Perceptual development: Visual, auditory, and speech perception in infancy

CHAPTER ONE

The development of basic visual abilities

Louise Hainline
Brooklyn College of the City University of New York, USA

INTRODUCTION

Our sensory systems serve as a kind of filter between us and the world. Information from the environment must cause a detectable response in one or more of the body's sensory systems before the brain can make any use of the information. Different species have evolved to rely more heavily on some sensory systems than others. Primates, and in particular, human primates have evolved to depend heavily on vision. In the adult human, it has been estimated that more than half of the brain deals in some way with the processing of visual information (e.g. Sereno, et al., 1995). To understand the world of the human infant, then, it is critical to understand how the visual system develops, as visual abilities form a vital link between the infants and objects and events around them.

Vision clearly undergoes substantial development from infancy to adulthood. Infants simply do not see as well as adults do, a fact that has been well-documented over the last 25 years. Acuity, colour vision, binocular vision, and other basic visual processes all are substantially poorer compared with adults, during the first year and even beyond. On the basis of this body of research, one might legitimately conclude that infants' poor vision would make them significantly visually handicapped.

It is therefore surprising when one observes young infants interacting with people and things around them. These interactions are often complex and impressively nuanced, with little indication that young infants are handicapped by their purported primitive visual abilities. The limitations in their behaviours seems more related to attentional and cognitive factors.
than to the lack of well-developed visual abilities. Indeed, infants faced with complex real-world stimuli appear to fare rather better than the typical laboratory study might lead one to think. Research in the Gibsonian tradition (e.g. Kellman, 1988, 1993; Spelke, 1988) demonstrates that infants actually are fairly competent when faced with visual stimuli that are more complex and regular than the isolated patches of light or sine wave gratings favoured by researchers working in the tradition of experimental sensory psychophysics. This paradox is one of the more interesting issues in the study of infant vision. Although infants may not, indeed, see as well as adults do, they normally see well enough to function effectively in their roles as infants. Vision does not limit an infant’s development, although for an infant with abnormal vision, it may be another matter.

But, in this chapter, the task is to try to give an account of how infants experience their visual world. The attempt to define what infants see evokes deep philosophical questions about the nature of our personal sensory experiences that have been pondered by philosophers and scientists for centuries. The critical question, of course, is how we can know how anyone else experiences events in the world. With verbally competent individuals, one has, at least, a social communication system that can help, but ultimately, we have to admit to the validity of solipsistic concerns about the nature of personal experience even for an observer who speaks our language. The problem is amplified when one is interested in the experiences of nonverbal organisms we would like to understand better, like infants and our pets.

Scientific experimentation offers some help, at least part of the way. The process of vision is usefully conceived of as a series of processing stages, with the results of early stages feeding into the processing of higher and generally more complex later stages. Using a kind of bottom-up logic, it is generally safe to assume if visual information does not get transferred from input to output at a lower level of processing, that information is not available for processing at higher stages. In other words, infants can see no better than their basic neural structures allow; for various reasons we will discuss, their effective vision could be worse if higher centres do not use the visual information provided by lower centres efficiently, but this issue can be addressed by appropriate research strategies to detect a mismatch between basic visual abilities and perceptual performance. The first stage of translation (or “transduction”) of information about environmental stimuli, in the case of vision, of physical dimensions of light energy into neural energy, is termed “sensation”. After the transduction stage, one may study the organisation of these sensory inputs, the processes of perception, and the interpretation and analysis of perceptual information by higher mental processes, generically typically termed “cognition”. The nature of these processes in infants will be dealt with in other chapters in this book. The purpose of this chapter is to describe the visual sensory abilities of infants as understood today, as a preliminary answer to the question of what things look like to a baby.

