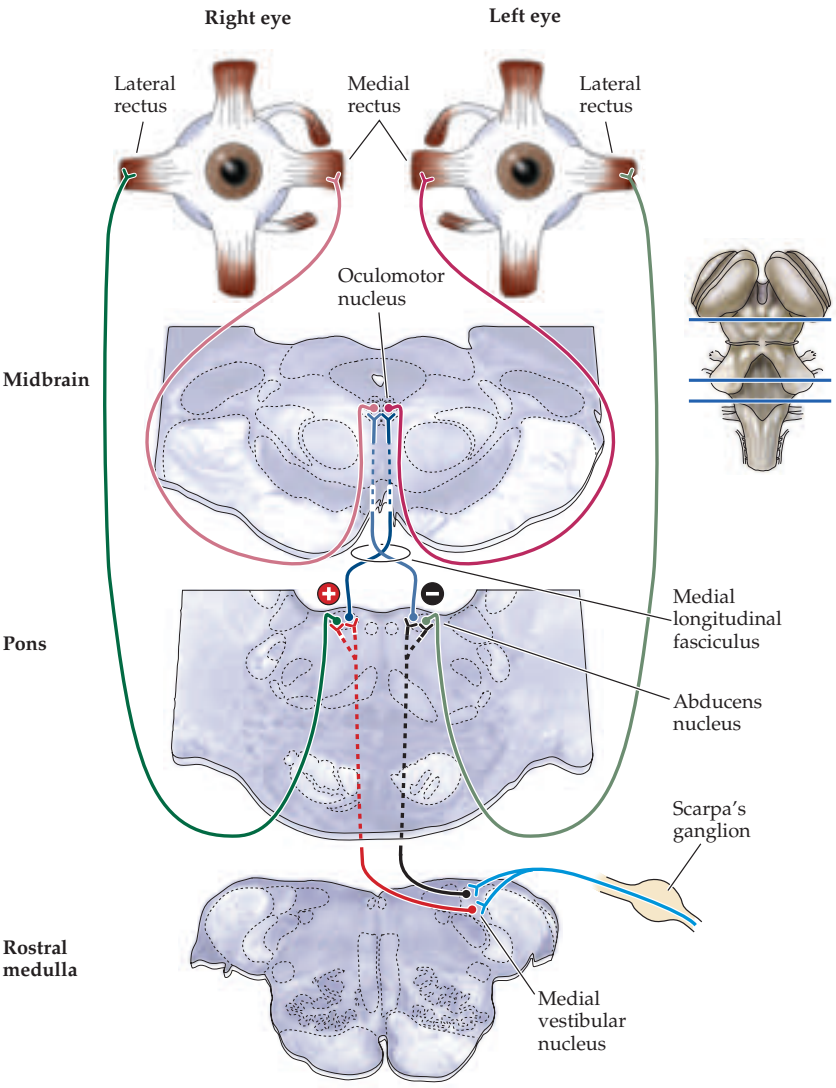


Figure 13.10 Connections underlying the vestibulo-ocular reflex. Projections of the vestibular nucleus to the nuclei of cranial nerves III (oculomotor) and VI (abducens). The connections to the oculomotor nucleus and to the contralateral abducens nucleus are excitatory (red), whereas the connections to ipsilateral abducens nucleus are inhibitory (black). There are connections from the oculomotor nucleus to the medial rectus of the left eye and from the adbucens nucleus to the lateral rectus of the right eye. This circuit moves the eyes to the right, that is, in the direction away from the left horizontal canal, when the head rotates to the left. Turning to the right, which causes increased activity in the right horizontal canal, has the opposite effect on eye movements. The projections from the right vestibular nucleus are omitted for clarity.

that causes the lateral rectus of the right eye to contract; the other is an excitatory projection that crosses the midline and ascends via the **medial longitudinal fasciculus** to the left oculomotor nucleus, where it activates neurons that cause the medial rectus of the left eye to contract. Finally, inhibitory neurons project from the medial vestibular nucleus to the left abducens nucleus, directly causing the motor drive on the lateral rectus of the left eye to decrease and also indirectly causing the right medial rectus to relax. The consequence of these several connections is that excitatory input from the horizontal canal on one side produces eye movements toward the opposite side. Therefore, turning the head to the left causes eye movements to the right.

