

BACKGROUND TO ACTIVE VISION

2.1 Introduction

In this chapter we shall discuss features of the visual and oculomotor systems that are particularly important for understanding active vision. The accounts of both systems will be highly selective and specific to our perspective. Many detailed reference works are available (Carpenter, 1988; Cronly-Dillon, 1991; Wurtz and Goldberg, 1989; are good sources).

We argued in Chapter 1 that the passive vision approach contains many pitfalls. While the existence of a fovea may be acknowledged at some point in such accounts, its importance is very often downplayed. Many discussions of visual perception make the implicit assumption that the starting point is a homogeneous 'retinal' image. We suggest this approach is misguided for at least three reasons. First, and most obvious, it neglects a basic feature of visual physiology and psychophysics, which is that the visual projections are organised so that the projections away from the central regions are given uniformly decreasing weighting. Second, it frequently leads to the assumption that the properties of foveal vision, for example faithful spatial projections, are found throughout the visual field. Third, accounts of visual perception starting from this basis frequently require supplementation with an attentional process such as a 'mental spotlight'. As we discuss in Chapter 3, we believe that this approach to visual attention is misguided.

Active vision takes as its starting point the inhomogeneity of the retina, seeing the fovea not simply as a region of high acuity, but as the location at which visual activity is centred. Moreover, vision away from the fovea must also be treated differently. Traditionally, vision away from the fovea is regarded as a degraded version of foveal vision, but serving the same purpose. In the active vision account, some visual representation is formed away from the fovea (although this representation turns out to be much less substantial than might be expected) but the major role of peripheral vision is to provide the appropriate information for subsequent orienting movements and foveal recognition.

2.2 The inhomogeneity of the visual projections

2.2.1 Introduction

In this section we shall largely be discussing the properties of vision away from the line of sight. For a single eye, the convention for specifying a location in peripheral vision is simple and straightforward. The *angle of eccentricity*, as shown in Fig. 2.1, measures the angle between the *visual axis* where the fovea is directed, and the peripheral location under consideration. Its complete specification involves both the angular distance from the fovea and the direction in the visual field, measured with reference to the axes up-down and left-right (or nasal-temporal). The term *perimetry* is used to describe the systematic measurement of peripheral vision throughout the visual field. A typical perimetric plot, as shown in Fig. 2.2, would show some visual property plotted with reference to the visual field. The projections are organised so that properties change gradually and systematically from the central fovea into the periphery, rather than with sudden transitions. This means that designation of subregions within the peripheral visual field, and even the designation of the foveal region itself, has no basis other than descriptive convenience. However it is often customary to delineate the *foveal* region, extending out to an angle of eccentricity of 1° , the *parafoveal* region from 1° to 5° , and the *peripheral* region encompassing the remainder of the visual field.

For most purposes in active vision, we can justify a treatment that is monocular. Vision evolved primarily as a distal sense. If the eyes are both directed to a point in a frontoparallel plane at 40 cm (a typical viewing distance for a VDU screen), other objects in this plane will have an eccentricity only about 1 per cent different between the two eyes. Of course when an activity involves viewing objects at different distances, considerations of retinal disparity come into play (Section 2.5.2).

Many visual functions show gradually declining ability as the stimuli are placed more eccentrically (Section 2.2.2). However, there are important exceptions.

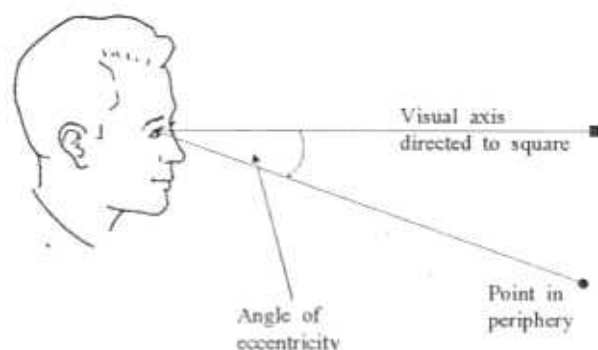


Figure 2.1 Demonstration of the angle of eccentricity.

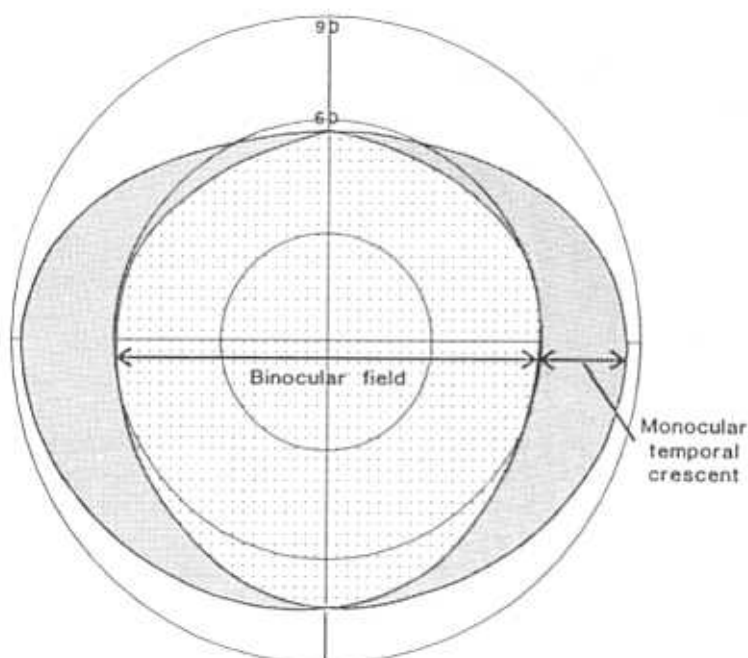


Figure 2.2 An example of a perimetric chart, in which a plot is made of some visual property at locations throughout the visual field. The centre of the plot corresponds to the foveal axis and the vertical and horizontal scales are the visual axes up/down and nasal/temporal respectively. This particular plot (from Henson, 1993) shows the extent of the visual field that is seen with each eye, and also the area of the binocular field, seen with both eyes.

Monitoring for change in the visual environment is a function of peripheral vision that has obvious evolutionary significance. Hence it is not surprising to find that some variables connected with temporal change, such as flicker and movement sensitivity, do not follow the general rule of declining abilities but actually demonstrate improved peripheral performance (Baker and Braddick, 1985).

2.2.2 Physiology of the visual projections

Anatomical descriptions of ocular structure have always provided an important launch point for visual science. Study of the retina with suitable microscopic techniques has yielded several basic facts about the specialised foveal region. First, the retinal surface is generally flat, but has a shallow pit (Fig. 2.3) of diameter about $1500\ \mu\text{m}$ coinciding with the area of acute vision. The thinning of the retinal layers occurs because, although the photoreceptors (cones) are present in their highest density within this region, the other visual cells of the retina (bipolar, horizontal, amacrine and ganglion cells) are displaced towards

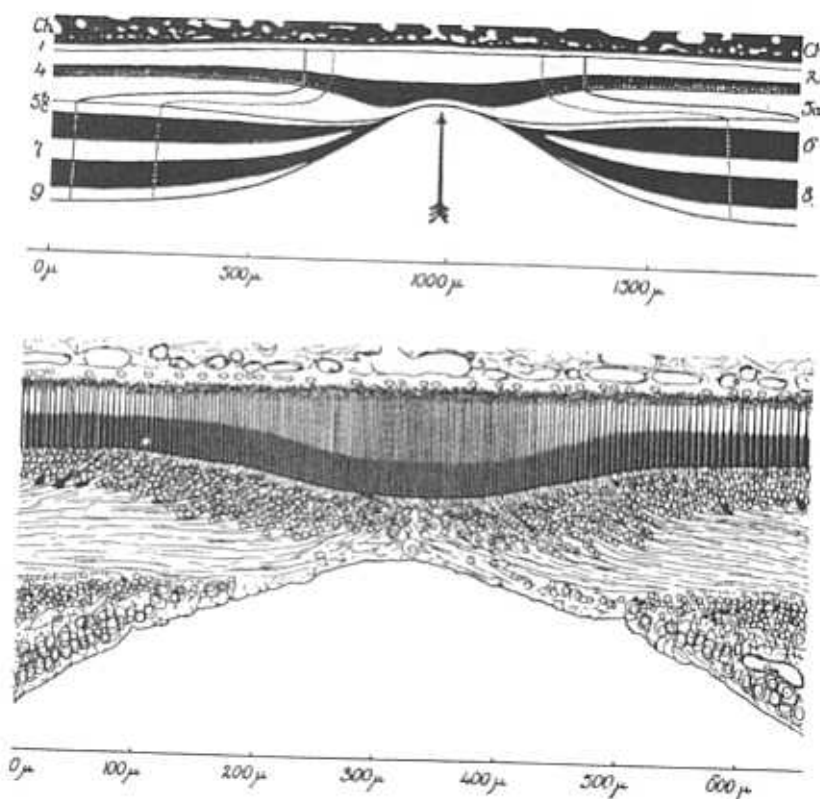


Figure 2.3 Cross-section of the human retina through the foveal area (from Polyak, 1957). Measurements on the retina are given in microns: 1 degree of visual angle corresponds approximately to 300 microns.

