1

Processing of Information in the Human Visual System

Frank Schaeffel

Sektion für Neurobiologie des Auges, Forschungsinstitut für Augenheilkunde, Universitätsklinikum Tübingen, Calwerstrasse 7/1, 72076 Tübingen, Germany

1.1 Preface

To gather as much necessary information as possible of the visual world, and neglect as much unnecessary information as possible, the visual system has undergone an impressive optimization in the course of evolution, which is fascinating in each detail that is examined. A few aspects will be described in this chapter. Similar limitations may exist in machine vision, and comparisons to the solutions developed in the visual system in the course of 5 billion years of evolution might provide some insights.

1.2 Design and Structure of the Eye

As in any camera, the *first step in vision* is the projection of the visual scene on an array of photodetectors. In the vertebrate *camera eye*, this is achieved by the cornea and lens in the eye (Figure 1.1) which transmit the light in the visible part of the spectrum, 400–780 nm, by 60–70%. Another 20–30% is lost as a result of scattering in the ocular media. Only about 10% is finally absorbed by the photoreceptor pigment [1]. Because of the content of proteins, both cornea and the lens absorb in the ultraviolet, and because of the water content, the transmission is blocked in the far infrared. The cornea consists of a *thick* central layer – the stroma – which is sandwiched between two semipermeable membranes (total thickness 0.5 mm). It is composed of collagen fibrils, with mucopolysaccharides filling the space between the fibrils. Water content is tightly regulated to 75–80%, and clouding occurs if it changes beyond these limits. The crystalline lens is built up from proteins, called crystallines, which are characterized by their water solubility. The proteins in the periphery have high solubility, but those in the center are largely insoluble. The vertebrate lens is characterized by its continuous growth throughout life, with the older cells residing in the central core, the nucleus. With age, the lens becomes increasingly rigid and immobile and the ability to change its shape and focal length to accommodate for close viewing distances disappears – a disturbing experience for people around 45 who now need reading

Handbook of Machine and Computer Vision: The Guide for Developers and Users,
Second Edition. Edited by Alexander Hornberg.

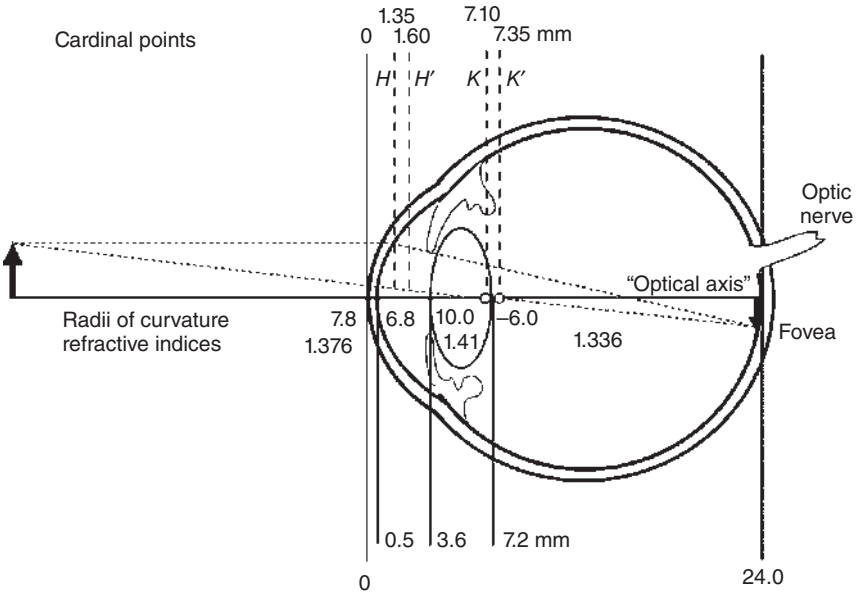


Figure 1.1 Dimensions and schematic optics of the left human eye, seen from above. The anterior corneal surface is traditionally set to coordinate zero. All positions are given in millimeters, relative to the anterior corneal surface (drawing not to scale). The refracting surfaces are approximated by spheres so that their radii of curvatures can be defined. The cardinal points of the optical system, shown on the top, are valid only for rays close to the optical axis (*Gaussian approximation*). The focal length of the eye in the vitreous (the posterior focal length) is 24.0 mm − *H* = 22.65 mm. The nodal points *K* and *K′* permit us to calculate the retinal image magnification. In the first approximation, the posterior nodal distance (*PND*, which is the distance *K* to the focal point at the retina) determines the linear distance on the retina for a given visual angle. In the human eye, this distance is about 24.0 mm − 7.3 mm = 16.7 mm. One degree in the visual field maps on the retina to 16.7 tan(1°) = 290 μm. Given that the foveal photoreceptors are 2 μm thick, 140 receptors are sampling 1° in the visual field, which leads to a maximum resolution of 70 cycles per degree. The schematic eye by Gullstrand represents an *average eye*. The variability in natural eyes is so large that it does not make sense to provide average numbers on dimensions with several digits. Refractive indices, however, are surprisingly similar among different eyes. The index of the lens (here homogenous model, *n* = 1.41) is not real but calculated to produce a lens power that makes the eye emmetropic. In a real eye, the lens has a gradient index (see text).

glasses (presbyopia). Accommodation is an active neuromuscular deformation of the crystalline lens that changes focal length from 53 mm to 32 mm in young adults.

Both media have higher refractive index than the water-like solutions in which they are embedded (tear film – on the corneal surface, aqueous – the liquid in the anterior chamber between the cornea and the lens, and vitreous humor – the gel-like material filling the vitreous chamber between the lens and the retina, Figure 1.1). Because of their almost spherical surfaces, the anterior cornea and both surfaces of the lens have positive refractive power with a combined optical focal length of 22.6 mm. This matches almost perfectly (with a tolerance 1/10 of a millimeter) the distance from the first principal plane (Figure 1.1, *H*) to the

photoreceptor layer in the retina. Accordingly, the projected image from a distant object is in focus when accommodation is relaxed. This optimal optical condition is called emmetropia but, in 30% of the population in industrialized countries, the eye has grown too long so that the image is in front of the retina even with accommodation completely relaxed (myopia).

The projected image is first analyzed by the retina in the back of the eye. For developmental reasons, the retina in all vertebrate eyes is *inverted*. This means that the photoreceptor cells, located at the backside of the retina, are pointing away from the incoming light. Therefore, the light has to pass through the retina (about a fifth of a millimeter thick) before it can be detected. To reduce scatter, the retina is highly translucent, and the nerve fibers that cross on the vitreal side, from where the light comes in, to the optic nerve head are not surrounded by myelin, a fat-containing sheet that normally insulates spiking axons (see below). Scattering in retinal tissue still seems to be a problem because, in the region of the highest spatial resolution, namely the fovea, the cells are pushed to the side. Accordingly, the fovea in the vertebrate eye can be recognized as a pit. However, many vertebrates do not have a fovea [2]; they have then lower visual acuity but their acuity can remain similar over large regions of the visual field, which is then usually either combined with high motion sensitivity (i.e., rabbit) or high light sensitivity at dusk (crepuscular mammals). It is striking that the retina in all vertebrates has a similar three-layered structure (Figure 1.9), with similar thickness. This makes it likely that the functional constraints were similar. The function of the retina will be described in the following.

The *optical axis* of the eye is not perfectly defined because the cornea and lens are not perfectly rotationally symmetrical and also are not centered on one axis. Nevertheless, even though one could imagine that the image quality is best close to the optical axis, it turns out that the human fovea is not centered in the globe (Figure 1.2). In fact, it is displaced to the temporal retina by an angle κ, ranging in different subjects from 0° to 11° but highly correlated in both eyes. Apparently, a few degrees away from the optical axis, the optical image quality is still good enough not to limit visual acuity in the fovea.

1.3 Optical Aberrations and Consequences for Visual Performance

One would imagine that the optical quality of the cornea and lens must limit the visual acuity since the biological material is mechanically not as stable and the surfaces are much more variable than in technical glass lenses. However, this is not true. In daylight, pupil sizes <2.5 mm constitute the optics of the human eye close to the diffraction limit (further improvement is physically not possible because of the wave properties of light). An eye is said to be diffraction-limited when the ratio of the area under its modulation transfer function (MTF; Figure 1.3) and the area under the diffraction-limited MTF (Strehl *ratio*) is higher than 0.8 (Marcos [3], Rayleigh *criterion*). With a 2 mm pupil, diffraction cuts off all spatial frequencies (SFs) higher than 62 cycles

