Understanding vision: theory, models, and data

This book is published by Oxford University Press, which owns the copyright, and must not be circulated or provided to any third party in any form.

Li Zhaoping
University College London
London
England
A very brief introduction of what is known about vision experimentally — figures and captions

Fig. 2.1: A: A neuron has a cell body (also called its soma), dendrites to receive inputs, and axons to send its outputs. This is a drawing (except for the red lines and printed words) by Cajal, a Spanish neuroscientist who died in 1934. B: A simple model of two neurons linked by a synaptic connection. C: A brain area can be modeled as a neural circuit comprising interacting neurons. It receives input from one brain area and sends output to another brain area. Neurons that do not send outputs to other brain areas are often called interneurons.
Fig. 2.2: Areas of the primate brain involved in vision. Shown is a monkey brain from the lateral view and medial view of its right hemisphere and from its unfolded representation. Cortical areas are about 1–3 mm thick and are folded like sheets to fit into the skull. In the digitally unfolded representation, for a clearer view, area V1 has been cut away from the other areas (notably area V2) in the unfolding process. Visual input enters the eye, is processed initially by the retina, and then is sent to the rest of the brain. The colored cortical areas are the ones exclusively or primarily focused on visual processing. Adapted with permission from Van Essen, D. C., Anderson, C. H., and Felleman, D. J., Information processing in the primate visual system: an integrated systems perspective, *Science*, 255 (5043): 419–23, copyright © 1992, AAAS.
Fig. 2.3: The hierarchy of levels of visual processing in the brain, simplified from information in Felleman and Van Essen (?), Bruce et al. (?), Shipp (?), and Schiller and Tehovnik (?). Various labeled areas can be located in the brain map in Fig. 2.2. V1 is also called the striate cortex. The gray shaded area encloses many of the visual areas collectively called the extrastriate cortex. The yellow shaded area outlines the areas carrying out gaze movements caused by visual inputs or other factors. Cortical areas downstream from V1 are differentially involved in processing “what” and “where” information about visual objects.
Fig. 2.4: A schematic of visual physiological experiments. Electrical activities in the neurons, including the action potentials (spikes), can be recorded while visual stimuli are presented to the animal.
Fig. 2.5: A schematic illustration of the retina and its neurons. Light enters the eye, and retinal neural responses are transmitted by the optic nerve to the rest of the brain. The bottom half of this figure is a zoomed-up view of a patch of the retina: visual input light passes through the retinal ganglion cells and other cell layers before hitting the rods and cones. Adapted with permission from Streilein, J. W., Ocular immune privilege: therapeutic opportunities from an experiment of nature, *Nature Reviews Immunology*, 3: 879–889, Fig. 1, copyright © 2003, Macmillian Publishers Ltd, http://www.nature.com/, and from Dyer, M. A. and Cepko, C. L., Regulating proliferation during retinal development, *Nature Reviews Immunology*, 2: 333–342, Fig. 4, copyright © 2001, Macmillian Publishers Ltd, http://www.nature.com/.
A very brief introduction of what is known about vision experimentally — figures and captions

$$\text{Response } O = \sum_x K(x)S(x)$$

Fig. 2.6: Schematic of how the response $O = \sum_x K(x)S(x)$ of a retinal ganglion depends linearly on the photoreceptor response signal $S(x)$. 
Fig. 2.7: A center-surround receptive field of a retinal ganglion cell modeled as a difference between two Gaussians. A: Gaussian excitatory field; B: larger Gaussian inhibitory field; and C: their combination as a center-surround shape of the receptive field. In each plot, the value of the receptive field \( K(x) \), or one of its Gaussian components, is visualized by the grayscale value of the pixel at image location \( x \). Parameters used are: \( \sigma_s/\sigma_c = 5 \), \( w_c/w_s = 1.1 \). D: Normalized contrast sensitivity \( g(k) \) versus spatial frequency \( k \) in the units of \( 1/\sigma_c \) for the receptive field in C. Also see Fig. ??.
Exposing a receptive field to an input of a sinusoidal wave

