BARLOW, H.B. & MOLLON, J.D. (Eds) (1982).
The Senses. Cambridge Texts in the Physiological Sciences, Vol. 3.
Cambridge University Press: Ch. 1, pp 1-33; Ch. 6, pp 102-113; Ch. 7
(with J.D. Mollon), pp 114-132; Ch. 8 (with J.M. Woodhouse), pp 133-164;
Ch. 20 (with J. Atkinson and O.Braddick), pp 448-469.

CHAPTER ONE

General principles: the senses considered as physical instruments

H. B. BARLOW

When a person falls ill, he is likely to become aware of it through his own senses, for he will experience something like a toothache, a pain in the abdomen, or a sudden chill. If the condition progresses to the point where he seeks help, his advisor, even while listening to the patient's story, will use his own senses to pick up what he can of the cause of the illness, and in very many cases what he sees, hears, feels, and smells will be enough for him to diagnose the condition. X-rays, laboratory tests, and monitoring devices have decreased what was previously a total reliance on direct sensory information, but the patient's sensations, and the messages from the doctor's sense organs, still provide the most important, potentially most reliable, and most up-to-date, sources of information about the illness.

Of course it is not just in sickness that the senses tell one about one's own body. The *interoceptors* constantly monitor its chemical and physical state, and these are key elements of the homeostatic mechanisms that stabilise the internal environment. Complex forms of life are completely dependent on accurate interoception. Similarly it is the *exteroceptors* that enable one to assess the surrounding environment, and so to comprehend the tasks one confronts. It is worth noting, however, that these so-called exteroceptors are really specialised interoceptors; they sense the outer world only by means of its physical and chemical influence on the special sense cells of the nose, the ears, or the eyes.

Thus the senses are the bodily mechanisms for gathering up-to-date information, and as such it is hard to exaggerate their importance. Wars are won and lost by the intelligence system, and if one considers the survival of individuals and species in more natural environments one can see at once that superiority of sight, hearing, or smell must confer an immense advantage in the competitive struggle for food resources, habitat, and mates. The interoceptive system, which monitors bodily functions, is less spectacular for its task is one of maintenance. But an army marches on its stomach, and faulty internal information, for instance to the supply department, can lose a war as decisively as erroneous intelligence.

This book describes what is known of the physiology of the senses: that is, the physical, chemical, and biological mechanisms of this information-gathering system. The principal methods for investigating these processes are very similar to those employed in other branches of physiology. One can examine the anatomical structure of the sense organs, and in some cases this gives immediate insight into their mode of operation, but this method is not as powerful as one might suppose when it is used by itself. The analogy between the structure of the eye and that of an image-forming device such as a camera seems obvious to us, but before image-forming devices were familiar, and before the image cast by the lens on to the retina had been directly observed, the anatomical cues were misinterpreted. The optic nerve was dissected down to the eye, where it was found to spread out over the retina. Following the retina round the vitreous humour one comes to the ciliary body suspending the lens, and it was thought that the 'quivering' of the lens as the light passed through it excited the sensitive terminals of the optic nerve. The apparent anatomical continuity of optic nerve and ciliary body was misleading, but the acceptance of this mistaken teaching emphasises the importance of functional models when the anatomy is interpreted, and the desirability of detecting and recording the physiological process at every possible point. Nowadays one can do this, for one can record the electrical responses of cells at almost every level in the visual, auditory, and other sensory pathways, and information from this source is prominent in this book. Much information has also been obtained from biochemical investigations, especially of the photosensitive pigments underlying vision. But sensory systems possess special interest as the objects of physiological study because one can receive guidance from two types of knowledge not available for other systems. In the first place, a branch of engineering has grown up over the past few decades that deals with the problems encountered in man-made instruments and information-gathering systems. Such knowledge tells one the physical limits and natural difficulties likely to be encountered in sensory systems, and this provides insight rather similar to that given by the law of conservation of energy and other thermodynamic principles in understanding metabolism. The next section of this chapter discusses some general aspects of information-gathering by the senses from this viewpoint.

Sensory systems are also of unique interest because they provide the information on which our conscious sensations are based, and this makes it possible to obtain guidance from a second source of knowledge. A human can be asked to report his sensations in experiments where the physical properties of the stimuli are accurately controlled. The study of how the subject's report varies with the physical parameters of the stimulus is called *psychophysics*. By this means an objective and rather complete account can be given of what information the human senses can and cannot convey to the brain. This overview of their performance is enormously useful when analysing their anatomical and physiological mechanisms, for one knows what these mechanisms must achieve with a precision that is often lacking in other systems. As a result of this body of psychophysical knowledge, and through applying some principles of communication engineering, we are beginning to understand the mechanism of the sensory parts of the brain as well or better than any other part.

1.1. PRINCIPLES OF OPERATION OF SENSORY PATHWAYS

Plan of sensory pathways

Fig. 1.1 shows a diagram of a peripheral sensory pathway. The sensory nerve fibre at the left has terminals in the skin at the top, and when these are appropriately stimulated they are depolarised, thus initiating action potentials that are propagated down the axon. When a stimulus is applied and held constant at a steady value these impulses are initiated after a short latency, and then reach a peak frequency which varies with intensity. In most sensory nerves the frequency then declines, even when the stimulus remains constant, a property which is known as *adaptation*. Many pathways also have a *maintained discharge* at a low rate which results from spontaneous depolarisation of the terminals occurring in the absence of a deliberately applied stimulus. Centrally the fibre terminates on second order neurons, and this part of the system will be considered later.

For such an information-gathering system to be effective there are certain definite requirements. It must have elements *selective* for different types of physical or chemical stimulus, these elements must respond *speedily*, they should be *sensitive* to small changes in the input, and they must respond *reliably* and not give spurious information. What we gain from the analogy with man-made instrumental systems is the knowledge that selectivity, speed, sensitivity and reliability cannot be improved indefinitely but are limited by physical factors. Furthermore they are usually interrelated, so that one of

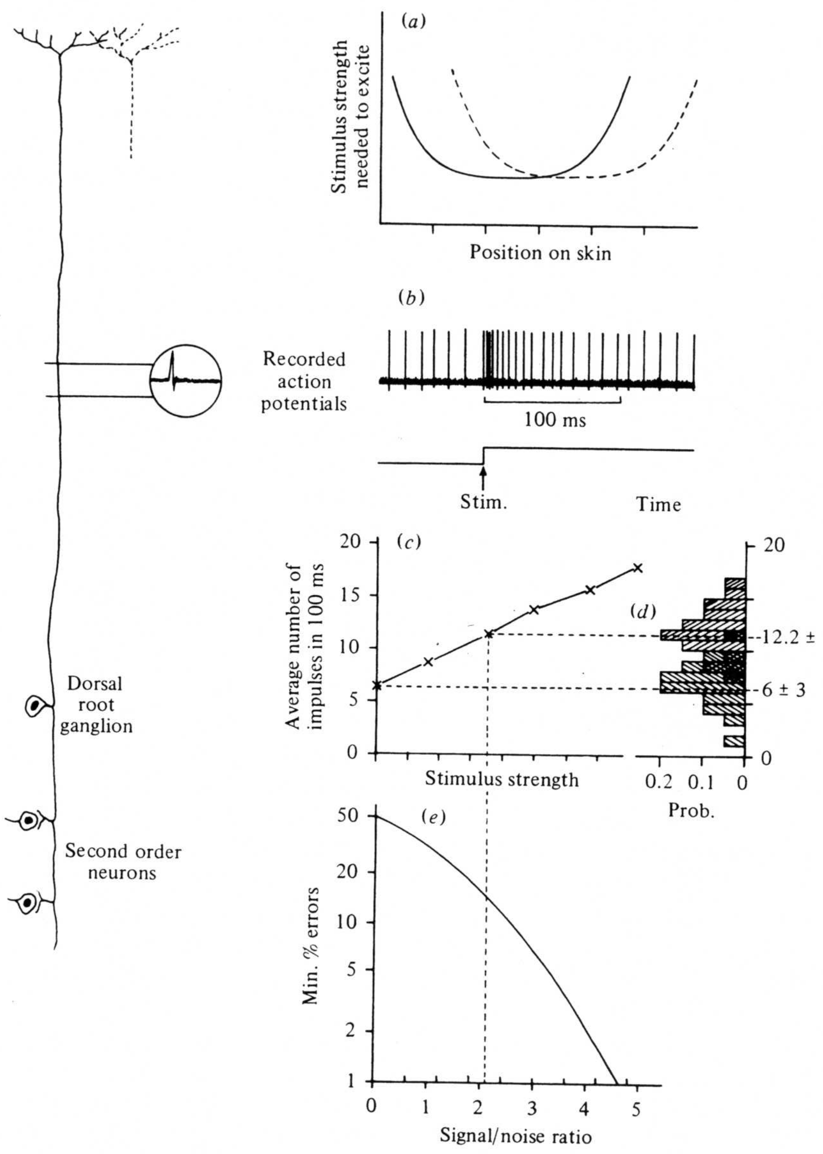

Fig. 1.1. Diagram of a peripheral sensory pathway. A nerve fibre running from sensory terminals in the skin to the synaptic endings on second order neurons in the spinal cord is shown to the left. On the right (*a*) illustrates the selectivity of the fibre for the position of the stimulus on the skin, (the dotted curve shows the corresponding curve for a neighbouring fibre); (*b*)

them cannot be improved without detriment to another. The physical factors limiting performance in living tissues are not necessarily the same as those encountered by the communications engineer, but it is a challenge to find out what the biological factors are, and how they affect performance. These aspects of sensory pathways will now be discussed, with special emphasis on how they can be measured and how they are interrelated.