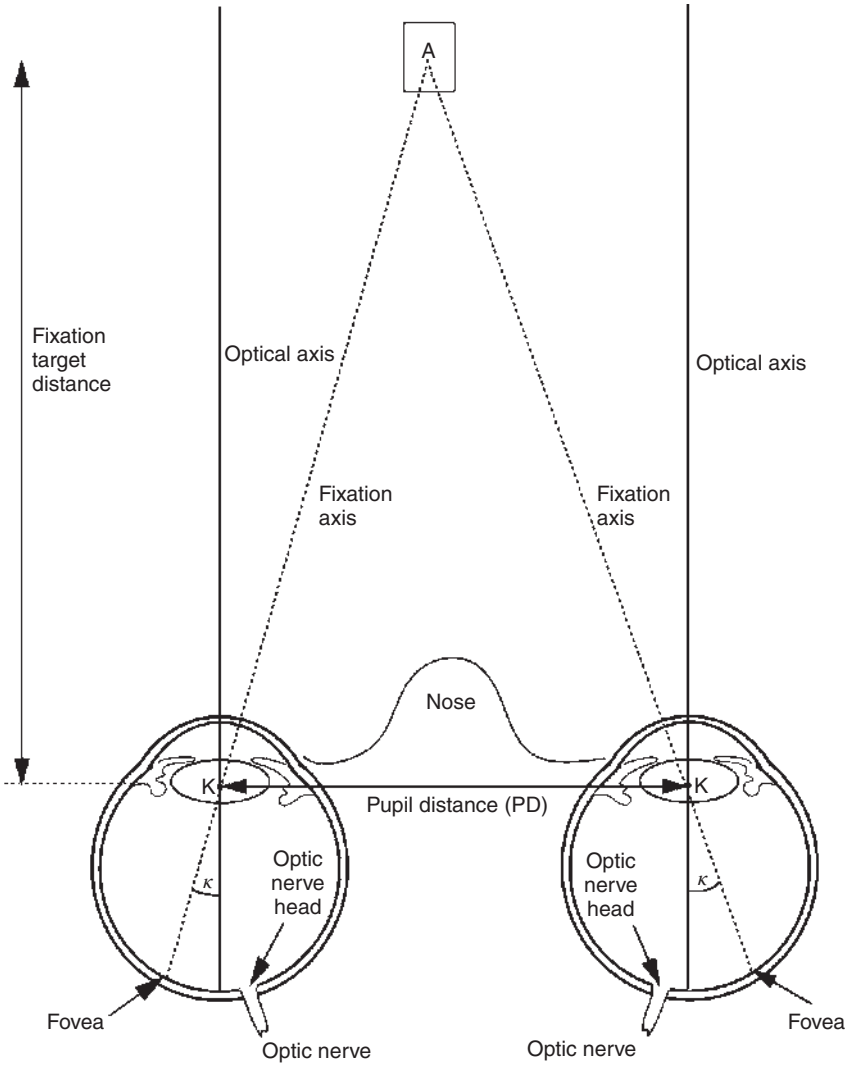


Figure 1.2 Binocular geometry of human eyes, seen from above. Since the fovea is temporally displaced with regard to the optical axis by the angle κ, the optical axes of the eyes do not reflect the direction of fixation. κ is highly variable among eyes, ranging from 0° to 11°, with an average of 3.5°. In the illustrated case, the fixation target is at a distance for which the optical axes happen to be parallel and straight. The distance of the fixation target for which this is true can be easily calculated: for an angle κ of 4°, and a pupil distance of 64 mm, this condition would be met if the fixation target is at 32 mm/tan(4°), or 457.6 mm. The optic nerve head (also called the *optic disk*, or *blind spot*, the position at which the axons of the retinal ganglion cells leave the eye) is nasally displaced relative to the optical axis. The respective angle is in a similar range as κ. Under natural viewing conditions, the fixation angles must be extremely precise since double vision will be experienced if the fixation lines do not exactly cross on the fixation target – the tolerance is only a few minutes of arc).

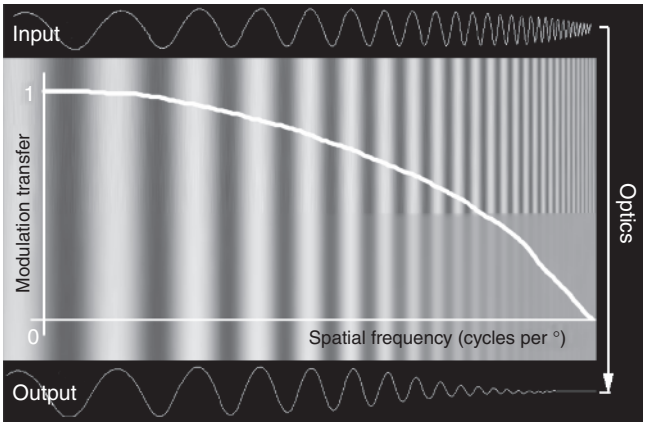


Figure 1.3 Spatial information in an image can be reconstructed as a linear superposition of sine wave components (spatial frequencies) with different amplitudes and phases (Fourier components). Low spatial frequencies (SFs) are generally available with high contrast in the natural visual environment, whereas the contrast declines for higher SFs, generally proportional to 1/SF (input). Because of optical imperfections and diffraction, the image on the retina does not retain the input contrast at high SFs. The decline of *modulation transfer*, the ratio of the output to input contrast, is described by the modulation transfer function (MTF, thick white line). At around 60 cycles per degree, the optical modulation transfer of the human eye reaches zero, with small pupil sizes due to diffraction and with larger pupils due to optical imperfections. These factors limit our contrast sensitivity at high spatial frequencies, even though the retina extracts surprisingly much information from the low-contrast images of high SFs, by building small receptive fields for foveal ganglion cells with antagonistic ON/OFF center/surround organization.

per degree – a limit that is very close to the maximum behavioral resolution achieved by human subjects. By the way, diffraction-limited optics is achieved only in some birds and primates, although it has been recently claimed that also an alert cat has diffraction-limited optics [4]. A number of tricks are used to reduce the aberrations that are inherent in spherical surfaces: the corneal surface is, in fact, clearly aspheric, flattening out to the periphery, and the vertebrate lens is always a *gradient index structure*, with the refractive index continuously increasing from the periphery to the center. Therefore, peripheral rays are bent less than central rays, which compensates for the steeper angles that the rays encounter if they hit a spherical surface in the periphery. The gradient index of the lens reduces its spherical aberration from more than 12 diopters (for an assumed homogenous lens) to less than 1 diopter (for a gradient index lens). Furthermore, the optical aberrations seem to be under tight control (although it remains uncertain whether this control is visual [5]). The remaining aberrations of the cornea and the lens tend to cancel each other, and this is true, at least, for astigmatism, horizontal coma, and spherical aberration [6, 7]. However, aberrations have also advantages: they increase the depth of field by 0.3 D, apparently without reducing visual acuity; there is no strong correlation between the amount of aberrations of subjects and their letter acuity. They also reduce the required precision of accommodation, in particular, during reading [8]. It is questionable whether optical correction of higher order aberrations by refractive surgery or individually designed spectacle lenses would further improve acuity

Figure 1.4 Spurious resolution. The modulation transfer function (Figure 1.3) shows oscillations beyond the cutoff spatial frequency, which show up in defocused gratings as contrast reversals. On the left, a circular grating shows the contrast reversals at the higher spatial frequencies in the center (top: in focus, below: defocused). On the right, the grating shown in Figure 1.3 was defocused. Note the lack of contrast at the first transition to zero contrast, and the repeated subsequent contrast reversals. Note also that defocusing has little effect on low spatial frequencies.

for high-contrast letters in young subjects, creating an *eagle's eye*. It is, however, clear that an extended MTF can enhance the contrast sensitivity at high SFs. Correcting aberrations might also be useful in older subjects, since it is known that monochromatic aberrations increase by a factor of 2 with age [9]. Aberrations may also be useful for other reasons; they could provide directionality cues for accommodation [10] and, perhaps, for emmetropization.

In a healthy emmetropic young eye, the optical MTF (Figure 1.3) appears to be adapted to the sampling interval of the photoreceptors.

Contrast modulation reaches zero at SFs of around 60 cycles per degree, and the foveal photoreceptor sampling interval is in the range of 2 μm. Since 1° in the visual field is mapped on a 0.29 mm linear distance on the retina, the highest detectable SF could be about 290/4 μm or about 70 cycles per degree. The MTF shows that the contrast of these high spatial frequencies in the retinal image approaches zero (Figure 1.3). With defocus, the MTF drops rapidly. Interestingly, it does not stop at zero modulation transfer, but rather continues to oscillate (although with low amplitude) around the abscissa. This gives rise to the so-called *spurious resolution*; defocused gratings can still be detected beyond the cutoff SF, although in part with reversed contrast (Figure 1.4).

The sampling interval of the photoreceptors increases rapidly over the first few degrees away from the fovea, and visual acuity declines (Figure 1.5), both because rods are added to the lattice, which increases the distances between individual cones and because their cone diameters increase. In addition, many cones converge on one ganglion cell. Furthermore, only the foveal cones have *private lines* to a single ganglion cell (Figure 1.9).

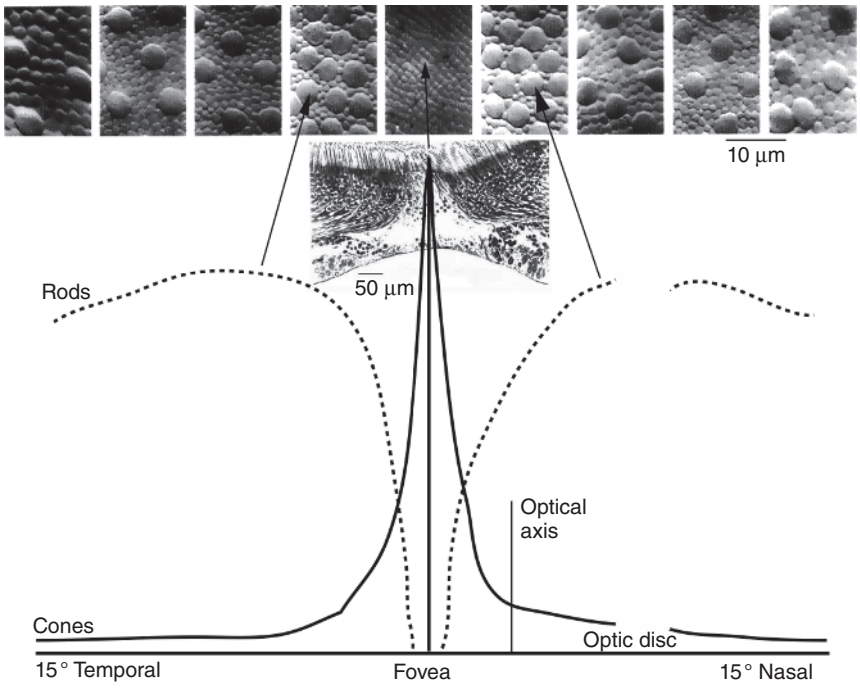


Figure 1.5 Regional specializations of the retina. The fovea is free from rods, and L and M cones are packed as tightly as possible (reaching a density of 200 000 mm^{-2} – histology on top replotted after [11]). In the fovea, the retinal layers are pushed to the side to reduce scattering of light that has reached the photoreceptors – resulting in the foveal pit. Rods reach a peak density of 130 000 mm^{-2} at 3° away from the fovea. Accordingly, a faint star can be seen only if it is not fixated. As a result of the drop in cone densities and due to increasing convergence of cone signals, visual acuity drops even faster: at 10°, visual acuity is only about 20% of the foveal peak. Angular positions relative to the fovea vary between individuals and are therefore approximate. (Adapted from Curcio *et al*. 1990 [11].)

Because the optical quality does not decline as fast in the periphery as the spatial resolution of the neural network, the retinal image is undersampled. If the receptor mosaic were regular, like in the fovea, stripes that are narrower than the resolution limit would cause spatial illusions (*Moiré* patterns, Figure 1.6). Since the receptor mosaic is not so regular in the peripheral retina, this causes just spatial noise. Moiré patterns are, however, visible in the fovea if a grating is imaged, which is beyond the resolution limit. This can be done by presenting two laser beams in the pupil, which show interference [12].