A CONTEXT FOR ASSESSING INFANT VISION

Excellent recent reviews of the research on infant vision can be found in Aslin (1987), Atkinson and Braddick (1989), Van Sluijters et al. (1990), and various chapters in Simons (1993) and Vital-Durand, Atkinson, and Braddick (1996). In most cases, individual visual abilities have been treated separately, with visual abilities summarised one by one (e.g. acuity, binocular vision, colour vision, etc.). Although one can understand the temptation to segment the processes of vision thus for expository reasons, this focus on separate visual abilities can be misleading, because effective use of visual information requires coordination among different functions and abilities. The image of the visually deficient infant may stem in part from the failure to consider the infant as an ensemble of mutually supporting functions, a system in which the whole is greater than the sum of the parts.

To take just one instance, it is unlikely that we can understand what the visual experiences of young infants are like without taking both sensory and motor behaviours into account. Adults are capable of acute vision because they are able to localise with rapid (saccadic) eye movements objects that are detected but not clearly seen in the periphery; adults see more clearly with the central portions of their retina (the sheet of neural tissue at the back of the eye, containing the photoreceptors, specialised cells that transduce light into neural signals; for humans the central portion of the retina, the fovea, is responsible for the highest-acuity vision). As a result, to inspect objects closely, they need to point their foveae at them directly. Once an adult is fixating an object, the two eyes’ foveae must be made to point at similar targets in the world, which is accomplished by a system of eye movements called “vergence”; we converge our eyes to look at closer targets, diverge to look at objects further away to maintain a single percept. Once we have positioned an object in corresponding locations on the two eyes, we need either to keep the eyes steady, “fixate”, so that the brain can analyse the detailed properties of a stationary object, such as form, colour and fine detail, or to follow a moving object with our eyes (through an eye movement termed “pursuit” – with either smooth movements, or a series of rapid, saccadic episodes) to keep the object of interest on our foveae. Development of vision requires a coordination of both the sensory aspects of vision (detection of form, motion, colour, pattern, etc.) with accurate eye movements. A good sensory apparatus without good oculomotor control will have its vision severely compro-
mised, and good oculomotor control cannot be achieved without adequate visual feedback from sensory levels. So, “seeing” for infants necessarily involves the development of well-balanced sensorimotor interactions.

Another point of discussion will be whether we should accept uncritically reports of reduced vision in young infants. Although there is no question that infants see more poorly than adults, the situation may not be as bad as it has been portrayed. Inherent in many of the methods in the tool kit used to test infants is the tendency to under- rather than overestimate infants’ capacities. Most of the paradigms that are currently used to study vision in infants depend on some level of interest and attentiveness toward the test stimuli. A legitimate concern with maintaining a desirable level of experimental control has meant that stimuli have been physically well-specified, but rather simple, varied along one dimension of interest at a time. Despite the merit of this rationale for designing stimuli, certain of the early “deficiencies” in vision may actually be reflections of infants’ disinterest in the austere stimuli chosen by researchers.

Even if we acknowledge that current methods yield underestimates, infant vision is still obviously inferior compared to adult vision. But which adults? In many aspects, everyday adult vision may also be far less acute and precise than we imagine, probably because many of the daily tasks requiring vision can be done with less full attention and the recruitment of less than the highest levels of performance. Indeed, in our own studies, we regularly find that adults, if they are naive and uninstructed, also perform at a level considerably below that commonly reported in the published literature for the practised adult subjects who are measured in most psychophysical experiments. Although an interesting observation about vision in its own right, this fact must change our perspective on the visual performance of infants, who necessarily are unpractised and uninstructed. Comparing infant performance to that of “world-class” adult subjects can give a misleading impression of infant vision; not all adults are equal.

Further, infants might not have much use for better, more adult-like vision, were it available to them. To put infant vision in some perspective, it is useful to ask what infants use their vision for. Adopting a functionalist stance, it could be said that the sensory “deficits” of early infancy may not really make much difference in the infant’s daily life. The most acute levels of adult vision depend on anatomical structures that are immature in young infants; but these structures support visual functions that are not of great use early in life (for example, extremely fine acuity for detecting and hunting distant prey, and its modern analogue, reading, or exquisitely fine stereoscopic discriminations of depth). Young babies simply do not need to be good hunters or be able to read the fine print in a contract.