Contrast sensitivity of a ganglion cell in cat

Fig. 2.8: Measuring a neuron’s response to spatial gratings. A: Illustration of a spatial center-surround receptive field $K(x)$ exposed to a sinusoidal wave $S(x)$. The neural response $O = \int dx S(x) K(x)$ is largest when the center-surround receptive field is exactly centered on the peak of the sinusoidal wave $S(x)$. In general the response is proportional to $\cos(\Delta \phi)$, the cosine of the phase value $\Delta \phi = \phi - \theta$ of the sinusoidal wave at the center location of the receptive field; see equation (??). B: Contrast sensitivity of a retinal ganglion cell in cat. Note the change with the mean light level to which the animal is adapted. The cat (whose visual acuity is typically 10 times worse than humans) was anesthetized, with a pupil diameter 3.5 mm, and the grating was drifting such that input temporal frequency was 1 Hz (cycle/second). Data from Enroth-Cugell, C. and Robson, J. G., The contrast sensitivity of retinal ganglion cells of the cat, *The Journal of Physiology*, 187 (3): 517–552, Fig. 15, 1966.
The 2D vector $g(k) = [g_c(k), -g_s(k)]^T$.

Fig. 2.11: A 2D vector $g(k)$, with the horizontal component $g_c(k)$ and vertical component $-g_s(k)$, forming an angle $\theta$ from the horizontal axis.
Fig. 2.12: Responses from a retinal ganglion cell in cat to inputs as recorded by T. Wiesel (?). Shown are three temporal traces of intracellular potentials in response to three different input patterns. The center input pattern in A is a disk of one degree diameter. Adapted with permission from Wiesel, T. N., Recording Inhibition and Excitation in the Cat’s Retinal Ganglion Cells with Intracellular Electrodes, *Nature*, 183 (4656): 264–265, Fig. 1, copyright © 1959, Macmillan Publishers Ltd, http://www.nature.com/.
Fig. 2.13: The impulse response function of a neuron modeled by equation (??).
A very brief introduction of what is known about vision experimentally — figures and captions

A: Human temporal contrast sensitivity

B: Contrast sensitivity of a monkey LGN neuron

Fig. 2.15: Sensitivities to luminance and chromatic temporal modulation for human observers (left) and monkey retinal ganglion cells (right). Note the following characteristics: the filter for the luminance signal prefers higher frequencies than the chromatic signal; M cells are more sensitive to the luminance signals than P cells; and P cells are more sensitive to the chromatic signal than M cells. Data from Lee, B., Pokorny, J., Smith, V., Martin, P., and Valberg, A., Luminance and chromatic modulation sensitivity of macaque ganglion cells and human observers. *Journal of the Optical Society of America, A*, 7 (12): 2223–2236, 1990.
Fig. 2.16: A: Spectral sensitivities of the cones as a function of the wavelength of light. B: Schematics of two retinal ganglion cells with center-surround color opponency in their receptive fields. These are called single-opponent cells since each subregion of a receptive field involves only one color type.
Fig. 2.17: A: The density of cones and rods in the human retina versus visual angle from the center of vision. Note that the sampling density of the cones drops dramatically with eccentricity. It is densest at the fovea, where there is no room for the rods. The density of rods peaks slightly off fovea. The optic disc is where axons from the ganglion cells exit the eye to form the optic nerve, causing a blind spot, which is not noticeable in binocular viewing when the blind location in one eye’s visual field is not blind in the other eye’s visual field. Data from Osterberg, G., Topography of the layer of rods and cones in the human retina, *Acta Ophthalmology*, 6, supplement 13: 1–102, 1935. B: Visual acuity drops dramatically with increasing eccentricity: all the letters are equally visible when one fixates at the center of the eye chart (?). Copyright ©Stuart Anstis, 2013.
Fig. 2.18: A schematic of the visual pathway from the retina to V1 via the LGN. Information from the two eyes is separated in separate layers within LGN, but combined in V1. Information from the left and right hemifields of the visual space are sent to V1 in the right and left hemispheres of the brains, respectively.
Fig. 2.19: The retinotopic map in V1, binocular and monocular visual space, and the definitions of eccentricity $e$ and azimuth $\theta$. A: Half of the visual field depicted in B is mapped to half of the V1 cortical area (in one hemisphere of the brain). More V1 surface area is devoted to central vision. B: The visual space mapped to cortical space in A. H.M. means horizontal meridian. Adapted from Vision Research, 24 (5), David C. Van Essen, William T. Newsome, and John H.R. Maunsell, The visual field representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability, pp. 429-48, figure 4, Copyright (1984), with permission from Elsevier.
ccvi | A very brief introduction of what is known about vision experimentally — figures and captions