Moiré patterns are explained from the Shannon's sampling theorem, which states that regularly spaced samples can be resolved only when the sampling rate is equal to or higher than twice the highest spatial frequency; that is, to resolve the samples, between each receptor that is stimulated, there must be one that is not stimulated. The highest SF that can be resolved by the photoreceptor mosaic (*Nyquist limit*) is half the sampling frequency. In the fovea, the highest possible spatial sampling is achieved. Higher photoreceptor densities are not possible for the following reason: Because the inner segments of the photoreceptors have

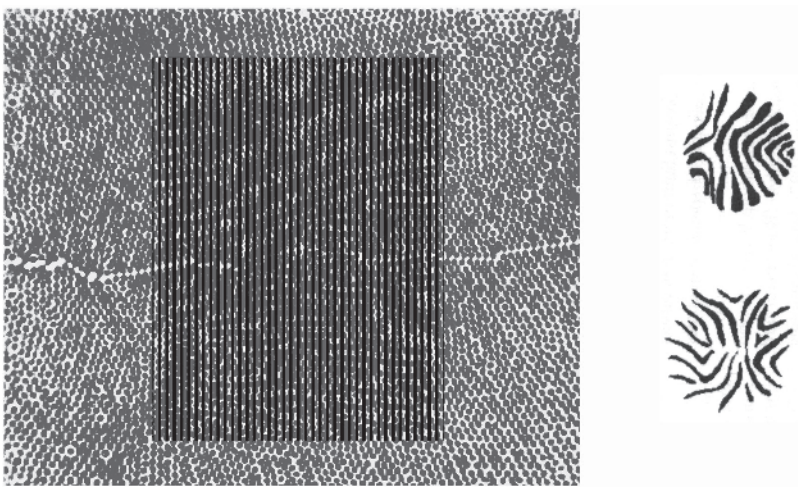
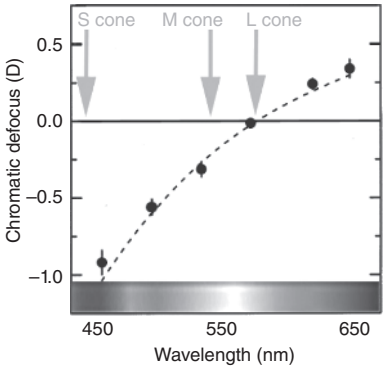


Figure 1.6 Aliasing (undersampling) and Moiré patterns. If a grating is imaged on the photoreceptor array, and the sampling interval of the receptors is larger than half the spatial wavelength of the grating, patterns appear. The photoreceptor lattice (left) is from a histological section of a monkey's retina. If laser interferometry is used to image fine gratings with spatial frequency beyond the resolution limit on the fovea of human subjects, the subjects see Moiré patterns, which are drawn on the right. (Adapted from Williams 1985 [12]. Reproduced with permission of Elsevier.)

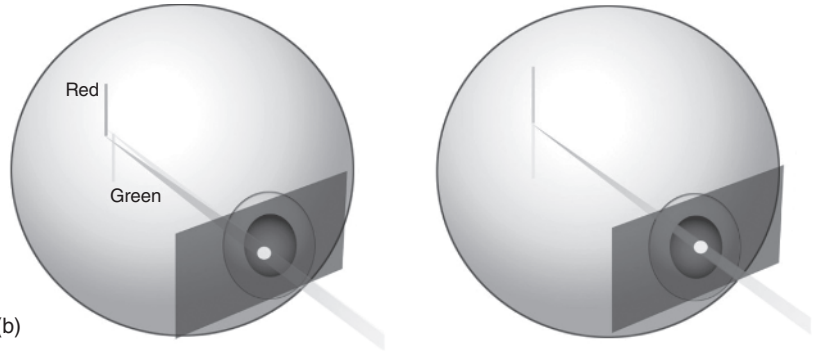
higher refractive indices than their surroundings, they act as light guides. But if they become very thin, they show properties of waveguides. When their diameter approaches the wavelength of light, energy starts to *leak out* [13], causing increased optical crosstalk with neighboring photoreceptors. As it is, about 5% of the energy is lost, which seems acceptable. But if the thickness (and the sampling interval) is further reduced to 1 μm, already 50% is lost.

Since the photoreceptor sampling interval cannot be decreased, the only way to increase visual acuity is to enlarge the globe and, accordingly, the PND and the retinal image. This solution was adopted in the eagle's eye, with an axial length of 36 mm, which is the highest spatial acuity in the animal kingdom (grating acuity 135 cycles per degree [14]).

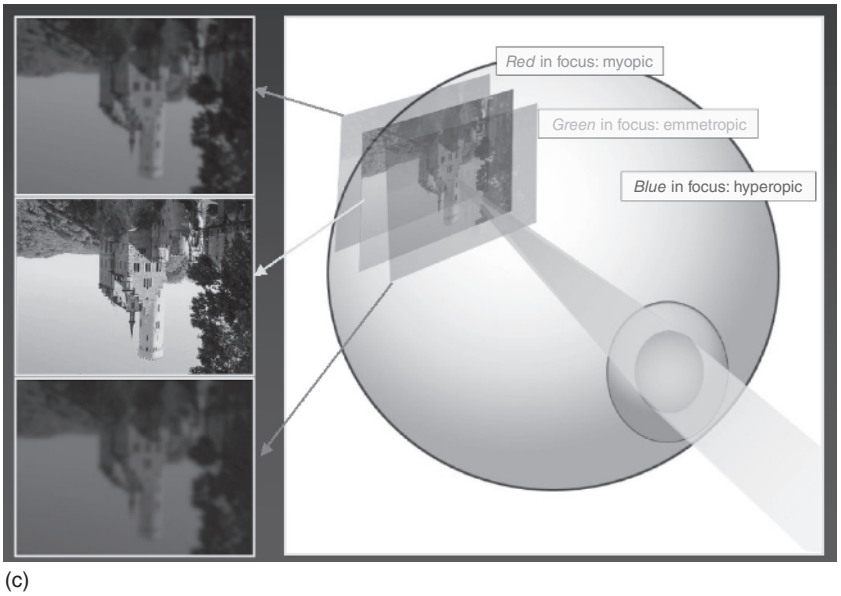
Figure 1.7 Chromatic aberration and some of its effects on vision. Because of the increase in the refractive indices of the ocular media with decreasing wavelength, the eyes become more myopic in the blue. (a) The chromatic aberration function shows that the chromatic defocus between L and M cones is quite small (about a quarter of a diopter) but close to 1 D for the S cone. (b) Because of transverse chromatic aberration, rays of different wavelengths that enter the pupil reach the retina normally not in the same position. If a red line and green line are imaged on the retina through selected parts of the pupil, and the subject can align them via a joystick, the *achromatic axis* can be determined. Light of different wavelengths entering the eye along the achromatic axis is imaged at the same retinal position (although with a wavelength-dependent focus). (c) Because of longitudinal chromatic aberration, light of different wavelengths is focused in different planes. Accordingly, myopic subjects (with too long eyes) see best in the red and hyperopic subjects (with too short eyes) best in the blue.


S cone
M cone
L cone
0.5
0.0
−0.5
−1.0
Chromatic defocus (D)
450
550
650
Wavelength (nm)

(a)


Red
Green

(b)


Red in focus: myopic
Green in focus: emmetropic
Blue in focus: hyperopic

(c)

1.4 Chromatic Aberration

In addition to monochromatic aberrations (those aberrations that persist in monochromatic light), there is also chromatic aberration that results from dispersion of the optical media, that is, due to the wavelength-dependent refractive index. In technical optical systems, lenses with different refractive indices are combined in such a way that the focal length does not vary much across the visible spectrum. In natural eyes, no attempt is made to optically balance chromatic aberration. Neural image processing makes it possible for us to be unaware of chromatic image degradation under normal viewing conditions, and, in addition, there are morphological adaptations in the retina. Inspection of the chromatic aberration function in the human eye (Marcos *et al*. [15]; Figure 1.7a) shows that a large dioptric change occurs in the blue end of the spectrum (about 1D from 570 to 450 nm), whereas the change is smaller in the red end (about 0.5 D from 570 to 680 nm). We have three cone photopigments, absorbing at *long* wavelengths (*L cones*, peak absorption typically at 565 nm), at *middle* wavelengths (*M cones*, typically at 535 nm), or at *short* wavelengths (*S cones*, typically at 440 nm). The dioptric difference between L and M cones is small (S0.2D) but the dioptric differences to the S cones are significant (>1D). It is, therefore, impossible to see sharply with all three cone types at the same time (Figure 1.7c). A *white* point of light is, therefore, imaged in the fovea as a circle with a diameter of up to 100 cone diameters, with a 6 mm pupil. Perhaps, as a consequence, the S cone system was removed from the highacuity region of the central 0.5° of the fovea (the foveola); one cannot focus them anyway and they would only occupy space that is better used to pack the M and L cones more densely, to achieve best sampling. High-acuity tasks are then performed only with the combined L and M cones. The *blue information* is continuously *filled in* because small eye movements enable the stimulation of the parafoveal S cones. Therefore, this scotoma is normally not visible, similar to the blind spot, where the optic nerve leaves the eye – surprising, given that the blind spot has 5× the diameter of the fovea. Nevertheless, a small blue spot viewed on a yellow background from a distance appears black, and a blue field of about 440 nm that is sinusoidally modulated at a few hertz makes the blue scotoma in the fovea visible as a star-shaped black pattern [16]. Also in the periphery, the S cones sample more coarsely than the L and M cones and reach a maximum spatial resolution of 5–6 cycles per degree in the parafoveal region at about 2° away from the fovea. The dispersion of the ocular media produces not only different focal planes for each wavelength (*longitudinal chromatic aberration*) but also different image magnifications (*transverse chromatic aberration*). Accordingly, a point on an object's surface is imaged in the different focal planes along a line only in the *achromatic axis* of the eye (which can be psychophysically determined; Figure 1.7b). Even a few degrees away from the achromatic axis, blue light emerging from an object point will be focused closer to achromatic axis than red light. In particular, since the fovea is usually neither in the optical axis nor in the achromatic axis (see above), the images for red and blue are also laterally displaced with respect to each other. With a difference in image magnification of 3%, a κ of 3.5°, and a linear image magnification of 290 μm

per degree, the linear displacement would be 3.5 μm × 290 μm × 0.03 μm or 30 μm, which is about the distance from one S cone to the next. Human subjects are not aware of this difference in magnification, and the rescaling of the *blue versus the red image* by neural processing seems to occur without effort.

1.5 Neural Adaptation to Monochromatic Aberrations

The neural image processor in the retina and cortex can relatively easily adapt to the changes of aberrations and field distortions. This can be seen in spectacle wearers. Even though spectacles, in particular progressive addition lenses, cause complex field distortions and additional aberrations, such as astigmatism and coma in the periphery, the wearer is usually not aware of these optical problems, already after a few days. The necessary neural image transformations are impressive. Nevertheless, it is realized that the individual visual system is best trained to the natural aberrations of the eye, even though it can learn to achieve a similar acuity if the same aberrations are experimentally rotated [17]. One of the underlying mechanisms is *contrast adaptation*. If all visible spatial frequencies are imaged on the retina with similar contrast (Figure 1.3), it can be seen that the contrast sensitivity varies with spatial frequency. The highest sensitivity is achieved at 5 cycles per second. However, the contrast sensitivity at each spatial frequency is continuously readjusted according to how much contrast is available at a given spatial frequency. If the contrast is low, the sensitivity increases, and vice versa. This way, the maximum information can be transmitted with a limited total channel capacity. Contrast adaptation also accounts for the striking observation that a defocused image (lacking high contrast at higher spatial frequencies) appears sharper after a while (Webster *et al*. [18]; see Movie, Nature Neuroscience[1]). If myopic subjects take their glasses off, they initially experience very poor visual acuity, but some improvement occurs over the first few minutes without glasses. These changes are not optical but neuronal. Contrast adaptation occurs both in the retina and the cortex [19].