Selectivity

Almost all sensory fibres show selectivity for the anatomical position of a stimulus. The schematic nerve fibre in Fig. 1.1 is distributed to only a portion of the skin and thus responds to physical stimuli applied in that region and no other. If all nerve fibres had the same receptive area, or *receptive field* as it is called, this system would convey no information at all on the whereabouts of an excitatory stimulus, so the transmission of such information entirely depends upon the fact that different nerve fibres exhibit different positional selectivities.

Sense organs are also selective for other characteristics of the stimulus. For instance in the skin, some respond to temperature change instead of mechanical deformation, others respond to products released from damaged cells and thereby signal injury. Qualitative selectivities of this sort lie behind the major modalities of skin sensation. In some cases the selectivity can be plotted as a graph, not of position as in Fig. 1.1, but of some other variable. Examples of these are the tuning curves of auditory nerve fibres (Fig. 14.16) or the spectral sensitivity curves of the different categories of cones

---

shows a train of action potentials resulting from application of a stimulus; (*c*) plots the average number of impulses in 100 ms following stimuli of varying strengths; (*d*) shows histograms of the frequency of obtaining 2 to 17 impulses in 100 ms when no stimulus was applied (lower peak), and following a moderate stimulus causing an average of 12.2 impulses (upper peak); both these distributions have a standard deviation of ±3 impulses. The reliability of detecting a stimulus depends on the signal/noise ratio, which is $(12.2-6)/3 = 2.1$ for a stimulus of the strength shown here. Because these distributions overlap, errors are unavoidable when judging whether a stimulus was present or not from the number of impulses in a pulse train, and (*e*) shows that for a signal/noise ratio of 2.1 the error rate would be at least 15% on average. A stronger stimulus would separate the distributions shown in (*d*), and the error rate would fall as shown in (*e*).

subserving colour vision (Fig. 9.3*c*). In the latter curve *wavelength* is plotted vertically and *sensitivity* is plotted horizontally; sensitivity is defined as the reciprocal of the energy required to produce some standard level of activity, hence these curves are rotated compared to the positional receptive field plotted in Fig. 1.1 and the auditory tuning curves in Fig. 14.16.

There are some important terms and concepts associated with selectivity curves like those shown in Figs. 9.3*c* and 14.16, because the degree of selectivity is the most important factor determining how a sensory pathway combines and separates the physical stimuli acting on the endings. To take the combining aspect first, it is a matter of practical importance to know how effectively mixtures of different spectral lights stimulate the eye. Suppose, for example, that a lamp for illuminating streets is made by mixing a narrow band of yellow from sodium, which has a wavelength near 589 nm, with a violet at 436 nm from mercury; will these lights act entirely independently, or will the eye be able to combine them? Under dark-adapted conditions, the visibility of a light depends solely on the total number of quanta absorbed in the photosensitive pigment rhodopsin (see Chapters 5 and 7). As a result the effectiveness of the combination can be calculated by multiplying the energy of each wavelength, E_λ, by a factor conventionally labelled V'_λ representing the relative effectiveness of light at that wavelength, and adding the products (i.e. calculating $E_{436} \times V'_{436} + E_{589} \times V'_{589}$). Similarly the effectiveness of a broad spectral band, or of white light containing all wavelengths, can be calculated by integrating the energy in each narrow band weighted by the relative sensitivity in that band, that is calculating $\int E_\lambda V'_\lambda \, \mathrm{d}\lambda$. Now one can see that a very broad curve of spectral selectivity will cause light over a broad band of wavelengths to be combined, with obvious advantage to the sensitivity of the system because the total amount of light energy utilised is increased.*

Now consider the spectral selectivities of the cones, which take over from the rod system at higher light levels. There are three different types, and this enables them to transmit information about the spectral composition of the light as well as the total quantity, so that objects such as ripe fruit can be distinguished by colour as well as by other characteristics. The spectral sensitivity curves of the three types differ from each other, and the effectiveness of the system in analysing and transmitting information about the spectral composition of the light depends on two properties; first the number of different classes of cone, and second the degree to which the spectral sensitivity curves differ from each other. In the case of the colour system there are only three different classes, and their spectral sensitivity curves overlap a great deal. A system with improved capacity to analyse the spectral composition of light would require more than three different spectral sensitivity curves; their peaks would be closer together, and it would help if they were narrower so that the overlap of the different spectral bands was not increased. In such a system the same amount of light would have to be divided among more receptors, and the narrower absorption curves would absorb less energy from white light, so this is an example where selectivity and sensitivity are opposed; one cannot improve the capacity of the system to separate, or resolve, without sacrificing sensitivity.

Many of the same considerations apply to the spatial resolution of the eye – its capacity to separate out the different parts of an image. High resolution requires a large number of nerve fibres each with a narrow receptive field, whereas if the resolution is poor the receptive fields can be large to achieve sensitivity and fewer of them are needed. The retina of a cat provides a beautiful example of this: in the periphery the ganglion cells are sparse and their receptive fields large, while towards the centre of the field of vision the ganglion cells are crowded together and also have smaller receptive fields, so that the amount of overlap between neighbouring receptive fields stays roughly constant.

Before leaving the topic of selectivity, a source of confusion that arises in discussions of resolving power must be discussed. If one were asked to measure the resolving power of the colour system, it would not be unreasonable to find out how small a difference of wavelength a person could discriminate. Such tests yield an astonishingly low figure, under 1 nm in the yellow region of the spectrum, which is sufficient to distinguish between the double lines of sodium near 589 nm! One might conclude from this fact that, in the visible spectrum stretching from 400 to 700 nm, there are about 300

---

[1] * When the photopic visibility function V_λ is substituted for the scotopic function V'_λ (see Figs. 5.8, 7.3) the resulting integral expresses the effectiveness of the radiant energy in stimulating the eye at the light levels where most visual tasks are performed. It is therefore the quantity that needs to be measured in many practical situations, such as determining if the light is good enough for reading easily, and for this reason it is used to define *photometric* units of measurement. These are based on the *lumen*, the unit of luminous flux which is given by $L = 680 \int_0^\infty E_\lambda V_\lambda \, \mathrm{d}\lambda$, where L is lumens, E_λ is in watts/nanometre, and V_λ is the photopic visibility function.

distinguishable hues; the actual number is rather less than this because a 1-nm difference cannot be detected across the whole spectrum, but about 200 hues are genuinely discriminable. One can be confused by the question 'Does the eye resolve colours into 3 classes, or 200?' There is no doubt that 3 is the more fundamental answer, for it indicates the number of separately variable aspects of the spectral distribution of light that the eye transmits to the brain; the excitation of a red, a green, or a blue cone can each vary while the excitation of the other two is kept constant. The figure of 200 discriminable hues results from different combinations of activity of these three types of cone in the same sort of way that 1000 different numbers can be represented by three decimal digits. One has to distinguish carefully between the number of classes of cone with separately variable outputs and the number of possible ways of combining those outputs in various proportions.

The same distinction is important in considering the spatial resolution of the eye; the position of a dot or line can be judged with a precision of a few seconds of arc, which is ten times better than the distance between the parts of the image that can be truly resolved and signalled separately. The improved positional accuracy can be achieved by the mathematical techniques of interpolating and averaging, and in long-evolved sensory systems there are many features that an engineer knowledgeable about instrument design can help one to understand. It is often remarkable to see how natural selection and evolution have led to designs that approach limits set by the physical nature of the tasks being performed.

There are some exceptions to the obvious good design of many sensory systems; in these the physiological facts point to arrangements that seem far from ideal. For instance the chemical senses use a peculiar method of signalling in which each nerve fibre has a mixed sensitivity to several distinct chemical modalities (see Chapter 19). Consequently it is a complex and difficult logical problem to make discriminations between different chemical stimuli, whereas it would seem much easier if each nerve fibre signalled only a single type of chemical. Perhaps we underestimate the ease with which the nervous system can perform the tasks of pattern recognition called for in this instance, and hence overestimate the advantages of keeping the modalities separate. Examples like this are intriguing because they point to areas of ignorance where it may be possible to make important discoveries.

Speed

The second desirable characteristic of a sensory system is speed of response. The great survival advantage of rapid action needs no emphasis, and a speedy sensory message is the first requirement to achieve this. Nerve fibres from the ear respond within a millisecond or less, and the same is true for fibres from the vibration-detecting Pacinian corpuscles, but many other systems respond at a surprisingly leisurely pace, and to understand this one must realise the advantage of a long summation time. Just as a large receptive field may have the advantage in sensitivity over a small one, so a mechanism that integrates in time and reacts to the total stimulus energy over a long period can be more sensitive than one that integrates over a short time only. The point can be illustrated by considering a pair of galvanometers of identical construction except for the strength of the return springs: the one with the weaker spring will be more sensitive, but the pointer will take longer to reach a steady reading. In Fig. 1.1 the latency and duration of the impulse discharge shown in the second line will set limits to the sensitivity of the system as well as determining how quickly the animal can react.