the periphery away from the pit. This leaves a thinner retinal layer, presumably improving the optical quality of the image on the photoreceptors. The diameter of this pit corresponds to a visual angle of 5° , according to Polyak (1957), and is thus somewhat greater than the usually accepted functional definition of the size of the fovea. The term *foveola* has been used to delineate the region in the very centre of the pit, although, as with the fovea itself, the boundary is arbitrary. In fact, cone density continues to increase to the very centre of the foveola, where the intercone spacing is about $2.5 \mu\text{m}$, decreasing to a value of $5 \mu\text{m}$ at 1° eccentricity (Hirsch and Curcio, 1989). Visual acuity appears to show a similar result, being best in the very centre of the fovea. In approximate correspondence with this depression, there is a region of retina that contains only cones. The rod-free region has a diameter of somewhat less than one degree (Hirsch and Curcio, 1989). A final related demonstration from the anatomists is that a region of yellow pigmentation is often observed over the

fovea. This yellow spot is termed the *macula* or *macula lutea*, *macular vision* being an alternative term for foveal vision.

The photoreceptors initiate the neural processing of the visual signal, which then proceeds through the retina to the ganglion cells and along the optic nerve to the visual centres of the brain. Local spatial interactions play a highly important role in the processes of adaptation and receptive field formation. However, a key feature of the visual projections is their topographic or retinotopic mapping, whereby neighbourhood relationships are maintained and the map of the retinal surface is reproduced in the ganglion cell and subsequent levels. Directional relationships within the map are maintained faithfully but a transformation occurs whereby more central regions are given an increasing proportion of the representation as the signal proceeds. From the retinal ganglion cell layer, the optic nerve sends the visual signal to the visual cortex, through the lateral geniculate nucleus of the thalamus. In primates, this is the principal projection pathway but a number of subsidiary pathways split off at the stage of the optic tract (following the partial crossover of fibres at the optic chiasm). The most substantial of these projections goes to the *superior colliculus*, a midbrain region of particular concern in active vision.

Figure 2.4 shows very schematically a remapping in which the central regions are disproportionately emphasized but the topology is maintained. It is widely accepted that the visual remapping has this general character. The magnification appears to come about because of transformations both in the cone $\rightarrow$ ganglion cell projections and also in the ganglion cell $\rightarrow$ striate cortex projections (Azzopardi and Cowey, 1993; Drasdo, 1991).

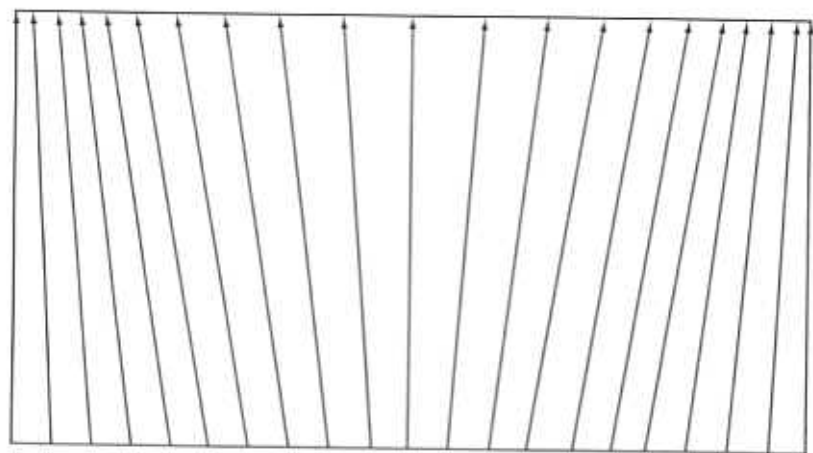


Figure 2.4 Schematic remapping in which topological relationships are maintained but increased magnification is given to the central region.

The *magnification factor* is the term used to describe the quantitative properties of the remapping and is defined as the distance on the cortical surface that corresponds to one degree of visual angle (Wilson *et al.*, 1990). For the purposes of this book, the absolute value of the magnification factor is of less significance than the manner in which the factor changes with eccentricity (E). The expression

$$M = M_f / (1 + E/E_s) \quad (2.1)$$

gives a reasonably accurate representation of experimental findings concerning the magnification factor. M_f is a constant (ca. 1 cm/deg) showing the value at the fovea. E_s gives the scaling factor and shows the eccentricity where magnification has fallen to half its foveal value. Estimates between 0.3° and 0.9° have been obtained (Wilson *et al.*, 1990), and it has been suggested that different values may apply to magnocellular and parvocellular systems (Section 2.3.1). An alternative description of the projection between retina and cortex has been proposed by Schwartz (1980), who notes that the projection can be well approximated by the following mathematically elegant transformation

$$\begin{aligned} u(r, \phi) &= \log r \\ v(r, \phi) &= \phi \end{aligned} \quad (2.2)$$

Here r and ϕ define a point in peripheral vision using radial co-ordinates while u and v describe the corresponding point in the cortical map using Cartesian co-ordinates.

2.2.3 Psychophysical performance in peripheral vision

Wertheim (1894) carried out a careful set of studies in which he plotted the ability to resolve a grating target presented at various positions in the visual periphery. His findings (Fig. 2.5) show that, for the range of values up to about 20 degrees in the near periphery, a surprisingly tight linear relationship between the size of the just resolvable grating and the angle of visual eccentricity. Similar results have been obtained by subsequent workers. It is possible, to a quite good approximation, to describe the decline in acuity by the following function

$$V_E = V_f / (1 + E/E_s) \quad (2.3)$$

where V_E is the acuity at eccentricity E , V_f is the acuity at the fovea and E_s is a scaling constant, which may be interpreted as the point at which the acuity has declined to one-half of its value at the fovea. For grating acuity, the constant is approximately 2.5° (Wilson *et al.*, 1990). As a rough approximation for some purposes, the constant component can be ignored and, as demonstrated by Anstis (1974), this leads to the simple but important approximate property of peripheral vision shown in Fig. 2.6.

Figure 2.7 also shows how acuity declines in the periphery, using a somewhat different procedure. In this example, the discrimination tested was the

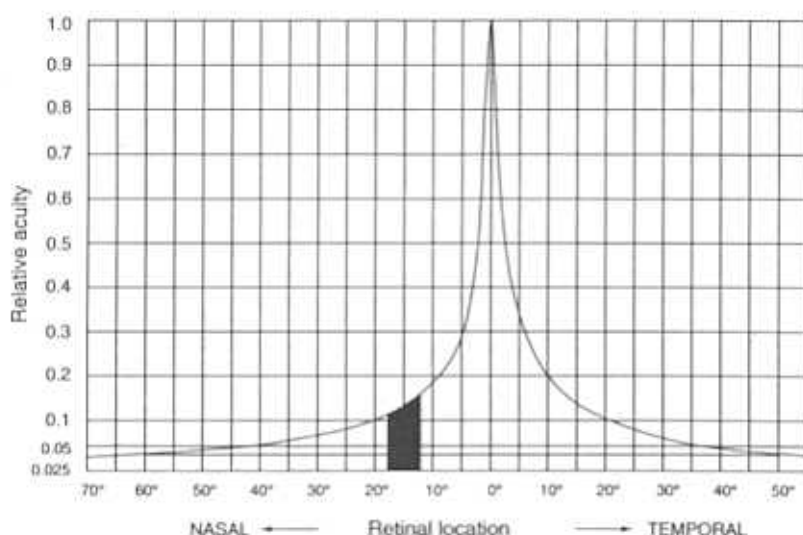


Figure 2.5 Measurement of grating acuity at various locations in the visual field. The black area shows the blind spot. Redrawn from Wertheim (1894).

identification of single letters, presented individually at different retinal locations. Two points are at once apparent. First, in a similar manner to the grating discrimination shown in Fig. 2.5, the discrimination ability declines gradually as the stimuli are placed more eccentrically. There is an extensive region over which partial, but imperfect, discrimination is possible. Second, discrimination is profoundly affected by the presence of 'irrelevant' surrounding contours. This is the phenomenon of *lateral masking*, discussed further below.