1.6 Optimizing Retinal Processing with Limited Cell Numbers, Space, and Energy

A striking observation is that only the visual system has extensive peripheral neural preprocessing, which starts already at the photoreceptors. Although the retina is a part of the brain, it is not immediately obvious why the nerves from the photoreceptors do not project directly to the central nervous system, as in other sensory organs. The reason is probably that the amount of information provided by the photoreceptors is just too much to be transmitted through the optic nerve (about 1 million fibers) without previous filtering [20].

There are about 125 million photoreceptors, and their information supply converges into 1 million ganglion cells, which send their axons through the optic

1 http://www.nature.com/neuro/journal/v5/n9/suppinfo/nn906_S1.html.

nerve to the brain (Figure 1.11). In the optic nerve, the visual information is more compressed than ever before, or after. It follows then why the optic nerve cannot be made any thicker, with more fibers, so that extensive information compression would be unnecessary. The reason appears to be that the visual cortex can process high-spatial-acuity information only from a small part of the visual field. As it is, the foveal region occupies already 50% of the cortical area, and processing the whole visual field of 180° would require a cortex perhaps *as large as a classroom*. On the other hand, confining high acuity to a small part of the visual field requires extensive scanning through eye (or head) movements. A thicker optic nerve would impair such eye movements and it would also increase the size of the blind spot. Eye movements occupy considerable processing capacity in the brain since they are extremely fast and precisely programmed (e.g., the tolerance for errors in the angular position of the eye with binocular foveal fixation is only a few minutes of arc – otherwise, one would see double images). But this seems to require still less capacity than required for extending the foveal area.

1.7 Adaptation to Different Light Levels

The first challenge that the retina has to deal with is the extreme range of ambient illuminances. Between a cloudy night and a sunny day at the beach, illuminance varies by a factor of about 8 log units. Without adaptation, a receptor (and also a charge-coupled device (CCD) photodetector or a film) can respond to 1.5 or 2 log units of brightness differences. This is usually sufficient because natural contrasts in a visual scene are rarely higher. But if the receptors are not able to shift and flatten their response curves during light adaptation (Figure 1.8), they would saturate very soon if the ambient illuminance increases, and the contrast of the image would decline to zero. So, the major role of adaptation is to prevent saturation.

It is clear that light/dark adaptation has to occur either before reception or in the photoreceptors. It is not possible to generate an image with spatial contrast if the photoreceptors are saturated. In the visual system, some adjustment occurs through the size of the pupil, which can vary from 2 to 8 mm in young subjects (a factor of 16, or little more than 1 log unit). It is clear that the remaining 8 log units have to be covered. In vertebrates, this is done by dividing the illuminance range into two parts, the scotopic part, where rod photoreceptors determine our vision, and the photopic part, where the cones take over. There is a range in between where both rods and cones respond, and this is called the mesopic range. Both rods and cones can shift their response curves from higher to lower sensitivity. This is done by changing the gain of the biochemical phototransduction cascade in the photoreceptor cells, which converts the energy of a photon of light that is caught by the photopigment into an electrical signal at the photoreceptor membrane. Adaptation occurs by changes in the intracellular calcium concentration, which, in turn, affects the gain of at least three steps in the cascade (Figure 1.8). Strikingly, photoreceptors hyperpolarize in response to a light excitation and basically show an inverted response, compared to other neurons.

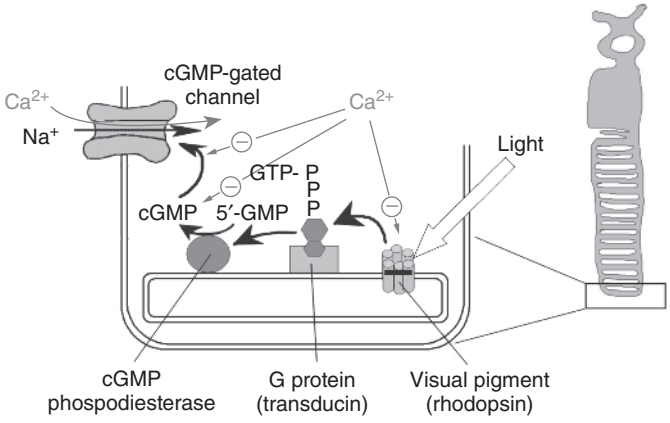


Figure 1.8 Principle of the phototransduction cascade and the role of calcium in light/dark adaptation in a rod photoreceptor. The pigment molecule embedded in the outer segment disk membrane of the photoreceptor, consisting of a protein (opsin) and retinal (an aldehyde), absorbs a photon and converts it into an activated state which can stimulate a G protein (transducin in rods). Transducin, in turn, activates an enzyme, cGMP phosphodiesterase, which catalyzes the breakdown of cGMP to 5′-GMP. cGMP has a key role in phototransduction. To open the *cGMP-gated cation channels*, three cGMP molecules have to bind to the channel protein. Therefore, if cGMP is removed by cGMP phosphodiesterase, the channels cannot be kept open. The Na^+ influx stops (which depolarizes the cell against its normal resting potential), and the membrane potential moves to its resting potential; this means hyperpolarization. When the channels are closed during illumination, the intracellular calcium levels decline. This removes the inhibitory effects of calcium on (i) the cGMP binding on the channel, (ii) the resynthesis pathway of cGMP, and (iii) the resynthesis of rhodopsin. All three steps reduce the gain of the phototransduction cascade (*light adaptation*). It should be noted that complete dark or light adaptation is slow: it takes up to 1 h.

This inverted response is energetically costly, since it means that *dark* represents the adequate stimulus, with a constant high release of the transmitter glutamate from the presynaptic terminals, with a high rate of resynthesis. Glutamate release is controlled by intracellular calcium, with high release at high calcium levels, and vice versa. If the light is switched on, the cation channels in the outer segment cell membrane close, which causes a constant influx of positive ions into the outer segment, and thereby effects its depolarization. Why the photoreceptors respond to light, the adequate stimulus, in an inverted manner, has not been convincingly explained.

The inverted response makes it necessary to reverse the voltage changes at a number of synapses in the retina. Normally, the signal for excitation of a spiking neuron is depolarization, not hyperpolarization. In fact, already at the first synapse in the retina, the photoreceptor terminals, the transmitter glutamate can either induce depolarization of the postsynaptic membrane (OFF bipolar cells) or hyperpolarization (ON bipolar cells), depending on the type of receptors that bind glutamate (OFF: ionotropic AMPA/Kainate receptor; ON: metabotropic mGluR6 receptor).

1.8 Rod and Cone Responses

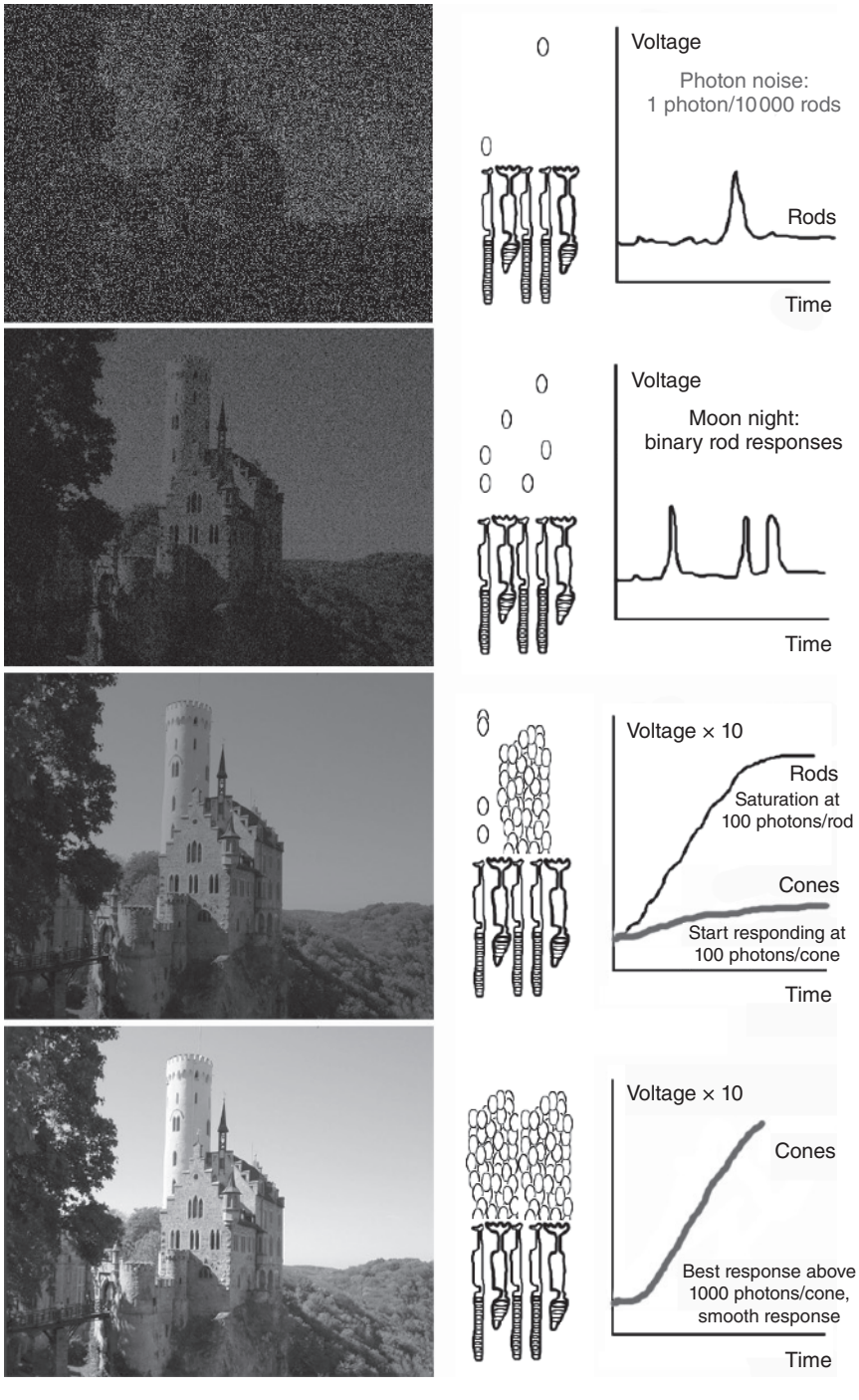
Rod photoreceptors can reach the maximum possible sensitivity: they show a significant transient hyperpolarization in response to the absorption of a single photon. Photons reach the retina in a star light like rain hits a paved road. Not every rod receives a photon and not every paved stone is hit by a rain drop. Accordingly, the image is *noisy* (Figure 1.9, top).