What this amounts to is the suggestion that visually normal infants have the level of visual functioning that is required for the things that infants need to do. Despite documented immaturities, infants’ vision is good enough to derive an accurate impression of the world in their immediate vicinity and to stimulate further visual, cognitive, and social developments. Even with their poor vision, it would be a mistake to see infants as “handicapped” functionally, given the functions appropriate for their age. We might even posit that higher levels of visual functioning could actually interfere with these tasks, if they add to the visual “noise” that needs to be filtered out in order for infants to attend to the most relevant information (Hainline & Abramov, 1992; Turkewitz & Kenny, 1982).

THE NEED FOR ASSESSMENT OF INFANT VISION

The observation that infants might not benefit from more acute vision does not detract from the need to study the development of visual abilities. On the contrary, it is important to understand visual development in early infancy and to evaluate infants’ vision earlier and on a much wider basis than has been customary. Visual development does not follow a fixed blueprint, and irregularities in the normal developmental sequence can have permanent effects on the final level of visual functioning. A well-documented fact about mammalian vision is that the development of the visual system depends greatly on the nature of the visual stimulation early in life; in other words, visual development is characterised by a series of sensitive periods for different visual functions, with the developmental process shaped by the nature of the inputs the system receives. Animals that are visually deprived when young often fail to develop a full range of visual abilities as they age (Atkinson & Braddick, 1988; Mitchell, 1988; Movshon & Van Sluyters, 1981), and humans appear to show similar developmental plasticity as vision develops during infancy (e.g. Archer, 1993; Atkinson, 1993; Held, 1981; Levi & Carkeet, 1993; Tychsen, 1993). The issue here is to delineate the normal sequence of some important visual milestones. Clinical assessment of vision and visual disabilities in infants is dealt with more extensively in Chapter 2, by Moseley et al.

DEVELOPMENT OF VISUAL STRUCTURES

The visual system consists of a sequence of stages for the translation of light energy in the world into neural impulses, which are processed at a series of higher and increasingly more complex levels in the nervous system. At the earliest stage, the relevant parts of the visual system are optical (cornea, lens, etc.); these structures form an adjustable optical system to bring images into focus on the retina at the back of the eye. Actually, as we have two eyes, at this stage the system consists of two
separate optical systems, forming separate images on each of the eyes' retinas. Thus, the quality of vision depends, first of all, on the ability to change the lens's focus appropriately (accommodation), and to rotate the eyes to point at a specific target location (vergence). Failure of accommodation results in blurred images. Failure of convergence results in the perception of double images (diplopia), or the suppression of information from one of the eyes. The brain is responsible for fusing the two retinal images into one coherent image; in most cases, the brain also uses the information from the two eyes to allow fine depth discriminations through stereopsis.

Because humans' eyes move to allow inspection of the details of the world, visual functioning depends critically on how well an individual is able to fixate and to make saccades and pursuit movements. In normal life, the head is not stabilised by laboratory artifacts, such as chin rests or bite bars; rather, some form of compensation is needed for movements of both head and body in order for vision to remain stable. The vestibulocular reflex serves to stabilise vision during self-produced motion, while smooth pursuit and optokinetic nystagmus assist in nulling the motion of the retinal image that results both from self-produced motion and from movement of objects past a stationary viewer. Problems or immaturities in any of these abilities can significantly compromise basic sensory as well as "higher" aspects of vision and visual information processing.

The image formed on the retina by the eye's optics consists of a pattern of energy that contains information about the colours and spatial arrangements of objects in the world. It is the job, first, of the cells in the retina, and, later, of cells at subsequent neural processing sites in the visual pathway to transduce and encode this pattern of energy into one that the brain can interpret. The visual system, in other words, is also a complex neural network for coding the spatial, temporal, and chromatic characteristics of successive visual images; it serves as an important "front end" for higher mental processing. In turn, both perceptual and higher cognitive processes modify the way in which basic optical, oculomotor, and sensory functioning proceeds (e.g. see Neisser, 1976). In the final analysis, seeing is not only a sensorimotor but also a cognitive process.