One obvious consequence of the decline in acuity is that certain discriminations become impossible when the stimuli are presented outside a certain central region. This region has been variously named the *stationary field* (Sanders, 1963), *conspicuity area* (Engel, 1971), *functional field of view* (Ikeda and Takeuchi, 1975), *useful field of view* (Bouma, 1978) or *visual lobe* (Courtney and Chan, 1986). Sanders (1963) also introduced the terms *eye field* and *head field* to denote the regions of the visual field where discriminations could be made by using eye movements alone and in association with head movements respectively (Section 4.1).

The conspicuity region is influenced by the specific task situation. Conspicuity areas shrink when the subject has a second foveal task, simultaneous with the peripheral task (Ikeda and Takeuchi, 1975), but are extended in a direction to which the subject is encouraged to direct covert attention (Engel, 1971). Conspicuity areas become particularly significant in tasks of visual search (Section 6.3.2).



Figure 2.6 A consequence of the variation in visual acuity with retinal eccentricity. When the centre of the display is fixated, each letter is at ten times its threshold legibility. This relationship applies, at least to a first approximation, regardless of viewing distance. From Anstis (1974).

2.2.4 Comparison of psychophysical and physiological measures

To what extent can the decline in visual abilities away from the fovea be *directly* attributable to the differential magnification in the visual pathways discussed in Section 2.2.1? Virsu and Rovamo (1979) suggested that different retinal patterns, which produce the *same* activation pattern on the visual cortex, will be equally discriminable. This implies that discriminability differences are all attributable to the differential magnification. As Virsu and Rovamo showed, this argument seems justifiable on the basis of the data for the case of certain discriminations, where the form of the decline in ability with eccentricity closely matches that of the cortical magnification factor. However in other cases, it is apparent that some further factor needs to be taken into account.

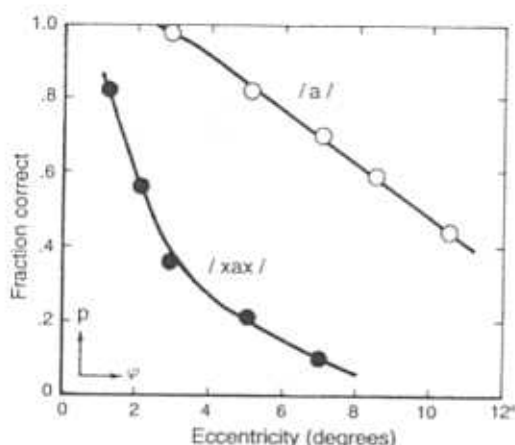


Figure 2.7 Lateral masking. Bouma (1978) required individuals to identify single alphabetic letters at various locations in the visual periphery. For the /a/ plot, the letter was presented in isolation. For the /xax/ plot, the letter was presented with flanking letter x's on either side. Even though the flanking letters were the same on every occasion, their presence profoundly affected the ability to identify the target letter.

There is an obvious similarity between eqs. 2.1 and 2.3, but the scaling factor is substantially different in the two cases. However, for another basic visual task, vernier alignment acuity, psychophysical experiments show that the value of E_v is much less than in the case of grating acuity, implying that vernier acuity is subject to a much greater relative degradation in the visual periphery. In this case, the scaling factor is close to that for the cortical magnification. Wilson *et al.*, (1990) suggest that vernier acuity shows cortical magnification limits whereas grating acuity is limited by the spacing of adjacent cones of the retina.

An important difference between the two forms of acuity is that vernier acuity requires judgement of a localisation difference, whereas detection of a grating can be done solely on the basis of a contrast difference between neighbouring regions. A recent study (Toet and Levi, 1992) of the phenomenon of lateral masking (Fig. 2.7) (study of the lateral masking effect) shows that the interference operates over an increasingly wide range, as stimuli are made more peripheral. Toet and Levi also noted considerable differences among individuals in the extent of lateral masking.

2.3 Parallel visual pathways

2.3.1 Magnocellular and parvocellular systems

Topographic mapping from the retina to the cortex was one fact underpinning the passive vision approach we have criticised in Chapter 1. For quite some time, the view held that the visual pathway transmitted a signal from the

retinal image in a monolithic way. One of the most important advances in visual science in recent years has been the appreciation of the existence of multiple types of nerve cell in the visual pathways. As early as 1966, Enroth-Cugell and Robson had demonstrated the presence of X- and Y-cells in cat retina but some time elapsed before the acceptance that a similar division occurred in the primate pathways.

The past two decades have seen the clarification of the distinction between the *magnocellular* (M) and the *parvocellular* (P) categories of cells in primate visual pathways. This nomenclature is based on a clear separation at the level of the lateral geniculate nucleus where the two cell types separate into distinct layers. However the separation of the cell types is also found in the retina, through to the cortex (Schiller and Logothetis, 1990). It is further suggested that the two cell types remain largely separated as two separate processing streams within the cortex (see next section) although there is considerable evidence of convergence from both streams in some cortical areas (Ferrera *et al.*, 1992; Maunsell *et al.*, 1990).

M and P cells are present in both central and peripheral retina, although the relative proportions differ. This has somewhat complicated the establishment of their properties, since the characteristics vary within each population, particularly between cells corresponding to different retinal regions. However, there is now general acceptance that the two cell types differ in a substantial number of ways, set out in Table 2.1. While the distinction between M and P cells is very widely accepted, the significance of the distinction has remained somewhat elusive. M cells have high contrast gain and fast response and thus

Table 2.1 Differential properties of cells in the magnocellular and parvocellular systems. For many of the properties (e.g. receptive field size), the property changes systematically with retinal eccentricity but at any particular eccentricity, the listed differentiation is found. Based on Kaplan *et al.* (1990) and Lennie (1993).

	Parvocellular system	Magnocellular system
Estimated number of cells (millions)	1.2	0.15
Percentage of ganglion cell total	80	10
Distribution on retina	Densest in fovea	Densest in fovea? (but more distributed)
Conduction velocity of axons	~6 m/sec	~15 m/sec
Response to stimulus onset	Tonic (sustained)	Phasic (transient)
Motion sensitivity	Lower	Higher
Receptive field size	Smaller	Larger (also periphery effects)
Spatial frequency sensitivity	Higher	Lower
Linear summation	Linear	Many non-linear
Contrast sensitivity	Lower	Higher (and saturates)
Contrast gain	Lower	Higher (8-10 times as high)
Colour sensitivity	Shown by many cells	None

are well suited for signalling the existence of a sudden change and it is thus very likely that they play a role in the dynamic processes of active vision. P cells, with more linear properties and small receptive field, seem well suited for signalling details of visual forms.

2.3.2 Visual processing in the cortex

Physiological studies of the cortex have shown the dominance and importance of the visual modality. Over much of the posterior half of the cortex, involving parts classified as occipital, parietal and temporal, cells are visually responsive and some degree of retinotopic mapping is retained. A large number (30+) of separate retinotopic maps, or visual areas, have been identified. Cells in each map possess different response properties and the analogy (Zeki, 1993), which considers the visual brain as an *atlas*, has some validity although it is certainly not the case that familiar visual properties such as colour and motion are exclusively differentiated into different areas.

