1

Rudiments of Human Visual System (HVS) Features

In this chapter, the basic facts about the processing of a visual signal by the retina and brain will be recalled. The purpose of this chapter is not only to provide an exhaustive treatise about the Human Visual System (HVS), but also to introduce some important concepts and formulae that will have a fundamental role in the development of the models described in Chapter 5. For complete details on these topics, see, for example, [FAI 05].

1.1. The retina

In Figure 1.1, a human eye and the cross-section of a retina are represented.

Several layers of neural cells constitute the retina, beginning with around 130 million photoreceptors (rods and cones) and ending with about 1 million ganglion cells. The specific processing that occurs in each type of cell is complex and not yet completely understood.

What we know for certain is that retinal cells may respond nonlinearly to stimuli and are connected via links called *synapses*, which are able to perform basic mathematical operations such as addition, subtraction, multiplication, division, amplification and gain control. Considered as a whole, these operations result in a clever and sophisticated modification of the visual input.

Among all retinal cells, the most important for our purposes are the photoreceptors (rods and cones), to which the next section is devoted.

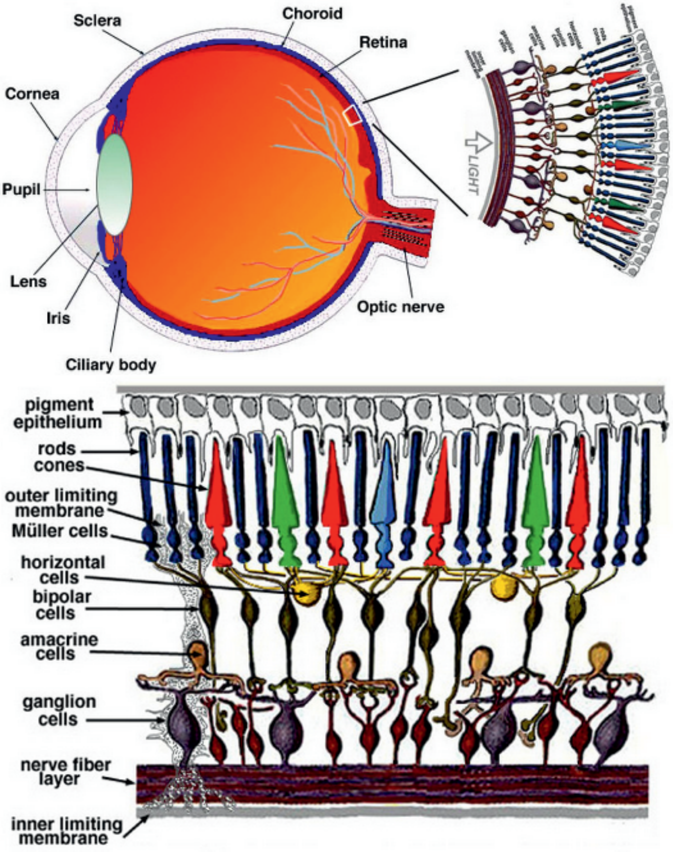

Figure 1.1. *Top: a human eye. Bottom: the cross-section of a human retina. Courtesy of [KOL 95]. For a color version of the figure, see www.iste.co.uk/provenzi/color.zip*

1.1.1. *Photoreceptors: rods and cones*

Rods and cones are labelled in this way because of their shape. Rods work in the so-called *scotopic region*, below 10^{-3} cd/m^2, while cones respond to luminance levels higher than 10 cd/m^2, a range called the *photopic region*. In the intermediate range, called the *mesopic region*, both rods and cones are activated, but their response is less efficient than when they work in isolation from each other. Henceforward, we will only consider photopic conditions and thus the properties of cones.

Color vision in the photopic region is possible, thanks to the existence of three types of cone receptors with peak spectral sensibilities distributed along the visual spectrum (see Figure 1.2). This is due to the existence of three slightly different molecular structures in each cone type, which are referred to as L, M and S cones. They refer to the *long*, *middle* and *short* wavelengths where cones have their maximal sensitivity at 560 nm, 530 nm and 420 nm, respectively.

The LMS cones can also be referred to as the RGB cones. Of course, RGB is the notation for monochromatic *red*, *green* and *blue*, but, as shown in Figure 1.2, this is an abuse of language, in particular because the L cones are gathered in the region of monochromatic green-yellow, not red.

