Spatial vision
Tim Meese

1 Introduction
To the layperson, vision is all about ‘seeing’ – we open our eyes and, effortlessly, we see the world ‘out there’. We can recognize objects and people, we can interact with them in sensible and useful ways and we can successfully navigate our way about the environment. Of course, the only information that we are really able to act upon is that which is encoded as neural activity by our nervous system. These visual codes, the essence of all of our perceptions, are constructed from information arriving from each of two directions. Image data arrive bottom-up from our sensory apparatus and knowledge-based rules and inferences arrive top-down from memory. This chapter focuses primarily on one aspect of the bottom-up process, sometimes called early vision: the processing and encoding of spatial information in the two-dimensional retinal image.

2 From retinal image to neural image
The job of early vision is to encode the retinal image ready for subsequent processing stages where more complex tasks such as object recognition are performed. Clearly, to be able to recognize an object it is important to be able to segment it from its background. As objects usually reflect different amounts of light from their background the identification of luminance boundaries in the retinal image might be a useful starting point. The initial stage of this process is performed very early in the visual pathway, by retinal ganglion cells (see Chapter 10, The retina).

Central to understanding this aspect of retinal processing is the concept of a cell’s receptive field. This is the region on the retina that, when stimulated, can produce a change in the retinal cell’s response. Not surprisingly, neighbouring retinal ganglion cells have neighbouring, though slightly overlapping, receptive fields. Importantly, the receptive fields of retinal ganglion cells have distinct ON and OFF regions like those shown in Figure 1. When light falls onto an ON region the response of the cell increases and when it falls onto an OFF region it decreases. This means that a cell with an ON-centre receptive field would respond most strongly to a bright spot on a dark background and a cell with an OFF-centre receptive field would respond most strongly to a dark spot on a bright background.

However, the construction of these receptive fields means that they respond to other stimuli, particularly luminance boundaries, in interesting and useful ways. Figure 2a (overleaf) is an image of a white square on a black background, and Figure 2b illustrates the set of responses of many ON-centre retinal ganglion cells whose receptive fields cover the image. In this figure, white represents a strong response, black represents no response and other grey levels represent intermediate responses. This representation of neural responses provides a convenient way of thinking about the two-dimensional distribution of responses of visual neurons and it is easy to see why it is sometimes called a neural image.

An interesting feature of the neural image in Figure 2b is that most regions of the image are mid-grey. This response level is the same as each cell’s spontaneous discharge (i.e. that produced in the absence of any stimulation), implying that the
image contains nothing of interest in these regions. The only responses that deviate from these are those close to the boundaries. On the (dark) outside of the square, the responses are very low and on the (light) inside they are high, giving the object boundaries a characteristic signature. Exactly why boundaries produce this signature is described in Box 1.

The neural image for OFF-centre retinal ganglion cells would look very much like that in Figure 2b, but the dark regions would be light and vice versa. In other words, the neural images for the two classes of cell would be of opposite polarity. Conceptually, these two images can be thought of as superimposed on top of one another. This is a clever feature of the retina because it means that image features are encoded by both increases and decreases in neural response effectively doubling the number of distinct response levels that can be usefully employed in the visual code.

3 Spatial frequency selectivity

Most images are much more complex than that in Figure 2 and consequently, the visual system employs an extra level of complexity to deal with it. The problem is illustrated by the cartoon of the cat in Figure 3. At one level of analysis we might conclude that the contour orientation midway along the cat's back is approximately horizontal and that this indicates the boundary of the cat's body. On the other hand, around the same general region, the orientations of the raised hairs are close to vertical. In other words, different visual information is being conveyed at different spatial scales: a coarse scale for information about general shape and a fine scale for details about the hairs. To solve this problem, vision operates at several different spatial scales using receptive fields of different sizes. In Figure 5 (overleaf), three different sized ON-centre receptive fields are shown superimposed on images containing luminance changes at three different spatial frequencies. As the spatial frequency of these images increases, the receptive fields that provide the best match to the images become smaller. The better the match between the receptive field structure and the image, the greater the response. For this reason, vision contains superimposed receptive fields of several different sizes (which are said to be 'tuned' to different spatial frequencies) to analyse the different image information that is conveyed at different spatial scales. This also means that we should think of vision as containing neural images at multiple spatial scales (or multiple resolutions) as well as ON- and OFF-centre images.