There are well-known and visually attractive diagrams such as those of Felleman and Van Essen (1991), which lay out the cortical areas and their interconnectivity patterns. For many purposes, this level of detail is overwhelming and considerable impact has been made by a simplificatory scheme first proposed by Ungerleider and Mishkin (1982). While acknowledging the multiplicity of interconnecting pathways between cortical visual areas, they suggested that two principal routes relaying the incoming information from visual cortex could be distinguished (Fig. 2.8). A *dorsal stream* runs from occipital to parietal cortex and a *ventral stream* runs from occipital to

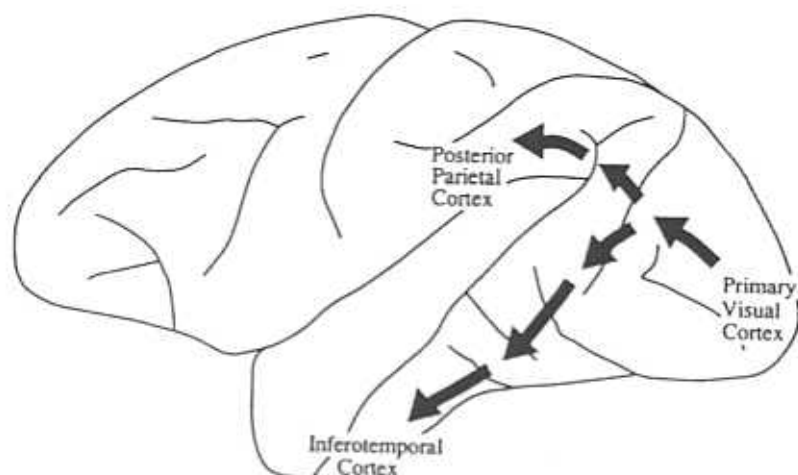


Figure 2.8 The dorsal and ventral streams. Two groupings of pathways that leave the primary visual cortex and can be traced through to parietal and temporal cortices respectively (from Ungerleider and Mishkin, 1982, as redrawn by Milner and Goodale, 1995).

temporal cortex. They suggested, on the basis of work that examined the differential effects of damage to the respective streams, that the ventral stream carried information for visual recognition, and the dorsal stream carried information relating to visuospatial awareness. Livingstone and Hubel (1987), in an influential paper, suggested that the M and P pathways, described above, map onto the cortical routes. This suggestion, however, has proved controversial (Merigan and Maunsell, 1993).

More recent work has modified and refined the original suggestion. Melvyn Goodale and David Milner have introduced a subtle amendment to the original distinction (Goodale and Milner, 1992; Milner and Goodale, 1995). In their modified scheme, the ventral stream supports Vision for Recognition in a similar way to the earlier account. However the dorsal stream, termed Vision for Action, provides for a series of direct vision-action links rather than any more reflective use of vision. In support of their revised position, they describe a patient, DF, who is able to carry out visuo-spatial tasks involving oriented objects (posting blocks through a slot) but has no awareness of the details of the process involved and cannot identify object orientations verbally. The relationship between vision and awareness is a complex topic of considerable current interest, and further discussion is given in Chapter 8. Recent physiological work has provided support for the position by showing that dorsal pathways project to motor areas of the frontal cortex, and their properties support visually guided actions in a rather direct way (Sakata *et al.*, 1997). The cells in the dorsal → motor route show a gradual transformation from a visual sensory signal to a motor output signal. We shall encounter a similar gradual transformation in connection with orienting saccades in Chapter 4.

How should the action/recognition distinction be linked to our passive/active vision distinction? The idea that vision operates in the support of action is a clear and welcome advance on the passive vision view. Making a more detailed link is not straightforward, first because the action/recognition distinction largely concentrates on the level of overall visual tasks rather than the subcomponents of these tasks, which is where active vision makes its contribution. There is evidence that areas of the dorsal pathway in the parietal lobe are very concerned with saccadic eye movements (Section 2.5.4). It might thus appear that orienting is simply another 'action' which vision can support. We suggest though that many visual activities involve intimate integration of action and recognition. The example of tea making, where action involves a number of utensils and substances, will receive some discussion in Chapter 7. In our analysis, we shall wish to argue that the orienting (looking) processes characteristics of active vision are indeed visual actions but also often intimately linked to recognition processes.

Another form of parallel processing occurs because the physiological pathways leaving the retina in fact project to several different brain regions. The geniculo-cortical route discussed above is, in primates, the largest and

most extensively studied of the visual pathways but visual information is also directed via several other pathways to brain centres such as the superior colliculus, sites in the pretectum, and elsewhere (Milner and Goodale, 1995). A traditional view of these pathways is that they are associated with 'reflex eye-movements'. Although much evidence supports this view, it turns out, rather surprisingly, that the fastest eye responses use cortical pathways (Miles, 1995, 1998).

2.4 The oculomotor system

The essence of active vision is continual sampling through gaze redirection. Although this can be achieved without the use of eye muscles, there is no doubt that the process is achieved most efficiently by using these muscles and that for many human activities, eye movements form the principal means of supporting active vision. In this section we discuss the different ways in which the eyes can be moved.

2.4.1 The muscles of the eye

Each eye is held in place by six extraocular muscles, grouped into opposing pairs (Fig. 2.9). Study of their arrangement and properties has long been of interest and of clinical concern to oculists and optometrists whose terminology is commonly used, although one may regret the passing of an even older terminology which included the term 'amatoris' for the lateral rectus muscle because of its employment in the furtive glances of flirting. Horizontal movements of the eye are achieved almost exclusively by the action of two

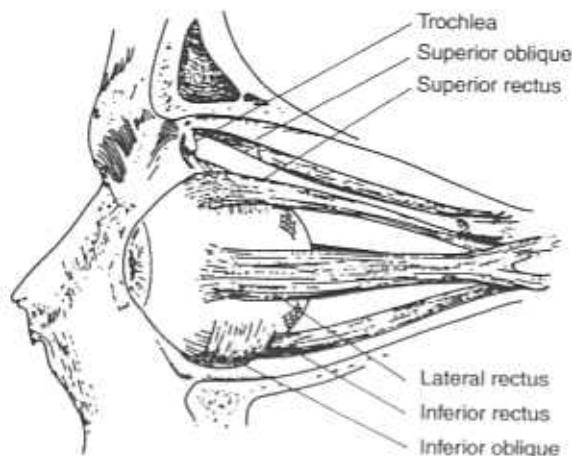


Figure 2.9 The six oculomotor muscles (from Howard, 1982). From the viewing position adopted, the lateral rectus is the closest muscle to the viewer but its corresponding antagonist; the medial rectus is occluded by the eyeball and so not shown.

muscles. These are the *lateral rectus* and the *medial rectus*, responsible for *abduction*, directing the eye outwards, and *adduction*, directing the eye inwards, respectively. Vertical movements are largely achieved by using the *superior rectus*, promoting upward *elevation* movements of the eyeball, and the *inferior rectus*, promoting downward *depressive* movements. The remaining pair of muscles, the *oblique* muscles do however make a partial contribution.

Rotations of the eye are customarily described with reference to the *primary position* of gaze, in which the eye is centrally placed in the socket (see Carpenter, 1988 for a more precise definition). *Secondary* positions refer to gaze directions achievable with a single vertical or horizontal rotation from the primary position and *tertiary* positions to all other gaze directions, that is all oblique directions. Note that these positions refer to the direction of the gaze axis only and for any gaze direction, the eye could, in principle, be in a number of different states, because of the freedom to rotate around the gaze axis (*torsional* movement). An early experimental finding was *Listing's Law*, which states that for each gaze direction, the eye has a unique position in the orbit irrespective of what combination of movements are used to achieve the gaze direction (strictly speaking, this is *Donders' Law*. Listing's Law also includes a specification of the particular position adopted). Listing's Law is a non-trivial result and has excited considerable recent interest since it appears to implicate a sophisticated neural and oculomotor mechanism (see e.g. Crawford and Vilis, 1995).

A loose interpretation of Listing's Law says that although the eye might, in principle, rotate arbitrarily in a torsional manner about its principal axis (assumed here to be the gaze axis), such arbitrary rotations effectively do not occur. A simple practical consequence is that for very many purposes, this degree of freedom can be ignored and eye movements specified adequately in terms of the horizontal and vertical components of rotations, or alternatively in terms of the *amplitude* and *direction* of rotation. However, it is also clear that under some circumstances significant torsional rotations of the eyes do occur. When the head rotates, some partially compensatory *countertorsion* is found (Howard, 1982) and when binocular vision is considered, *cyclotorsion* movements, oppositely directed in the two eyes, are also important (Howard and Rogers, 1995).

