From light to optic nerve: Optimization of the front-end of visual systems

Arthur E. Burgess
Radiology Dept., Brigham and Women's Hospital
Harvard Medical School. 75 Francis St. Boston, MA 02115

ABSTRACT

Blindness in nature is (almost always) fatal. In biology and physiology one finds many situations where nature has obtained neat solutions to problems, solutions that are very nearly the best possible. Many of the design parameters for the eye are not arbitrarily selected, but are constrained to a narrow range of values by physics and information theory considerations. As Helmholtz (1868) mentioned more than a century ago "The eye has every possible defect that can be found in an optical instrument and even some which are peculiar to itself; but they are all so interacted, that the inexactness of the image which results from their presence very little exceeds, under ordinary conditions of illumination, the limits which are set to the delicacy of sensation by the dimensions of the retinal cones." Helmholtz was particularly prescient in his reference to cone dimension because, as we will see, many eye properties are completely determined once cone diameter is selected. The ideas presented in this paper are based on the working assumption that the eye does the best possible job within physical limits. This idea originated with Horace Barlow more than 40 years ago. One excellent reference is the proceedings {*Vision: Coding and Efficiency* (Blakemore, 1990)} of a conference organized to honour Barlow's (nominal) retirement with presentations by his many collaborators over the years. The list includes practically everyone referenced in this paper, which explores the design and optimization of the optics of the eye, retinal transduction and coding of visual data.

OVERVIEW OF PAPER

The design problem can be divided into a number of steps - each with its own optimization considerations. It will be seen that each step leads inexorably to a beautifully optimized system.

a: Select wavelength range, photo-transduction and high gain, low noise pre-amplifier mechanisms.
b: Select retinal matrix material and size of photoreceptors - using waveguide physics.
c: Consider daylight and night operation - resolution by day and sensitivity by night.
d: Select f# (focal length/pupil diameter) - based on photoreceptor size and Nyquist sampling theorem
e: Select simple or compound lens and focal length - depends on body size.
f: Select retinal data encoding method (spatial, temporal, and chromatic) to overcome optic nerve channel capacity limitations, taking into account photon fluctuations, scene statistics and neural noise.

1. LIGHT SPECTRUM CONSIDERATIONS

Summary:
Thermal noise problems at long wavelengths. Problem of UV damage to cells at short wavelengths. 500 nm is the best spectrum peak wavelength at night and 550 nm in the day for terrestrial organisms. There is some biological variation in the peak wavelength for specialized purposes.

Transduction. Sunlight at the surface of the earth covers a spectral range of wavelengths, λ, from 300 nm to more 4000 nm. The human visual system has a narrower tuning, 400 nm to 700 nm, peaked at about 500 nm, as shown in figure 1. Most primates have a similar spectral response. One common misconception is that the sensitivity peak is matched to the solar spectrum peak - not so. The maximum of a continuous emission spectrum has no fixed peak - the peak position depends on the spectral coordinates (wavelength, frequency) and the dependent variable (energy, photon flux) used to plot the graph. In any event, as can be seen in figure 1, the solar spectrum is nearly flat over the human spectral sensitivity range. The basic transducer element in visual systems is a visual pigment consisting of an apoprotein, opsin, covalently linked to 11-cis-retinal (or in rare circumstances 11-cis-dehydroretinal). The retinal molecule undergoes a shape change when a photon is absorbed. This in turn leads to a conductivity change in the membranes of the photoreceptive cells - the result is a very high gain, low-noise multi-stage amplification process (as is good engineering practice). The exact wavelength of the absorption peak is determined by the which one of many opsin molecules is attached to the basic retinal chain. The most fundamental

SPIE Vol. 3340 • 0277-786X/98/$10.00

consideration is we live in a world at an absolute temperature of about 300K without refrigerated sensory systems. The activation energy required to initiate the change in shape of the retinal molecule can be delivered either by a photon or by thermal agitation. It would be disastrous to have a high noise level due to thermal activation. Hence, the long wavelength cutoff for visual sensitivity is determined as follows (Vos and van Norren, 1984). Let E (= hc/λ) be the energy needed to active one transducer. The number of spontaneous thermal excitations per second for one rhodopsin molecule with a mean lifetime, $t \approx 10^{-13}$ sec, of a given energy distribution is $q \approx \tau^{-1} \exp(-E/kT)\Sigma[(E/kT)^p/p!]$, where p is the index number for vibrational modes and E is about 70 kT. This gives a single molecule false alarm about once every 10^{10} seconds ($\approx$ 300 years). For the entire monkey rod ($\approx 4x10^9$ transducer molecules), the observed false alarm rate is about 1 every 180 seconds (Baylor, 1987). Human rods have a similar number. The cone false alarm rate is about one per second. The variation of relative rod false alarm rate that would occur if the sensitivity peak were shifted to longer wavelength is shown by the extremely steep upward line in figure 1. A shift of 200 nm would increase the rate by a factor of 10^7. One lesson from these calculations is how fast the function e^{-ax} changes if 'x' is doubled when 'a' is a large number.

The short wavelength spectral sensitivity cut-off is possibly determined by the risk of molecular damage in the retina from ultraviolet light during daylight hours (Ham et al., 1982). Many animals do have spectral sensitivity well down into the UV region but some also show evidence of retinal damage from extended exposure to high levels of sunlight (Lythgoe,1979). Damage is prevented in primates by a protective pigment in the lens, which increases light absorption dramatically below 400 nm, to ensure that the number of ultraviolet (UV) photons falling on the retina is very low. The curve labeled "relative retinal damage rate" in figure 1 is actually the measured lens pigment absorption factor (from Ham et al.).

Considering both thermal noise and retina damage constraints, the best compromise choice is a spectral peak around 500 nm. The human night sensitivity curve [scotopic, $V(\lambda)$] is the shaded region in figure 1. Thermal noise is not as important a consideration in daylight because of the much higher photon flux, so a peak at 550 nm is a better choice. Also, the best choice of spectral peak for different species can vary slightly from the above values depending on lifestyle details - fish, for example, live in a variety of spectral environments that are depth dependent. There is still one open question to which no one has been able to answer. Why is the retinal molecule and the transformation from 11-cis to all-trans the basis for photoreception in most, if not all, visual systems?