In a similar fashion, head turns in other planes activate other semicircular canals, causing other appropriate compensatory eye movements. Thus, the VOR also plays an important role in vertical gaze stabilization in response to

the linear vertical head oscillations that accompany locomotion, and in response to vertical angular accelerations of the head, as can occur when riding on a swing. The rostrocaudal set of cranial nerve nuclei involved in the VOR (i.e., the vestibular, abducens, and oculomotor nuclei), as well as the VOR's persistence in the unconscious state, make this reflex especially useful for detecting brainstem damage in the comatose patient (see Box C).

Loss of the VOR can have severe consequences. A patient with vestibular damage finds it difficult or impossible to fixate on visual targets while the head is moving, a condition called **oscillopsia** ("bouncing vision"). If the damage is unilateral, the patient usually recovers the ability to fixate objects during head movements. However, a patient with bilateral loss of vestibular function has the persistent and disturbing sense that the world is moving when the head moves. The underlying problem in such cases is that information about head and body movements normally generated by the vestibular organs is not available to the oculomotor centers, so that compensatory eye movements cannot be made.

Descending projections from the vestibular nuclei are essential for postural adjustments of the head, mediated by the vestibulo-cervical reflex (VCR), and body, mediated by the vestibulo-spinal reflex (VSR). As with the VOR, these postural reflexes are extremely fast, in part due to the small number of synapses interposed between the vestibular organ and the relevant motor neurons (Box D). Like the VOR, the VCR and the VSR are both compromised in patients with bilateral vestibular damage. Such patients exhibit diminished head and postural stability, resulting in gait deviations; they also have difficulty balancing. These balance defects become more pronounced in low light or while walking on uneven surfaces, indicating that balance normally is the product of vestibular, visual, and proprioceptive inputs.

The anatomical substrate for the VCR involves the medial vestibular nucleus, axons from which descend in the medial longitudinal fasciculus to reach the upper cervical levels of the spinal cord (Figure 13.11). This pathway regulates head position by reflex activity of neck muscles in response to stimulation of the semicircular canals from rotational accelerations of the head. For example, during a downward pitch of the body (e.g., tripping), the superior canals are activated and the head muscles reflexively pull the head up. The dorsal flexion of the head initiates other reflexes, such as forelimb extension and hindlimb flexion, to stabilize the body and protect against a fall (see Chapter 16).

The VSR is mediated by a combination of pathways, including the lateral and medial vestibulospinal tracts and the reticulospinal tract. The inputs from the otolith organs project mainly to the lateral vestibular nucleus, which in turn sends axons in the lateral vestibulospinal tract to the spinal cord (see Figure 13.11). These axons terminate monosynaptically on extensor motor neurons, and they disynaptically inhibit flexor motor neurons; the net result is a powerful excitatory influence on the extensor (antigravity) muscles. When hair cells in the otolith organs are activated, signals reach the medial part of the ventral horn. By activating the ipsilateral pool of motor neurons innervating extensor muscles in the trunk and limbs, this pathway mediates balance and the maintenance of upright posture.

Decerebrate rigidity, characterized by rigid extension of the limbs, arises when the brainstem is transected above the level of the vestibular nucleus. Decerebrate rigidity in experimental animals is relieved when the vestibular nuclei are lesioned, underscoring the importance of the vestibular system to the maintenance of muscle tone. The tonic activation of extensor muscles in