2.4.2 Classification of eye movements

Vision is important in a wide variety of situations, from watchmaking to slalom skiing. The evolutionary demand of such tasks, or their forerunners, has resulted in a complex set of oculomotor control processes. These can be separated into a set of distinct categories. A landmark article by Walls (1962), proposed an evolutionary history for the different types of eye movement in an article that was both erudite and entertaining.

Walls proposed that, paradoxically, the extraocular muscles did not evolve to move the eyes so much as to keep them still with respect to the visual environment as the organism moved. Two fundamental systems promote visual

stabilisation in this way, the *vestibulo-ocular* and *optokinetic* reflexes (VOR and OKR). In the case of the former, the stabilising signal is derived from the vestibular organs of the inner ear; in the case of the latter, from an extensive pattern of coherent optic flow on the retina. These systems have become elaborated to provide a wonderfully effective way for vision to operate from a stable viewing platform (for details see Carpenter, 1988, 1991; Miles, 1995). Investigation of the stabilising reflexes often makes use of continuing steady stimulation, either by subjecting the observer to continuous body rotation (rotating chair) or to continuous whole field rotation (rotating drum). This results in a characteristic *nystagmus* movement of the eyes. The eyes move in a sawtooth pattern with a *slow phase* in which the eyes are kept stably aligned with the visual surroundings followed by a *fast phase* rapid movement in the opposite direction. These repeat to produce the nystagmus pattern. The sharp movements of the fast phase minimise the time that vision stability is disrupted and, importantly for the current theme, were the probable evolutionary precursors of the saccadic mechanism by which rapid movements of the eye could occur more generally.

VOR and OKR are essentially involuntary and automatic. This contrasts with the remaining eye movement types, all of which might be considered to show rudimentary volition. The *saccadic*, *pursuit* and *vergence* systems can all be described in terms of target selection, which in turn is likely to be tied to the motivational state of the perceiver and to higher cognitive processes. The saccadic system rotates the eye so that a selected target can be brought on to the fovea. The pursuit system, often termed smooth pursuit, allows a selected target that is in motion to be followed smoothly with the eyes. The vergence system maintains both eyes on a target that moves in depth, or makes an appropriate adjustment of the directions of the two eyes to a new target at a different depth. The saccadic system uses fast, stereotyped, jump-like movements, and typically rotates the eyes for a brief period at speeds up to several hundred degrees per second. These movements are clearly differentiated from pursuit movements that are continued movements of the eyes, generally at speeds well under $100^\circ \text{ sec}^{-1}$. Both saccadic and pursuit movements are essentially *conjugate* with the two eyes rotating equally. Vergence movements are classically (but see Section 2.5.2) described as continuous movements in which the eyes move in a *disjunctive* manner (in opposite directions) with maximum speeds under $20^\circ \text{ sec}^{-1}$.

Both pursuit and vergence systems can operate in a closed loop manner to maintain the eye or eyes aligned onto a moving target. However in both cases, there is an initial component that operates in an open loop manner and facilitates fast target acquisition (Bussetini *et al.*, 1996; Semmlow *et al.*, 1994) and a corresponding fast suppression of the visual stabilisation systems (Lisberger, 1990).

For stationary observers viewing stationary scenes, no stimulation to drive VOR, optokinetic responses or pursuit is present. Thus the expectation is that

the eye movement pattern will consist of saccades only, separated by periods where the eye is stationary, together with vergence movements to the extent that parts of the scene are at different viewing distances. For many purposes, only the saccadic movements are significant and these situations form the principal subject of much of the remainder of the book. The term *fixation* is given to the stationary periods between saccades (the term is used both as a generic description of the act and in the noun form *a fixation* that describes each instance). For most practical purposes, it can be assumed that the eye is stationary during fixations but close examination of fixation shows it to be a dynamic state in which the eye makes continuous miniature movements (Ditchburn, 1973). A typical record of such movements is shown in Fig. 2.10 and shows slow irregular *drift* movements of a few degrees per second, together with more rapid irregular movements termed *tremor*. Occasionally, small, jump-like, movements occur which have been termed *microsaccades*. It should be noted that Fig. 2.10 was obtained by asking an observer to maintain their eyes continuously viewing the same target location for an extended period.

These movements have the effect of jiggling the retina with respect to the retinal image. The relative motion can be eliminated with a *stabilised retinal image* technique in which, for every movement of the eye, an equal but opposing movement of the display being viewed occurs. Such image stabilisation leads to a dramatic 'fading' and entire loss of vision (Ditchburn and Ginsborg, 1952; Riggs and Ratliff, 1952). This result initially led to the idea that the details of the miniature eye movements might be of fundamental importance for vision and a period of intense study ensued (see Ditchburn, 1973). The original use of 'involuntary' for these fixation eye movements was shown to

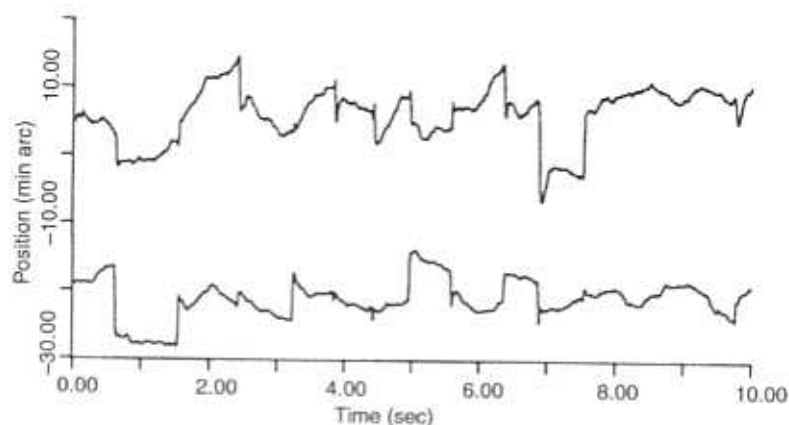


Figure 2.10 Record of fixational eye movements (the terms 'physiological nystagmus' and 'miniature eye-movements' are equivalent), showing drift, tremor and microsaccades. The upper and lower traces show the vertical and the horizontal components of the movement. The traces have been displaced vertically, for clarity. From De Bie (1986).

be inappropriate by the demonstration of some higher-level input into the fixation mechanism. The incidence of microsaccades could be changed with instructions (Steinman *et al.*, 1967), and directed drift movement was found to occur in anticipation of a subsequent target following a saccade (Kowler and Steinman, 1979). Interest in the topic waned with the appreciation that, particularly in head free situations, quite substantial retinal image movement was usual (Steinman *et al.*, 1982). Nevertheless, no thoroughly worked out neurophysiological account of visual loss under stabilisation has emerged, and the phenomenon offers a challenge to some thinking in the passive vision tradition.

2.5 Saccadic eye movements

Saccadic eye movements are a ubiquitous feature of vision. Credit for the recognition that the eye moves in a series of jerky jumps should be given to the group of nineteenth century French ophthalmologists amongst whom Javal (1878, 1879) was a prominent figure. The term saccade can be traced to Javal's work and its incorporation into the English language credited to another influential early investigator, Raymond Dodge, one of many Americans who profited from a spell of study in the German laboratories (Dodge, 1900; Dodge and Cline, 1901; Erdmann and Dodge, 1898). A fascinating historical account of the early work on eye movements may be found in Tatler and Wade (2002).

In most visual activities (see Chapters 5–7) we move our eyes by making saccades 3–4 times each second. Simple calculation shows that we must make many tens of thousands of saccades each day and many billions over the course of a lifetime. They can be made voluntarily (for exceptions, see Section 5.8.2, Section 8.3 and Section 8.5) but for the most part operate well below the level of conscious awareness. In everyday activity, most saccades are only a few degrees in size (Bahill *et al.*, 1975a; Land *et al.*, 1999). However, particularly during active tasks, a small number of very large saccades also occur and a recent estimate of the average saccade size during an everyday task (tea-making) is 18–20° (Land *et al.*, 1999).