Sensitivity

It has already been pointed out that high spatial resolution requires small receptive fields and such small fields can collect only a small amount of energy from the applied stimulus. A speedy response requires a short integration time, which also reduces the available energy. High sensitivity of the basic receptor mechanism – the capacity to respond to a small amount of energy – is thus a key requirement for attaining good resolution and quick responses. We shall see later that high sensitivity is also needed for *reliability*. In addition, high sensitivity may itself have protective value because it might, for instance, enable an animal to hear or smell a predator at a greater distance. In the case of vision, it may also enable an animal to hunt or graze for extra hours at dawn and dusk.

It is worth noting, however, that high sensitivity is not always advantageous. A hyper-sensitive skin would be distracting and might lead to constant grooming behaviour or scratching, and if one's ears were too sensitive one might be deafened by the throb of the pulse and roar of blood flowing in the artery of the basilar membrane, or by molecules bombarding the tympanic membrane under Brownian motion. These thoughts make one reconsider what is meant by sensitivity: what is it that limits the process of gathering *important* information?

The third figure from the top in fig. 1.1 shows the result to be expected when the average number of impulses occurring in a sensory nerve in, say, 100 ms, is plotted against the intensity of an applied stimulus. Note first that some impulses occur *before* the stimulus is applied. This is the *maintained discharge*, which occurs in the absence of stimulation, and its rate varies greatly in different systems, and in different types of nerve fibre. For individual optic nerve fibres it varies from over 50 s^{-1} to under 5 s^{-1}, but is rarely completely absent. Auditory nerve fibres also usually have a maintained discharge, as do those from vestibular organs. Others, such as those from Pacinian corpuscles and many touch receptors, are quiet until excited.

When the sense organ is stimulated more strongly the number of impulses increases, linearly at first but usually at a lower rate as the stimulus intensity is raised further. Sensitivity can be defined by the slope of this increase, that is by the average number of extra impulses elicited per unit of applied energy. Notice, however, that there is something missing from such a definition, useful though it is. If the nerve is quiet before stimulation, a single extra impulse is easily distinguished and is obviously a significant event. But if there is an irregular maintained discharge of, say 60 impulses s^{-1}, then in a counting period of 100 ms there will be about 6 impulses even without any stimulation, and a single extra impulse may easily pass unnoticed.

Reliability and noise

Communication engineers are of course familiar with the situation in which unwanted disturbances obscure the detection of the signal of interest, and it has come about that these unwanted disturbances are always called 'noise', even with a video television signal or when considering erroneous operations in a computer, because the problem was first encountered with acoustic signals. For sensitivity, what is important is not the amplitude of output signal achieved by a given input signal, but the relationship between the output signal and the unwanted signals, or noise level. This concept of signal/noise ratio can now be incorporated into the discussion of a sensory pathway.

If one goes back to the case where the maintained discharge was obscuring the detection of an extra impulse, a moment's consideration will show that it is not the average *rate* of discharge that matters, but its *variability*. If exactly 6 impulses occurred every 100 ms, then when 7 or more impulses occurred we should know that a stimulus had been applied. But if 6 impulses is only an average, and the number is sometimes 3, sometimes 9, then one will need many extra impulses before one can reliably detect a change, for the number must be raised beyond the upper limit of the maintained discharge.

To the right of the plot of the average number of impulses, two histograms are shown (Fig. 1.1*d*). These represent the actual numbers of impulses counted in the 20 trials used to determine two of the average numbers. The lower one is for the maintained discharge, and it shows that the number ranged from 2 to 12 in the hypothetical example. The average was 6, but what is more important is the variability. The spread of the numbers is best expressed by their standard deviation, $(\Sigma(\bar{x}-x)^2/(n-1))^{\frac{1}{2}}$, which is 3.0 impulses in this case. This is the 'noise' of the neural message over the 100-ms period, and it means that 3 extra impulses are required to achieve a signal/noise ratio of 3/3, or unity.

The upper histogram shows the distribution of the numbers of impulses obtained on the 20 trials used to determine the response, which averaged 12.2 impulses in 100 ms. This is an increase of 6.2 over the maintained discharge and corresponds to a signal/noise ratio of 6.2/3, or 2.1. Note that, although the means of the two histograms are quite well separated, the tails of the distributions overlap. If a response of 9 impulses occurred, this could be an unusually high value of the maintained discharge, or it could be an unusually low value resulting from a stimulus whose mean response is 12 impulses. It follows that a response of this value cannot be assigned with certainty to either histogram. When it occurs it is impossible to know whether a stimulus was applied or not; the best that could be done would to be choose some criterion, say of 10 impulses, and decide that when this value or more occurred, the stimulus was present. But even then there would be a proportion of 'false alarms', when non-existent stimuli appeared to occur, and a proportion of 'misses', when genuine stimuli remained undetected.

The lowest part of Fig. 1.1 shows how this proportion of errors (false alarms and misses) changes with stimulus strength, assuming that in each case the criterion has been adjusted to minimise the total number of errors. Even with zero signals, half the answers are right by guesswork, and the proportion of errors decreases as the stimulus is increased. The scale of stimulus strength is the same as in the figure showing response amplitude, directly above, but since the variability of the maintained discharge (noise), as well as the number of extra impulses (signal), is known for each stimulus strength we can add another scale, the signal/noise ratio. The important point is that the number of errors bears the same relation to signal/noise ratio in all communication systems. This statement would require careful qualification to be exact and universal, but as an approximation one can see that signal/noise ratios of 2 or 3 are needed to reduce the errors to about 10%, and ratios of 4 or 5 to reduce them to 1%. It should also be noted that an analysis of the errors made by human subjects when performing a sensory task enables one to derive a figure, referred to as d' by signal detection theorists, that is a lower limit to the signal/noise ratio in the sensory pathway involved. This method provides the firmest means of connecting sensations and physiological mechanisms.

One can now see both why sensitivity is important for reliability, and why the simple definition of sensitivity is inadequate. If more impulses were obtained per unit of stimulus energy, this would separate the two distributions shown in Fig. 1.1*d* and at first sight would appear to improve reliability. But there would be no gain if, as a result of increasing responsiveness to the stimulus, the responsiveness to the unwanted disturbances causing the fluctuating maintained discharge had also increased, so that the breadth of the distributions increased in proportion to their separation. Whether an increase of sensitivity would necessarily cause such changes depends upon the source of the maintained discharge. It is clear that the origin of noise in sensory systems is an important question.

1.2. TRANSFORMATIONS OF SCALE AND INTENSITY

The senses have been considered so far as a set of physiological mechanisms for gathering up-to-date information, but as we follow the sensory messages centrally we would like to be able to explain the ways in which this information is utilised. How does a tennis player predict the future position of the ball so that he can return it? How does my secretary decipher my handwriting? What goes on between the ears and the mouth of a simultaneous translator? It must be realised at the outset that we can only go a very short distance towards answers to questions like these, but several clues have already come from communication engineers who handle electronic information, and it is likely that we shall get more from those who try to program computers to perform tasks of a complexity comparable to the ones mentioned above.

The ideas from electronic engineers are mainly concerned with the advantages obtained from mathematical transformations of the quantities conveying the information. To take a simple example, it is well known that the pitch of a note is determined by the fundamental frequency, F, of the source of the sound. The natural musical intervals of octave, fifth, fourth, major third and minor third,[*] correspond to simple multiples of F, namely $2F$, $3F/2$, $4F/3$, $5F/4$, $6F/5$. Now define P as the logarithm of F ($P = \log F$). In terms of P, the natural intervals correspond to constant additions, namely 0.301 for the octave, 0.176 for the fifth, and so on. Addition is a simpler operation than multiplication, so if one is interested in musical intervals there is something to be said for dealing with the transformed quantity P instead of the actual frequency F. It is therefore interesting to find that the keys of a piano are placed according to P; that is the position of a note on the piano, measured from the lowest note, is nearly proportional to the logarithm of its frequency; likewise in musical notation the vertical position of a note follows P, not F. Also, if you look at a graphical representation of the frequency response of a Hi-Fi set, you will almost always find that frequency has been plotted on a logarithmic scale. A physiological counterpart can be found in the position of maximum response of the basilar membrane.

There are perhaps even greater advantages in expressing the intensity of a sensory stimulus on a logarithmic scale, because the physical values of stimuli tend to vary over enormous ranges; furthermore relative values are usually more important than absolute intensities. Thus an acoustical frequency–response curve usually has response given in decibels, which is a logarithmic scale. It used to be thought that subjective sensory intensities were, in general, proportional to the log of the physical intensity, and this is still a useful idea to have in mind. It is certainly generally true that the added stimulus (ΔI below) required to elicit the smallest perceptible change in sensation corresponds, very roughly, to a constant fractional increase (k) in the physical stimulus (I). This can be expressed in a

[*] These are called the Pythagorean intervals and it was known in Pythagoras' time that they corresponded to lengths of a vibrating string bearing to each other the ratio of small integers. In the equal-tempered scale used on modern fixed-pitch instruments like the piano, only the octave is strictly accurate. On this scale a semitone interval is the ratio of frequencies corresponding to the twelfth root of 2: repeated 12 times this gives the octave, 2, but for the fifth (7 semitones) the result is 1.4983 instead of 1.50, for the major third it is 1.26 instead of 1.25, and for the minor third 1.189 instead of 1.2.

form known as Weber's law: $\Delta I = kI$. If it is then assumed that such a threshold change in sensation corresponds to the addition of a constant quantity of sensation ΔS, one obtains $\Delta S/\Delta I = 1/kI$ and when this is integrated one gets the so-called Fechner law: $S \propto \log I$. But this treatment can be criticised on many grounds: for instance it neglects important temporal factors in adapting to new stimulus levels, and it is far from clear that increments of sensation can be integrated to predict the value of a steady level of sensation. Furthermore Weber's law has not been found to hold at all accurately, so these relations have not proved to be useful in physiology except as approximate general guides.