Fig. 2.20: A schematic of how three on-center retinal ganglion neurons, or LGN neurons, feeding into a V1 cell could make a V1 cell tuned to a bright oriented bar, according to Hubel and Weisel.
Fig. 2.21: The Gabor filter $K(x, y) \propto \exp \left( -\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} \right) \cos(kx + \pi/4)$ in A is tuned to the vertical orientation. B shows the frequency tuning of this filter; note that the peaks are on the horizontal axis. The brightness levels in both pictures scale with the numerical values of the functions in space $(x, y)$ or in frequency $(k_x, k_y)$.
Quadrature model of motion direction selectivity

Fig. 2.22: Using a quadrature model to build a spatiotemporal receptive field tuned to motion direction by summing two filters not selective to direction. Each of the three grayscale images plots a spatiotemporal receptive field, with space and time along horizontal and vertical axes, respectively. The two images at the top are space-time separable and not direction selective; each is the product of a spatial filter (the curve to its bottom) and a temporal filter (the curve to its left). The receptive field in the grayscale image at the bottom is made by superposing the top two grayscale images. The blue and red curves superposed in each graph, to the bottom and left of the bottom grayscale image, depict two filters which are quadratures of each other; see equations (??–??).
Fig. 2.23: Disparity tuning and ocular dominance of neural receptive fields. A: Each object in the visual scene has a disparity. The visual angles $x_L$ and $x_R$ give image locations on the retina. B: A schematic of the receptive fields for the left and right eyes, respectively, for a single V1 cell. The two receptive fields (RFs) can differ in both amplitude and phase. This single neuron is said to be left eye dominant since the RF for left eye input has a higher amplitude. The peak location $x_R$ of its on-region in the right eye is right shifted relative to its counterpart $x_L$ in the left eye. C: Ocular dominance columns revealed by optical imaging (copyright © 2013, Anna Roe). Black and white stripes mark V1 cortical surface locations for neurons preferring inputs from one and the other eye, respectively. Each column is about 400 $\mu$m thick in monkeys; they abruptly stop at the border between V1 and V2, in the upper part of the image, since V2 neurons are mostly binocular.
A very brief introduction of what is known about vision experimentally — figures and captions

**Response = 0.5**

**Response = 0**  
**Response = 0.3**

**A:** Classical receptive field  
**B:** Surround alone  
**C:** Iso-orientation suppression

**D:** Random surround suppression  
**E:** Cross-orientation suppression  
**F:** Colinear facilitation

**Response = 0.5**  
**Response = 0.8**  
**Response multiple**

Fig. 2.24: A schematic of typical types of contextual influence on the response of a V1 neuron. The CRF is marked by the dashed oval, which is not part of the visual stimuli. In A, the neuron responds to its optimal stimulus, a vertical bar, within the CRF. The surrounding vertical bars outside the CRF do not excite the cell (in B) but can suppress the response to the vertical bar within the CRF by up to about 80% (in C)—iso-orientation suppression. This suppression is weaker when the contextual bars are randomly oriented (D); it is weakest when the contextual bars are orthogonal to the central bar (E)—cross-orientation suppression. F: Colinear facilitation—when the contextual bars are aligned with a central, optimally oriented, low contrast bar, the response level could be a multiple of that when the contextual bars are absent.
Fig. 2.25: Examples of visual inputs to which some visual cortical neurons in monkey are selective. Posterior IT and anterior IT are also referred to in the literature as areas TEO and TE respectively (?). Reproduced from *Journal of Neurophysiology*, Kobatake, E. and Tanaka, K., Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex, 71 (3): 856–867, Fig. 11, copyright © 1994, The American Physiological Society, with permission.

<table>
<thead>
<tr>
<th>V2</th>
<th>V4</th>
<th>posterior IT</th>
<th>anterior IT</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Fig. 2.26: V2 neurons respond to illusory contours and can signal border ownership, as observed by von der Heydt and colleagues. A–K sketch the visual inputs used to probe V2 neural properties. The red ellipse in each plot marks the classical receptive field (CRF) of a V2 neuron of interest and is not part of the visual input. In A and D, the CRF contains a vertical contour or surface border defined by actual contrast in the image; in B, E, and F, the CRF contains only an illusory vertical border or contour. A neuron tuned to vertical orientation can be activated by these real or illusory contours or borders. However, the neuron will not be activated at its CRF position in C and G, since the CRF contains no real or illusory contour. In H–K, the images contained within the receptive field are identical, depicting a transition from a white region from the left to a darker region on the right. This transition may signal a border to a surface to the left of the border (in H and I) or a border to a surface to the right of the border (in J and K). A V2 neuron preferring a border to a surface on the left of the border will be more activated by H and I than by J and K.
A: Moving gratings and plaids