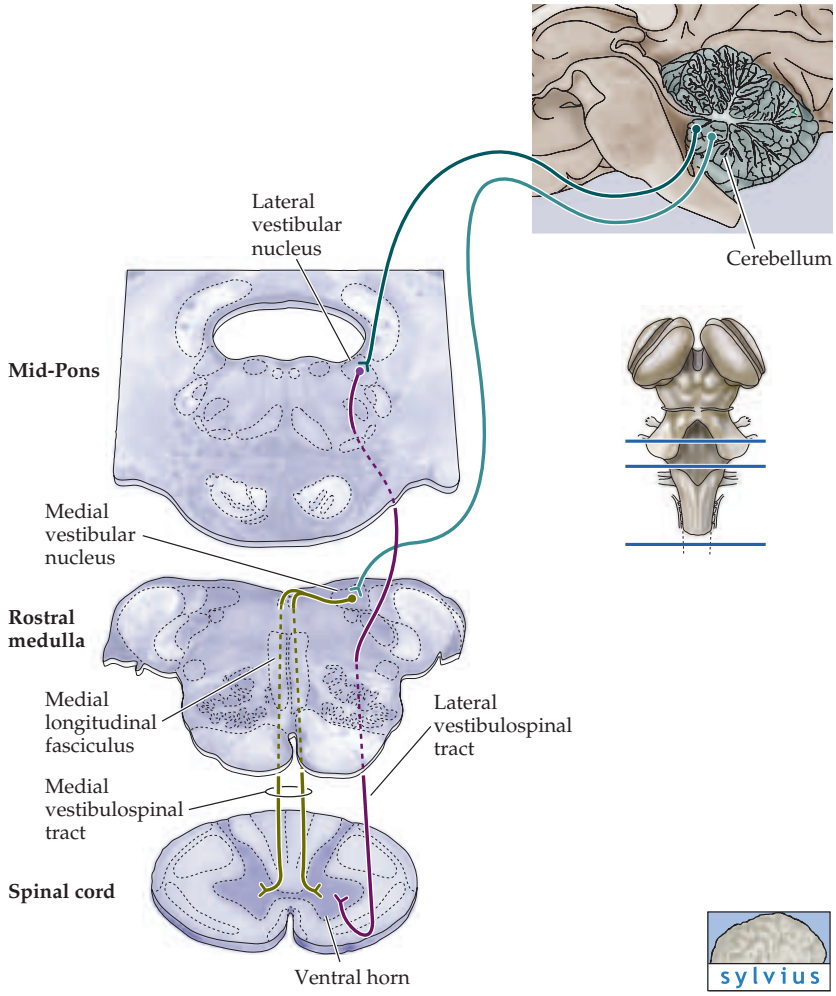


Figure 13.11 Descending projections from the medial and lateral vestibular nuclei to the spinal cord underlie the VCR and VSR. The medial vestibular nuclei project bilaterally in the medial longitudinal fasciculus to reach the medial part of the ventral horns and mediate head reflexes in response to activation of semicircular canals. The lateral vestibular nucleus sends axons via the lateral vestibular tract to contact anterior horn cells innervating the axial and proximal limb muscles. Neurons in the lateral vestibular nucleus receive input from the cerebellum, allowing the cerebellum to influence posture and equilibrium.

decerebrate rigidity suggests further that the vestibulospinal pathway is normally suppressed by descending projections from higher levels of the brain, especially the cerebral cortex (see also Chapter 16).

Vestibular Pathways to the Thalamus and Cortex

In addition to these several descending projections, the superior and lateral vestibular nuclei send axons to the ventral posterior nuclear complex of the thalamus, which in turn projects to two cortical areas relevant to vestibular

Box D
Mauthner Cells in Fish

A primary function of the vestibular system is to provide information about the direction and speed of ongoing movements, ultimately enabling rapid, coordinated reflexes to compensate for both self-induced and externally generated forces. One of the most impressive and speediest vestibular-mediated reflexes is the tail-flip escape behavior of fish (and larval amphibians), a stereotyped response that allows a potential prey to elude its predators (Figure A; tap on the side of a fish tank if you want to observe the reflex). In response to a perceived risk, fish flick their tail and are thus propelled laterally away from the approaching threat.

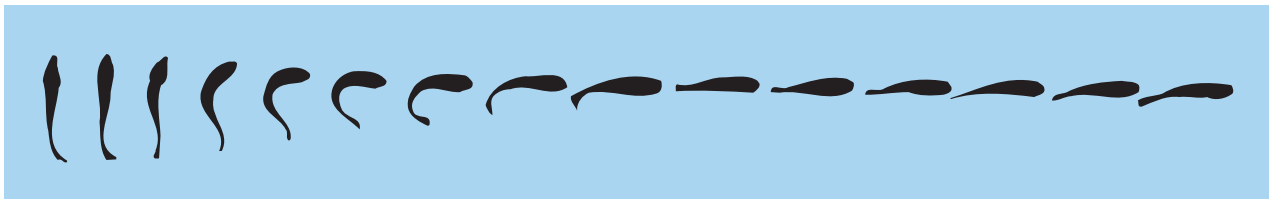
The circuitry underlying the tail-flip escape reflex includes a pair of giant medullary neurons called Mauthner cells, their vestibular inputs, and the spinal cord motor neurons to which the Mauthner cells project. (In most fish, there is one pair of Mauthner cells in a stereotypic location. Thus, these cells can be consistently visualized and studied from animal to animal.) Movements in the water, such as might be caused by an approaching predator, excite saccular hair cells in the vestibular labyrinth. These receptor potentials are transmitted via the central processes of vestibular ganglion cells in cranial nerve VIII to the two Mauthner cells in the brainstem. As in the vestibulo-spinal pathway in humans, the Mauthner cells project directly to spinal motor neurons. The small number of synapses intervening between the receptor cells and the motor neurons is one of the ways that this circuit has been optimized for speed by natural selection, an arrangement evident in humans as well. The large size of the Mauthner axons is another; the axons from these cells in a goldfish are about 50 µm in diameter.