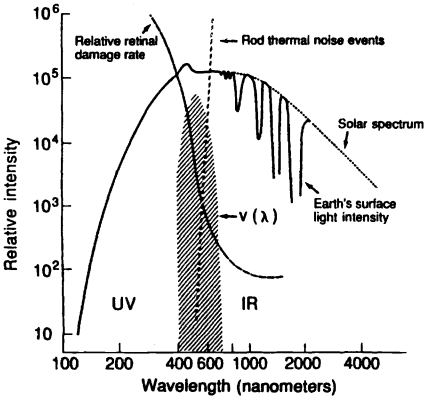

Fig. 1. Spectral Effects

Amplification. To quote Schnapf and Baylor (1987); "One might naively think that the [photoreceptor] cell would be dormant in the absence of light; in reality, however, the cell is abuzz with activity." There is a difference in ion concentrations across the photoreceptor surface membrane and a flow of sodium and potassium ions. The net loop flow is called the "dark current". When a rod or cone absorbs light, the influx of sodium is blocked for a short time. The dark current is reduced and a negative polarization of the cell interior occurs. The graded potential change (up to -30 mV) propagates to nearby "signal processing" cells. The following summary is based on McNaughton (1990). The isomerized rhodopsin molecule activates up to 500 transductin molecules during random diffusion. Each transductin molecule switches on one phosphodiesterase molecule which in turn breaks down about 500 molecules of 3',5'-cyclic guanosine monophosphate (cGMP). The cGMP molecule interacts cooperatively to open membrane channels that allow passage of about 15 ions during an opening lifetime. The transient reduction in cGMP due to photon absorption blocks some channel opening events. The total gain (about $4x10^6$) occurs in three stages and the noise level is intrinsically low. There are 3 dark noise sources. Thermal noise, a low frequency (1/f) noise that is not well understood and also some high frequency noise due to random opening and closing of light-sensitive channels - but both seem to be removed from the signal before it is transmitted to the brain. The variability in response to steady light is little more than that due to photon statistics.

2. EYE OPTICAL DESIGN

Summary:
a: The photoreceptor waveguide should have light confined to its interior -> diameter, p ≈ 2 to 4 λ.
b: Select operation mode: day (angular resolution -> best f#), night (light sensitivity-> low f#).
c: Matching the optics to the retinal grain by optimum (Nyquist) sampling -> best f# in 4 to 8 range.
d: Angular resolution, Δθ, needs vary inversely with body height (H) -> focal length, f = αH.
e: Simple or compound eye? eye diameter scaling is (d α $\Delta\theta^{-1}$) and (d α $\Delta\theta^{-2}$) respectively.
f: Variety in eye position (front, side) and view angles based on survival needs.
g: Maximize photon transmission through eye media to retina and photon use at retina (e.g. tapitum in cat).

Examples: human and cat eyes. (figure 2) The cornea and lens focus incident light on the retina. Most of the refraction is from the cornea and refractive index gradients in the lens. Lens thickness variation provides adaptive focusing. The (circular) human pupil diameter varies (2 to 8 mm) to control the amount of light entering the eye. The cat has a slit pupil, which allows a much great range of pupil area, and a lens near the middle of the eye to lower the f#. Largest pupil sizes, focal lengths and lowest f# are (human; 8 & 17 mm, f/2) and (cat ; 14 & 12.5 mm, f/0.9). The intraocular media provides mechanical support.

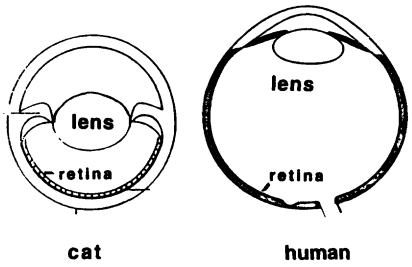

Fig. 2. Cross -sections of cat and human eyes.

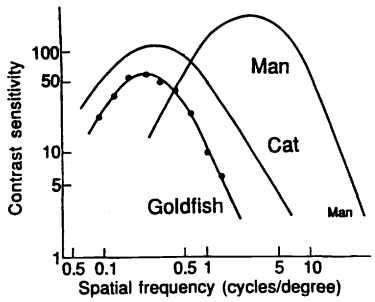

Fig. 3. Daylight contrast sensitivity examples.

Photoreceptors - typically consist of a long cylinder with a sequence of plates perpendicular to the cylindrical axis. The photoreceptor boundary acts as a waveguide. The design of this waveguide must ensure that a significant fraction of the light is confined to the interior of the photoreceptor - to avoid "cross-talk" between receptors. The required diameter, p, of this waveguide is completely determined by the wavelength of light and the refractive indices, n_1 and n_2, of the interior and exterior of the waveguide (Snyder, 1975). The waveguide equation, $V=(\pi p/\lambda)[(n_1)^2 - (n_2)^2]^{1/2}$, with 50% confinement at V=2, shows that the minimum feasible diameter for the photoreceptors is on the order of 2 to 4 wavelengths (1 - 2 microns). The uncertainty is due to variation in interior refractive index, n_1 and the difficulty in estimating its value. The 1 to 2 micron photoreceptor minimum diameter holds everywhere in nature, for all species and for both compound and simple eyes.

f-number (f#). The first step is to consider the nature of what is to be imaged. Optical systems can be divided into 2 classes. One, like astronomical telescopes, is designed to image point sources and the sensitivity criteria is the number of photons that fall within the Airy disk diameter. So telescopes have large apertures and focal length is not important as long as the disc is much larger than the grain of the recording medium. The second class is devices, such as cameras, designed to image extended sources (where f# is the quantity to optimize). The visual world consists mainly of extended objects so eyes are members of the second class. For day (high light level) operation the design goal is high resolution - ultimately limited by diffraction. The diameter of the Airy disc is d = 2.44λf#, independent of focal length. The Nyquist theorem gives an optimum of 5 photoreceptors (samples) across the Airy disc, so the best f# choice is f# = 5p/2.44λ =2p/λ = 4 to 7. Birds have a minimum f# of 4. Humans are diffraction-limited for pupil diameter about 2.4 mm (f# = 7). However, it is very difficult to build a diffraction-limited lens using the full lens diameter. Makers of cheap cameras build a large diameter lens and then stop it down with an smaller pupil to obtain good optical quality. Biological optical systems are similar.