Two separating gratings moving, in apertures

Noncoherent motion of two gratings

Coherent motion of a single plaid

Fig. 2.27: MT properties and perception. A: Moving gratings and patterns to probe MT neurons. Superposing two drifting gratings (top, seen through apertures), each moving in the direction indicated by the arrow, gives rise to the percept either of a single pattern moving upwards or two transparent gratings moving in their respective directions. Manipulating the intersection points of the two gratings in the superposition can bias this percept, as demonstrated in the two example superpositions. Some MT cells (pattern cells) are tuned to the moving direction of the (single) whole pattern, whereas others (component cells) are tuned to the directions of the component gratings (?). Furthermore, some MT neurons behave more like a pattern or component cell, respectively, according to whether the perception is more like a coherent pattern or otherwise (?). B: Dots moving in the preferred direction of an MT neuron are also effective stimuli, but the responses are typically suppressed when adding another group of dots moving in the opposite direction in the same depth plane (?). Separating the two groups of dots in depth can release the suppression (?). When an input is consistent with more than one depth order (bottom), the responses of some neurons in MT are correlated with what the animal perceives, according to the preferred directions and disparities of the neurons concerned (?).
Fig. 2.28: A V4 lesion makes a monkey unable to perform the task in A, even though it can still do this task when the foreground bars are thicker or distinct in color (>). It also makes a monkey unable to detect the odd object in the left ring in B, when the square is insufficiently salient; although performance in the right ring, in which the square is most salient, remains sound (>).
Fig. 2.29: Brain circuit for the control of eye movements. This schematic summarizes and simplifies information in the literature (?, ?, ?, ?).
A: Two designs of stimulus and task to study the effects of top-down attention

- **Attention inside or outside the RF, to preferred or non-preferred stimulus**

  - A stimulus outside the RF
  - Two visual stimuli inside the RF, one preferred, one non-preferred

- **Attention inside or outside the RF**

  - A stimulus outside the RF
  - A visual stimulus inside the RF

B: Various effects of top-down attention on neural responses to input contrast or input feature (e.g., orientation and motion direction) in extrastriate cortices

- **Scales responses**
- **Changes effective input contrasts**
- **Scales feature tuning curves**
- **Sharpen feature tuning curves**

Fig. 2.30: Effects of top-down attention on extrastriate (e.g., in V2, V4, MT, MST) neural responses in behaving monkeys. A: Sketches of two designs to compare neural responses to the same inputs in different attentional states. Monkeys fixate on the central fixation point (cross) while doing a task based on a stimulus (e.g., bar, grating, moving dots) inside or outside the receptive field (RF, marked by the dashed rectangle) of the neuron whose response is being measured. Their attention is assumed to be covertly directed to the task-relevant stimulus. In the left design, two stimuli are inside the RF, one (preferred) evokes a higher response than the other if each is presented by itself within the RF. In the right design, only one stimulus is within the RF. Attentional modulation is manifest in the difference between the responses when attention is inside versus outside the RF; it is also manifest in the difference between the responses when attention is to the preferred (e.g., preferred orientation or motion direction) versus the non-preferred stimulus (inside or outside the RF). B: Four examples of possible effects of top-down attention on neural response curves to input contrast and input features. In each plot, the red and black curves are two response curves to the same visual input when attention is directed respectively to two different targets, as schematized in A. Attention can scale the response curves (in the first and third plot), act as if the external input contrast is changed (the second plot), or change the shape of the feature tuning curves (the fourth plot).
Fig. 2.31: Influences on the responses of V1 neurons of top-down attention, task, and experience. A: the contextual influence of the bars surrounding the RF on the response of the V1 neuron (with that RF) depends on the task the monkey is performing (?). The monkey is trained to do two tasks, a bisection task involving deciding whether the left or right neighbor is closer to the central bar, and a vernier task requiring a decision as to whether the central bar is to the left or right of the top and bottom bars (which are vertically aligned). During the bisection task, the neuron’s response to the central bar is typically more strongly modulated by the task-relevant distance between this bar and the center of mass of the left and right neighbors, and less modulated by the task-irrelevant distance between the central bar and the center of mass of the top and bottom neighbors. During the vernier task, the reverse holds. B: A V1 neuron’s response to the central ball in each image does not depend on whether the surrounding balls have the same shading direction as the central ball. This ceases being true if the monkey has had sufficient experience in the task of detecting an odd-ball by its shading (?). The shading direction of the context only affects the longer latency responses after the initial response to the image.