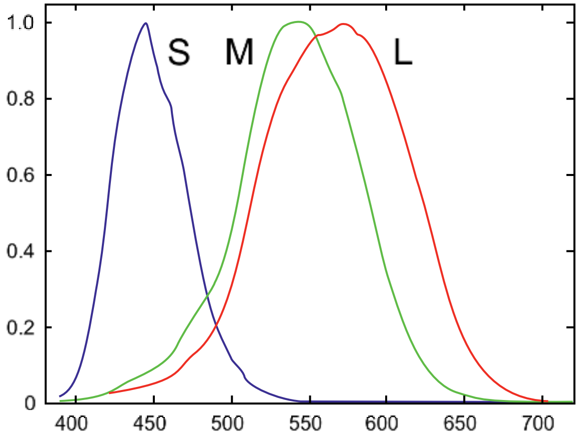

Figure 1.2. *The normalized spectral sensitivity functions of the LMS cones. Courtesy of [KOL 95]. For a color version of the figure, see www.iste.co.uk/provenzi/color.zip*

Note that the spectral sensibilities of the three cone types are broadly overlapping, in particular those of the L and M cones. This constitutes a substantial difference with respect to most physical imaging systems, in particular digital cameras (see, e.g., [JIA 13]), where sensor responses are only slightly overlapping.

Finally, it must be noted that the distribution of cones in the retina is not uniform: S cones are relatively sparse and completely absent in the *fovea*, the central part of the retina with the highest density of L and M cones.

1.2. Adaptation and photo-electrical response of receptors

Light adaptation is the name used to describe the fact that the HVS is able to adapt to different light intensities in order to enable detail perception over a range of 10 orders of magnitude.

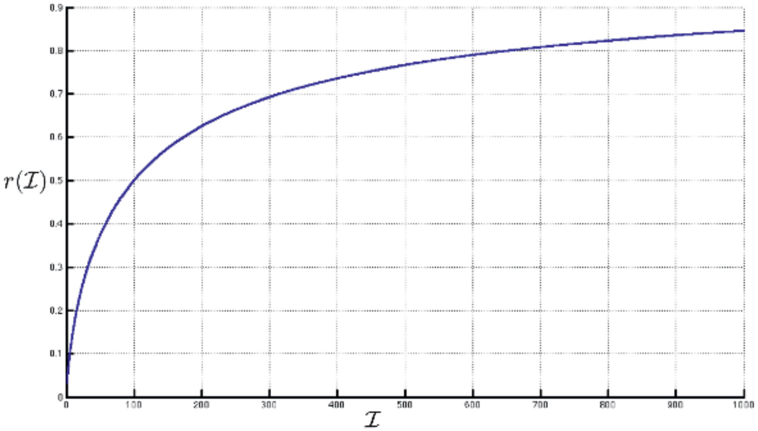

Figure 1.3. *Compressive effect of Michaelis-Menten's response in arbitrary units between 1 and 1000 and with* $\gamma = 0.74$*. The semi-saturation constant* $\mathcal{I}_S$ *has been arbitrarily set to 100 and* $r(\mathcal{I}_S) = 1/2$

Before reaching a photoreceptor, rod or cone, light intensity is reduced by the cornea, crystalline lens, the humors and the macula. Moreover, when a light photon is absorbed by a photoreceptor, a transduction occurs: the electromagnetic energy carried by the photon is passed to the photoreceptor, which changes the electric potential of its membrane. The empirical law that describes the photoreceptor transduction is known as *Michaelis-Menten's equation* [SHA 84] (or *Naka-Rushton's equation* when $\gamma = 1$):

$$r(I) = \frac{\Delta V}{\Delta V_{\max}} = \frac{\mathcal{I}^\gamma}{\mathcal{I}^\gamma + \mathcal{I}_S^\gamma}, \qquad [1.1]$$

where $\Delta V_{\max}$ is the highest difference of potential that the membrane can handle, γ is a constant (measured as 0.74 for the rhesus monkey), $\mathcal{I}$ is light intensity and $\mathcal{I}_S$ is the value at which the photoreceptor response is half maximal, called the *semi-saturation level*. Note that, as previously mentioned, each type of cone is most sensitive over a particular waveband, thus the value of the semi-saturation constant $\mathcal{I}_S$ can change for the three types of cones.