The probability of catching a certain number of photons during the integration time is described by Poisson statistics, and the standard deviation is the square root of the number of photons caught per unit time. Therefore, the signal-to-noise ratio can be calculated as the square root of the number of photons divided by the number of photons. If 100 photons are absorbed during integration time of the rod (about 200 ms), only changes in contrast that are larger than 10% can be detected. Low-contrast detection requires many photons: to distinguish contrasts of 1%, at least 10 000 photons must be absorbed during the integration time [20].

Another limiting factor is thermal noise . The rhodopsin molecule has an average lifetime of 300 years at 37 °C, but when it decays, rod photoreceptors cannot distinguish between a photon absorption and spontaneous decay. Because of the abundance of rhodopsin in the photoreceptors, 10^6 decays occur spontaneously each second. Thermal noise matches photon noise in a clear star night. If rhodopsin decays only 1 log unit faster, our threshold sensitivity would rise by 3 log units.

With increasing ambient illuminance, rods continue to give binary responses (either hyperpolarization or not) over the first 3 log units [20]. If brightness further increases, in the mesopic range, the degree of hyperpolarization of the rod membrane increases linearly with the number of photons caught during the integration time, up to about 20 photons. At about 100 photons, the rods saturate, although adaptation can reduce their sensitivity so that a graded response is possible up to 1000 photons. Cones take over at 100 photons per integration time (here only a few milliseconds), and only at this number their

Figure 1.9 Photon responses of rods and cones. From complete darkness to a moon-lit night, rods respond to single photons: their signals are binary (either *yes* or *no*). Because not every rod can catch a photon (here illustrated as small white ellipses), and because photons come in randomly as predicted by a Poisson distribution, the image appears noisy and has low spatial resolution and little contrast. Even if it is 1000 times brighter (a bright moon-lit night), rods do not catch several photons during their integration time of 100–200 ms and they cannot summate responses. Until up to 100 photons per integration time, they show linear summation, but their response curve is still corrupted by single photon events. Beyond 100 photons per integration time, rods show light adaptation (see Figure 1.8). At 1000 photons/integration time, they are saturated and silent. Cones take over, and they work best above 1000 photons per integration time. Because cone gather their signal from so many photons, photon noise is not important and their response to brightness changes is smooth and gradual. If the number of photon rises further, the sensitivity of the cone phototransduction cascade is reduced by light adaptation, and their response curve is shifted to higher light levels. Similar to rods, they can respond over a range of about 4 log units of ambient illuminance change.


Voltage
Photon noise:
1 photon/10 000 rods
Rods
Time
Voltage
Moon night:
binary rod responses
Time
Voltage × 10
Rods
Saturation at
100 photons/rod
Cones
Start responding at
100 photons/cone
Time
Voltage × 10
Cones
Best response above
1000 photons/cone,
smooth response
Time

response rises above the dark noise. They work best at 1000 photons or more. In summary, rods' responses are always corrupted by photon noise, whereas cones respond in a smooth, graded manner, and their responses contain more bits (Figure 1.9, bottom). Generally, the response function of sensory organs is logarithmic, which means that for a weak stimulus a small change in stimulus strength is detectable, whereas for a strong stimulus the change must be larger to be detected (Weber's law: detectable stimulus strength difference proportional to stimulus strength).

Since rods are hunting *each photon*, they occupy, in most mammals, as much territory of the retinal area as possible (>95%), whereas cones, which are not limited by photon noise, are densely packed only in the fovea where they permit high spatial sampling. Rods were left out because they would increase the foveal sampling intervals of the M and L cones.

To match the *information channel capacity* of rods and cones to the available information, the rod axons are thin and the terminal synapses small, with only about 80 transmitter vesicles released per second, and with only one region of transmitter release (*active zone*), whereas the cone axon is thick, with up to 1500 vesicles released per second, and a synapse (*cone pedicle*) which is perhaps the most complicated synapse in the whole nervous system. It makes contacts with several hundred other retinal neurons (horizontal and bipolar cells) and has a typical appearance (the *cone pedicle*). The signal is diverged into 10 parallel channels [20], 5 ON- and 5 OFF-type bipolar cells. Rather than using one broadband *super channel*, 10 parallel channels are used, each with a different bandwidth. The separation into different channels is thought to occur because (i) different aspects of visual processing can be separated into different pathways already at an early level and (ii) such a *broadband superchannel* cannot be made because of the limited range of possible spike frequencies (see the next section).

1.9 Spiking and Coding

In electronic devices, electrical signals can be transmitted either in an analog or a binary manner. In fact, both types are also realized in the nervous system. Short-distance signal transmission, like through the dendrites to the cell body of neurons, occurs via an electronically propagating depolarization of the membrane. For long-distance travel, the electronic signals become too degraded and lose their reliability. Therefore, binary coding (*0 or 1*), as used in action potentials, is used, which is much more resistant to degradation. At the root of the neuron's axon, and along the axon, voltage-dependent sodium channels are expressed, which are necessary for the generation of action potentials (*spikes*). Action potentials are rapid and transient depolarizations of the membrane, which travel with high speed (up to 120 m s^{-1}) along the fibers. The level of excitation is now encoded in the frequency of the spikes. Because of a recovery phase after each spike of about 2 ms, the maximum frequency is limited to about 500 Hz. This means that the dynamic range is limited. Because, even with a constant stimulus, the spike frequency displays some noise, the response functions of the neurons have no *steps*. It turns out that noisy signals are a common feature in the nervous system, and high precision is achieved by parallel

channels, if necessary (probability summation), or by temporal summation. The signal-to-noise ratio, when a difference in stimulus strength should be detected, is determined by the standard deviation of the firing rate. If the firing rate varies little with constant stimulation (small standard deviation), a small change would be detectable, but if it varies much (large standard deviation), only large differences are detected. The signal-to-noise ratio can be calculated from the differences in spike rates at both stimulus strengths divided by the standard deviation of spike rates (assuming that the standard deviation is similar in both cases). Summation of several channels is not always possible: sometimes decisions need to be made based on the signals from only two cells, for instance, when two spots are resolved that are at the resolution limit of the foveal cones and, accordingly, two retinal ganglion cells.

In the retina, most neuron are nonspiking. This is possible because the distances are short and the signals can be finely graded – *the retina is a tonic machine*. Only at the output side, mostly in ganglion cells (but also in a few amacrine cells), the signals are converted into spikes, which can then travel down the long axons, a few centimeters, to the first relay station, the lateral geniculus (LGN, Figure 1.11). Since ganglion cells can be excited and inhibited (since they can generate ON or OFF responses), it is necessary that they have a baseline spontaneous spike activity (ranging from 5 to 200 Hz). It is clear that those ganglion cells with high spontaneous activity can encode smaller changes in stimulus strength than those with low activity.

1.10 Temporal and Spatial Performance

The temporal resolution of the retina is ultimately limited by the integration time of the photoreceptors. The integration time, in turn, determines the light sensitivity of the receptors. Rods have longer integration time (200 ms) and, accordingly, have lower flicker fusion frequencies (up to 10 Hz). Cones, with integrations times of 20 ms, can resolve stroboscopic flicker light of up to 55 Hz. But under normal viewing conditions, the flicker fusion frequency is much lower, 16–20 Hz. If it were 55 Hz, watching TV would be disturbing. The European TV or video format has a frame rate of 25 Hz, but two frames with half vertical resolution, presented alternatingly at 50 Hz, are interlaced to prevent the flicker from being seen.

The complete description of the eye's spatial performance is the contrast sensitivity function. This function describes the contrast sensitivity (1/contrast threshold) as a function of the SF. It is clear that contrast sensitivity must decline with increasing SF, just based on the optics of the eye: the higher the SF, the less the contrast preserved in its retinal image, due to aberrations and diffraction (the MTF, Figure 1.3). Even if the neural processor in the retina has the same contrast sensitivity, its sensitivity would decline in a psychophysical measurement. On the other hand, it is not trivial that contrast sensitivity also declines in the low-SF range. This decline is determined by neural processing: there are no such large ON/OFF receptive fields to provide high sensitivity to low SFs. This represents probably an adaptation on the abundance of low SFs in natural scenes. It has been shown that the energy at SFs falls off with about 1/SF [21]. The peak of

the contrast sensitivity function moves to lower SFs with declining retinal illuminance. At daylight, the peak contrast sensitivity is at 5 cycles per degree. Here, brightness differences of only 1/200 (0.5% spatial frequencies of 50–60 cycles per degree. However, due to the striking feature of the optical transfer function (the first Bessel function), the MTF shows a number of phase reversals beyond the *cutoff frequency* at which the contrast first declines to zero (Figure 1.4). Therefore, it may be possible to detect a grating even though its spatial frequency is higher than the cutoff frequency, even though it may have reversed contrast. For this reason, grating acuity is not the best measure of spatial vision, in particular with defocus [22].

1.11 ON/OFF Structure, Division of the Whole Illuminance Amplitude

Perhaps because no important information is contained in absolute brightness values in the visual environment, the visual system has confined its processing almost exclusively to differences – spatial and temporal contrasts. Recording from neurons along the visual pathways shows that the cells have structured *receptive fields* – defined angular positions in the visual field where they respond to the stimulation. The receptive fields are initially circular (retina, lateral geniculate (LGN), striated cortex) but become later elongated at higher cortical areas, and finally larger and may even cover the whole visual field (e.g., neurons that recognize highly specific features, like a face). Receptive fields in the retina and LGN are organized in an ON center – OFF surround structure, or vice versa. If a small spot of light is projected on to the center of an ON center ganglion cell, the cell fires vigorously, but if the surround also is illuminated, the response returns to baseline activity. If only the surround is illuminated, the cell's activity is reduced below the resting level. From the structure of the receptive fields, it is already clear that homogenously illuminated surfaces without structure are poor stimuli for our visual system; *the inside of a form does not excite our brain* (David Hubel).

Dividing the processing into ON and OFF channels, starting from an intermediate activity level, has also the advantage that the dynamic range of the responses can be expanded. It is surprising that OFF ganglion cells have smaller dendritic fields and denser sampling arrays than the ON cells. This asymmetry seems to correspond to an asymmetric distribution of negative and positive contrasts in natural images [23].

