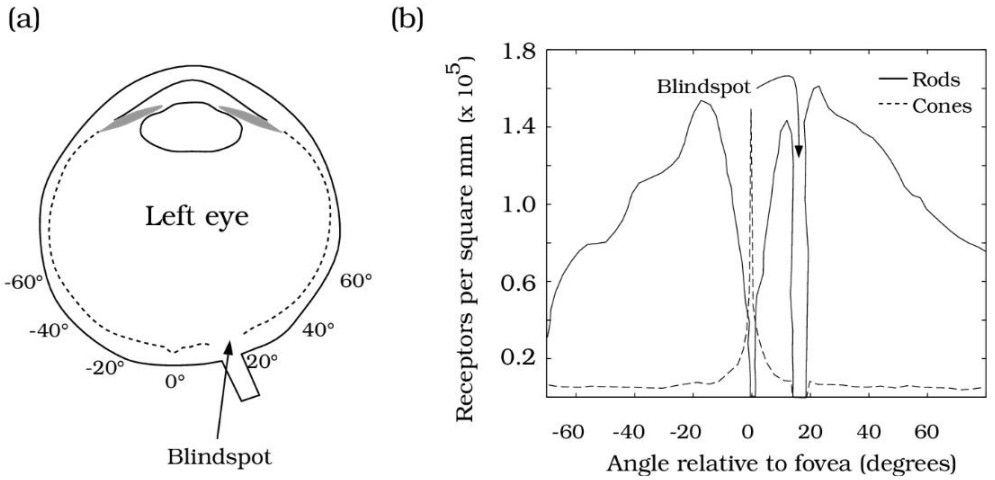
Retina

We didn't quite finish up the discussion of photoreceptors last lecture, so let's do that now.

Let's consider why we see better in the direction in which "we are looking" than we do in the periphery. There are a few reasons for this. One of the main reasons is that the density of cones is so much greater in the center of the field of view. More pixels per degree of visual angle means more image information. Cone density falls off quickly to about 10 degree away from the center and byond that it remains roughly constant.

In the periphery, the density of rods is much higher than the cones, and indeed the density of rods in the periphery is comparable to the density of cones in the fovea. Does this mean that we can see as well in the periphery at night as we can in the fovea during the day? Obviously that can't be, and the reason is that the rods are much less reliable sensors. They are noisy since they operate under low lighting conditions where the image "signal" is relatively small and so any noise has a more significant effect.



The cell densities are plotted in cells per mm^2. You should be able to say what a mm^2 corresponds to (roughly). See Exercises. The exercises also discuss the blindspot shown in the figure.

Ok, we are done with the photoreceptors for now. Let's consider other cells in the retina. The retina consists of several layers of cells. The first layer contains the photoreceptor cells, and is followed by three layers which perform computations to encode the image. The cells in these initial four layers have continuous responses. See slides.

The cells in the fifth layer are called the retinal *ganglion cells*. They are quite different from the other cells in the retina since they need to transmit their response to the brain. They do so by sending spikes. I discussed spikes in the introductory lecture (0). Let me return to them briefly now and repeat some of the points I made back then.

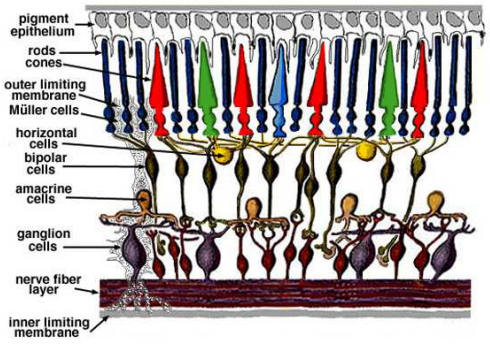


Fig. 2. Simple diagram of the organization of the retina.

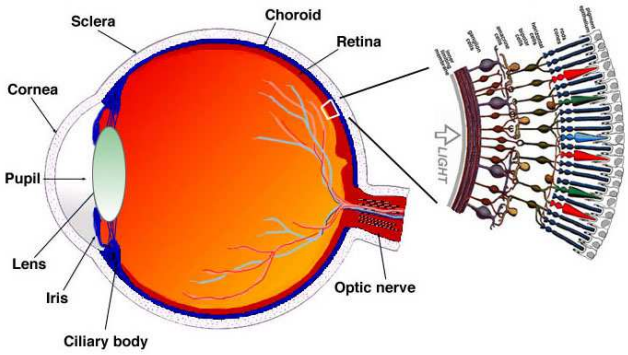


Fig. 1.1. A drawing of a section through the human eye with a schematic enlargement of the retina.