The photo-electrical response of photoreceptors, along with other phenomena occurring mainly in the retina, is considered one of the main explanations for the property of *adaptation to the average luminance level* of the HVS. In fact, after the photoreceptors transduction, the dynamic range is centered in $r(\mathcal{I}_S) = 1/2$, as can be seen in Figure 1.3, which shows the nonlinear compressive behavior of Michaelis-Menten's response. The adaptation property of the HVS is crucial: without it, the operational range of our vision would be much narrower and sight as we know it would be impossible.

1.3. Spatial locality of vision

Transduction curves shown in Figure 1.3 represent the very first stage of visual processing. The electrical signals generated by the photoreceptors are processed by the retinal neurons, synapses and ganglion cells, until they are then finally transmitted to the brain via the optic nerve. In the brain, the visual signal is processed in several zones, each of which is devoted to processing different characteristics, e.g. shape, orientation, spatial frequency, size, color, motion [ZEK 93].

Our present understanding of post-photoreceptors physiological operating principles is far from being precise: not only the brain, but also retinal functions still present some unknown features. Without entering the very complicated analysis of post-photoreceptor physiology, what is important to underline here is that the signals transmitted from the photoreceptors to higher levels of the visual path are not simple point-wise representations, but they consist of sophisticated combinations of receptors responses to photons coming from different parts of the visual scene. In fact, even when we fix a single point, our eyes are constantly moving and capturing light information from all over the visual scene. These movements are called *saccadics*, and they are the fastest of our body.

A conventional nomenclature has been introduced to rigorously define the local neighborhood of a point in a visual field (see, e.g., [HUN 14] and [FAI 05]):

– Stimulus: the visual element corresponding to foveal vision, about 2° of angular extension;

– Proximal field: the closest environment of the stimulus, it extends isotropically for about 2° from its edge;

– Background: extends isotropically for about 10° from the edge of the proximal field;

– Surround: a field that lies outside the background;

– Adapting field: the total environment of the stimulus considered the proximal field, the background and the surround, until the limit of vision in all directions.

The components of the adaptive field are shown in Figure 1.4.

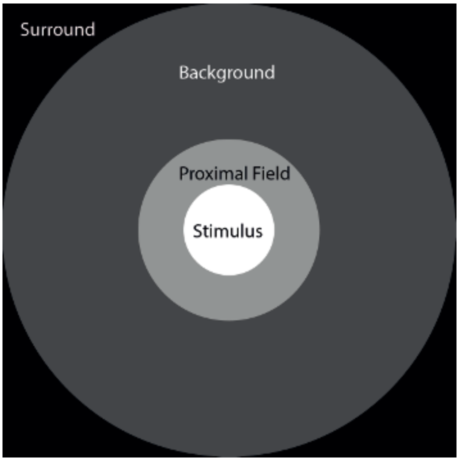

Figure 1.4. *Components of the adaptive field*

1.4. Local contrast enhancement

The eye’s optical system and the response of photoreceptors strongly reduce the range of light intensity that can be processed. To *compensate* this reduction, the HVS has developed a system to enhance contrast perception already in the retina, in particular with the typical lateral inhibition phenomenon exhibited by ganglion cells [CRE 87, CRE 90, HUR 90, ZAI 99, MCC 11].

Studies about ganglion cells revealed that the electric potential of their membrane spontaneously oscillates at a base rate while at rest, an event called *firing of action potential*. *Excitation* of retinal ganglion cells results in an increased firing rate while *inhibition* corresponds to a depressed rate of firing. For this reason, the magnitude of the signal is represented in terms of the number of spikes of voltage per second fired by the cell rather than by the difference of voltage across the cell membrane.

To represent the physiological properties of ganglion cells, it is useful to consider the concept of *receptive field*, which is a *graphical representation of the area in the visual field to which a given cell responds*. The positive or negative response is indicated in the receptive field, as shown in Figure 1.5, which represents a prototypical receptive field for ganglion cells.