A rather different type of transformation is the topographical distortion in the mapping of sensory surfaces on to the cerebral cortex: in man, the cortical area for the thumb is many times larger than that for the knee, while in the pig the snout area is larger than that for the whole front leg. Again, the cortical representation of the central part of the visual field is enormously enlarged relative to the periphery, and interestingly enough an approximately logarithmic relation is also encountered here: in the monkey the distance on the cortical surface to the representation of a point in the visual field at eccentricity $e°$ is nearly proportional to $\log e$. This may be the direct consequence of the high resolution at the fovea, for this would require proportionately more central machinery which would necessarily occupy more space. However the particular mapping that is found has an interesting property, for it is approximately 'conformal'. Lines at a fixed orientation in the visual field such as vertical have an orientation in the map that is not constant but varies with position in the visual field. However, in spite of this distortion the angle between two short lines lying close to each other is unchanged; as a consequence the shape of an L, for instance, is preserved, and this may have functional importance.

The transforms so far considered are monotonic distortions of the physical scales encountered in specifying sensory stimuli. Much more radical transformations are known which would change the whole scheme of representation of the sensory stimuli. Though we do not know for certain what reorganisations of sensory information *actually* occur in the brain, it is important to realise that they can occur, and the next section describes the Fourier transformation. This has long been recognised as important in hearing, and the application of the same ideas to vision has attracted much attention.

1.3. FOURIER TRANSFORMS AND CONTRAST SENSITIVITY FUNCTIONS

The basic idea is that a quantity such as sound pressure, which varies in time, can be expressed as a sum of unending sinusoidal waves of constant amplitudes which persist indefinitely. Leading mathematicians disbelieved Fourier's theorem when he proposed it in 1807 and it is not easy to accept the fact that one can represent a transient event, such as a single sharp click, by a set of superimposed sine waves that have persisted indefinitely in the past and will persist indefinitely into the future. These aspects lie at the root of the utility and power of the Fourier transform, and perhaps the radical change of time scale also explains its biological relevance, but before returning to this theme some illustrations will be given.

Fig. 1.2 shows a sine wave at top, and a square wave of the same frequency at bottom. In between are the waveforms created by adding up the first three and the first eight terms of the series:

$$\sin \Omega t + (\sin 3\,\Omega t)/3 + (\sin 5\,\Omega t)/5 + \ldots + (\sin n\,\Omega t)/n$$

where Ω is the frequency in radians s^{-1} and the frequency in cycles s^{-1} is $F = \Omega/2\pi$. Notice that only odd harmonics are involved. Even with only three terms, up to $5\,\Omega$, one can see that an approach to the square wave is being made. The third figure is a good approximation, and here the highest term has a frequency 15 times the fundamental Ω, and an amplitude 1/15 as great. The square wave at the bottom has all the higher terms; notice that the amplitude of the square wave is actually less than that of the fundamental at the top; the ratio of their peaks is $\pi/4$.

In this example only odd harmonics were involved because only these happen to be required to synthesise a square wave. Changing the coefficients (1/3, 1/5, 1/7, etc.) would obviously change the shape of waveform that was formed, and the variety of synthesisable waveforms can be further increased in three important ways.

(1) The even harmonics (2Ω, 4Ω, etc) can be included: this will have the slightly unexpected effect of making the top and bottom halves of the waveforms unlike each other. For instance in order to synthesise a rectangular wave with the flat tops shorter (or longer) than the flat bottoms one would require even harmonics.

(2) Cosine as well as sine terms can be included: both terms are necessary when the waveforms to be synthesised do not have mirror symmetry about a vertical axis.

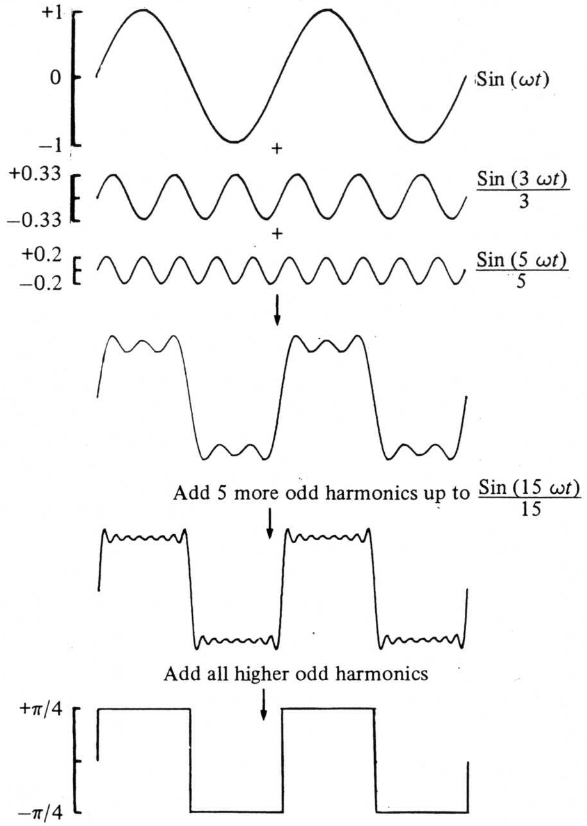

Fig. 1.2. Synthesis of a square wave (bottom) from sine waves. Any repeating waveform can be represented as the sum of the sine and cosine waves of appropriate amplitudes at the repetition frequency of the waveform and integral multiples of it. For a square wave, the coefficients defining the amplitudes of the sine waves are 1, 1/3, 1/5, 1/7, ... for frequencies 1 ×, 3 ×, 5 ×, 7 ×, ... the repetition frequency; all others (i.e. cosine waves and all even multiples of the fundamental) are zero. The figure shows the result of adding the first three and the first eight terms of the series. Notice that the square wave synthesised by the full series has an amplitude 0.785 (i.e. $\pi/4$) times that of the fundamental sine wave.

(3) The value of the fundamental F ($F = \Omega/2\pi$) can be lowered, and the number of harmonics increased. In Fig. 1.2 the frequency chosen for F was that of the original square wave that was to be synthesised, but exactly the same approximations could have been achieved by choosing another frequency, say 1/10 of F, and then picking out tenth, thirtieth, fiftieth, etc., harmonics of appropriate amplitude. By lowering the fundamental a large number of harmonics are introduced that are not required for that particular square wave and have zero coefficients; they are available, however, and they could be used in an interesting way, for they allow the synthesis of a waveform that repeats every $10/F$ sec, instead of every $1/F$ sec as with the square wave in Fig. 1.2. Thus a single cycle could be synthesised, with nothing happening for the next nine cycles, and then a repeat of the single cycle. If this process were carried to the limit, one could represent an event which *never* repeated, but this would require that F be zero and the harmonics would have to be spaced at infinitesimal intervals. It is only in this sense that one can represent a single click by a set of unending sine waves, but that is remarkable enough.

There is an obvious parallel between the mathematical representation of a complex waveform as a set of superimposed sine waves, and the musician's ability to decompose a chord into its component notes. This parallel should not be pushed too far, but the important lesson of the Fourier transform is that one can represent a rapidly changing sound pressure as the superposition of functions, in this case sine waves, that have a fixed, unchanging, description. Sine waves of constant amplitude are not, however, the only possible set of functions, called 'basis functions', that one can use for such a decomposition, and for the ear the most realistic basis functions would be sine waves that die away after half a dozen or so vibrations. Since these do not persist indefinitely, transformation using them does not give a static, unchanging, representation of a waveform, but one that changes in time. However, the important point is that it changes more slowly than the sound pressure itself; the time scale of the representation has been slowed down to a point where nerve fibres can transmit it and the nervous system react to it. That is presumably the main biological advantage derived from analysing sounds in terms of frequency rather than as instantaneous sound pressure.