2.5.1 Characteristics of saccades

Saccadic eye movements are *stereotyped* and they are *ballistic*. The *trajectory* of the saccade refers to the exact details of the way that the eye rotates. Figure 2.11 shows typical trajectories for horizontal saccades of various sizes. The eye is initially stationary. At a quite well defined point, it begins to accelerate, reaches a maximum velocity, and then decelerates rapidly to bring the eye to rest in its new position. The angular rotation is referred to as the *amplitude* of the saccade. The stereotypy of saccades is shown by the fact that every time a saccade of the same amplitude occurs, the same trajectory is followed closely. The *duration* and *maximum velocity* of saccades are measures readily obtained

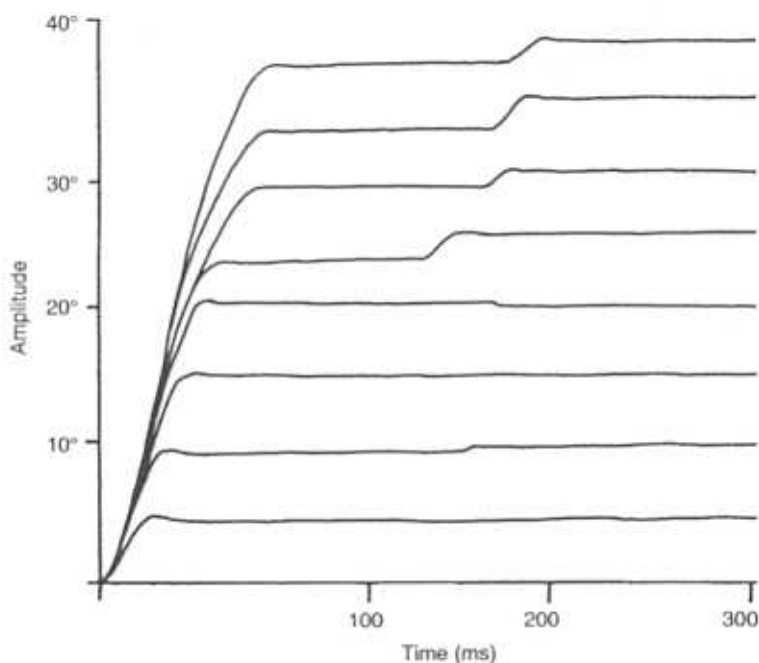


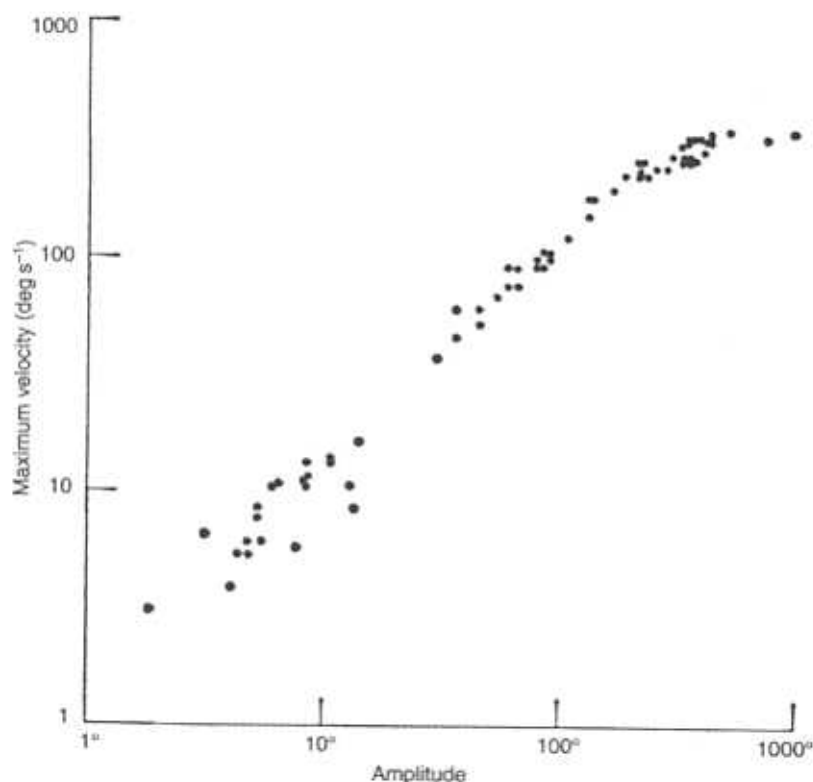
Figure 2.11 Time course of horizontal saccadic eye movements of various sizes (from Robinson, 1964).

from the trajectory. Plots such as that of Fig. 2.12 show how these parameters vary little for saccades of a particular amplitude, but depend systematically on the saccade amplitude. The term *main sequence* has been adopted for such plots following an imaginative analogy with an astrophysical relationship (Bahill *et al.*, 1975*b*). For saccade duration, the main sequence is well described (Carpenter, 1988) by the expression

$$T_S = 2.2A_S + 21 \quad (2.4)$$

where T_S is the saccade duration in ms, and A_S is the saccade amplitude in degrees.

Some qualifications must be made to the basic picture. The oculomotor system is a biological system and such systems are invariably characterised by greater variability than mechanical ones. Saccades can occur which are slower (occasionally also faster) than the main sequence. Although the oculomotor system does not appear to be subject to fatigue through repeated use, slower saccades than predicted by the main sequence are sometimes found in states of drowsiness and also occur as a result of some drugs such as benzodiazepines (Glue, 1991). An attempt to slow saccade trajectories by a biofeedback training procedure showed some success (Findlay and Crawford, 1986). Slow saccades can also occur as a result of brain damage (Zee *et al.*, 1976). The termination



a moderate degree of curvature. This curvature is systematic (Viviani *et al.*, 1977: Fig. 2.13). One situation in which curvature might be expected is when an oblique trajectory combines a horizontal component and a vertical component of different amplitudes. If such a movement came about through a simple additive combination of the movements made for each component, then it would be expected that the shorter amplitude component would have shorter duration (because of the main sequence). The curvature found with oblique saccades does not show this pattern; the shorter component shows *stretching* to match it to the other component (Van Gisbergen *et al.*, 1985).

A ballistic movement cannot be modified by new information once it is initiated. Saccadic eye movements have this character. This is shown from studies of two-step tracking (Section 4.4.2), which shows that visual information arriving less than about 70 ms prior to the start of a saccade cannot modify the movement. If the saccade goal is modified immediately prior to this deadline, curved saccade trajectories may be obtained which show clear target-seeking properties (Van Gisbergen *et al.*, 1987). Such curved goal-seeking saccades have been observed when brain damage causes saccade slowing (Zee *et al.*, 1976). They are of some significance because they demonstrate that an internalised goal seeking process operates as part of the saccade generation mechanism. Nevertheless, such goal seeking trajectories are almost entirely absent for *small* saccades (Findlay and Harris, 1984), suggesting that the goal is predetermined at the outset.

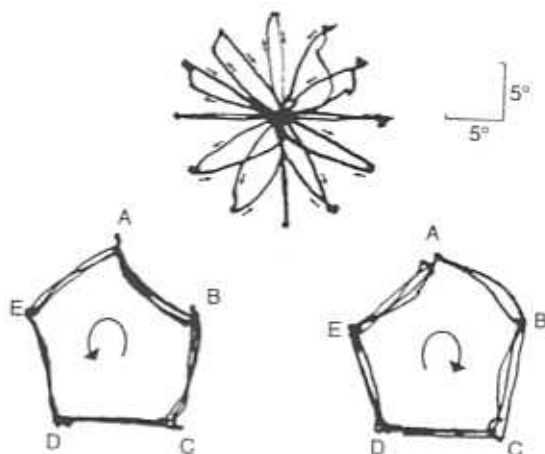


Figure 2.13 Plots of the trajectories of saccadic eye movements showing that each movement is associated with a systematic curvature. The top traces show a set of saccades made from a central point to and from a series of locations on a clock face. The bottom traces show scanning around the points A-E in a counterclockwise (left trace) and a clockwise (right trace) direction. From Viviani *et al.* (1977).