1.12 Consequences of the Rod and Cone Diversity on Retinal Wiring

Since the illuminance range is divided by the visual system into a scotopic and a photopic range, and both are not used at the same time, it would be a waste to use separate lines in the rod and cone pathway. In fact, there is no rod OFF biploar cell at all, and the rod ON pathway has no individual ganglion cells. Rather, the rods are piggy-packed on the cone circuitry at low light levels via the A2 amacrine

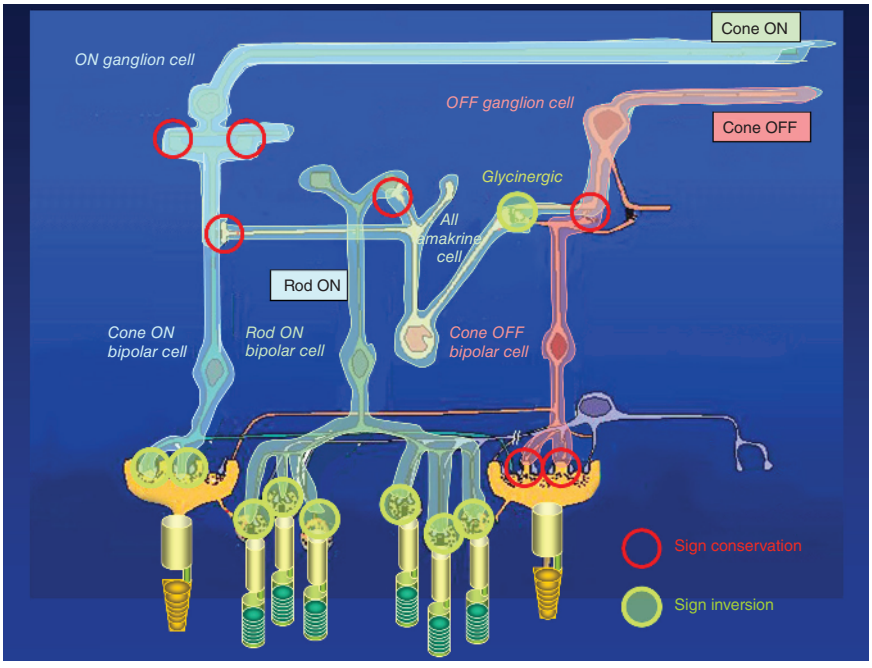


Figure 1.10 Rod and cone pathways and ON/OFF channels. To make the system more sensitive to differences rather than to absolute brightness, the image on the retina is analyzed by an ON/OFF system, that is, cells that respond preferably by changes in brightness in either direction. The division into these two major channels occurs already at the first synapse, the photoreceptor *endfoot*. Because the photoreceptors can hyperpolarize only in response to illumination, the subsequent cells must be either depolarized (*excited*) or hyperpolarized (*inhibited*). This means that the signal must either be inverted (*ON channel*) or conserved (*OFF channel*). It is shown how the signals change their signs along the processing pathway. Since rods and cones respond to different illuminance ranges, it would be a waste of space and energy to give both of them separate lines. In fact, the rods have only the first cells, the rod ON bipolar cell, which then *jumps* on the cone pathways via the *All amacrine cell*; they are not used by the cones at low light. Rods do not have their own OFF bipolar cell. Cones, with the need for coding small differences in brightness with high resolution and with large information content, have two separate lines (*ON* and *OFF*) to increase information capacity.

cells (Figure 1.10) (at least 40 amacrine cell types have been classified on the basis of their morphological appearance, their transmitters, and their electrical responses). Since the information contained in the cone signals is much richer (since it is not limited by photon noise), the neurons that carry their information have thicker axons and much more synapses. They also have higher spike rates and contain more mitochondria – the energy sources of the cell.

1.13 Motion Sensitivity in the Retina

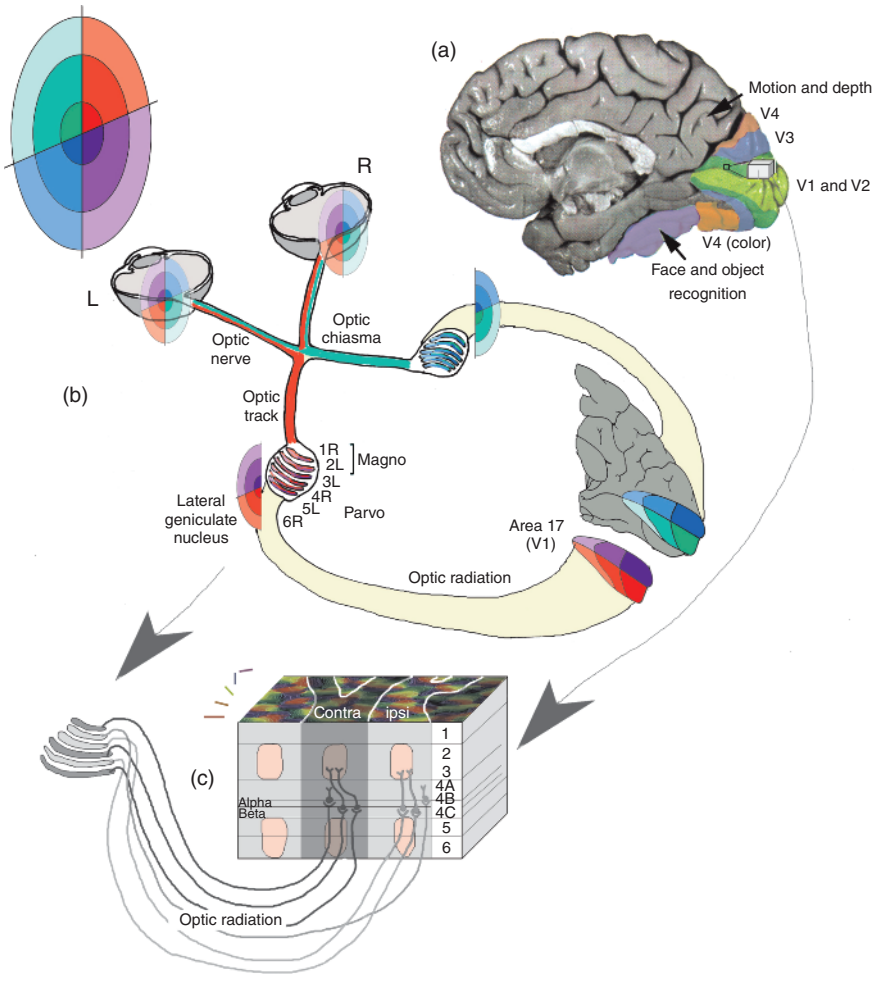
The ability to detect motion, and in particular its direction, is present already not only at the amacrine cell level in the retina but also in many neurons of the cerebral cortex. Typically, motion-selective cells fire strongly when an edge moves across their receptive field in the *preferred direction*, and they are

Figure 1.11 Feed-forward projections from the eyes to the brain and topographic mapping. In each eye, the visual field on the left and right of the fovea (the cut goes right through the fovea!) projects on to different cortical hemispheres: the ipsilateral retina projects on to the ipsilateral visual cortex, and the contralateral retina crosses the contralateral cortex (*hemi-field crossing* in the optic chiasma). The first synapse of the retinal ganglion cells is in the lateral geniculate nucleus (LGN), but information from the left (L) and right (R) eye remains strictly separated. The LGN consists of six layers; layers 1 and 2 are primarily occupied by the magnocellular pathway, and 3–6 by the parvocellular pathway. Information from both eyes comes first together in the visual cortex, area 17, layers 2 and 3, and a strict topographic projection is preserved (follow the color-coded maps of the visual field areas (b)). The wiring in A17 (c) has been extensively studied, in particular by the Nobel Prize winners Hubel and Wiesel (1981). The input from the LGN ends in layer 4C alpha (magnocellular) and 4C beta (parvocellular) and layers 1–3 (koniocellular). These cells project further into the *blobs*, cytochromoxidase-rich peg-shaped regions (pink spots in (c)). The innervation has a remarkable repetitive pattern; parallel to the cortical surface, the preferred orientation for bars presented in the receptive field of the cells shifts continuously in angle (illustrated by color-coded orientation angles on top of the tissue segment shown in (c)). Furthermore, the regions where the contra or ipsilateral eye has input into layer 4 interchange in a striking pattern. A17 is the cortical input layer with mostly *simple cells*, that is, cells that respond to bars and edges with defined directions of movements. At higher centers (a), two streams can be identified on the basis of single cell recordings and functional imaging with new imaging techniques (functional magnetic resonance imaging, fMRI): a *dorsal stream*, concerned about motion and depth (*where?* stream) and a ventral stream concerned about object features, shape color, structure (*what? stream*). Feedback projections are not shown, and only the major projections are shown.

inhibited or even silent when the motion is in the *null direction*. The illusion of motion can be provoked by stimulating briefly at two different positions in the receptive field, with a short time delay, and this illusion appears to work both during electrophysiological recordings at the cellular level and psychophysically. It was concluded that excitation evoked by motion in the preferred direction must reach the ganglion cell before inhibition can cancel it; and inhibition evoked by motion in the null direction must arrive before excitation can cancel it. The asymmetric signal transmission speed was assumed to result from morphologically asymmetric input of the so-called starburst amacrine cells to the directionally sensitive ganglion cells, although neither the developmental mechanism for the asymmetric connections nor the role of the involved transmitters, acetylcholine, and gamma-aminobutyric acid (GABA) is completely understood [24].

1.14 Visual Information Processing in Higher Centers

About 15 types of ganglion cells can be classified in the retina, and they are characterized by different morphologies, bandwidths, and response characteristics. The underlying hypothesis is that each part of the visual field is sampled by a group of ganglion cells that process different aspects of the visual information – how many aspects there are is not exactly known but the number should be between 3 and 20 [25].



1.14.1 Morphology

Researchers have always attempted to divide the visual pathways into functionally different channels. This is more successful in the initial steps, up to the primary cortex, but the separation of the pathways becomes more diffuse in higher centers. Three pathways were identified: the magnocellular, parvocellular, and koniocellular pathways. The magnocellular pathway is basically a luminance channel with low spatial acuity, large receptive fields of ganglion cells, high motion sensitivity, and high contrast sensitivity under scotopic conditions but with little or no spectral opponency. It is relayed in the LGN in the two basal layers 1 and 2 (1: contralateral eye, 2: ipsilateral eye (Figure 1.11)), and makes 10% of the LGN population. The parvocellular pathway is the high spatial acuity channel, with low temporal resolution and smaller receptive fields of ganglion cells, low contrast sensitivity under scotopic conditions, but with color opponency. It is relayed in layers 3–6 (layers 4 and 6 contralateral, 3 and 5 ipsilateral) and makes

80% of the LGN population. The koniocellular pathway is specific for blue–yellow opponency and large receptive fields of ganglion cells, but no distinct projection to the LGN layers (*intercalated projection*), making 10% of the optic nerve fibers. The separation of these pathways is preserved to the primary visual cortex, also called V1 or area A17. Here, the M cells project onto layer 4Cα and the P cells onto layer 4Cβ. The koniocellular pathway feeds into the upper layers (1–3), and there into the cytochrome oxidase-rich regions, the *blobs* (Figure 1.11).