Fourier-type transformations are not confined to one-dimensional variables like sound pressure, but can be employed on two-dimen-

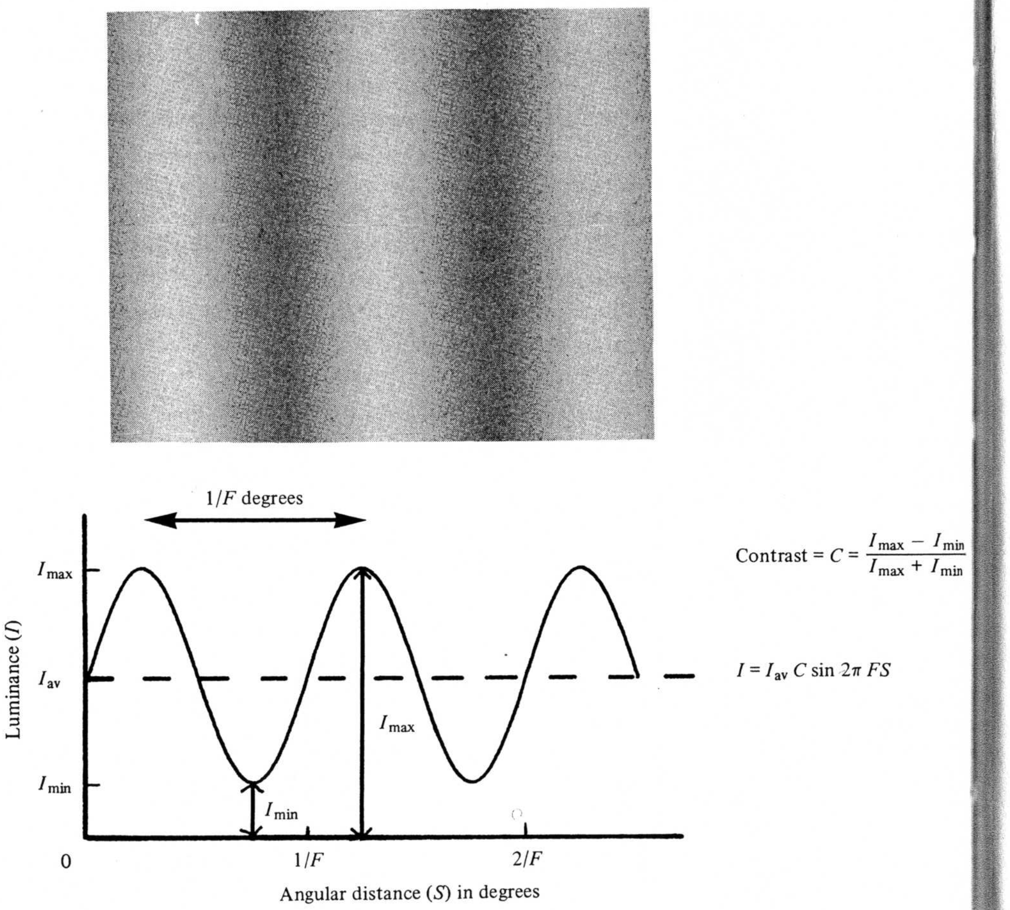

Fig. 1.3. A spatial sinusoid (top) and a cross-section of the luminance along it. The amplitude of modulation is expressed as *contrast*, given by $(I_{max} - I_{min})/(I_{max} + I_{min})$. The frequency, F, is usually expressed in cycles per deg, so one period of the grating is $1/F$ deg. At 57.3 cm from the eye 1 cm subtends 1 deg. At 6 m this grating would have spatial frequency 3.6 cycles deg^{-1}, and you could detect a contrast of about 0.003 (contrast sensitivity $1/0.003 = 333$, see Fig. 8.1).

sional variables like the pattern of luminances that make a picture. A photograph can in theory be represented by the sum of a harmonic set of sinusoidal spatial waves that are uniform over the whole picture area and run across it in a variety of directions. (These wave patterns must not of course be confused with the wave nature of light itself, where frequency determines colour.) A spatial sinusoid is shown at the top of Fig. 1.3, with the light luminances along a cross-section below. Note that luminance must always be positive, so these spatial sinusoids are modulations of an average luminance. It is convenient to refer to the *contrast C* of such a grating, which is the maximum deviation from the average luminance divided by the average luminance; this is given by

$$C = (I_{max} - I_{min})/(I_{max} + I_{min}).$$

The frequency of a spatial sinusoid in a visual experiment is usually given in cycles deg^{-1}. This is simply the number of complete cycles of the grating that subtends 1 degree at the eye; the grating shown in Fig. 1.3, would have a frequency of 3.6 cycles deg^{-1} when viewed from a distance of 6 m.

Since any picture can be broken down into sinusoidal components like this it is instructive to see how sinusoidal gratings are handled by an optical system. Fig. 1.4 shows this for the human eye in a form known as the *Modulation Transfer Function* (MTF). Suppose one looks at a sinusoidal grating, such as that of Fig. 1.3 but with a contrast of 1.0. If the optics are good one naturally expects the image on the retina to preserve the full contrast, but this is not the case; the ordinates show the contrast in the image for different spatial frequencies, and it will be seen that the high spatial frequencies suffer severe loss of contrast, as shown by the decline of the curve to the right. Indeed there is a total loss of contrast for spatial frequencies above a certain limiting value, known as the *limiting resolution*: spatial sinusoids of frequency above this value are completely flattened out.

One might think that an MTF like that of Fig. 1.4 represents a poor performance, but this is not so at all. The dotted line shows the best possible performance for an optical system with 2 mm entrance pupil, the diameter for which the eye's optical performance is best. The limiting performance is set by diffraction of light at the pupil; the period (in radians) of this *cut-off frequency* is actually given by λ/D, where λ is the wavelength of light and D the diameter of the pupil. The eye does not perform as well as this perfect instrument, but it is not much worse and the major part of the loss at high frequencies is unavoidable.

Fig. 1.4 was obtained by an objective technique described in Chapter 3 in which the quality of the image falling on the retina is determined from the small fraction of light that is reflected back out

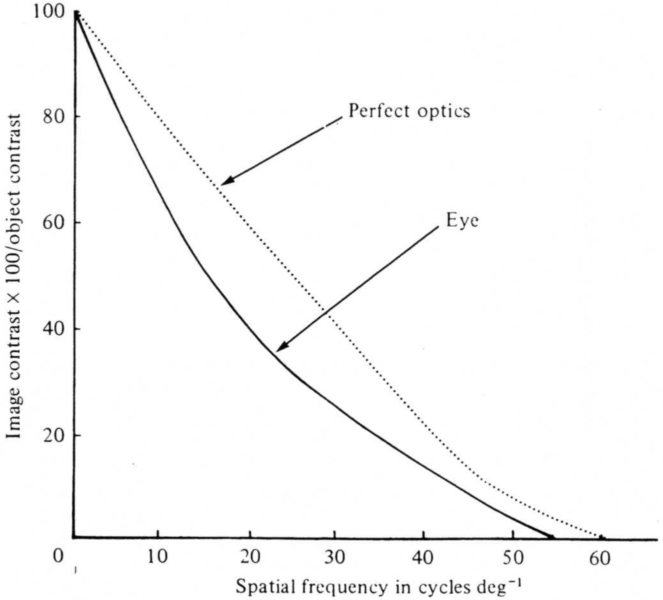

Fig. 1.4. The modulation transfer function (MTF) of the human eye with 2 mm pupil compared with a diffraction-limited optical system of the same aperture. The curve shows the contrast in the image formed of a spatial sinusoid of 100% contrast. Spatial frequencies of both object and image are expressed in cycles deg^{-1}. It will be seen that there is a progressive loss of modulation in the image as the spatial frequency is increased. At 62 cycles deg^{-1} for perfect optics, all contrast is lost; this is the cut-off frequency resulting from diffraction for this aperture. The eye loses all modulation just below this limit at 55 cycles deg^{-1}. (From Campbell & Gubisch, 1966.)

of the eye again. One would expect the optical loss of contrast at high spatial frequencies to make them less easily visible, and this will be discussed in Chapter 8, where Fig. 8.1 shows a *contrast sensitivity function*. To determine this a subject looks at a spatial sinusoid of variable contrast which is adjusted until it is barely distinguishable from a uniform, unmodulated grey. The contrast sensitivity is the reciprocal of this setting, so if at threshold $(I_{max}-I_{min})/(I_{max}+I_{min}) = 0.01$, the contrast sensitivity is 100. Note that in this figure the scales for both frequency and contrast are logarithmic, whereas linear scales were used in Fig. 1.4.

Modulation transfer functions are a very good way of characterising the performance of physical instruments such as camera lenses and

television systems, because they give such a complete description of their performance. To a large extent the same is true for contrast sensitivity functions, and for this reason sensory thresholds have been measured for many types of sinusoidal stimulation. The reader will come across curves characterising the eye's ability to detect flickering lights (temporal contrast sensitivity curves), lights of varying wavelength (spectral sensitivity curves), how the ear detects sounds of different frequency (audiograms), and how the vibration of skin is sensed. The technique has also been used on isolated components of sensory pathways such as the retinal ganglion cells, or cortical neurons in visual, somatic or auditory sensory areas. However, a word of warning may be needed when interpreting these curves because it is only in linear systems* that knowledge of the frequency response gives complete information about the response to an arbitrary stimulus: the brain is certainly not a linear system, so the Fourier transform does not necessarily give important insight into how the brain interprets sensory messages. Nevertheless the theoretical possibility of decomposing pictures into spatial sinusoids of different frequency and orientation introduces the idea of representing sensory messages in unfamiliar, non-obvious ways. If a picture, or a sound, can be represented as the sum of a set of sinusoids, we need no longer think of the representations in our brains as *copies* of the picture or sound; we must be prepared for surprises and subtlety in the transformations of sensory messages in the nervous system.

1.4. SELECTIVITY FOR PATTERN; TRIGGER FEATURES

Some of these surprises were encountered when sensory pathways were investigated by recording the activity of single neurons more central than those connected to the receptors themselves, for such neurons were found to have unexpected properties resulting from the complicated interactions that occur at the synapses leading to second and higher order neurons. The important new property achieved as a result of these synaptic interactions is selectivity for *pattern*, rather than for simple physical or chemical characteristics of the stimulus. Moreover, much of the complexity of the mechanism is concerned

---

* The linearity referred to is between stimulus strength S, and response magnitude R, i.e. $R = KS + \text{Constant}$ for the condition to hold. When this is true, 'superposition' holds, so that the added response ΔR resulting from an added stimulus ΔS is the same regardless of any steady level of stimulation.

with achieving *invariant* responsiveness to a specific pattern, in the face of variations in the physical attributes of the stimulus and its position.