Many animals also operate at night. Then vision is limited by photon noise rather than resolution. So photon collection of photons and concentration on the retina should be optimized. Retinal luminance is proportional to $(f^{\#})^{-2}$ so it is an advantage to have a low $f^{\#}$ and optical quality can be sacrificed in order to use the full entrance pupil. Humans have a lowest $f^{\#}$ of 2; owls, 1.3; cats, 0.9; and some nocturnal insects as low as 0.5. Another desirable feature in low light is efficient use of the photons that enter the eye. For humans; about 2/3 of photons entering the eye reach the retinal surface, about 3/4 of these enter rods, about 1/2 of those are absorbed by rhodopsin molecules, and about 2/3 of absorptions give rise to neural excitation (Barlow, 1981). In many animals the fraction of photon collection at the retina is increased by the presence of a reflective layer (tapetum) directly behind the photoreceptors, which ensures two chances for absorption. The net photon use is about 1/6 for humans and 1/4 for cats (which have a tapetum).

Focal length. As we know from photographs, angular resolution improves when camera focal length increases. The resolution needs of an animal will be determined by survival tasks. Kirschfeld (1976) suggested that the space important to a particular species ought to be measured using its body length as a unit of distance. If this is so - then larger animals would need a better angular resolution to see further and resolution angle would be inversely proportional to body size. Biological variation of angular resolution with body height is illustrated in figure 4 adapted from a marvelous paper by Kirschfeld. Most animals and insects fall within a comparatively narrow band. At the high resolution end one finds birds of prey such as falcons and eagles and predatory mammals. At the low end one encounters a variety of insects. Hence, one concludes that eye focal length ought to be proportional to body size. Note that your cat may be legally blind according to the definition applied to humans (acuity less than 20/200), rats certainly are.

Compound or simple eye? For simple lens eyes, focal length is inversely proportional to resolution angle (fig. 5). For compound eyes, radius is proportional to $(\text{res. angle})^{-2}$ because both the size and number of elements must be increased. Humans would need 30 meter compound eyes to achieve our foveal resolution over the entire eye and 1 meter if resolution varied (Kirschfeld, 1976). Why do insects have compound eyes? It is not feasible to have an image-forming eye less than about 0.5 mm in diameter because photoreceptor length of about 0.2 mm is needed to ensure adequate photon absorption. At 0.5 mm diameter a simple eye would suffer severe aberration off the lens axis, whereas most compound eyes have fixed resolution over the entire visual field. So, it would seem that for small species opt for the latter.

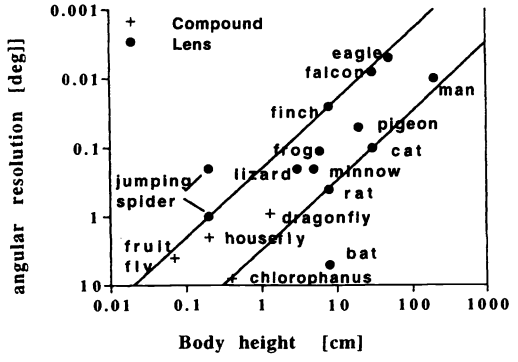

Fig 4. Variation of resolution with body size for example species: both simple and compound eyes. Data is from Kirschfeld (1976) with permission.

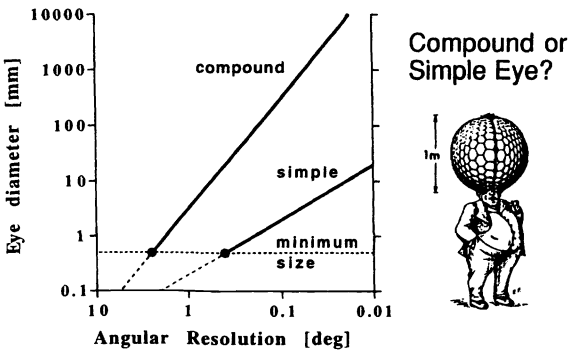

Fig. 5. Variation in the necessary eye size to achieve a desired angular resolution. From Kirschfeld (1976) with permission.

Other species variations. Life in water presents problems for eye design - the following is based on Denny (1993). Terrestrial animals have relatively large refraction at the corneal surface because of the high refractive index difference between air and tissue. That is followed by a variable focus lens which is thin to minimize spherical aberration. In water there is a much smaller corneal refractive index difference, so fish have a smaller radius cornea and a (more refractive) nearly spherical lens with a radially graded refractive index to minimize spherical aberration. Some species have to see both in and out of water; e.g. surface

feeding fish, diving mammals, birds and reptiles. A variety of eye designs have evolved to solve these problems. There is also variation due to lifestyle on land; hunters have eyes in the front and limited high acuity angles of view for stereoscopic vision while prey tend to have eyes placed on the sides to obtain a 360 degree view angle. The range of "best" acuity visual angles found in nature is large. "Best" being in a relative sense, since acuity falls off steadily for some maximum value for simple eyes (more on this below). Some animals such as primates have a very rapid decrease from a high central acuity while others (such as rats) start with a lower maximum and have a much shallower gradient. Some birds of prey have an extra design wrinkle - foveal pits shaped like the inside of an ice cream cone. These pits allow cones to be placed at an angle to the incoming light and give an closer cone packing (shorter sampling distance - by a factor of 2) in the direction perpendicular to the light flux. So eagles and hawks have a visual acuity about twice that of humans with eyes of about the same size.

This concludes the discussion of the design of the optics itself. It is interesting to note that the design of the visual system has a number of tight constraints so that parameter selection is far from arbitrary. One presumes that the arrival at nearly optimum values for the design parameters has been mediated by natural selection in which animals with visual systems of suboptimal design would not have as high a survival rate as those with better designs.