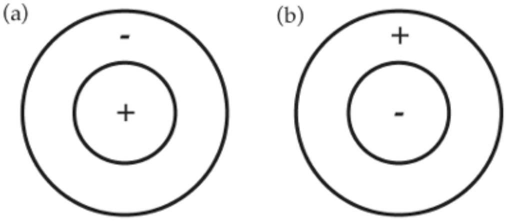

Figure 1.5. *Center-surround receptive fields: (a) on-center; (b) off-center*

Receptive fields illustrate *center-surround antagonism*: the receptive field in Figure 1.5 (a), called *on-center*, describes a positive central response, surrounded by a negative surround response, while Figure 1.5 (b), called *off-center*, shows a ganglion cell response of opposite polarity.

The excitation-inhibition processing can explain some *local contrast enhancement* effects as Mach bands, see Figure 1.6: If we focus on the vertical band on the right the adjacent gray level appears lighter, if we focus on the vertical band on the left the adjacent gray level appears darker, in spite of the fact that the luminance value in each vertical gray band is constant.

Consider Figure 1.7: in situation 1, excitation and inhibition fire signals equally, so a uniform patch is perceived; in situation 2, part of the inhibition component of the receptive field is activated by a region of highest luminance and so it prevails, generating a sensation of darker gray; on the contrary, in situation 3, part of the inhibition component of the receptive field is activated by a region of lowest luminance and so it is dominated by the excitation component, which produces a sensation of lighter gray.

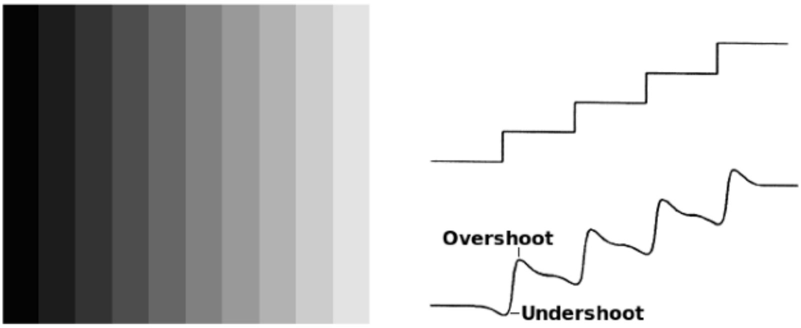

Figure 1.6. Left*: Mach bands effect.* Right*: real (above) and apparent (below) luminance pattern*

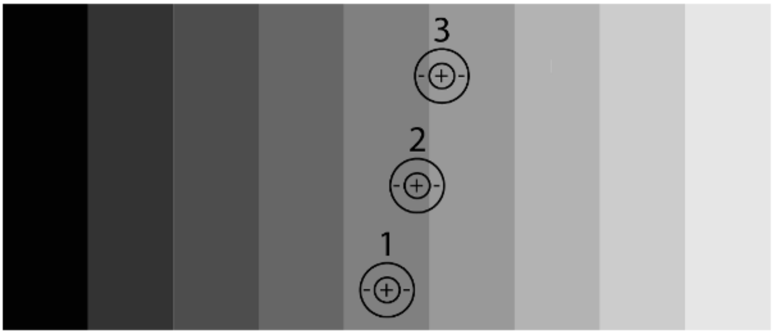

Figure 1.7. *Excitation-inhibition explanation of the Mach band effect*

Another well-known local contrast enhancement effect is simultaneous contrast, first discussed by Chevreul in 1839 in his essay "*La loi du contraste simultané des couleurs*", as shown in Figure 1.8.

Figure 1.8. *Simultaneous contrast: even though the two inner squares have the same luminosity, our perception is different, due to the difference in the left and right background*

The influence of context on the perception of a point decreases with distance, as discussed, for example, in [WAL 48, RUD 04], where the authors developed a clever psychophysical experiment to measure induction of achromatic stimuli. We will discuss this method in Chapter 3.

1.5. Physical vs. perceived light intensity contrast: Weber-Fechner's law

In the previous section, we have seen that the HVS rearranges spatial information in order to amplify the response to edges present in a visual scene. This phenomenon concerns *spatial contrast*. In this section, we recall the HVS response to *light intensity contrast*, without considering the spatial distribution across the scene.

Psychophysics is the science that aims to model the magnitude of human perception in response to external stimuli in a mathematically rigorous way. The German physicist E. Weber, with results obtained in the second half of the 19th Century, was one of the first scientists in history to develop psychophysical experiments to test perception of light intensity contrast.