In fact, this introduces a completely different way of thinking about vision. An astonishing fact is that all images can be constructed by adding together a sufficiently large number of sine-wave gratings such as those in Figure 5, at different spatial frequencies and different orientations. For example, an image containing vertical black and white stripes with sharp boundaries can be produced by summing together a series of vertical sine waves with fuzzy boundaries (see Box 2). Because all images are made from sine waves, it is useful to know how the visual system processes sine-wave stimuli. Indeed, by characterizing vision in terms of its response to different sine-wave stimuli, it is possible to make reasonable predictions as to how vision processes any one of an infinite number of natural images constructed from those sine waves.

Figure 3 A cat. The orientation of the hairs on the cat's back are vertical, but the orientation of the back itself is approximately horizontal.
Box 1 Responses of retinal ganglion cells to luminance borders

Figure 4 shows a horizontal cross section of the light level of the image in Figure 2a, the receptive field locations of a selection of retinal ganglion cells and the pattern of responses of these cells. Receptive fields are also shown in cross section. For each cell, the contribution from its excitatory centre is matched to that from its inhibitory surround, meaning that when stimulation is the same in both regions the effects cancel out.

For cells 1 and 2 there is no stimulation in any part of their receptive fields and so these cells respond with spontaneous discharge. For cell 3, however, some light from the central square falls into its inhibitory surround. This means that the cell's spontaneous discharge is slightly inhibited and so its response is less than for the previous two cells. For cell 4, even more light falls into the inhibitory surround and so its response drops even further. For cell 5, although more light falls into the inhibitory surround, which would have the effect of reducing the response still further, light is also falling into half of the excitatory centre. Because the same proportion of both centre and surround are stimulated, the two effects cancel out and the result is spontaneous discharge. For cell 6, much of the inhibitory surround is stimulated but all of the excitatory centre is stimulated so the net contribution is one of excitation and the response is greater than spontaneous discharge. For cell 7, the level of excitation is the same as that for cell 6, but the level of inhibition is greater and so the response decreases.

Finally, for cells 8 and 9, light fills the entire receptive field and so the contributions from the two regions are equal and the response is the same as spontaneous discharge.

Figure 4 A cross section of the luminance profile of the image in Figure 2a is shown at the top of the figure. The locations and receptive field profiles of nine different retinal ganglion cells (1–9) are shown in the centre, and the responses of these cells are shown at the bottom. Note that the response profile at the bottom of the figure is a horizontal cross section through the response profile in Figure 2b.
Figure 5 Three vertical sine-wave gratings whose spatial frequency increases from left to right. Superimposed on each are the receptive fields of three ON-centre retinal ganglion cells.

Box 2 Synthesizing a square wave from a series of sine waves

The left-hand side of Figure 6 shows an image of a square wave and, below it, the first few in the series of sine waves from which it can be constructed.

At first glance it is not at all clear how this is going to work. For example, the square wave has sharp edges, the sine waves do not, and the square wave has flat plateaus, whereas the sine waves have ripples. The construction process is illustrated in the remaining two columns of Figure 6. The centre column shows a set of sine-wave gratings, with decreasing amplitude and increasing spatial frequency from top to bottom. In the right-hand column is shown a square-wave grating, with the sine waves progressively superimposed upon it. As the sine waves of increasingly higher spatial frequency are included in the summing process, the result is an image profile with increasingly sharp edges and increasingly flat plateaus, which looks more and more like the square wave. If the process is continued indefinitely, adding higher and higher spatial frequencies, then a perfect square wave is obtained. Of course, a square wave can be generated only with the correct combinations of sine waves at the correct spatial frequencies and amplitudes; different combinations of different sine waves will produce different images.
Figure 6  The construction of a square wave (top) from sine waves of different frequencies. The symbol \( q \) represents spatial frequency, thus a sine wave of \( 3q \) has a spatial frequency three times that of a sine wave of \( 1q \).
4 The contrast sensitivity function

The visual system’s sensitivity to (vertical) sine waves is illustrated psychophysically by the contrast sensitivity function (CSF) shown in Figure 7. The CSF is generated by measuring the lowest contrasts at which sine-wave gratings of different spatial frequencies can be just detected. A popular way of doing this is to use the two-interval forced choice technique (2IFC) described in Box 3. The CSF illustrates the range of spatial frequencies to which the human visual system responds and has a characteristic inverted-U shape. In other words, vision attenuates high and low spatial frequencies in the image. The CSF is sometimes referred to as the window of visibility because for an image to be visible it must contain sine-wave components that fall within the bounds of the CSF. In principle it would be possible to construct images from sine waves that fall entirely outside the CSF. For human observers, these images would be invisible.