Retinal ganglion cells

These properties of pattern selectivity and invariance of response are best illustrated by examples from the visual system, because it was there that the importance of pattern excitation was first discovered, and where the analysis has progressed furthest. The technique is to place a recording microelectrode in the ganglion cell layer (see Fig. 2.2*b*) and move it about until the action potentials from a single ganglion cell can be picked up separately from those of its neighbours. A search is then made for a region in the visual field where excitation by light causes a response. Most of the ganglion cells have a maintained discharge rate of 5 to 60 impulses s^{-1}, and the effect of stimulation is to increase or decrease the firing rate. The first step is usually to turn a small spot of light on and off and mark the positions in the visual field from which responses are elicited, thus mapping the *receptive field* of the particular ganglion cell one has isolated. Fig. 1.5 shows examples of such maps; a plus sign means that a spot turned on at that position increases the firing rate, and a minus sign means that it decreases the firing rate; when a decrease occurs when the spot is turned 'on', there is usually a rebound above the maintained discharge rate when it is turned 'off'. An '0' means that no change occurred with a spot at that position, and this would also be so outside the 0s. It is usually found that plus (or 'on') and minus (or 'off') regions inhibit each other so that a light turned either on or off simultaneously on a pair of such regions causes a smaller response than the same stimulus applied to reach region separately.

Now consider the maps at the top of Fig. 1.5. To the left a roughly circular region of the visual field subtending about 1° at the eye has plus signs in it, and this is surrounded by minus signs. The field shown to the right has the opposite arrangement, minus signs surrounded by plus signs. Receptive fields with this concentric arrangement of regions that respond in the opposite sense are very common in the retinae of all species, and the same arrangement is found in second and higher order neurons of other modalities such as touch and hearing. It was first discovered by Hartline in his investigation of the compound eye of the primitive crustacean *Limulus*, and is called *lateral inhibition*. It has the effect of making the cell compare the amount of light at one position with the amount at neighbouring

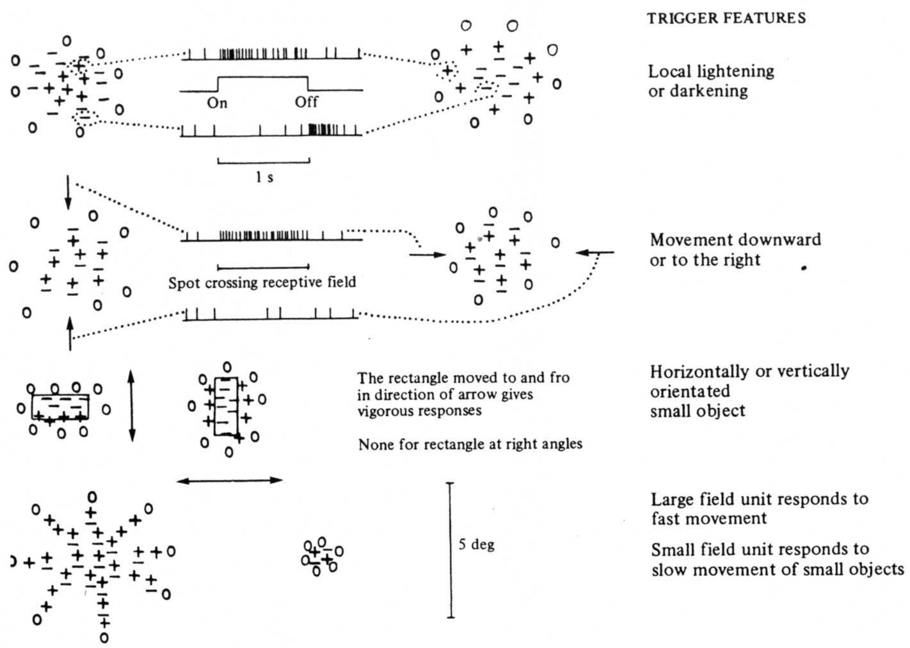

Fig. 1.5. Receptive fields and trigger features of five types of ganglion cell found in the retina of rabbits. + and − signs mean that a stationary spot of light turned on at that position causes an increase (+) or decrease (−) of the maintained discharge; when it is turned off the opposite change occurs, namely a decrease (at +) or an increase (at −). Examples of these responses with indications of the receptive field regions that would produce them in an on-centre and off-centre ganglion cell are shown in the centre. No responses result from stimulating at positions marked 0, or outside the ring of 0s. Ganglion cells with receptive fields in which 'on' (+) and 'off' (−) zones are concentrically arranged, as in the top pair, are commonly found in many species, and they respond best to small areas that are lighter or darker than the surrounding region in the visual field. However, maps of the receptive field made with stationary spots often fail to reveal important pattern-selective properties of the ganglion cell. When mapped in this way the second pair of ganglion cells responded at both 'on' and 'off' throughout their receptive fields, but testing with a moving spot showed that the left-hand one responded best to downward movement and not at all to upward movement, while the right hand one responded to rightward movement and not to leftward movement; this is indicated in the second pair of responses in the centre. The ganglion cells with the other four receptive fields mapped here also responded selectively to specific trigger features (see text).

regions. The signal from a cell with an 'on' centre thus indicates that the illumination of a particular region in the visual field is greater than the illumination of its neighbours; the cell with minus signs at its centre correspondingly signals that the illumination is less than in neighbouring regions, i.e. it is darker. It is important to realise that the retina is densely covered with ganglion cells of these two types, and that their receptive fields overlap and collectively cover the whole visual field.

The map of the receptive field is a useful preliminary stage in analysing high-order neurons in sensory pathways and one can proceed in three ways. The first is to analyse the cellular mechanisms whereby selectivity is brought about, and some success has been achieved along this path by recording intracellularly from the cells of the retina (see Chapter 6).

The second way to proceed is to make a map in the Fourier or frequency domain: instead of determining the sensitivity to spots of light in different spatial positions the sensitivity to spatial sinusoids of different frequencies and orientations is measured. This approach is described in the chapter on spatial resolution and analysis.

The third approach is to make explorations of the response properties of the neurons when faced with the kind of visual stimuli encountered in real life, in the hope of discovering what part these neurons ordinarily play in perception. These concentric cells respond very vigorously to white (for the on-centre type) or black (for off-centre type) spots of appropriate size moved into their receptive field centres. They also respond well to white or black bars of about the same width as their field centres, and to black–white edges crossing their fields, but they do not respond well to uniform illumination. It is probably correct to regard them as performing the first step in emphasising borders and edges in the visual image. However, one could have reached this conclusion equally well by inspection of the receptive field map, or even from the contrast sensitivity function; the importance of further qualitative exploration in revealing the pattern selectivity of cells only becomes fully evident with the other types of retinal ganglion cell. The retina of the rabbit provides some good examples.

Consider the receptive field map for the unit shown in the second row in Fig. 1.5. When explored with a spot turned on and off it gives a brief discharge at all the points marked with a plus and minus sign, and no responses elsewhere. It was only by exploring with a *moving* spot that the striking result shown by the records was found. A small spot of light moved downwards through the receptive field gives a vigorous burst of impulses sustained for the whole time the spot moves in the field; however, when the spot retraces its path in the opposite direction, only a few impulses occur. Other examples of this class of cell respond preferentially to motion to the right, as shown in a second receptive field map. Yet others prefer movement in other directions. As with all classes of ganglion cell there are large numbers of them, and together these cover the whole visual field for all four directions of movement with their overlapping receptive fields.

Invariance of pattern selectivity

The directional selectivity of such retinal ganglion cells persists in the face of large changes in the physical properties of the stimulus (see Fig. 1.6). First one can change the overall luminance level by a factor of 10000 or more, and the retinal ganglion cell will persist in responding to an object moving in one direction but not in the reverse. Second, one can keep the background illumination constant and vary the contrast of the spot; the ganglion cell responds preferentially to the same direction of movement, regardless of the brightness of the spot, and it even persists in that preference when a dark spot is used instead of a light one. Finally a small stimulus spot can be moved to and fro almost anywhere within the receptive field and the cell will respond to the same direction of movement at all positions. Thus it can be shown that a ganglion cell's directional selectivity is invariant over a wide range of luminances and contrasts, and also over the limited range of positions in the visual field covered by its individual receptive field.

One naturally starts off by thinking that the physical properties of the light stimulus must be the important variables determining the response of a retinal ganglion cell. Fig. 1.6 shows that this is too simple a view, for a response occurs only when the spatio-temporal *pattern* of stimulation characteristic of movement in a particular direction occurs; the light intensities and the position are relatively unimportant.