3. RETINAL FUNCTION

Summary
a: Aim: convert continuous data to discrete samples in space and time; encode and transmit to cortex
b: Nyquist-sampled input rate at the human retina is $5x10^8$ spatial samples over the entire field of view.
c: Temporal sampling need is limited by photon capture rate and depends on light levels.
d: Estimated total data input rate for the entire visual field is about 10^{10} samples/sec. [10^{11} bits/sec.]
e: There are about 10^9 to 10^{10} processing cells and $1.5x10^6$ output (ganglion) cells in the retina.
f: The data handling capacity of the cortex is limited so data rates must be reduced at early stages.

Input data rates. The function of the retina is to transduce a continuous 2D luminance pattern and convert it to discrete electrical pulse data for transmission to the brain - where most visual analysis is done in higher animals. The first question to be answered is how finely the sampling and quantization ought to be done. A typical human eye has an angular field of view of about 160° (horizontal) by 175° (vertical direction). The Nyquist sampling angle at diffraction-limited operation (f/7) is 0.01 degrees which would give a total of $5x10^8$ spatial samples over the entire retina. The required combination of temporal sampling and amplitude quantization can not be determined by a similar fundamental analysis. Barlow (1981) observed that "The eye as a whole works over a vast range of luminance, from roughly 10^{-7} cd/m^2 for the lowest visible extended surface to 10^4 cd/m^2 for the brightest patches that one commonly encounters. ... If one allows for changes in pupil diameter and for optical losses ... the corresponding photon capture rates are 10^{-3} photons/sec. per rod and $2x10^5$ photons/sec. for cones". The external scene can have a instantaneous luminance dynamic range as high as 1000/1 and the RMS contrast of scenes as high as 0.4 [Laughlin (1983)]. In addition, the eyes are continually moving.

One approach might be to start with an empirical human minimum reaction time of about 0.2 seconds and claim that a temporal sampling rate of 4 per reaction interval (20 per second) is adequate. This gives 10^{10} samples/sec. for the whole retina. Amplitude quantization per sample can then be estimated. The photoreceptor output, v, is proportional to the photon count; $v = an$, with standard deviation, $\sigma_v = (av)^{1/2}$. The appropriate adaptive quantization transformation is square-root [$x(v) = (v/a)^{1/2}$] which gives σ_x equal to unity. We then must select a quantization step, q, for the transformed data. Let N and N/R be the largest and smallest number of photons per photoreceptor per sampling time interval (50 msec.).The number of steps required is (for large R) equal to $N^{1/2}/q$, independent of R. The best choice of step size is to keep the q/σ ratio between 1 and 0.5 (Burgess, 1985). For $N = 10^4$, the number of required steps is 100 to 200 (7-8 bits). This gives an estimated maximum spatio-temporal data rate of about 10^{11} bits per second.

An alternative approach is to base the combined temporal and amplitude coding on the maximum of $2x10^5$ photons/sec., which gives estimate (for a square-root quantization scale) of between 500 and 1000 levels

per second. Since there is no particular result to prefer either amplitude or temporal variation over the other, one could use the principle of equipartition of channel capacity and devote half (on a ratio scale) to each - would gives about 30 temporal samples/sec. and 30 amplitude levels per sample. This approach also gives an estimated maximum spatio-temporal data rate of about 10^{11} bits per second.

Overview of the visual system. Cell counts variation in the visual system are shown in figure 6. Unfortunately, the rest of the visual system has a limited processing capability - certainly not enough to handle the full retinal data rate. So some economy of data handling must be introduced at an early stage. The optic nerve has a limited channel capacity and a large fraction (20 to 50%) of the brain is devoted to visual analysis after all the economies introduced by retinal encoding. Retina design features are based on a number of strategies. We do not really need both high spatial and high temporal resolution over the entire visual field. Fast temporal response in the periphery can trigger reflexive eye movement and direct the fovea to the direction of interest. Example visual reaction times are 30 msec. for the fly and 125-200 msec. for man (Land, 1981). At high light levels we can perceive flickering lights up to frequencies of about 80 Hz in the periphery and about 25 Hertz at the fovea. Other problems are to code intensity, colour, and temporal variations. The solution is to use three types of cones (L='Red', M='Green', S='Blue') with antagonistic, circular receptive fields (center/surround, +/- and -/+) and both transient and sustained temporal responses to encode spatio-temporal differences from local means. Analog voltage variations are not suitable for long distance transmission over neural fibers because of rapid attenuation. The physiological solution is to use action potentials - short, two level (on/off) pulses of a fraction of a millisecond in duration and maximum pulse firing rate of about 800 per second - constrained by membrane diffusion time constants. The pulse rate is modulated to represent the amplitude of the peripheral stimulus. The analog to pulse encoding conversion is done by the ganglion cells which act as signal drivers for the optic nerve fibers (which preserve retinal location information for higher level mapping registration). There are many classes of ganglion cells in the retina, 13 in the cat for example (Vaney and Hughes, 1990), but a few types tend to dominate in a particular animal. Individual optic nerve fibers code location in the retina and the number of fibers limits the 2D spatial bandwidth of the system. The ganglion cell firing rate encodes the instantaneous light amplitude at a location and the maximum firing rate limits the range of amplitude coding and temporal bandwidth. There is a lower limit to fiber size because the pulse transmission velocity is inversely proportional to the size, so small fibers would give very sluggish response. There is an upper limit to the number of fibers because of the limited volume of the head. Of course, one might ask why the visual cortex not at the front of the head, instead of the back, to reduce transmission distance.

Retinal cells. A very large number (10^9-10^{10}) of interconnecting cells (hundreds of types) define ganglion cell dendrite (receptive) fields. It should be noted that light must pass through this (transparent) maze of interconnecting fibers and cells before reaching photoreceptors. Even the photoreceptors are 'backwards' - with the transduction region behind the light-insensitive part. This may seem strange from an engineering point of view. Why is this arrangement biologically useful? One possibility is the need for efficient cooling of photoreceptors (Robson, 1998). The image of the sun frequently falls somewhere on the retina and most of this light is absorbed by photoreceptors. Cooling is provided by blood flow and blood vessels are not transparent to light yet must be close to photoreceptors. A good design would be to have blood vessels at the back of the retina and photoreceptors immediately in front of them. The logical place for processing cells is then in front of the photoreceptors.