The optimization for speed and direction in the escape reflex also is reflected in the synapses vestibular nerve afferents make on each Mauthner cell (Figure B). These connections are electrical synapses that allow rapid and faithful transmission of the vestibular signal.

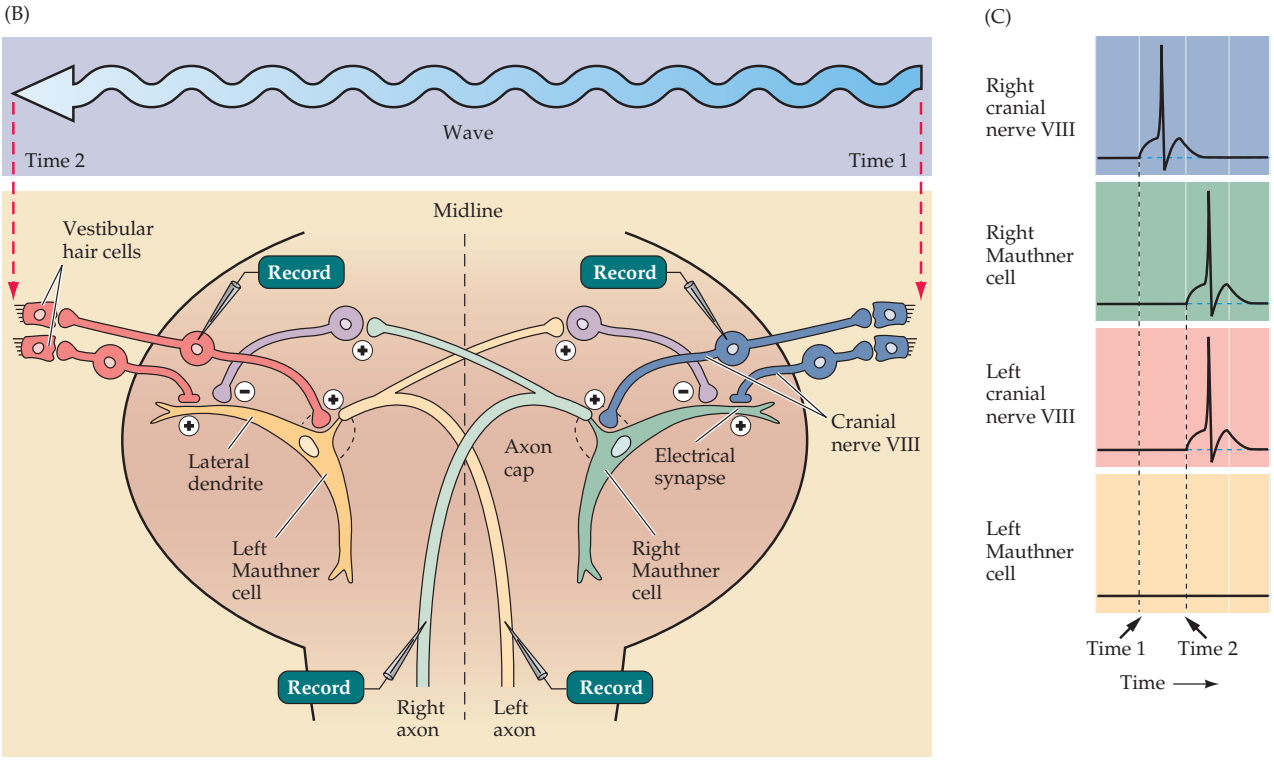
An appropriate direction for escape is promoted by two features: (1) each Mauthner cell projects only to contralateral motor neurons; and (2) a local network of bilaterally projecting interneurons inhibits activity in the Mauthner cell away from the side on which the vestibular activity originates. In this way, the Mauthner cell on one side faithfully generates action potentials that command contractions of contralateral tail musculature, thus moving the fish out of the path of the oncoming predator. Conversely, the Mauthner cell on the opposite side is silenced by the local inhibitory network during the response (Figure C).