Infant retina

The visual system of the human infant is structurally immature compared to the level of development seen in older children and adults. This immaturity begins at the retina, particularly in its area of highest acuity, the fovea (Abramov et al., 1982; Hendrickson & Yuodelis, 1984; Yuodelis & Hendrickson, 1986). Figure 1.1A is a photograph through a light microscope of a section from a region just outside the fovea of an adult retina; Fig. 1.1B is a stylised drawing (based partly on electron microscopy) of the cell types and interconnections shown in panel A. In these figures, the front of the eye is in the direction of the top of the page, so that light passes, as indicated, through three layers of cells before reaching the outer segments of the receptors; the outer segments contain the photopigments that absorb and transduce the incident light in the image formed by the eye's optics. It is clear from Fig. 1.1B that there is considerable neural interaction and processing of the responses of the receptors before the information is transmitted to the brain by the axons of the ganglion cells, which together form the optic nerve. Figure 1.1C is a section through an adult fovea, which shows some of the features that are unique to the fovea. First, the receptors are more slender and are more densely packed together. Second, the inner nuclear and ganglion cell layers are much less dense and even absent from the very centre, which creates the pit-like appearance of the fovea. However, this apparent reduction in cell density is misleading: the ganglion and inner nuclear cells associated with the foveal receptors are in fact more numerous than in peripheral regions— the bodies of these cells have simply migrated to the edges of the fovea so that light does not have to pass through them, which would reduce the sharpness of the image on the fovea.

Figures 1.1D and 1.1E show sections from the retina of a human newborn. Figure 1.1D is from a region just outside the fovea, which is just beyond the range of this photograph on the right. The left-hand portion of this section appears to be much like the adult peripheral retina in Fig. 1.1A, and only on the right, as the fovea is approached, are any marked immaturities encountered: the density of receptor cells decreases, the available space for the outer segments of the receptors becomes progressively narrower, and there are other signs of immaturity, particularly in the fovea. Figure 1.1E shows the neonate's fovea, in which the above abnormalities are even more pronounced. In particularly, the morphology of the receptors is very different, with the outer segments of the cones quite short and thick, rather than elongated and slender as in the normal adult and the infant periphery. In addition, the cells in the inner nuclear and ganglion cell layers have not yet migrated to the edges of the foveal pit. At this point we know that these neonatal immaturities have mostly disappeared by the end of the first year, but we still lack detailed information of the time course of many of the changes, especially during the first two to three months from birth when visual performance is changing very rapidly. The structural immaturity of the fovea is mirrored at successive levels in the visual pathway (e.g. Garey & De Courten, 1983; Hendrickson, 1993; Hickey, 1977; Huttenlocher, 1979; see Daw, 1995, for a particularly complete and clear discussion of visual neural development in humans).

Given the striking state of the anatomy at birth, it has not been difficult
1. THE DEVELOPMENT OF BASIC VISUAL ABILITIES

Spatial contrast sensitivity

Discrimination among objects in the world depends on spatial vision, that is, the ability to resolve the components of the image on the retina. According to Linear Systems Analysis, based on Fourier's theorem, any such image can be mathematically described by a linear summation of a specific set of basis functions (see Levine & Shefrin, 1990, for an introductory treatment; also, De Valois & De Valois, 1988, and Graham, 1989). It is common to use sine waves as the basis functions; each sinusoidal component of the particular set is described by its spatial frequency (measured as cycles per degree of visual angle), phase (starting point relative to the other sinusoids), and amplitude (power); in most cases the results of the analysis are given as a power spectrum, which plots the amplitude of each sinusoidal component versus spatial frequency.

Like all optical systems, the visual system selectively filters different spatial frequencies due to imperfections in its optical structures. These factors are most obvious at the acuity limit, which is a measure of the finest pattern (i.e., highest spatial frequency) that the eye can resolve when the pattern consists of extreme dark/light transitions (i.e., high contrast). However, as the retinal image also contains many lower-frequency components, measures of only acuity provide incomplete descriptions of visual performance. A more complete measure of spatial vision is the spatial...
contrast sensitivity function (csf), on which acuity is only one high-frequency point.