Figure 7 Typical contrast sensitivity function (CSF). Note the inverted U-shape of the CSF.

So, the CSF characterizes human spatial vision and tells us what can and cannot be seen by the visual system, but why does it have the shape that it does? The low spatial frequency attenuation is readily explained by the inhibitory surrounds of retinal ganglion cells that we have met already. For example, in Figure 5, the cell with the medium-sized receptive field would not respond very well to the low-frequency grating (the one on the left) because both its excitatory centre and its inhibitory surround are stimulated by the light bar of the sine-wave grating. Of course, in Figure 5 this is not a problem because there is a cell with a larger receptive field that will respond to this spatial frequency. However, there will clearly be a lower spatial frequency to which this cell would not respond and this will always be true, no matter how large the receptive field.

To understand the high spatial frequency attenuation we must turn our attention to the very earliest stages of vision. In order for the retinal image to be converted to a neural image it must be sampled by light receptors in the retina. To be able to represent a high spatial frequency grating correctly these receptors must be packed very closely together so as to convey the rapidly changing light levels across the retina. The physical size of light receptors imposes a limit on their packing density and so this imposes a limit on the highest spatial frequency that can be correctly encoded.
Box 3 Two-interval forced choice (2IFC)

This psychophysical technique is widely used in vision science in a variety of experimental contexts. Its use is illustrated here by considering how to measure the contrast detection threshold of a sine-wave grating. A single experimental trial consists of two temporal intervals (two brief presentations of a stimulus separated in time) each signalled by an auditory beep. One of the intervals, chosen at random, contains a sine wave and the other contains no stimulus, just a blank display with the same mean luminance as that in the test interval. The observer has to decide in which interval the test stimulus was contained and indicates his or her response by pressing one of two buttons. If the observer could see the stimulus then the response would be correct, whereas if they could not, they would have to guess. If the observer guesses, then the probability of pressing the correct button would be 0.5 (i.e., 50% correct). By performing many trials at a range of contrasts it is possible to generate a psychometric function such as that shown in Figure 8. Contrast detection threshold is then taken as being the contrast level associated with some criterion level of performance, such as 75% correct. In the figure this gives a detection threshold of 1%, or 0.01, and the reciprocal of this number gives sensitivity, in this case 100.

![Psychometric function](image)

Figure 8 Psychometric function generated in a two-interval forced choice (2IFC) experiment.

However, there is a problem, because as explained in Box 4 (overleaf), spatial frequencies that are higher than the limit imposed by the sampling density result in the encoding of spurious low spatial frequency components known as aliases. In other words, we might expect that very high spatial frequencies can in fact be detected by vision (i.e., they can be discriminated from mean luminance; see Box 4) but that they would look like low spatial frequencies. This state of affairs is clearly unfortunate because not only would vision be removing some information from the image (the high spatial frequencies), but by misrepresenting it as low spatial frequencies it would introduce distortions: sine-wave components that are seen but are not present in the original image.
Box 4 Sampling and aliasing

Figure 9 shows the pattern of responses of retinal receptors for sine-wave gratings of different spatial frequencies. For the lower spatial frequencies, the pattern of responses captures the luminance changes contained in the image. However, for spatial frequencies that are so high that they are sampled less than twice per stimulus cycle, the pattern of responses does not correctly convey the original image information. This spurious pattern of responses is similar to that which would occur for a stimulus containing lower spatial frequencies. These spurious components are called aliases.