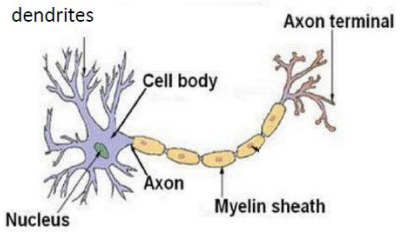
Responses of a neuron: continuous vs. discrete

Cells in the retina respond to images and they encode (in some sense) what is in the image. But what does it means to say that a neuron in the retina has some response to an image? Let's distinguish to aspects of "response". The first is what an experimenter can measure from a single neuron, namely an electical potential difference across the neuron's cell membrane (more on that below). The second is what a neuron communicates to a neighboring neuron, namely it releases hormones (neurotransmitters) that are picked up by the neighbor; these neurotransmitters in turn affect the responses of neighboring neurons.

Here is a bit more detail. First, the figure below shows the basic structure of nerve cell (or neuron) such as a retinal ganglion cell. It has a cell body, and the cell body has branches coming out of it. These branches are called *dendrites*. When neurotransmitters are released from neighboring cells in the retina, these neurotransmitters may bind to the dendrites, which causes a change in the electrical potential across the neuron's cell membrane. As in the case of photoreceptors, the change is that membrane channels will open or close, allowing ions such as potassium and sodium to travel in or out of the cell. The net effect is that the concentration of ions inside versus outside the cell will vary over time, and thus there may be a difference in electrical potential across the cell membrane. This is what an experimenter typically measures, when studying the state and response or a single neuron.

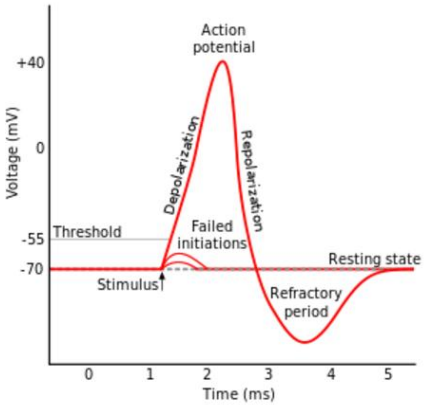
The resting (average) potential of a cell is typically about -70 mV (millivolts), namely the inside of the cell has more negative charge than the outside. If the potential difference is above -70 mV, then we say that the cell is *depolarized* (i.e. closer to 0), and if the potential differences is below -70 mV then we say that the cell is *hyperpolarized* (i.e. further from 0).

The communication between cells occurs at a location called a *synapse*. The cell releasing the neurotransmitter is the pre-synaptic cell and the cell receiving (binding) the neurotransmitter is the post-synaptic cell. These neurotransmitters can be either depolarizing or hyperpolarizing. As we will see next, a depolarizing neurotransmitter is *excitatory* and a hyperpolarizing neurotransmitter is *inhibitory*.



Spikes (Action Potentials)

How do cells communicate over long distances? For example, how does a retinal ganglion cell communicate its response to the rest of the brain? The basic mechanism for long distance signally is called a *spike* or *action potential*, which is sudden and large depolarization of the cell membrane. See figure below.



An action potential is triggered when the cell membrane reaches a certain depolarization threshold, which causes it to depolarize further and even become positive. The action potential is propagated over a distance as a single wave (spike) along a special part of the cell called the cell *axon*. Think of an axon as a long wire. (In fact it is tube wrapped in a fatty insulator.) A cell typically can "spike" up a rate of up to 200 times per second.

There is much to say about spikes but let's just consider a few important facts for now. First, for a given cell, every spike has the same shape. (See sketch above.). The information carried by spikes is purely in the timing of the spikes, not the shape. There has been much effort in the past few decades to understand exactly how much the timing matters. On the one hand, the initiation of a spike depends on a somewhat noisy signal (namely, binding of neurotransmitters from neighboring cells) and so it is difficult to imagine how the exact timing could be reproducable and hence reliable. On the other hand, some computations do require precise timing, as we'll see later in teh course when we study the auditory system.