To avoid unwanted biases, Weber worked in a very constrained setting: a dark-adapted human observer was put in a dim room in front of a white screen on which a narrow beam of light was shone in the center of the visual field. The light intensity $\mathcal{I}$ of the beam was increased very slowly and the observer was asked to tell whether he/she could perceive an intensity change. The least perceptible intensity change $\Delta\mathcal{I}$ is called JND for *Just Noticeable Difference*.

Weber found out that the JND increased proportionally with the light intensity[1], i.e. $\Delta\mathcal{I}/\mathcal{I} = K$ or $\Delta\mathcal{I} = K\mathcal{I}$. K is called *Weber's constant* and the relationship between $\Delta\mathcal{I}$ and $\mathcal{I}$ is called Weber's law.

Weber's law says that, as we increase the background light $\mathcal{I}$, the difference $\Delta\mathcal{I}$ must increase proportionally in order to be able to appreciate $\mathcal{I} + \Delta\mathcal{I}$ as being different from $\mathcal{I}$. This partially explains why dark areas of a visual scene are more sensitive to noise, and thus why it is more important to perform denoising in dark areas of digital images rather than in bright ones. This last consideration is a practical application of a psychophysical phenomenon.

The founder of psychophysics, the German experimental psychologist G. Fechner, gave the following interpretation of Weber's law: he introduced the adimensional quantity $s(I)$ called light *sensation* and stated that the difference of sensation $\Delta s(I)$ is proportional to a slightly modified Weber's ratio, i.e.

$$\Delta s(\mathcal{I}) = k\frac{\Delta\mathcal{I}}{n + \mathcal{I}}, \tag{1.2}$$

where $k > 0$ is a constant and $n > 0$ is a quantity, often interpreted as *internal noise* in the visual mechanism. Fechner assumed that this finite difference equation could be valid for arbitrarily small differences, which of course is not possible because of the very definition of JND. However, following Fechner's hypothesis, the finite-difference equation [1.2] becomes a differential equation [WYS 82]:

$$ds(\mathcal{I}) = k\frac{d\mathcal{I}}{n + \mathcal{I}}. \tag{1.3}$$

By integrating both sides from $\mathcal{I}_0$, the lowest threshold of perceivable light intensity, defined by $s(\mathcal{I}_0) = 0$ and $s(\mathcal{I}_0 + \varepsilon) > 0$ for all $\varepsilon > 0$, to a generic value of $\mathcal{I}$, we obtain

$$\begin{aligned}\int_{\mathcal{I}_0}^{\mathcal{I}} ds(\mathcal{I}) = k\int_{\mathcal{I}_0}^{\mathcal{I}} \frac{d\mathcal{I}}{n + \mathcal{I}} \iff\quad & s(\mathcal{I}) - s(\mathcal{I}_0) \\ &= k[\log(n + \mathcal{I}) - \log(n + \mathcal{I}_0)],\end{aligned} \tag{1.4}$$

---

[1] Weber's law is approximately valid not only for the visual sense, but also for all the other senses, with different values of Weber's constant.

which can be rewritten as follows:

$$s(\mathcal{I}) = k \log\left(\frac{n+\mathcal{I}}{n+\mathcal{I}_0}\right) = s_0 + k\log(n+\mathcal{I}), \qquad [1.5]$$

where $s_0 = -k\log(n+\mathcal{I}_0)$. This last formula is called *Weber-Fechner's law* and it states that *the sensation of light differences*, in the very constrained context of Weber's experiment, *grows as the logarithm of the light intensity*.

We must stress the *limitations of Weber-Fechner's law*:

1) First, it is only valid for very simple visual scenes, such as those considered by Weber in his experiments. As discussed in the previous section, the presence of a non-trivial spatial context introduces significant modifications in visual perception;

2) Second, even for very simple visual scenes, Fechner's assumption about the possibility to maintain the validity of Weber's law passing from finite to infinitesimal light intensity differences is correct only for a limited range of light intensity between the minimum and the maximum perceivable light. As we approach these extreme situations, this assumption fails dramatically due to strong nonlinearities in the visual mechanism and Weber-Fechner's law does not hold anymore.

The last phenomenological feature of human vision that is worth recalling is color constancy; however, its role in image processing is so important that it will be separately discussed in the following chapter.