The following summary of human retinal anatomy is from Wandell (1995). Highest acuity is in the rod and capillary-free foveola (diameter = 0.3 mm, 1°) which has only 50,000 ganglion cells, each with a single cone contributing to the receptive field core and many contributing to the surround. Cone density drops very rapidly in the first 10° relative to the fovea (diameter = 1.5 mm, 5°) to a constant value over the rest of the retina. Cone diameter increases (1.5 ->10 μm) away from the fovea. Outside the fovea there is more than one cone in the central part of a receptive field. Rod density increases rapidly outside the fovea to a maximum at 20° and then decreases slowly. Hundreds of rods contribute to a receptive field. There are two main classes of ganglion cells in the human retina: midget cells (70%) and parasol cells (10%) that seem to transmit complete and independent representations of the visual scene to the cortex. These are

referred to as the parvocellular (P) pathway (from midget cells) and the magnocellular (M) pathway (from parasol cells). The two pathways differ significantly in the manner that they encode information. Midget dendritic field sizes increase smoothly from 5 μm in the fovea to 100 μm in the periphery and have a maximum spatial frequency response of 60 cycles/degree. Parasol dendritic field sizes increase from 30 μm near the fovea to 400 μm with best response of 20 cycles/degree. Investigation of the two pathways is still in its early days. The M pathway seems to provide better response to low spatial frequency, high temporal frequency variations (e.g. motion perception). The two pathways also have different contrast-response curves (fig. 9).

4. RETINAL DESIGN

Summary
a: Select acuity variation over the visual field to fit survival needs.
b: Optimize sampling and coding for day or night operation; based on consideration of scene spatio-temporal statistics and three noise sources (photon, thermal and neural).

Biological Variation in Acuity. Insects and lower animals have smaller data handling problems because of slower resolution needs. Their visual acuity varies little over the visual field. In addition much more analysis and decision capability is built into early stages of their visual systems. Higher level animals tend to have must of their analysis done at cortical levels. One strategy used in such animals to reduce optic nerve transmission and cortical data handling requirements is to vary spatial resolution over the visual field. There are many patterns of acuity inhomogeneity in nature with the selection apparently based on survival tasks (Smythe, 1975). For example (figure 7), birds living in a forest habitat have a single fovea; birds of prey have two fovea connected by a narrow high resolution band; birds that live at sea and in the open plains have broad horizontal fovea that extend across the entire retina. Similar acuity variation is found in mammals. For example, plains animals such as antelopes and cheetahs have highest angular resolution along their visual horizon.

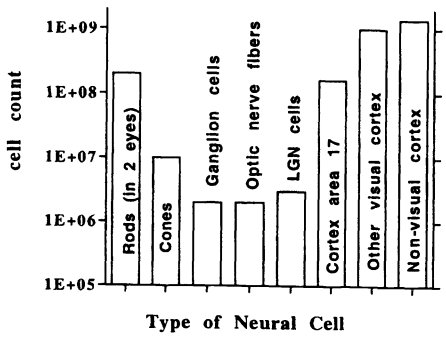

Fig. 6. ß Numbers of different types of cells at various stages (including both eyes).
Data from Wandell (1995)

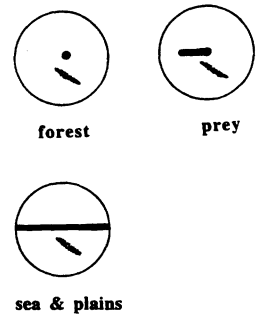

Fig. 7. Three bird fovea examples The diagonal, feather-like structure is the pecten (highly vascular).

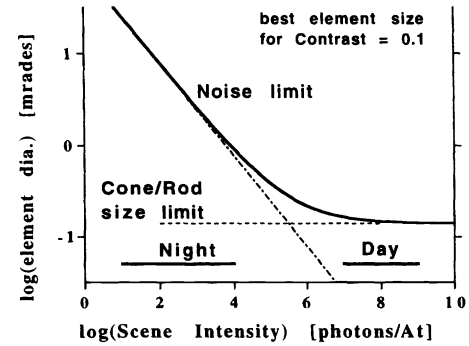

Fig. 8. Variation in optimum size for retinal data acquisition elements with luminance.
Data from Snyder et al.(1977).

Data Acquisition Element Size. At low light levels it is advantageous to connect one ganglion cell to many photoreceptors to create one "data acquisition element". The optimum element size depends of scene intensity and contrast. The following is a summary of Snyder et al. (1977). Their information theory-based analysis was done for a hexagonal lattice of non-overlapping idealized photon acquisition elements that tile the visual field. The quantity to be optimized is the spatial information capacity of the eye [based the maximum number of different pictures that can be constructed for the entire field]. Each element is assumed to be a linear photon counter with a fixed integration time of 0.03 seconds. At high light levels the best element size equals the photoreceptor size. At low light levels, because of photon fluctuations, the

optimum strategy is to have large elements that pool the outputs of individual photoreceptors. The best element size also depends of scene contrast. An example plot for 10% contrast is shown in figure 8 with average night and day scene intensities indicated by the horizontal bars. The human cone threshold corresponds to log(relative intensity) of 3 and rod saturation is at 7. The basic message is that an optimum data acquisition system would be adaptive.

Data Encoding #1. (Correlations). Image element intensity values are usually not the most economical transmission coding method - for three reasons. (1) Scene amplitude histograms are very broad and the data dynamic range can be reduced by sending differences. (2) There is usually considerable correlation between neighboring sample values in space and in time and therefore data redundancy. (3) Neural transmission channels are usually degraded by various types of noise, so additional encoding before transmission would ensure that the effects of the anticipated noise are minimized. Different strategies are needed by day and night because of the marked difference in photon flux rate.