2.5.2 Combining saccadic movements with pursuit and vergence

It was stated (Section 2.4.2) that the saccadic, pursuit and vergence systems were regarded as separate systems. This claim will now be analysed in more detail. Figure 2.14 shows a plot of eye position as an observer tracks an object moving in a smooth course with reversal of direction at the end of a fixed period. This shows a clear separation of following movements and faster velocity saccades. If a target that an observer is asked to follow commences a regular but unpredictable movement, the eye commences pursuit after a short delay (the pursuit latency). Shortly afterwards, a saccade in the direction of the target movement usually occurs and the term *catch-up* saccade is used for saccades occurring during smooth pursuit. This situation and similar ones were the subject of a set of classical experiments (Rashbass, 1961; Westheimer, 1954) that supported the separation of a pursuit system, driven by target movement, from a saccadic system, driven by target position (retinal error). Subsequent work has shown this to be a useful generalisation but there is some cross talk between the systems. Thus the pursuit system shows some response to whether the eyes are lagging or leading the target (Wyatt and Pola, 1981) and the saccadic system shows the ability to program saccades to the anticipated future position of a moving target, taking into account subsequent target movement (Newsome *et al.*, 1985).

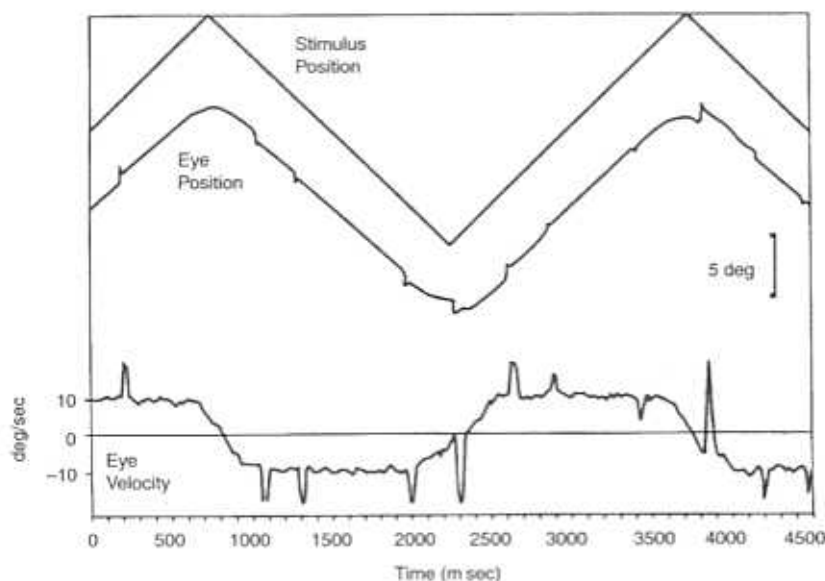


Figure 2.14 Eye movements of an individual tracking a spot moving smoothly in a horizontal left $\rightarrow$ right $\rightarrow$ left $\rightarrow$ right regular sequence. The eye position trace shows periods of smooth movement, interrupted by occasional small saccades. The saccades show up clearly as brief peaks on the lower trace of eye velocity. The record also shows the phenomenon of anticipatory changes in pursuit direction prior to direction reversal of the target. From Boman and Hotson (1992).

Early studies of the vergence system measured the response to unpredictable target steps or movements in depth (Rashbass and Westheimer, 1961). The view that emerged from these studies was that the vergence system moved the eyes slowly, with retinal target disparity being the principal input. A critical situation is that of *asymmetric vergence* where a target steps to a new position, differing from its previous one in both distance and direction. Following records reported by Yarbus (1967) it was long believed that the response to this situation was a rapid conjugate movement of the eyes to the appropriate target direction, followed by a much slower disjunctive vergence movement which gradually

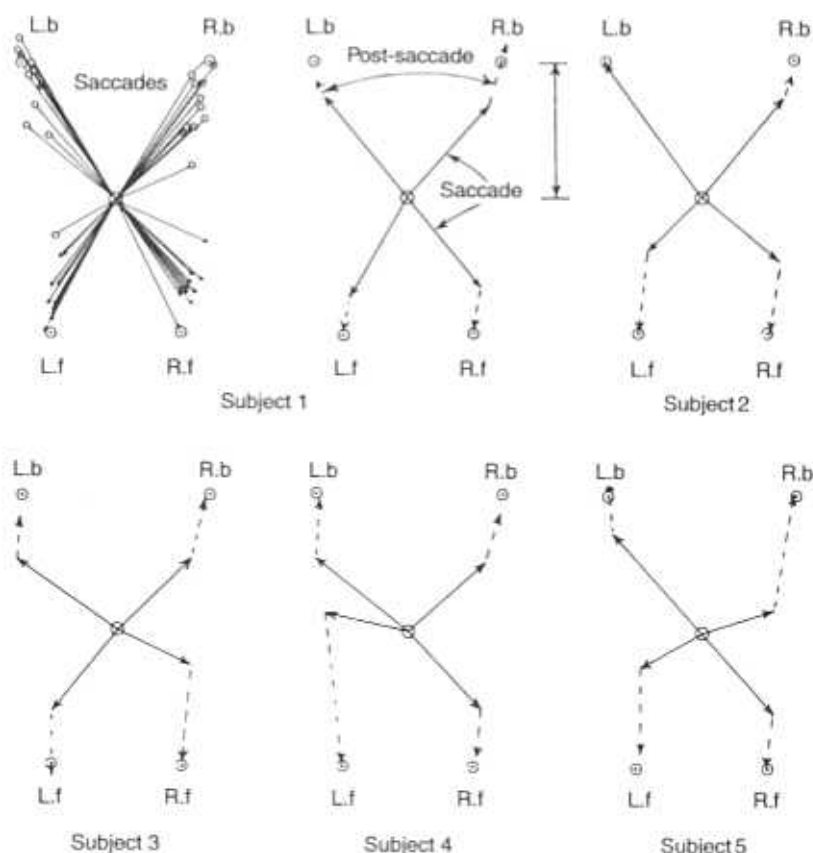


Figure 2.15 Schematic illustration of the fast and slow components of fixation changes in depth (asymmetric vergence). The plots show the movement in the horizontal plane of the point of intersection of the visual axes of both eyes. The subject initially fixated the display centre and a target appeared unpredictably in one of four positions, left behind (L.b), left front (L.f), etc. The top left panel shows a set of individual records, the others show averaged data from five subjects. The solid lines point to the eye position immediately after the saccade and the dashed lines to the position 1 second later. After Enright (1986).

brought the two eyes into alignment in the appropriate target depth plane. It has subsequently emerged that a much more integrated pattern is often found.

Enright (1984, 1986) recorded saccades in the asymmetric vergence situation. He showed that saccade movements under these conditions were unequal in the two eyes with the disconjugacy acting to bring the eyes together onto the target (Fig. 2.15). This finding was confirmed by Erkelens *et al.* (1989). These results are disturbing for the classical picture of separate saccadic and vergence subsystems and require a major modification of its basic postulates. Two possibilities have been suggested. In the first, a distinction between a conjugate saccadic system and a disjunctive vergence system is maintained but the vergence system is boosted so as to speed up during saccades. In the second, the idea that saccades are conjugate movements is abandoned and the alternative proposed that saccades in each eye are programmed separately. These positions are fiercely debated (Mays, 1998; Zhou and King, 1998) and since both can predict the behavioural findings, resolution will depend on a full understanding of the brain processing pathways for both conjugate and disjunctive movements. A further well-established phenomenon is that, during the course of a saccade, a period of transient divergence is found since the abducting eye moves more rapidly than the adducing eye (Zee *et al.*, 1992).

2.5.3 Saccadic suppression

Saccades have always been somewhat troublesome for the passive theory of vision. As we emphasized in Chapter 1, this theory assumes that the aim of vision is to create a stable mental representation of the visual world. How then can we move our eyes and maintain a stable 'world-picture' in the face of the changes, which are evidently taking place on the retina? The regularity with which this question is addressed in texts of vision, to the exclusion of other questions concerning the role of saccades in vision, is indicative of the pervasive nature of the passive vision view.