Receptive Field of Retina Ganglion Cell

Cells in the visual system (at least those in the early processing stages) typically respond to images in a restricted region. For a given cell, we refer to the set of visual directions that the cell is sensitive to as its *receptive field*. Photoreceptors obviously have a very small receptive field, since they pretty much only respond when light strikes them directly. For other cells in the retina, the response of one cell will affect another neighboring cell. The result is that cells can respond over a wider range of visual directions, by being indirectly affect by responses from other cells.

Just like the density of rods and cones varies over the retina, there is variation in the retinal ganglion cells across the retina. In particular, the sizes of retinal ganglion cells is smallest in the fovea and increases in the periphery. (See plot in slides.) This increase in receptive field size roughly follows the decrease in density in the cones.

What information do the spikes from each retinal ganglion cell encode about the retinal image? The ganglion cells do not simply encode a pixel by pixel copy of the LMS photoreceptor image. Rather, they pre-process the image to make some aspects of the image more explicit. Indeed all layers of the retina contribute to this pre-processing. Rather than looking at the detailed circuits in the various layers of the network, let's look at some simple models of the what image transformations are being computed.

The simplest model is that the visual system encodes sums and differences of LMS (cone) response values in local neighborhoods. That is, after the LMS cones measure the light arriving at the retina at each location (x, y), subsequent layers of cells in the retina compute weighted sums and differences of the LMS responses. We'll look at a few types of these sums and differences: spectral, spatial, temporal, and combinations of all these. Today we'll just discuss spectral and spatial.

Spectral sums and differences, and color opponency

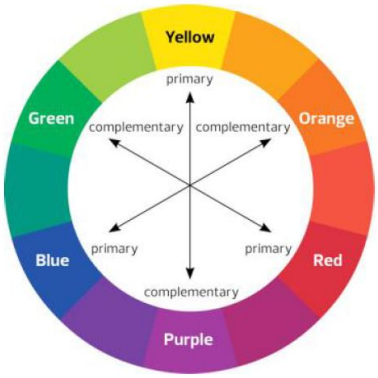
From many experiments over the years, neuroscientists have learned that the retina transforms the LMS measurement as follows: "L + M" measures the overall physical brightness in the medium and long wavelengths, "L-M" compares the long and medium wavelength response, and "L + M - S" compares the medium/long overall intensity to the short wavelength intensity. These arithmetic expressions should not be taken literally right now, since I haven't defined what exactly L, M, S mean here in terms of numerical values. (e.g. What are the units?) Rather, for now, just think of them symbolically: e.g. L+M-S means that there are cells who response increases when the image in the receptive field of the cell has stronger L or M components and the response decreases when the cell's receptive field has a stronger S component.

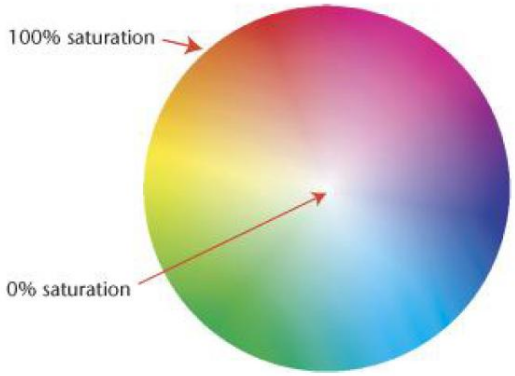
Measuring *differences* in cell responses is called *opponency*. L - M is called red-green opponency. L + M - S is called yellow-blue opponency. The reason $L + M$ is called "yellow" is that if you mix together two lights that appear red and green, then you get a light that appears yellow. e.g. an image pixel with RGB value (1,1,0) appears yellow.

Color opponency is a very old idea and can be expressed in many ways. For example, in school you may have learned about primary colors and secondary colors and how to use them. (See ASIDEs below.) In vision science, the idea of opponency goes back to Hering in the late 1800's. One of the key observations is that some colors seem to be in-between other colors, e.g. we perceive orange as reddish yellow, as if both red and yellow are both *in* orange. Similarly, we perceive cyan as blueish green, and we perceive purple as reddish blue. However, we cannot perceive a color to

be blueish yellow, or reddish green. These pairs of colors oppose each other in some fundamental sense. These observations are believed to be the direct perceptual consequency of an underlying opponency circuitry, namely computing the LMS differences mentioned above.

If you took art class in school, then you are familiar with the idea of color opponency already. You learned about primary and secondary colors and how they related to color mixing. You also learned about "complementary" colors" and how colors can be arranged on a wheel (red, orange, yellow, green, blue, purpose) and how there are special relationships between colors that are opposite each other on the wheel. (See below left.) I am not going to attempt to explain color art theory in this course; I just want to mention that there are connections to color opponency.