By Day. Figure 8 results show that photon fluctuations are not a problem in daylight and so the optimum data acquisition element size is one cone. Human foveal cones have angular diameter of about 0.01° (0.2 milliradians). One popular view is that nerve fibers have a very limited dynamic range and cells are intrinsically noisy. If this view is true, then cells should encode output to minimize the range of values and render fine detail detectable in noise injected by other cells in the processing mosaic. Srinivasan, Laughlin and Dubs (1982) suggest that the retinal strategy is linear predictive coding based on a combination of scene statistics (in both space and time) and neural noise levels. The cell sends the difference between the actual signal found at the central photoreceptor and the local mean intensity. The profile of the receptive field is derived from statistical estimation theory. The definition of 'local' depends on the correlation properties of the scene and the neural SNR. The local region size should increase as the width of the scene autocorrelation function increases. If the neural SNR is high then the local region can be small. If the neural SNR is small then the region ought to be rather large. This strategy has two benefits: (1) It devotes the entire dynamic range to encode a small intensity range and (2) it reduces data redundancy. Srinivasan et al. gave a very nice demonstration that the actual receptive fields used by some insects correspond very well with the very best choice of receptive field design based on these engineering considerations. There is evidence that the fly eye adaptively adjusts its encoding to local visual scene properties (Laughlin, 1990).

By night. As was seen in Figure 8, at night the limiting factor is low photon flux and the resulting input noise. Scene statistics are secondary. So the optimum strategy is to combine rod outputs to increase the size of a "data acquisition" element. Human rod receptive fields cover a range of diameters from 0.1° to 0.8° (2 to 14 milliradians) which nicely matches the element size range suggested in figure 8 for night vision. The final limitation to sensitivity is the intrinsic neural fluctuation noise in the retina. Here again, the cat's level is significantly lower than that of the human. In addition, the cat has a factor of 2.6 advantage in light collection because of a lower $f^{\#}$ and a tapetum, so a cat's intrinsic sensitivity to light at night is about 10 times as high as that of a human. One question that has intrigued scientists for a long time is whether we can see single photons at night. The evidence for this will be rather indirect. Data recorded from cat ganglion cells in response to brief flashes of light suggest that they are responsive to single photons. Attempts have been made to determine whether the same is true for humans. The results are still controversial, however there is a suggestion that we in fact need two photon coincidence (within 200 msec) on the retina to provide a perceptual effect.

Data Encoding #2 (Amplitude coding). Once again, there are different optimum strategies for day and night because of the different statistical considerations. The amplitudes that are coded represent differences in luminance (from local mean) normalized by local mean. Again, the principles of data communication tell us that the best transformation is that which gives a uniform probability of firing over the scene dynamic range. During the day, the optimum transformation is based on the cumulative probability distribution of scene contrasts - which is similar to "histogram equalization". This is illustrated in figure 9 with data from the fly (Laughlin, 1983). The dots are measured L-cell response as a function of contrast. The dashed curve is the transfer function of the typical chemical synapse. The solid curve is the cumulative probability

of encountering a particular contrast level in a random selection of natural scenes that were obtained using a scanner with similar spectral and angular sensitivity to the fly's eye. Scene measurements suggest that static scene contrast histograms are bi-exponential and spatial contrast variation is scale invariant with a radial spatial frequency power spectrum $S(f) \approx Af^{-2}$ (Ruderman and Bialek, 1994 for example). This hypothesis was also tested by Srinivasan et al. in insect vision and they found a very nice agreement between the theory and biology. A daylight amplitude coding example for a primate is shown in figure 10. The data (from Wandell) are contrast response curves for magnocellular and parvocellular neurons on the lateral geniculate nucleus (LGN) - which acts as a "relay station" between the retina and the visual cortex. The curves through the data are power law functions. At night the main concern is optimum absolute amplitude (rather than contrast) quantization given high photon noise. As was mentioned above, the best strategy is a square-root transformation to allow quantization steps of equal statistical significance. There is evidence for such a result in the cat ganglion cell for small diameter stimuli but the situation is more complex for large stimuli (Barlow and Levick, 1976).

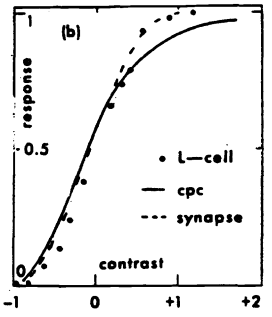

Fig. 9. Response versus contrast for the fly's eye L-cells (dots), synapse transfer function (dashes) and the cumulative probability curve for scenes. From Laughlin (1983) with permission.

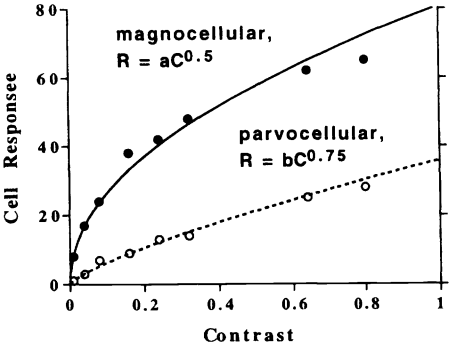

Fig 10. Contrast responses of primate magno -cellular and parvocellar pathways (at LGN). Data of Shapley (1990). From Wandell (p 126) with permission.

Defects. If retinal transduction and coding is done properly, we ought not to be aware of its existence under ordinary circumstances. This is similar to the situation of long distance telephone calls. The telephone handset (imperfectly) converts sound pressure variations into analog electrical signals for short distance transmission. The telephone company uses a variety of encoding strategies for long distance transmission over microwave, fiberoptic, and satellite relay links. If the job is done well we should not be aware of the defects and the limitations imposed by technology.