A partial answer to the question comes from the finding that visual thresholds are elevated during the course of saccadic movements. An informal demonstration of this can be experienced by obtaining a retinal after-image and moving the eyes around with the eyes closed. Latour (1962) used a visual probe to measure the ability to detect brief faint light flashes. His results (Fig. 2.16) showed a decrease in threshold that commenced some time before the start of the actual movement. This threshold decrease is a central phenomenon, shown by the suppression of visual phosphenes (Riggs *et al.*, 1974). However its magnitude is relatively small. It has been recognised that other contributions come from the way the visual system handles full field retinal motion. As well as the obvious blurring that results, a masking mechanism comes into play (Matin, 1974). It has also been suggested that a similar masking mechanism is important in preventing information from one fixation interfering with that from the subsequent one (Breitmeyer, 1980). Detection of visual motion occurring during the course of saccades is particularly poor

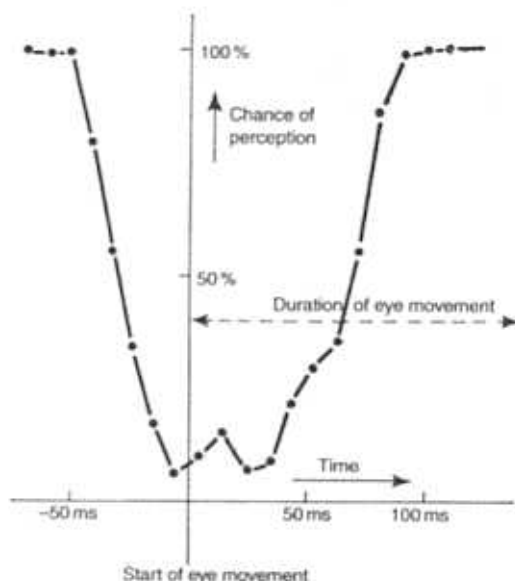


Figure 2.16 Time course of the suppression of visual threshold for lights flashed during a saccadic eye movement (from Latour, 1962).

(Bridgeman, 1983). The term *saccadic suppression* is used to describe the loss of vision resulting from these processes.

A series of recent studies by Burr and colleagues (Burr and Morrone, 1996; Burr *et al.*, 1994; Ross *et al.*, 2000) has supported the idea that saccadic suppression is primarily occurring in the magnocellular system and very little suppression (as opposed to smear blurring) occurs when discriminations can be carried out exclusively with the parvocellular system (e.g. high spatial frequency gratings).

2.5.4 Physiological pathways for saccadic eye movements

Using a similar approach to that of tracing visual input pathways into the brain, it has been possible for neurophysiologists to identify a set of brain areas concerned with oculomotor output. We shall concentrate here on the cortical areas involved in saccadic orienting movements (see Section 4.3.1 and Section 4.4.5 for further details of the immediate pre-motor mechanisms). Such areas are characterised by two properties. First, stimulation of each area will produce orienting movements of the eyes and second, electrical recording shows that cells in the area discharge prior to the production of a saccadic eye movement (Schall, 1991). Further confirmation comes from two other approaches. Careful study of the way in which the saccadic system is affected in patients with damage to cortical areas of the brain (Pierrot-Deseilligny *et al.*, 1991) has been used (see Chapter 8) together with lesion studies on

animals. Recently, direct investigations of cortical activity have become possible using PET and fMRI techniques (Corbetta, 1998).

The emergent picture shows clearly that there are multiple parallel routes involved in the generation of saccades. Figure 2.17 shows a diagram of the main areas and their interconnectivity. The areas of cortex that are most intimately linked to saccadic eye movements are the area LIP of posterior parietal cortex and the frontal eye field region of the pre-motor frontal cortex (FEF). Saccades can be elicited by electrical stimulation in each of these areas. Studies using the lesion technique have shown that no single pathway is essential. However, the combined loss of both the superior colliculus (SC) and FEF renders an animal unable to make saccades (Schiller *et al.*, 1980), attributable to the fact that these centres form parallel output pathways. There are direct projections from FEF and SC to the brainstem saccade generators although the direct FEF pathway seems to be of subsidiary importance (Hanes and Wurtz, 2001). Loss of ability to make saccades occurs also

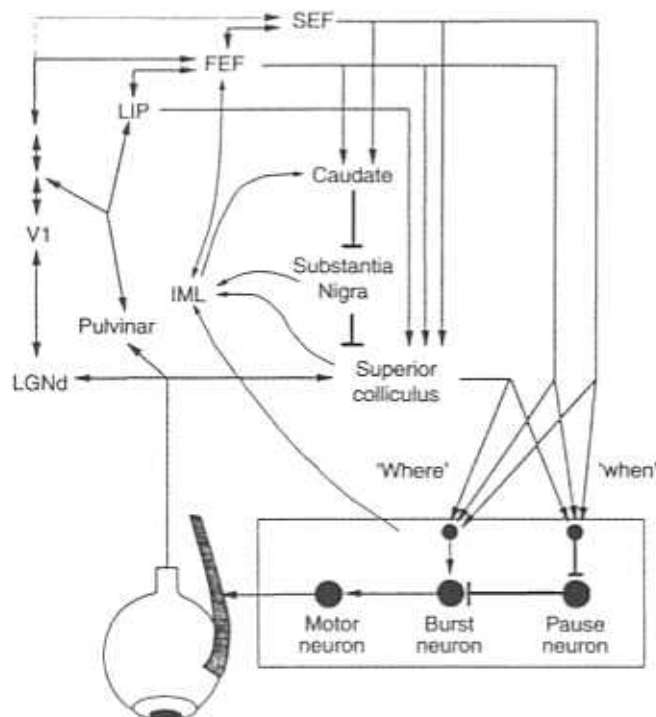


Figure 2.17 Schematic diagram of the oculomotor output pathways involved in generating saccadic eye movements. LGNd – dorsal part of the lateral geniculate nucleus, V1, LIP, FEF, SEF are the cortical areas Visual 1, lateral intraparietal area, frontal eye field and supplementary eye field respectively. IML is the internal medullary lamina of the thalamus. From Schall (1995).

with lesions to both SC and occipital cortex (V1), presumably because no visual input pathway is available (Mohler and Wurtz, 1977). It is probable that the saccade related regions in the parietal and occipital lobes send their signal through the SC, since stimulation of these regions, in contrast to stimulation of FEF, no longer elicits saccades following SC ablation (Schiller, 1998). A further important pathway links FEF and SC via the caudate and the pars reticulata of the substantia nigra (Hikosaka and Wurtz, 1983; Hikosaka *et al.*, 2000).

Recent work has attempted to go beyond the simple identification of brain regions involved in saccade generation, to detailed discussion of the computational mechanisms. The most well elaborated instance concerns the superior colliculus (Section 4.3.2). This region, and also the cortical regions (FEF, LIP) which project to it, contain motor maps, such that, for example, the orienting saccade generated by electrical stimulation is dependent on the exact locus of stimulation. Within several such maps, lateral inhibition has been shown to operate as a selection process to enhance processing in one direction at the expense of neighbouring directions (inferior temporal cortex: Chelazzi *et al.*, 1993; frontal eye fields: Schall and Hanes, 1993; superior colliculus: Glimcher and Sparks, 1992). The areas can then be considered to operate as salience maps in which the selection of a saccade target is achieved. This concept is developed further in connection with visual search in Chapter 6.

2.6 Summary

This chapter has attempted to digest information about the visual and the oculomotor systems that is of particular importance for active vision. We adhere strongly to the principle that active vision is a sub-area of neuroscience and thus its study must be grounded in neurobiological principles. Of course we appreciate that many of the topics that feature in texts about vision (colour, depth, motion, spatial frequency, cortical areas) have been almost entirely ignored. Likewise, only the saccadic part of the oculomotor armamentarium has received any detailed consideration. The approach of the chapter has been 'outside-in', in the sense that the visual and oculomotor system have been treated principally as fixed entities which operate to interface with the environment. We have thus adopted the time-honoured approach of ignoring plasticity, learning and development (some redress occurs in Section 4.7), recognising that, for many purposes, the assumption of biological hardwiring is a productive one. We have also not yet made much reference to an active perceiver, having dealt only with the processes operative within such a perceiver. In the next chapter we begin the link between neurobiological and cognitive accounts through consideration of the topic of attention.

ACTIVE VISION: THE PSYCHOLOGY OF LOOKING AND SEEING

JOHN M. FINDLAY,

*Centre for Vision and Visual Cognition,
Department of Psychology, University of Durham*

and

IAIN D. GILCHRIST,

*Department of Experimental Psychology,
University of Bristol*

OXFORD
UNIVERSITY PRESS

(2003)