Hue, saturation, value (HSV)

The color signal in an image is a 3D vector (LMS) and there are many ways to encode these 3D vectors. One of the common ways to distinguish colors from each other is based on the *relative* amounts of the spectrum at different wavelengths versus the *total* amount of light in the spectrum. In LMS theory, the former concerns the two difference channels (L-M, L+M-S) and the latter concerns the L+M channel. If one thinks of a color circle, then the points on the edge of the circle define colors that are as pure as can be, and points in the interior of the circle (see right above) correspond to a mix of pure colors with a neutral color (white or grey). By using a polar coordinate system for points in the circle and its interior, one can sweep out a range of colors. The angle or direction from the center of the circle defines the (maximally) pure color – often called the *hue*. The distance from the center is the purity – often called the *saturation*.

The polar coordinate system accounts for two of the three dimensions of LMS color space. The third color dimension is often called the `value`, or lightness, or luminosity. (These terms all have specific technical definitions in color science, but the details don't concern us.) The specific case of saturation equal to 0 is the center of the color circle. In this case, the values can range from black to grey to white. Think of this third dimension as coming out of the page in the figure above right.

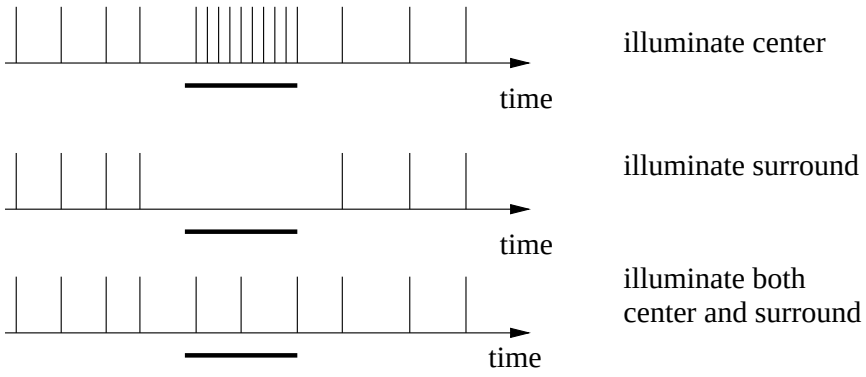
If you have used color pickers in MS paint or Powerpoint to select colors, then you will be familiar with these terms. I encourage you to experiment for a few minutes and see how RGB values gives rise to different HSV (or HSL) codes. At the very least, see the slides for an example.

Spatial sums and differences: Lateral inhibition

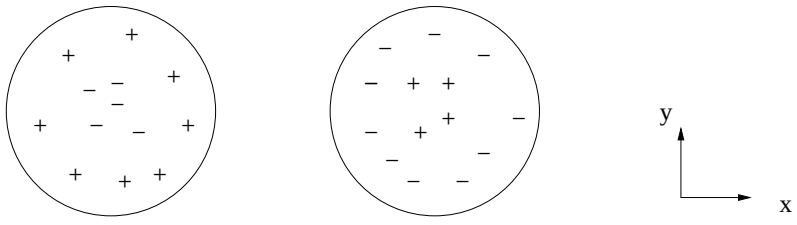
As mentioned earlier today, because interactions in the retina are spatially localized, each ganglion cell can respond to only a spatially restricted region in the retina: the receptive field. Interestingly, a cell does not have a uniform response to light over its receptive field, but rather it computes sums and differences of the light intensity over different parts of the receptive field. They also compute temporal sums and differences, but I won't mention that today.

In 1950s, a researcher named Steve Kuffler for the first time measured spike trains from single ganglion cells of the cat retina. He recorded from single cells over time, while shining a tiny spot of light on the retina. He carried out these experiments in a very dark room, so that the only light shining on the retina was the tiny spot. He found that for each retinal ganglion cell there was small region of the retina that affected the spike rate of that cell, i.e. the receptive field.

Kuffler found many ganglion cells for which the firing rate increased when the tiny spot of light shone on a particular region. This is called *ON region* for that cell. He also found that surrounding this ON region was an annulus (ring) shaped region in which the tiny spot of light *decreased* the firing rate of the cell. This surrounding region is now called the *OFF region*. Because these cells were excited by light in the center and inhibited by light in the surround, these cells are called *ON center/OFF surround*.