5. COLOUR

To this point we have been considering transduction and coding of monochromatic light signals. Now we shall turn to the world of colour which involves only cone transduction (not rods). How do cones send colour information to the brain? Historically there have been two theories. The first, known as the trichromatic theory, was proposed by Young in 1800, forgotten, and then revived by Helmholtz in the 1850's. According to this theory there are three visual receptor types that are sensitive respectively to red (R), green (G) and blue (B) light. This is what is found physiologically. The spectral sensitivity curves (fig. 11) are broad and peak wavelengths are: blue (S-cones) at 440 nm, green (M-cones) at 540 nm, and red (L-cones) at 580 nm. This theory agrees with much experimental data. One might ask why we use three colors rather than four or five to represent our colored world? A possible answer is that spectral sensitivity curves are broad relative to range of available wavelengths of light. Applying the Nyquist sampling theorem, Barlow, (1982) showed that three samples completely describe all the colour discriminations that we could possibly make. Some birds have oil droplets at the top of their cones which markedly reduce spectral widths. The birds are able to make use of four or five colour samples and presumably can make superior colour discriminations.

The trichromatic theory had deficiencies and a rival theory (also based on 3 signals) was proposed by Hering in 1878 - now known as the opponent-processing theory. One signal describes luminance variation and the other two are based on four primary colors arranged as opposing pairs using (R-G) and (B-Y) differences. The Hering theory was not met with a high degree of enthusiasm and lay dormant for about 60 years, but through brilliant experimental work it is now recognized that the theory has a great deal to offer. One might then ask which theory is correct. The answer is that both theories appear to be correct. The R, G, and B signals are transformed to the opponent coding scheme at the ganglion cell level for very good information theory reasons. There is a large degree of overlap in the sensitivity curves and therefore redundancy in the output signals. The redundancy in the photoreceptor colour can be removed by a transformation to another coordinate system in which the components are uncorrelated (orthogonal). The B-Y, R-G, and luminance outputs are shown in figure 12. Buchsbaum and Gottschalk (1984) demonstrated that this is, in fact, an optimum (minimum redundancy) coding strategy. Their mathematical predictions agree rather well with the human experimental data.

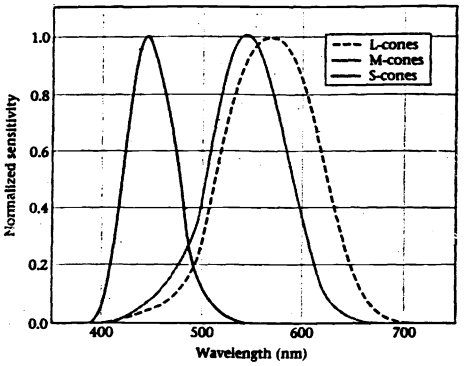

Fig. 11. Absorption spectra for human cones: S (Blue), M (Green), and L (Red).
From Wandell, p48 (with permission)

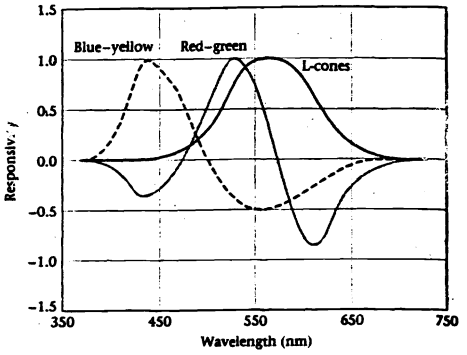

Fig. 12. Spectral responsivity of a set of decorrelated colour sensors under mean daylight conditions.
From Wandell, p326 (with permission)

6. THE BRAIN

Visual input data is transmitted to area 17 of the visual cortex. The functions and detailed cellular anatomy of this area has been beautifully mapped by Hubel, Weisel and others over the last 40 years. This work suggests that the visual cortex is constructing a multidimensional feature space that gives highly redundant representations of visual data from both eyes. The human visual cortex, by the way, has an total area of about 1/3 of a credit card. Output from area 17 goes to cortical higher levels. There is also a large amount of feedback - with mapping of higher area cells back to lower area cells. The physiological wiring and the function of these cells and mappings is not yet well understood. However it would appear that at higher levels there is a synthesis of input visual data with a prior knowledge stored in cortical memory and then interpretation of the scene. It is clear from present work that operation on visual data occupies a very significant fraction of the brain. A map of the cortical sheet shows that visual function occupies nearly 50% of the human cortex (Barlow, 1981). There are about 10^{10} neurons in the cortex, 10^{18} synaptic connections and the capability of doing about 10^{16} computations per second! (Watson, 1997)

7. OVERALL EFFICIENCY

Between 1940 and 1960, a number of people set out to determine the absolute performance of the human eye at low light levels. By this, they meant the determination of the fraction of photons entering the eye that were actually used in visual signal detection tasks. They found efficiencies in the order of 5%. In the years following 1950, there were a number of physiological determinations of photon losses in the eye. These measurements suggested an efficiency between the cornea and the optic nerve on the order of 15%. Therefore one might ask what is the source of the disagreement? Actually, they not determining just the efficiency of the components of the eye up to the optic nerve; they were measuring the efficiency of the entire visual system from the cornea to the highest decision centers of the brain. Barlow (1977) measured the efficiency of the central decision making processes using easily visible random dots displayed for a

short time on a CRT and found efficiencies in the range from 25% to 50% without any significant suggestion of increased sensitivity to particular spatial patterns.

Since 1980 there have been a number of investigations of statistical decision efficiency for a variety of tasks using noisy gray scale images. The results can be summarized as follows (Burgess, 1990). The efficiencies seem to range from 10 to 75% when display conditions are optimized. The experimental results suggest that humans act as suboptimal Bayesian decision makers - selecting the most likely decision alternative given the available prior information together with the new image data. However we suffer from a variety of shortcomings. These include transduction loses due to the receptive fields and encoding schemes at the retina. This means that humans cannot do detection and discrimination tasks with accuracy limited solely by photon fluctuations. There is a limit to the range in space and time over which we can appreciate correlations (i.e. integrate). Another shortcoming is the existence of internal noise, which can arise from any source of non-reproducible behaviour. This includes physical sources of noise such as neural fluctuations; psychological variations due to variations in attention, motivation; cognitive aspects such as inexact application of prior knowledge to the data collection task (for example imprecise and variable comparisons between the expected signals and the existing signals) and finally - variability in applying decision criteria. In spite of the presence of these limitations, it should be observed that human visual signal detection efficiency is rather high. One need only contemplate the problem of designing a multistage system operating at a total efficiency of 50% to realize that the performance efficiency at each stage in the system must be very good. For example, if one were to construct a 10-stage system to operate at an overall efficiency of 50% then each stage in the system would have to operate at about 93% efficiency. Therefore one ought to be very impressed with the performance of the visual system.