One needs a term to describe the pattern of stimulation that would normally excite such a cell, and this is sometimes called its *trigger feature*. In the rabbit's retina further exploration reveals ganglion cells with a wide variety of trigger features. For instance velocity of movement is an important variable for some cells which respond vigorously to objects moving so rapidly that they fail to excite other cells. These are called 'fast movement detectors' (see Fig. 1.5). In the

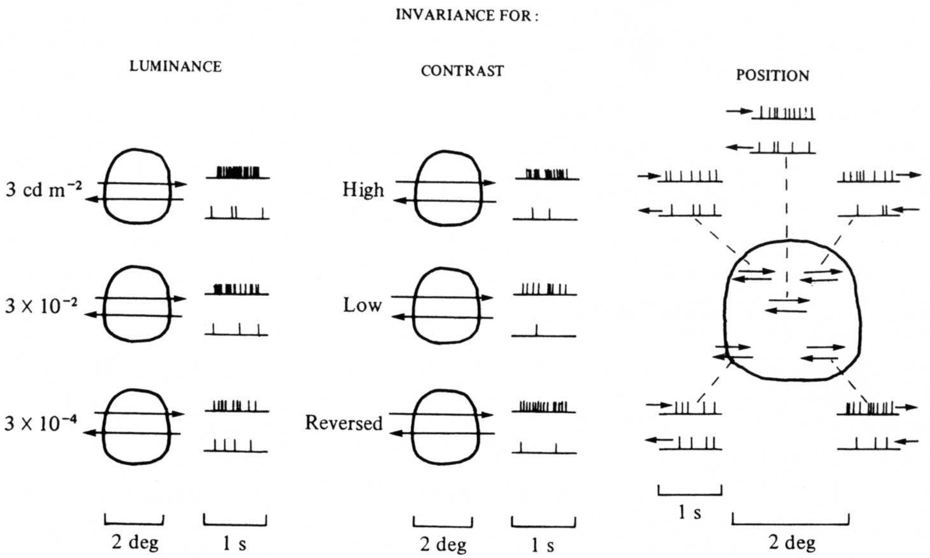

Fig. 1.6. Invariance of pattern selectivity over luminance, contrast, and position. This retinal ganglion cell responded well to rightward movement and poorly or not at all to leftward movement over a wide range of luminances, contrasts, and at most positions within its receptive field. Only the outline of the receptive field is shown here. The left column shows responses for two directions of movement at three different luminances. The centre column shows them for spots of high, low and reversed contrast. The right hand part shows responses for to-and-fro motion confined to different subdivisions of the receptive field. In all cases, rightward movement causes a vigorous response, leftward movement little or none.

region of the rabbit's retina that is normally aligned with the horizon one can find cells that respond best to elongated dark or light objects, and also cells which refuse to respond to anything but very slowly moving objects.

This list of trigger features of ganglion cells found in the rabbit is not complete. Some known types have been omitted and there are types of cell, for instance those subserving colour vision, which we are confident the rabbit possesses, but which have only recently been recorded from neurophysiologically. Incomplete though it is, the above description illustrates the extraordinary diversity of pattern-selective elements that may be found at a low level in the visual pathway of a single species. Knowledge of the *anatomical* diversity of ganglion cell types was actually available almost 100 years ago

from Ramon y Cajal's histological work on the retina, but only now are we beginning to correlate the physiology with the anatomy. A start has also been made in uncovering the mechanisms whereby pattern selectivity is achieved, but here too there is a long way to go.

Knowledge of the diverse types of ganglion cells and trigger features is less complete in other species. The most prominent ganglion cells in the cat retina are the concentric type, which were in fact first discovered there by Kuffler before the work on rabbit was done. Much later it was found by testing with sinusoidal gratings that these concentric cells are not all alike but belong to two subtypes called X or sustained, and Y or transient. The properties of these subtypes are discussed further in Chapter 8. Later still a host of other types, collectively called W-cells, were discovered, and some of these appear to be similar to those in the rabbit; their properties, however, are less well known because they are harder to record from. Colour opponent cells are prominent in monkey (see Chapter 9) and the division into sustained and transient types probably also holds; it is not known whether the retinas of monkey or man possess the more strongly pattern-selective elements of the rabbit.

Neurons at higher levels in the visual pathway

Hubel and Wiesel pioneered the study of neurons at more central points in visual pathways, starting with cats and then going on to Rhesus monkeys. At the level of the lateral geniculate nucleus (see anatomical plan of visual pathway, Fig. 2.5) the predominant type of neuron has a concentrically arranged receptive field as at the retina, and we know from subsequent work that the division into sustained (X) and transient (Y) subtypes is maintained. Fig. 1.7 shows these receptive fields diagrammatically. At the next level, in the primary area of visual cortex (area 17), new properties appear; the single cells here require an elongated stimulus at a particular orientation for optimum response. Different cells require different orientations, and in addition the stimulus often has to be moved in the right direction and has to be about the right size. Furthermore this is the first point in the pathway where cells are found that can be stimulated through either eye.

As well as describing these new properties of orientational selectivity, directional selectivity, size selectivity and binocularity, Hubel and Wiesel divided the cells into two classes they called 'simple' and 'complex'. The main basis for the distinction was whether a map of the receptive field, with its 'on' zones and 'off' zones, gave evidence

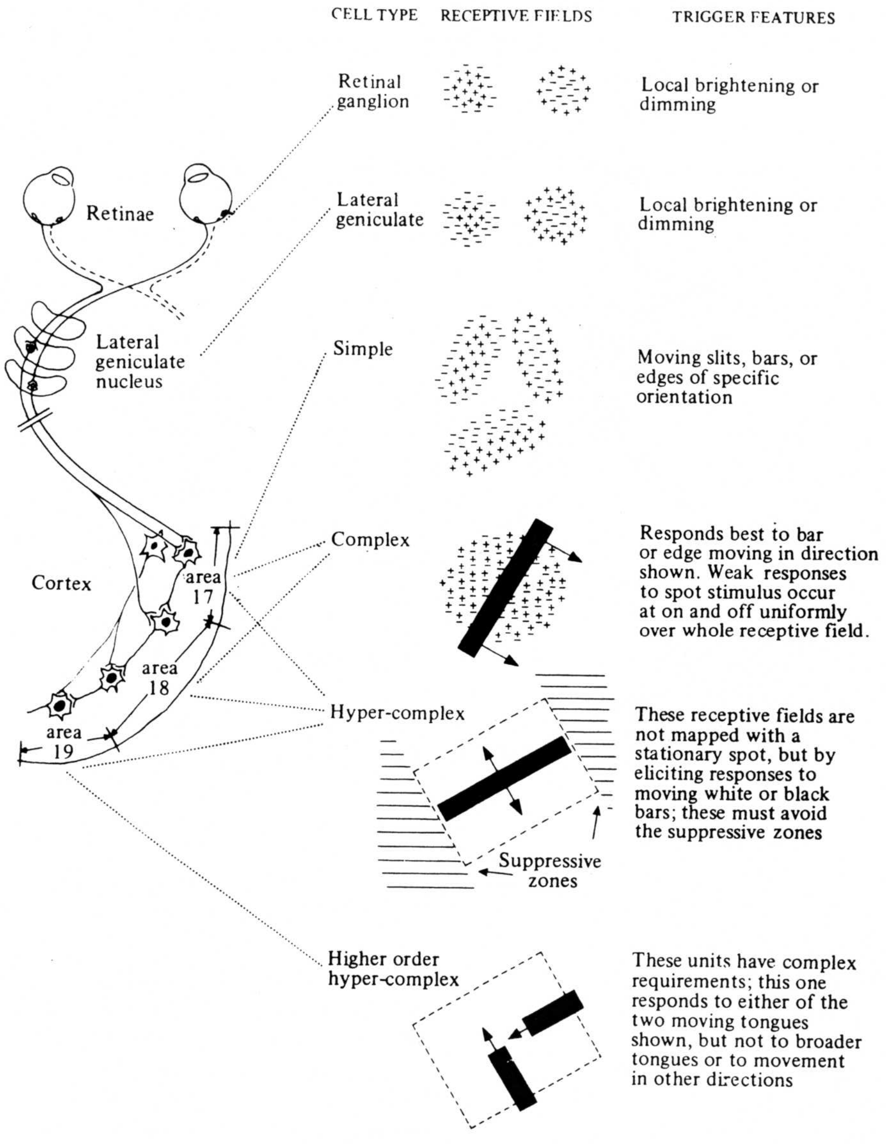

Fig. 1.7. Cell types and trigger features at different positions in the visual pathway of the cat, shown diagrammatically. Monkey cortex is mainly similar (see text). The elongated receptive fields in area 17 are sensitive to bars and edges of the appropriate orientation and size. This pattern selectivity results from the excitatory and inhibitory connections to cortical neurons made by lateral geniculate neurons (LGN) with concentric receptive fields. The more complex properties of cells in areas 18 and 19 result from connections made to these areas from cells in area 17, though there is also a direct input to area 18 from LGN.

of their pattern-selective properties. This was the case for simple cells; for instance an elongated 'on' zone showed the orientation and approximate size of the bright bar that would optimally stimulate it. Complex cells, however, often gave both 'on' and 'off' responses throughout their receptive fields, and even when this was not so their orientational or directional selectivity could not be predicted from the map alone (compare the directionally selective cells of the rabbit's retina).

At first Hubel and Wiesel thought that the complex cells received their input solely from the simple cells, but it is now thought that there is a direct input from the lateral geniculate nucleus (LGN). Thus it seems likely that there are parallel X and Y pathways from the retina through to the primary cortex. This discussion will be continued in Chapter 8; the significance of this dual system may be that the X system is concerned with analysis of form, whereas the Y or transient system is concerned with movement and hence with the precise timing of visual stimuli. One must also remember the W system in cats; these neurons are thought to project to the cortex but their influence has not yet been worked out.

Next to area 17 lie areas 18 and 19. These are often called secondary visual areas, though in the cat area 18 receives a direct projection of Y type fibres from the LGN. These areas have neurons with more complex type of selectivity than those of area 17, and two of these are shown in Fig. 1.7.

The situation in the monkey is broadly similar to that in the cat; two differences worth noticing are the reduced direct projection from the LGN to area 18, and the presence of an extra stage of densely packed cells with concentric receptive fields in a subdivision of layer IV of area 17. These cells are all monocular, and it is only in other layers that binocular neurons are found.