Collaterals of M and P cells also project onto layer 6 from where they project back to the LGN; feedback is one of the common features in the visual system: apparently, the selected input can be shaped, depending on the demand of the unit that sends the feedback. For instance, feedback seems to enhance the resolution of an earlier level to the pattern isolated by the higher level [26], for instance, by enhancing the inhibitory surround in the receptive field of a cell. The feedback seems to be extremely well developed: only 10% of the input in the LGN comes from the periphery, the retina, while the corticofugal feedback connections make up to 30% of its input.

1.14.2 Functional Aspects – Receptive Field Structures and Cortical Modules

The receptive fields of retinal ganglion cells and cells in the LGN are concentric with a typical ON-center/OFF surround structure, or vice versa. At the level of V1, receptive fields become elongated, and they respond best to elongated moving slits, bars, or edges at a particular orientation, but the excitatory and inhibitory regions are no longer concentric (*simple cells*). Different cells require different orientations, and the response can be improved if the stimulus moves in the preferred direction. These cells can now be stimulated through either eye; in fact, the primary visual cortex is the first level at which binocularly driven cells are found. There are also *complex cells* that are selective for the position and orientation of an edge, but they have no longer excitatory or inhibitory regions. In the topographic representations of the visual field in the visual system, the receptive field sizes generally increase with the distance from the fovea. They also increase with hierarchic level of the brain area, and the stimuli that are necessary to excite the cells become increasingly specific. A common view is (Figure 1.10) that there are two major *streams* of information processing from the visual cortex: a dorsal stream that processes predominantly the *where* aspects of an object (location in space, depth, movement), and a ventral stream that processes the *what* aspects (shape, color, details). The P stream is assumed to feed predominantly into the *what* stream, and the M stream more into the *where* pathway. A current view is that the ventral *what* stream is actually responsible for *seeing*, whereas the dorsal *where* stream is only necessary for the direction of attention and for the control of visually guided movements [27]. Along the ventral *what* stream, cells that are extremely specific for certain features of the visual stimuli are found: for instance, they may respond only to faces or even to facial expressions, and this happens largely independently from shading and the visual angle of presentation – a demanding task also in machine vision. The receptive fields of cells that respond to faces may cover the entire visual field, but already in area MT, the major motion processing center in the *dorsal stream*, the receptive field sizes are 10 times as large as in V1.

Hubel and Wiesel won the Nobel Prize (1981) also for their discovery of the modular organization of the striate cortex (Figure 1.11c). There are topographically arranged units with 0.5 mm diameter, which contain the following subunits: (i) a column of cells with a defined orientation selectivity (*orientation column*); (ii) peg-shaped structures that extend through layers 2 and 3 and stain heavily for cytochrome oxidase, a mitochondrial enzyme linked to metabolic activity (*blobs*, assumed to be involved in color processing, with predominant input from the koniocellular stream); and (iii) two ocular dominance columns, with preferential input from either eye (*ocular dominance column*). These units repeat one after the other, and the preferred orientation (see Figure 1.11c) smoothly rotates until a 180° reversal is achieved after about 3/4 mm. A unit with a complete set of orientations has been termed a *hypercolumn*.

Topographic representation of the visual world occurs in the visual system at many levels: first, certainly, in the retina, but then also in the LGN, in the different cortical layers, the superior colliculus and the motion processing center area MT (V4), and others.

1.15 Effects of Attention

Most observations suggest that attention alters the sensitivity of neurons without affecting their stimulus preferences [28]. Only some neurons in V1 are modulated by attention; others ignore it, and some respond exclusively when the stimulus is attended. The influence of attention increases with cortical hierarchy, perhaps also with increasing feature specificity of the cells. Effects of attention can be measured, for instance, by training a monkey to fixate a red or green spot. A neuron, for instance in area V4, is recorded, and the receptive field is mapped. A red or green stimulus, or nothing (e.g., a bar) appears in the receptive field. The monkey is trained to respond if the stimulus color matches the color of the fixation spot or not. In this kind of experiment, the responses of the V4 cell can be compared with different levels of attention and different stimuli.

Attention may also change the synchrony of neuronal signals in the visual cortex, although it is not yet studied how synchrony affects the strength of the neural responses. If attention can shape the responses of neurons to better performance, the question arises why all neurons are not at maximum sensitivity at all times. Presumably, the sensitivity is set to produce an appropriate balance between false alarms about the presence of a stimulus and failure to detect it [28].

1.16 Color Vision, Color Constancy, and Color Contrast

Both the location of the absorption peak of the photopigment and the number of photons that arrive determine the probability that a photon is caught by the photoreceptor. These two variables cannot be separated because the photoreceptor is only a *photon counter*. For the same reason, a single photoreceptor type also cannot provide information on the light intensity. Only the sum of the responses from several receptor types can provide this information, whereas the differences

between their responses provide information on the spectral composition of the light reflected from an object and its *color*. For this purpose, photoreceptors have photopigments with different spectral absorptions. Most mammals, including male new-world primates, are dichromatic, indicating that they have only two cone pigments. Only old-world monkeys and humans are regularly trichromatic, but fish, reptiles, and birds may even be tetrachromatic.

Perhaps because the spectral absorbance curves of the photopigments are wide, and there is particularly much overlap in the spectra of the L and M cones with high correlations in the signals, fine wavelength discrimination (as present in our visual system, best performance at about 550 and 470 nm with a detection of only 1 or 2 nm difference) can be achieved only by antagonistic circuitry. Already at the level of the ganglion cells, the initial three spectral sensitivities of the cones are recombined into three mechanisms: (i) a luminance channel, which consists of the added L + M signals, (ii) an L − M color opponent channel where the difference between the signals from L and M cones is taken to compute the red–green component of a stimulus; and (iii) an S − (L + M) channel where the sum of the L and M cone signals is subtracted from the S cone signal to compute the blue–yellow variation of the stimulus. These three channels represent the *cardinal directions in the color space* pathways.

The so-called magnocellular pathway (large *magnocellular* ganglion cells projecting on to the bottom two layers in the LGN) is most sensitive to luminance information, with high contrast sensitivity, low spatial resolution, and no color sensitivity. In the parvocellular pathway (small *parvocellular* ganglion cells, projecting to the upper four layers of the LGN), the red–green information is transmitted, and in the koniocellular pathway (intermediate ganglien cell sizes to all LGH layers), the blue–yellow information is transmitted.

Up to the LGN, it seems to be possible to define cell classes with different functions, but later, in the cortex, this becomes increasingly diffuse. Considerable efforts have been devoted to link the structural differences to functional differences, looking for *the cell type* that processes *color or form, or motion*, or locating *the brain area* that processes a certain aspect of a stimulus. However, the more experiments are done, the less clear becomes the link between the function, position, and structure, and it seems as if most cells in the central visual system have access to most features of a stimulus and that there is no complete segregation of processing aspects. These so-called multiplexing properties are found in cortical neurons as early as in V1.

Also, color processing does not seem to occur independently of the form [29]. For instance, patients with normal photopigments but with loss of color vision due to accidental damage of the cortex (*acquired achromatopsia*) may see objects without color but may still have near-normal sensitivity to chromatic gratings [30].

The color of an object is determined by the proportion of light that it reflects at a given wavelength, which is described by the spectral reflectance function. Color vision is confounded by the spectral composition of a light source. For instance, if the light source includes more light of long wavelengths, the L cones absorb

more photons and this should cause a reddish impression of the scene. However, this does not usually happen because the effect of illumination is successfully compensated by the visual system (*color constancy*). Color constancy is not only locally controlled at the receptoral level, since the spectral composition of the light at distant regions in the visual field changes the local spectral sensitivity function.

Analogous to luminance contrasts, which were best detected by ON center/OFF surround cells (or vice versa), color contrasts would be best detected by cells that have antagonistic red–green mechanisms (e.g., +L −M) both in the center and surround, and measure the differences between the center and surround (*double opponent cells*). Such cells, with concentric receptive fields, have been found in the primary visual cortex of monkeys. If their receptive fields are larger, they could also mediate color constancy over extended regions in the visual field.

1.17 Depth Perception

The visual system uses several independent cues to determine the distance of objects in depth. These can be divided into monocular and binocular cues. Monocular depth estimations are typically possible from motion parallax, familiar size, shading, perspective, and – to a minor extent and only for close distances – from the level of accommodation necessary to focus an object. The major depth cue, however, is binocular and results from the fact that both eyes see an object at a different angle. Accordingly, the retinal images best match at the fixation point, that is, the images in the fovea. The peripheral parts of the images do not match; they show *disparity* (Figure 1.3). Rather than producing the impression of double images (*diplopia*), the cortex has some tolerance to these non-corresponding images and can still place them together. But the differences are recognized, which provides a highly sensitive mechanism for depth detection, called *stereopsis*. The two most striking features of stereopsis are that (i) it does not require object recognition but rather works also on Julesz's random dot patterns, and (ii) the difference between the images in both retinas is detected that are smaller than the diameter of a photoreceptor: stereopsis involves a *hyperacuity*. This is necessary, for instance, to insert a thread through a needle's eye. Considering that the visual processing may be based on displacements of less than a photoreceptor's diameter, it is even more striking that displacement of *blue versus red images*, equivalent to about 15 photoreceptor diameters (Figure 1.7), remains undetected.

Disparity-sensitive neurons have been extensively recorded not only in the visual cortex in V1 but also in V2 and the motion areas in the dorsal stream. Cells can be classified as *near cells*, which respond best to crossed disparities, and *far cells* that respond best to uncrossed disparities. Both these types are most frequent in the motion areas, whereas cells that respond best to zero disparity (*zero-disparity cells*) are abundant in areas V1 and V2 (Figure 1.11).