(A) Bird's-eye view of the sequential body orientations of a fish engaging in a tail-flip escape behavior, with time progressing from left to right. This behavior is largely mediated by vestibular inputs to Mauthner cells.

(A)



sensations (Figure 13.12). One of these cortical targets is just posterior to the primary somatosensory cortex, near the representation of the face; the other is at the transition between the somatic sensory cortex and the motor cortex (Brodmann's area 3a; see Chapter 8). Electrophysiological studies of individual neurons in these areas show that the relevant cells respond to proprioceptive and visual stimuli as well as to vestibular stimuli. Many of these neurons are activated by moving visual stimuli as well as by rotation of the body (even with the eyes closed), suggesting that these cortical regions are involved in the perception of body orientation in extrapersonal space. Con-



(B) Diagram of synaptic events in the Mauthner cells of a fish in response to a disturbance in the water coming from the right. (C) Complementary responses of the right and left Mauthner cells mediating the escape response. Times 1 and 2 correspond to those indicated in Figure B. (After Furshpan and Furukuwa, 1962.)

The Mauthner cells in fish are analogous to the reticulospinal and vestibulospinal pathways that control balance, posture, and orienting movements in mammals. The equivalent behavioral responses in humans are evident in a friendly game of tag, or more serious endeavors.

References

EATON, R. C., R. A. BOMBARDIERI AND D. L. MEYER (1977) The Mauthner-initiated startle response in teleost fish. J. Exp. Biol. 66: 65–81.

FURSHPAN, E. J. AND T. FURUKAWA (1962) Intracellular and extracellular responses of the several regions of the Mauthner cell of the goldfish. J. Neurophysiol. 25:732–771.

JONTES, J. D., J. BUCHANAN AND S. J. SMITH (2000) Growth cone and dendrite dynamics in zebrafish embryos: Early events in synaptogenesis imaged *in vivo*. Nature Neurosci. 3: 231–237.

O'MALLEY, D. M., Y. H. KAO AND J. R. FETCHO (1996) Imaging the functional organization of zebrafish hindbrain segments during escape behaviors. Neuron 17: 1145–1155.

sistent with this interpretation, patients with lesions of the right parietal cortex suffer altered perception of personal and extra-personal space, as discussed in greater detail in Chapter 25.

Summary

The vestibular system provides information about the motion and position of the body in space. The sensory receptor cells of the vestibular system are located in the otolith organs and the semicircular canals of the inner ear. The