It has long been known that there is an orderly mapping of the retina on to area 17, so that the cells in one part of the cortex have their receptive fields lying close to each other in one part of the visual field. Hubel and Wiesel discovered an orderly microstructure within this mapping. Cells in the monocular lamina of layer IV are arranged in alternating strips about 0.5 mm wide, one strip connecting to one eye, the other to the other. Furthermore, cells of different orientation selectivities are not arranged at random: they lie in orderly sequences of gradually changing orientation, a shift through 180° occurring in about 0.5 mm. Some of this microstructure is pictured in Chapter 2, Fig. 2.6.

This work on the cortex has led to interesting work on development, for the properties of the cortex at birth in monkeys, or at eye-opening in kittens, are not altogether like those of adults. Furthermore deprivation of vision during the first three months by rearing in darkness or by suturing the eyelids interferes with the development of the cortex, and this sensitivity to deprivation of normal visual input during early life is thought to be the causative factor in human *amblyopia*, and also the much commoner condition of *stereoblindness*. These conditions are discussed in Chapters 10 and 20, and the role of area 17 in the preliminary sorting of visual stimuli according to their disparity is also discussed in Chapter 10.

It is natural to ask 'What is the role of the primary visual cortex, area 17, in the processing of sensory information?'. The primary visual cortex certainly acts as a redistribution centre, and this is the clue to its six-layered lamination, for the cells are largely segregated according to the destination of their axons. Thus many lamina V I cells send their axons to the thalamus, including the LGN, whereas many lamina V cells send axons to the superior colliculus, the mid-brain visual centre. Lamina IV contains the granule cells on which incoming fibres from the LGN mainly terminate, and these cells and those of laminae II and III send axons to other regions of the cortex. These other regions include secondary visual areas 18, 19 and 20 (also called V II, V III and V IV), and also the contralateral visual cortices by way of the corpus callosum.

Unfortunately there is no agreement about what is achieved by the selectivity of cortical neurons and the redistribution described above. In contrast, the pattern-selective properties of the rabbit retina make sense almost intuitively, for a small herbivore that is popular fodder for many carnivorous predators *needs* special ganglion cells to detect the slowly stalking beast approaching over the horizon, or the suddenly swooping hawk descending from above. But the specificity of many of these retinal cells is as great or greater than that of the cortical neurons, and it is achieved after two or three synapses instead of five or six for most neurons in area 17.

The more orderly arrangement of cells according to eye preference and orientation preference suggests a more methodical analysis of the information that reaches the cortex, and perhaps the most plausible hypothesis is that its role is to detect *linking features* such as motion, colour, disparity, and texture which are local properties of each part of the image. The multiple secondary visual areas (V II, V III, V IV) surrounding the primary cortex, and to which V I relays information, would then have the task of using this information to identify the parts of the visual scene that possess a common linking feature and thus probably belong to an object that is to be recognised. Such segregation of *figure* from *ground* was a familiar theme of the Gestalt school of perceptual psychologists who attacked the problem of how we see objects as a whole rather than as a multiplicity of separate parts. It is a tempting hypothesis that the primary cortex detects the characteristics (linking features) that allow this synthesis to be achieved, but it will require more experimental knowledge of the secondary areas to put the hypothesis to the test.

Finally it is worth remarking that the cortex, which is thought to be responsible for man's intellectual pre-eminence, is a remarkably homogeneous structure throughout its extent. In the visual system we have progressed further than anywhere else towards discovering the mechanisms of some simple examples of pattern selectivity, such as the detection of motion, orientation, and disparity. However this is only a beginning and there is much more to be discovered before we gain the insight into its overall operation that the uniformity of structure of the cortex tempts one to believe may be attainable.

1.5 CONSCIOUS PERCEPTION

Some account must be given of the relation between the information reaching the brain and our conscious perceptions, if only to dispel the natural belief that the relationship is a simple one. Of course stimulating sense organs does usually result in a conscious perception, but the relation should not be regarded as a direct, causal one. The occurrence of an earthquake may result in the appearance of headlines in the newspaper, but the causal chain is complicated and involves so many conditional factors that it would be a bit absurd to say that the earthquake caused the headline. The same is true for sensory excitation causing perception. Conscious perceptions are best regarded as *interpretations*, made in the light of previous experience, of the evidence provided by our sense organs. This interpretation occurs unconsciously and the existence of this step is apt to be denied, for one instinctively places great reliance on the validity and directness of perceptions. But one cannot persist in this denial after experiencing illusions (see Chapters 7 and 12), and those such as the tilt after-effect (Fig. 7.1) and others requiring pre-adaptation (Fig. 8.7) give a very direct demonstration of the importance of immediately previous experience; it becomes a little easier to accept that a life-time of previous experience must influence what one perceives.

Our perceptions, then, are not always valid and they are not the direct appreciation of the environment; they are interpretations of sensory messages, and this has important consequences. For instance two people will often given different reports when they witness the same scene, not because one is a liar, unobservant, or perverse, but simply because the past experiences of the two people are different and hence their interpretations in the light of it lead to different results; in other words, they genuinely have different perceptions of the scene. One need not cease to accept that 'seeing is believing', but one comes to realise that seeing is *only* believing, and beliefs are based on prejudice as well as fact.

It was said at the beginning of this chapter that the patient's and the doctor's senses provide the most important, potentially the most reliable, and the most up-to-date information about an illness. That is true not only for the diagnosis of an illness, but for all situations where you use the evidence of your own senses and those of others. However, one needs to be aware of the complexity of the sensory pathways and the fallibility of perception in order to make optimal use of this evidence.

1.6. SUGGESTIONS FOR FURTHER READING

General references

Wiener, N. (1948) *Cybernetics, or control and communication in the animal and the machine.* The Technology Press. New York: John Wiley. (A look at biological control mechanisms by the mathematician who invented cybernetics; inspiring but not very accurate.)

Cherry, C. (1957) *On human communication.* New York: Science Editions Inc., MIT Press and John Wiley. (An engineer's viewpoint on human communication; discursive but instructive.)

Pierce, J. R. (1962) *Symbols, signals and noise.* London: Hutchinson. (Another engineering viewpoint biased towards psychology.)

Woodward, P. M. (1953) *Probability and information theory with applications to Radar.* Pergamon Press. (A very clear and concise account of the basics of information theory.)

Swets, J. A. (ed.) (1964) *Signal detection and recognition by human observers.* New York: John Wiley. (A collection of early papers on signal detection theory.)

Laming, D. (1973) *Mathematical Psychology.* London and New York: Academic Press. (Mathematical rigour applied to problems of decision and choice.)

Barlow, H. B. (1972) *Perception*, **1**, 371–94. (A speculative attempt to account for the performance of complex perceptual tasks in terms of known properties of nerve cells.)

Mollon, J. D. (1977) Neurons and neural codes; Neural analysis. Chapters 3 and 4, in *The Perceptual World*, ed. K. Von Fieandt & I. K. Moustgaard. London and New York: Academic Press. (Discussions of the relation between nervous activity and perception.)

Specific topics

Photometry. Le Grand, Y. (1968) *Light, colour and vision.* London: Chapman and Hall.

Human signal/noise discriminations. McNicol, D. (1972) *A primer of signal detection theory.* London: Allen and Unwin.

Reliability and noise of nerve responses. Barlow, H. B. & Levick, W. R. (1969) *Journal of Physiology*, **200**, 1.

Physical limits of vision. Barlow, H. B. (1964) in *Photophysiology*, ed. A. C. Giese. New York: Academic Press.

Cortical maps. Adrian, E. D. (1947) *The physical background of perception.* Oxford: Clarendon Press.

Maps of visual field on cortex. Cowey, A. (1979) *Quarterly Journal of Experimental Psychology*, **31**, 1.

Fourier transforms. Bracewell, R. (1965) *The Fourier transform and its applications.* New York: McGraw-Hill.

Modulation transfer function. Hopkins, H. H. (1962) *Proceedings of the Physical Society (London)* **79**, 889.

Optics of the human eye. Campbell, F. W. & Gubisch, R. W. (1966) *Journal of Physiology*, **186**, 558.

Receptive fields and trigger features in rabbit retina. Barlow, H. B., Hill, R. M. & Levick, W. R. (1964) *Journal of Physiology* **173**, 377.

Invariant properties of trigger features. Maturana, H. R., Lettvin, J. Y., McCulloch, W. S. & Pitts, W. H. (1960) *Journal of General Physiology* **43**, Suppl. 2, Mechanisms of Vision, 129.

Histology of retina. Dowling, J. E. (1970) *Investigative Ophthalmology*, **9**, 655.

Sustained or X and transient or Y ganglion cells Lennie, P. (1980) *Vision Research*, **20**, 561.

Responses of cells and organisation of visual cortex. Hubel, D. H. & Wiesel, T. N. (1977) *Proceedings of the Royal Society* B **198**, 1.

Mechanisms of pattern selectivity in retina. Barlow, H. B. & Levick, W. R. (1965) *Journal of Physiology* **178**, 477.

Perception as interpretation. Gregory, R. L. (1973) (with Gombrich, E. H.) in *Illusion in Nature and Art.* London: Duckworth.

Linking features and Gestalt perception. Barlow, H. B. (1981) *Proceedings of the Royal Society* B, **212**, 1–34.