1.18 Adaptation in the Visual System to Color, Spatial, and Temporal Contrast

One of the most striking features of neural processing in our visual system is that the gains for all aspects of vision are continuously adapted. The most immediate adaptation is light/dark adaptation, independently in each photoreceptor, but this adaptation in cones modifies their relative weight and, therefore, also color vision. Not only the receptors adapt but also higher processing steps in the visual system (review [31]). A few examples are that there are (i) motion adaptation (after the train stops, the environment appears to move in the opposite direction); (ii) tilt adaptation (after one looks at tilted bars, vertical bars appear to be tilted in the opposite direction); (iii) contrast adaptation (after prolonged viewing of a high-contrast grating, the sensitivity for detection of similar gratings is severely reduced – this also affects the impression of *sharpness* of an image); (iv) adaptation to scaling (if one looks at a face in which the distance between the eyes is artificially reduced, the *control face* appears to have a larger interocular distance, the *face-distortion aftereffect*; and (v) adaptation to optical aberrations and field distortions of the image on the retina, which are well known to spectacle wearers. Recovery in most of these adaptations is generally in the range of seconds, but extended exposure may also cause extended changes in perception. A most striking example here is that wearing red or green spectacles for 3 or 4 days will shift the ratio of the red–green (L–M cone) weighting also for several days [32]. That the weighting of the different cone inputs can be extremely shifted (perhaps also adapted) could explain why subjects with very different L/M cone ratios (naturally occurring variability: 0.25 : 1 to 9 : 1) can all have normal color vision. Selective adaptation of color channels can also nicely be seen at the homepage of Professor Michael Bach, http://www.michaelbach.de/ot/col_rapidAfterimage/.

1.19 Conclusions

The most striking features of the natural visual system are the extreme plasticity of image processing and the apparent optimal use of energy, space, and cell numbers.

1) To prevent saturation of photoreceptors over a range of possible ambient illuminances of, at least, 8 log units, in the presence of only about 2 log units of simultaneous contrast in natural scenes, their response curves are shifted by altering the gain of the phototransduction cascade. Furthermore, the retina divides the entire illuminance range into two, the low-illuminance range, where rods respond, and the high-illuminance range, where cones take over. To save energy and wire volume in the retina, both photoreceptors use the same circuitry, and their inputs to these circuits are automatically switched over by synaptic plasticity, although, as in most cases in nature, with a smooth transition.

2) Cable capacity is matched to information content, and multiple parallel channels with different bandwidths are used, if necessary, in the case of cones to make it possible to transmit all relevant information.
3) There is extensive image preprocessing directly in the initial light sensor, namely the retina. This is necessary because 125 million photoreceptors converge into an optic nerve with only about 1 million lines and also because the information from the cones is too much to be fully processed by the cortex. For the same reason, high-acuity information is processed only in the central 1° of the visual field, and this limitation is compensated by eye movements. The spatial resolution here is as good as physically possible for the given eye size of 24 mm, since diffraction at the pupil and the waveguide properties of the cones preclude denser spatial sampling. The preprocessing in the retina focuses on temporal and spatial brightness differences, and spatial and temporal bandpass filtering, by building antagonistic receptive field structures of the output cells of the retina (ON/OFF, color opponency, motion selectivity with preferred direction), to enhance sensitivity to small changes. By matching the sensitivity to the available stimulus strength, maximum information is extracted; for example, contrast sensitivity is continuously adapted at each spatial frequency to make optimal use of the stimulus contrast.
4) Rescaling the retinal image, or compensating for local distortions or optical aberrations, seems to occur largely without effort – a most impressive performance.
5) Depth is determined from several monocular cues, but the most powerful mechanism is stereopsis, derived from binocular disparity. Stereopsis does not require object recognition and works on random dot stereograms; it is impressive how corresponding dots in the two retinal images are identified, and it demonstrates extensive parallel comparisons over a wide range of the visual field.
6) In the cortex, different aspects of the image are initially analyzed at each position in the visual field (small receptive fields of the respective neurons). Later, the trigger features of the cells become more and more specific and the receptive field sizes increase. Cells can be recorded in higher cortical centers that are selective for faces and expressions, and that retain their selectivity for these stimuli at different illuminations. How this information is extracted is not yet clear.
7) An unsolved problem is how and where all the separately processed features finally converge to provide the complete picture of a visual object (the *binding problem*).
8) It is interesting to note that there is an extensive description of the neural responses in the visual system (i.e., retinal ganglion cells) to stimulation with light spots and simple patterns, but no one would dare to describe the visual scene given only a recording of the optic nerve trains. *Natural vision has not been the focus of much research* [33]. It seems that there is need for more research on the responses of the cells in the visual system with natural stimulation.

Acknowledgements

I am grateful to Annette Werner, Howard C. Howland, Marita Feldkaemper, and Mahmound Youness for reading and commenting on an earlier version of this manuscript.

References

Material that is covered by most common text books on the visual system is not referenced.

1 Rodieck, R.W. (1973) *The Vertebrate Retina*, Freeman, San Francisco, CA.
2 Hughes, A. (1977) The topography of vision in mammals of contrasting life styles: comparative optics and retinal organization, in *Handbook of Sensory Physiology*, vol. **VII/5**, Part A (ed. F Crecitelli), Springer-Verlag, Berlin, pp. 615–637.
3 Marcos, S. (2003) Image quality of the human eye. *Int. Ophthalmol. Clin.*, **43**, 43–62.
4 Huxlin, K.R., Yoon, G., Nagy, J., Poster, J., and Williams, D. (2004) Monochromatic ocular wavefront aberrations in the awake-behaving cat. *Vis. Res.*, **44**, 2159–2169.
5 Howland, H.C. (2005) Allometry and scaling of wave aberrations of eyes. *Vis. Res.*, **45**, 1091–1093.
6 Artal, P., Guirao, A., Berrio, E., and Williams, D.R. (2001) Compensation of corneal aberrations by the internal optics of the human eye. *J. Vis.*, **1**, 1–8.
7 Kelly, J.C., Mihahsi, T., and Howland, H.C. (2004) Compensation of corneal horizontal/vertical astigmatism, lateral coma, and spherical aberration by the internal optics of the eye. *J. Vis.*, **4**, 262–271.
8 Collins, M.J., Buehren, T., and Iskander, D.R. (2006) Retinal image quality, reading and myopia. *Vis. Res.*, **46**, 196–215.
9 Guirao, A., Gonzalez, C., Redono, M., Geraghty, E., Norrby, S., and Artal, P. (1999) Average optical performance of the human eye as a function of age in a normal population. *Invest. Ophthalmol. Vis. Sci.*, **40**, 203–213.
10 Wilson, B.J., Decker, K.E., and Roorda, A. (1992) Monochromatic aberrations provide an odd-error cue to focus direction. *J. Opt. Soc. Am. A*, **19**, 833–839.
11 Curcio, C.A., Sloan, K.R., Kalina, R.E., and Hendrikson, A.E. (1990) Human photoreceptor topography. *J. Comput. Neurol.*, **292**, 497–523.
12 Williams, D.R. (1985) Aliasing in human foveal vision. *Vis. Res.*, **25**, 195–205.
13 Kirschfeld, K. (1983) Are photoreceptors optimal? *Trends Neurosci.*, **6**, 97–101.
14 Reymond, L. (1985) Spatial visual acuity of the eagle Aquila audax: a behavioral, optical and anatomical investigation. *Vis. Res.*, **25**, 1477–1491.
15 Marcos, S., Burns, S.A., Moreno-Barriusop, E., and Navarro, R. (1999) A new approach to the study of ocular chromatic aberrations. *Vis. Res.*, **39**, 4309–4323.

16 Magnussen, S., Spillmann, L., Sturzel, F., and Werner, J.S. (2001) Filling-in of the foveal blue scotoma. *Vis. Res.*, **41**, 2961–2971.
17 Artal, P., Fernandez, E.J., Singer, B., Manzamera, S., and Williams, D.R. (2004) Neural compensation for the eye's optical aberrations. *J. Vis.*, **4**, 281–287.
18 Webster, M.A., Georgeson, M.A., and Webster, S.M. (2002) Neural adjustments to image blur. *Nat. Neurosci.*, **5**, 839–840.
19 Heinrichs, T.S. and Bach, M. (2001) Contrast adaptation in human retina and cortex. *Invest. Ophthalmol. Vis. Sci.*, **42**, 2721–2727.
20 Sterling, P. (2003) How retinal circuits optimize the transfer of visual information, in *The Visual Neurosciences*, vol. **1** (eds LM Chalupa and JS Werner), MIT Press, Cambridge, MA, pp. 234–259.
21 Field, D.J. and Brady, N. (1997) Visual sensitivity, blur and sources of variability in the amplitude spectra of antural scenes. *Vis. Res.*, **37**, 3367–3383.
22 Gislen, A. and Gislen, L. (2004) On the optical theory of underwater vision in humans. *J. Opt. Soc. Am. A*, **21**, 2061–2064.
23 Ratliff, L., Sterling, P., and Balasubramanian, V. (2005) Negative contrasts predominate in natural images. *Invest. Ophthalmol. Vis. Sci.*, **46** (Suppl.), 4685 (ARVO abstract).
24 Sterling, P. (2002) How neurons compute direction. *Nature*, **420**, 375–376.
25 Kaplan, E. (2003) The M, P, and K pathways of the primate visual system, in *The Visual Neurosciences*, vol. **1** (eds LM Chalupa and JS Werner), MIT Press, Cambridge, MA, pp. 481–493.
26 Silito, A.M. and Jones, H.E. (2003) Feedback systems in visual processing, in *The Visual Neurosciences*, vol. **1** (eds LM Chalupa and JS Werner), MIT Press, Cambridge, MA, pp. 609–624.
27 Movshon, J.A. (2005) Parallel visual cortical processing in primates. *Invest. Ophthalmol. Vis. Sci.*, **46** (Suppl.), 3584 (ARVO abstract).
28 Maunsell, J.H.R. (2003) The role of attention in visual cerebral cortex, in *The Visual Neurosciences*, vol. **2** (eds LM Chalupa and JS Werner), MIT Press, Cambridge, MA, pp. 1538–1545.
29 Werner, A. (2003) The spatial tuning of chromatic adaptation. *Vis. Res.*, **43**, 1611–1623.
30 Gegenfurtner, K. (2003) The processing of color in extrastriate cortex, in *The Visual Neurosciences*, vol. **2** (eds LM Chalupa and JS Werner), MIT Press, Cambridge, MA, pp. 1017–1028.
31 Webster, M.A. (2003) Pattern-selective adaptation in color and form perception, in *The Visual Neurosciences*, vol. **2** (eds LM Chalupa and JS Werner), MIT Press, Cambridge, MA, pp. 936–947.
32 Neitz, J., Carroll, J., Yamauchi, Y., Neitz, M., and Williams, D.R. (2002) Color perception is mediated by a plastic neural mechanisms that is adjustable in adults. *Neuron*, **35**, 783–792.
33 Meister, M. and Berry, M.J. II, (1999) The neural code of the retina. *Neuron*, **22**, 435–450.