Since we started with a quote from Helmholtz, it seems only fitting to give him the last word. It is evident that he took a Bayesian inference view of perception. The language in the following quotation is archaic, but it is clear that Helmholtz recommends that we think of our perceptions as mental representations of the object most likely to explain the sensory input [Wandell, p282]. The Helmholtz quote is: "The general rule determining the ideas of vision that are formed whenever an impression is made on the eye, is that *such objects are always imagined as being present in the field of vision as would have to be there in order to produce the same impression on the nervous mechanism.*" [Italics in the original]

This brief presentation about visual systems is just one part of the marvelous story of the whys and wherefores of comparative physiology. For those who want to pursue the topic much further I can highly recommend books by Denny, Wandell and Withers - there, of course, many others that can be read with delight. One last personal note is that if I had my life to live over, given what I now know, certainly one decision would have been different - I would have become a physiologist.

8. ACKNOWLEDGMENTS

I would like to thank Simon Laughlin and John Robson (physiologists) for their many constructive suggestions on a draft of this manuscript and their patience with the gross oversimplifications of a mere physicist. John had a particularly nice observation - he called the approach used in this paper "inverse optimization". Given an optimum solution - determine the problem and the constraints.

9. REFERENCES

DL Ruderman and W Bialek. Statistics of Natural Images: Scaling in the woods. Phys. Rev. Letters. 73, 814-817 (1994)

HB Barlow. The Efficiency of Detecting Changes in Density of Random Dot Patterns. Vision Res. 18, 637-650 (1978)

HB Barlow. Critical Limiting Factors in the Design of the Eye and Visual Cortex. Ferrier Lecture. Proc.Roy.Soc.B212, 1-34 (1981)

HB Barlow. What causes trichromaticity? A theoretical analysis using comb-filtered spectra. Vision Res. 22, 635-43 (1982)

DA Baylor. Photoreceptor Signals and Vision - Proctor Lecture.

Invest. Opthal. and Visual Sci. 28, 34-49 (1987).
AE Burgess, Effect of Quantization Noise on Visual Signal Detection in Noisy Images. J. Opt. Soc. Am. A2, 1424-1428 (1985).
AE Burgess, High Level Visual Decision Efficiencies. in Vision: Coding and Efficiency, C. Blakemore (edit.), Cambridge Univ. Press (1990)
G.Buchsbaum and A.Gottschalk. Trichromacy, Opponent Colour Coding and Optimum Colour Information Transmission in the Retina. Proc.Roy.Soc. (London) B220, 89-113 (1983)
HJA Dartnell. The interpretation of Spectral Sensitivity Curves. Brit. Med. Bull. 9, 24-30 (1953)
MW Denny. Air and Water: The Biology and Physics of Life's Media. Princeton Univ. Press, Princeton NJ (1993)
WT Ham, HA Mueller, JJ Ruffolo, D Guerry III, and RK Guerry. Action Spectrum for Retinal Injury from Near-ultraviolet Radiation in the Aphakic Monkey. Am. J. Opthal. 93, 299-306 (1982).
H. von Helmholtz.The Recent Progress of the Theory of Vision. (1868) Republished in: Popular Lectures by Hermann von Helmholtz. Dover, NY (1962)
K Kirschfeld. The Resolution of Lens and Compound Eyes. In: Neural Principles in Vision. Edit.R.Weiler, Springer, Berlin (1976)
MF Land. Optics and Vision in Invertebrates. In: Handbook of Sensory Physiology, Vol.VII/6B. Edit.H.Autrum. Springer-Verlag, Berlin (1981)
SB Laughlin, Matching Coding to Scenes to Enhance Efficiency. In Physical and Biological Processing of Images, OJ Braddick and AC Sleigh (edit.), Springer-Verlag Berlin (1983)
SB Laughlin, Coding Efficiency and Visual Processing. In Vision: Coding and Efficiency, C. Blakemore (edit.), Cambridge Univ. Press (1990)
JN Lythgoe; Ecology of Vision, Clarendon Press, Oxford (1979)
PA McNaughton. The light response of photoreceptors. In Vision: Coding and Efficiency, C. Blakemore (edit.), Cambridge Univ. Press (1990)
J Robson, personal communication (1998)
JL Schnapf and DA Baylor. How Photoreceptor Cells Respond to Light. Scientific American 256, 40-47 (April, 1987)
R. Shapley. Visual sensitivity and parallel retinocortical channels. Ann. Rev. Psy. 41,635-658((1990)
RH Smythe. Vision in the Animal World. St. Martin's Press (1975)
AW Snyder, Photoreceptor optics - theoretical principles. In: Photoreceptor Optics. AW Snyder and R Menzel (edits) Springer. NY (1975)
AW Snyder, SB Laughlin and DG Stavenga. Information Capacity of Eyes. Vision Res. 17, 1163-1175(1977)
AW Snyder, TJ Boosmaier and A Hughes. The Theory of Comparative Eye Design. In Vision: Coding and Efficiency, C. Blakemore (edit.), Cambridge Univ. Press (1990)
MV Srinivasan, SB Laughlin and A Dubs. Predictive Coding: A Fresh View of Inhibition in the Retina. Proc.Roy.Soc.B216, 427-459 (1982)
A.J.Van Doorn, W.A.Van de Grind and J.J.Koenderink, (Editors). Limits in Perception, VNU Science Press, Utrecht (1984)
BA Wandell. Foundations of Vision. Sinauer Associates Ltd. Sunderland, MA (1995)
A Watson. Why can't a computer be more like a brain. Science 277, 1934-1936 (1997)
PC Withers. Comparative Animal Physiology. Saunders College Pub., NY (1992)