Kuffler also found retinal ganglion cells that had the opposite property, namely there was a central round region in which the the cell's response decreased when the tiny spot of light was shone there, and a surrounding annulus region in which the response increased when light was shone there. These cells are called *OFF center/ON surround*.



One can model the cell's response behavior by assigning weights to the different points in the receptive field. The ganglion cell's response is the sum of the weighted intensities over the receptive

fields. For now, we just to think of the L+M channel. In Assignment 1, you will think about difference channels too.

Note that because the intensities in the surround have the opposite effect as the intensities in the center, you can think of the image in the surround as inhibiting the response to the image in the center. This local spatial inhibition (or opponency) is often called *latera inhibition*.

DOG model

One model for achieving the center-surround effect is to suppose that there is one mechanism for excitement over a neighborhood and that the effect falls off with distance from the center of the receptive field, and that there is a different mechanism for inhibition that also falls off with distance. If the excitation were to come from a small neighborhood and be strong in that neighborhood and if the inhibition were to come from a larger neighborhood and be weaker over that neighbhorhood, then this would naturally lead to an ON-center/OFF-surround receptive field.

Rodieck and Stone (1965) proposed a specific model which was based on the 2D Gaussian function:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{1}$$

A 2D Gaussian is just the product of two 1D Gaussians,

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$$

$$G(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2} \ .$$

The 2-D Gaussian is *radially symmetric* in the sense that it only depends on the squared radius x^2+y^2. Note that this Gaussian is centered at $(0,0)$ but more generally it could be centered at any (x_0, y_0) by shifting. Also, note that we are ignoring the time dimension.

The *difference of Gaussian* function is then defined:

$$DOG(x,y,\sigma_1,\sigma_2) \ = \ G_1(x,y,\sigma_1,\sigma_2) \ - \ G_2(x,y,\sigma_1,\sigma_2)$$

and again it is centered at $(0,0)$. Here 1 and 2 are the center and surround, i.e. $\sigma_1 < \sigma_2$. This center would be ON-center and OFF-surround. To obtain OFF-center ON-surround, one would use $\sigma_1 > \sigma_2$.

Finally, the response of a retinal ganglion cell whose receptive field is centered at (x_0, y_0) depends on the inner product of the DOG with the image

$$L(x_0,y_0) \equiv \sum_{x,y} \ DOG(x-x_0, y-y_0) \ I(x,y)$$

where L stands for "linear". However, the response of the cell (e.g. firing rate of a retinal ganglion cell) isn't exactly modelled by L. For example, cells have a maximum firing rate, so if we were to increase the image intensity, then eventually the response would saturate. Also, cell's cannot have negative responses. So if the image were positive only in the "negative part" of the DOG function, then the model would give a negative number for L, which wouldn't make sense as a response. To convert the L into a meaningful response, we would need to set the response to 0 when the L values

are negative. One can model these non-linear mappings from L to a response in several ways, for example, using a sigmoidal shape curve, or by *half-wave rectifying*, namely setting all negative L values to 0. See slides.

A related point is that we need both ON-center OFF-surround cells and an OFF-center ON-surround cells. Depending on the image, an ON-center OFF-surround cells can have an L value that is either positive or negative. In the case it is negative, the cell would have no response and so the information about the image would be lost. Having an OFF-center ON-surround cell at that same location would have a positive L value, namely the negative of the negative value of L of the first cell. So as long as both types of cells are around, no information will be lost. (Of course, we still have the issue of saturating to bright images. The only way to deal with that is adaptation, as discussed last lecture.)

Cross-correlation

To understand retinal processing of images, we want to know not just the response of a single cell to the images, but also also the responses of a family of cells that all have the same receptive field shape. For this, one defines the cross correlation of two functions, in this case DOG and I by:

$$DOG \otimes I(x_0, y_0) \equiv \sum_{x,y} DOG(x, y)\ I(x_0 + x, y_0 + y)\ = \sum_{u,v} DOG(u - x_0, v - y_0)\ I(u, v)$$

where I used a change of variables $x_0 + x = u$ and $y_0 + y = v$.

Think of the DOG as a template, and imagine sliding that template across the image. See slides. The formulas above says the template is at (x_0, y_0). But you should think of (x_0, y_0) one of many positions. So we are thinking of cells at many different positions and we are thinking of the responses of a *population* of cells that all have the same receptive field weighting function, namely a DOG of some particular σ_1 and σ_2 that define a center and surround size.