Figure 9 Aliasing due to under-sampling. Each panel shows the pattern of receptor responses to a different sine-wave grating stimulus. The sampling frequency of the receptors in the retina is fixed. (a) The sampling frequency is more than twice the spatial frequency of the input. (b) The sampling frequency is exactly twice the spatial frequency of the input. (c) The sampling frequency is less than twice the spatial frequency of the input. Only in (c) does the pattern of responses incorrectly convey the spatial frequency of the input. This spurious response pattern of responses is known as an alias.
Fortunately, the visual system has evolved a neat trick for rectifying this problem. The optics of the eye are not perfect but blur the image. As shown in Figure 6, sharp image structure is conveyed by high spatial frequencies and so blurring the image actually removes the very high spatial frequency content. In other words, a substantial part of the high spatial frequency attenuation of the CSF is due to optical deficiencies but this provides the benefit of removing the high spatial frequency components that would have resulted in aliasing. For this reason, the eye’s optics can be thought of as acting as an anti-aliasing filter.

5 Primary visual cortex

So far, we have seen that spatial vision responds to a broad range of spatial frequencies (the CSF) and that retinal processing operates at different spatial scales. As we move up the primary visual pathway from the retina to the lateral geniculate nucleus and then to the primary visual cortex, this operation becomes increasingly refined. For example, although retinal ganglion cells are tuned to spatial frequency, each cell does in fact respond to a rather broad range of spatial frequencies. Cells in the lateral geniculate nucleus are very similar to those in the retina but each one responds to a narrower range of spatial frequencies and cortical cells respond to a narrower range still. In fact, cortical cells are so selective that if the spatial frequency of a sine-wave grating at a cortical cell’s preferred spatial frequency is increased by a factor of three, then the cell would respond weakly or not at all. This means that the CSF can be thought of as the envelope of the sensitivities of a collection of cells each tuned to a different spatial frequency as illustrated in Figure 10.

![Figure 10](image_url)

There is good supporting evidence for this from experiments using a psychophysical paradigm known as adaptation. In this paradigm, an observer first inspects a high contrast sine-wave grating at a particular spatial frequency (known as an adapting grating) for a couple of minutes. This prolonged stimulation fatigues the cortical cells that respond to the adapting grating; and when the CSF is subsequently measured, this shows up as a reduction in sensitivity. Most importantly, however, sensitivity is not reduced for all spatial frequencies, just those around the spatial frequency of the adapting grating. This implies that the unaffected spatial frequencies are detected by cells that do not respond to the adapting grating, meaning that there must be cells that are tuned for several different spatial frequencies.
All this leads to the surprising conclusion that different spatial frequency components of the square wave in Figure 6 are processed by different cells in the primary visual cortex. Again, there is good psychophysical evidence for this. A low contrast square-wave grating is detected only when its lowest spatial frequency component (the one with the greatest amplitude in Figure 6) reaches its own detection threshold. Furthermore, the square wave is only discriminable from a sine-wave grating when its sine-wave component with the next highest amplitude (the $3\pi$ component in Figure 6) reaches its own detection threshold.

As we have learned already, one reason that vision operates at multiple spatial scales is so that it can read the different image information that the different scales convey. However, in the case of the square wave, for example, there seems to be little benefit in this. Although the details are still far from clear, it seems that where appropriate, subsequent visual processes are involved in recombing spatial frequency information across spatial scale.

One particularly important feature of cortical cells is that, in marked contrast to those encountered earlier in the primary visual pathway, they are selective for orientation. (Chapter 12, From retina to cortex). Receptive fields of a class of cortical cells known as simple cells are similar to those shown in Figure 11a and can be generated by wiring together an array of LGN cells in the way shown in Figure 11b. Their elongated structure means that, unlike retinal ganglion cells, a stimulus such as a sine-wave grating would have to be at a particular orientation for the cell to respond. This makes these cells ideally suited to coding the orientation of image structure. For example, retinal ganglion cells highlight both the vertical and horizontal borders in Figure 2, whereas the neural images for vertical and horizontal simple cells only pick out their preferred orientations, as shown in Figure 12. Thus,

**Figure 11** Receptive fields of cortical simple cells. (a) Simple cells have oriented receptive fields of various sizes, orientations and polarity, a small selection of which are shown. Some simple cells have more or less than the number of sub-regions shown. (b) The receptive field of a simple cell is approximately equivalent to summing together the receptive fields of an array of ON- or OFF-centre cells. This is achieved by making neural connections from an array of cells in the LGN onto a single simple cell in the cortex.
orientation adds yet another dimension to the neural image representations in vision. How this rather fragmented view of the world is useful to spatial vision is illustrated in the next section by considering the details of orientation coding in the cortex.