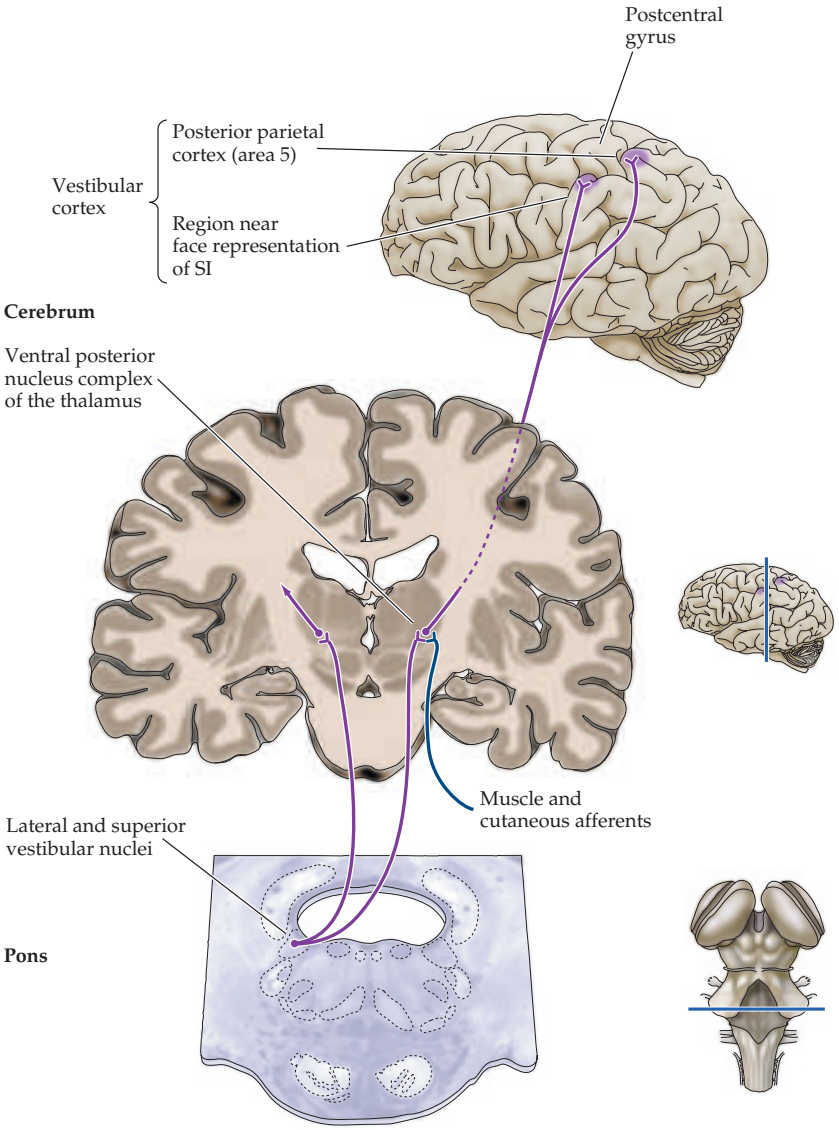


Figure 13.12 Thalamocortical pathways carrying vestibular information. The lateral and superior vestibular nuclei project to the thalamus. From the thalamus, the vestibular neurons project to the vicinity of the central sulcus near the face representation. Sensory inputs from the muscles and skin also converge on thalamic neurons receiving vestibular input (see Chapter 9).

otolith organs provide information necessary for postural adjustments of the somatic musculature, particularly the axial musculature, when the head tilts in various directions or undergoes linear accelerations. This information represents linear forces acting on the head that arise through static effects of gravity or from translational movements. The semicircular canals, in contrast, provide information about rotational accelerations of the head. This latter information generates reflex movements that adjust the eyes, head, and body during motor activities. Among the best studied of these reflexes are eye movements that compensate for head movements, thereby stabilizing the visual scene when the head moves. Input from all the vestibular organs is integrated with input from the visual and somatic sensory systems to provide perceptions of body position and orientation in space.

Additional Reading

Reviews

BENSON, A. (1982) The vestibular sensory system. In *The Senses*, H. B. Barlow and J. D. Mollon (eds.). New York: Cambridge University Press.

BRANDT, T. (1991) Man in motion: Historical and clinical aspects of vestibular function. A review. Brain 114: 2159–2174.

FURMAN, J. M. AND R. W. BALOH (1992) Otolith-ocular testing in human subjects. Ann. New York Acad. Sci. 656: 431–451.

GOLDBERG, J. M. (1991) The vestibular end organs: Morphological and physiological diversity of afferents. Curr. Opin. Neurobiol. 1: 229–235.

GOLDBERG, J. M. AND C. FERNANDEZ (1984) The vestibular system. In *Handbook of Physiology*, Section 1: *The Nervous System*, Volume III: *Sensory Processes*, Part II, J. M. Brookhart, V. B. Mountcastle, I. Darian-Smith and S. R. Geiger (eds.). Bethesda, MD: American Physiological Society.

HESS, B. J. (2001) Vestibular signals in self-orientation and eye movement control. News Physiolog. Sci. 16: 234–238.

RAPHAN, T. AND B. COHEN. (2002) The vestibulo-ocular reflex in three dimensions. Exp. Brain Res. 145: 1–27.

Important Original Papers

GOLDBERG, J. M. AND C. FERNANDEZ (1971) Physiology of peripheral neurons innervating semicircular canals of the squirrel monkey, Parts 1, 2, 3. J. Neurophysiol. 34: 635–684.

GOLDBERG, J. M. AND C. FERNANDEZ (1976) Physiology of peripheral neurons innervating otolith organs of the squirrel monkey, Parts 1, 2, 3. J. Neurophysiol. 39: 970–1008.

LINDEMAN, H. H. (1973) Anatomy of the otolith organs. Adv. Oto.-Rhino.-Laryng. 20: 405–433.

Books

BALOH, R. W. AND V. HONRUBIA (2001) *Clinical Neurophysiology of the Vestibular System*, 3rd Ed. New York: Oxford University Press.

BALOH, R. W. (1998) *Dizziness, Hearing Loss, and Tinnitus.* Philadelphia: F. A. Davis Company.