6 Population coding

The first thing to realize is that the response of a single simple cell cannot provide unambiguous information about the stimulus that stimulated it. To see this, consider the simple cell whose receptive field is shown in Figure 11b. It would respond strongly to a high contrast vertical grating. Importantly, however, the response of the cell could be reduced in at least two very different ways: first, by reducing stimulus contrast, and second by changing stimulus orientation. Thus, by inspecting the response rate of a single cell, later stages of the visual system could not know whether the stimulus was at the cell’s preferred orientation but of low contrast, or of high contrast but oriented away from the cell’s preferred orientation.

The solution to the problem is to consider the distribution of activity across cells that look at the same region of the image but through receptive fields at different orientations. Figure 13 (overleaf) shows the pattern of responses that would occur for a vertical grating at two different contrasts. Although the response of each cell changes with stimulus contrast, the shape of the distribution across the population of cells does not and in both cases the peak of the distribution indicates the orientation of the stimulus. This is known as a population code because a population of cells are required to encode the stimulus attribute, in this case, orientation. A psychophysical phenomenon known as the tilt after-effect (TAE) suggests that this is in fact the way that orientation is encoded in human vision.

The TAE is another important psychophysical phenomenon that has been revealed using the adaptation paradigm that we met earlier. If an observer adapts to a slightly tilted grating (oriented at say 15° from vertical), a subsequently presented vertical test grating appears tilted in a direction opposite from that of the adapter. As the perceived orientation of the grating is different from the physical orientation of the test image on the retina, the visual system must have made an orientation coding error, but how does this relate to the earlier idea about a population code? Figure 14b shows the response distribution (population code) for a vertical grating. The peak of the distribution is for a cell with a vertical receptive field and so this is the orientation that is seen. Now suppose that the observer adapts to a grating oriented at 15°.
**Figure 13** Population coding. Response distributions are shown for a population of simple cells for vertical gratings with high and low contrast.

**Figure 14** The tilt after-effect.
(a) An adapting grating oriented at 15° and the population response.
(b) A vertical test grating, distribution of fatigue and unadapted population response for the vertical test grating.
(c) Perceived orientation of a vertical test grating after adapting to the grating in (a). The distorted population code is shown after taking into consideration the effect of fatigue.

You might be able to demonstrate the TAE for yourself. Look at the grating stimulus in (a) for about 30 seconds and then transfer your gaze to the vertical grating below it. You should notice that the grating now appears tilted in the same direction as the grating in (c), although the magnitude of the after-effect might not be as large as that illustrated.
The population response to the adapter is shown in Figure 14a. A consequence of adaptation is that the cells that responded to the adapter become fatigued according to their level of excitation by the adapter. This leads to the distribution of fatigue illustrated in Figure 14b. The adapter is now removed and a vertical grating is presented. Normally this stimulus would produce the response distribution shown in Figure 14b, but because of fatigue, the response distribution is distorted. This is shown in Figure 14c and was calculated simply by summing the expected response distribution with the fatigue distribution (the two distributions in Figure 14b).

Because the peak of this distribution is oriented anticlockwise from vertical this is the way that the stimulus is seen.

7 Summary

Our visual perception of the world is the end result of extensive neural processing that starts in the retina by decomposing the retinal image into a collection of parallel neural images of different polarity and multiple spatial scales (different spatial frequency bands). A crucial feature at the very front end of the visual system is that an imperfect optical system blurs the image, removing very high spatial frequency components that would otherwise cause visual distortions due to under-sampling by retinal receptors. The overall effect of this blurring and the properties of spatial processing by retinal ganglion cells are characterized by the contrast sensitivity function: the window through which we view the world. At the level of the primary visual cortex, spatial frequency tuning is tighter than at earlier stages and the representation of visual information has further branched into neural images for multiple orientations. The spatial decomposition of the image in the visual cortex achieves at least two things. First, images convey different information at different spatial scales and so a useful starting point for spatial analysis is the independent representation of different spatial frequency bands in the image. Second, it provides a direct means by which individual stimulus features such as orientation (as illustrated here) and size can be measured. This is done using population codes in which image features are represented by the distribution of activity across a set of cells whose receptive fields look at the same location in the image but with selectivity for different stimulus attributes such as orientation. The end result of all this early visual processing is the transformation of a two-dimensional array of light (the retinal image) into multiple neural representations that encode useful image features for subsequent stages such as object recognition.