The stimuli typically used to measure the spatial contrast sensitivity function (csf) are gratings with alternating lighter and darker bars, each pair of light and dark bars constituting one cycle of the grating. Starting with the edge of a light bar, the luminance of the pattern gradually increases, then gradually decreases, and so on across the grating, to produce a luminance profile that varies sinusoidally across the grating (Fig. 1.2a). Although sine waves are not per se "ecologically valid" stimuli, quantitative approaches to vision argue that detection and recognition of objects (at least while stationary) can be predicted from the functions measured with such stimuli; this is discussed further later. In an experiment to measure the spatial csf, the spatial frequency of the grating (number of grating cycles per degree of visual angle) is varied systematically, and for each grating, the contrast needed for threshold detection is found. The normal csf has a characteristic inverted U-shape when contrast sensitivity (i.e. reciprocal of threshold contrast) is plotted as a function of spatial frequency (see Fig. 1.2b). At higher spatial frequencies, as the acuity limit is approached, optical factors impose a fundamental limit; the eye's optics simply cannot form an image of an extremely fine grating—the light and dark bars are so blurred that the image is a spatially uniform field. However, the drop in sensitivity at lower spatial frequencies cannot be optical in origin in a human eye and is caused, in all likelihood, by neural factors such as lateral inhibition, that is, a type of neural interaction in which stimulation of one area reduces responses to stimulation of a neighbouring area (e.g. Ratliff, 1965). The exact location of the spatial csf with respect to the spatial frequency and contrast sensitivity axes depends greatly on the conditions under which the function is measured. For example, changes in the mean luminance shift the absolute location of the curve, with higher luminances increasing sensitivity particularly at high spatial frequencies. The method used to measure the function also influences the form and location of the curve (Graham, 1989).

Infant spatial csf

The early studies of spatial resolution in infants concentrated on acuity (see Dobson, 1993; Dobson & Teller, 1978), which develops rapidly over the first year of life. Although very important, these studies cannot provide all the information inherent in measures of complete spatial csfs. The methods that have been used to obtain csfs from infants include forced-choice preferential looking (fpl), visual evoked potentials (vep), and a method we have developed that is based on recordings of subjects' eye movements, which we term eye movement-voting (emv). Fpl is derived

**Fig. 1.2.** (a) Contrast sensitivity is typically measured by presenting patterns whose luminance profile across the stimulus increases and decreases according to a sine function, as illustrated in this example. Contrast is a measure describing the difference between the highest and lowest luminances compared to the mean. (b) An example of an adult spatial contrast sensitivity function (csf), based on data presented by Van Nes and Bouman (1987). Spatial frequency, on the abscissa, is a measure of the fineness of the pattern; high spatial frequencies have many cycles of light-to-dark transitions in the same spatial extent for which a low-frequency grating has only a few such transitions. Contrast sensitivity is the reciprocal of the contrast at threshold, that is, the minimum contrast required for the detection of that spatial frequency; high contrast sensitivity implies that a low contrast is enough for detection, whereas low contrast sensitivity implies that a high level of contrast is needed to detect the pattern. (From Hainline & Abramov, 1992.)
from Fantz' (1956) preferential looking paradigm, with the variant that the dependent measure is not seconds of looking but a judgement by a person observing the infant about the position of a stimulus being viewed by the infant (Teller, 1979). Visibility and position of the stimuli are manipulated across trials to derive a threshold estimate. VEP methods use electrodes attached to the scalp to record the EEG during presentation of the stimulus. Because the visually evoked response is very small and the record very noisy, stimuli are presented in some repetitive fashion and successive responses are averaged. Usually this is done by alternating between two versions of the grating stimulus such that a pattern is always present, but the bright bars of one replace the dark bars of the other; this counterphase reversal of the pattern is at some fixed temporal rate and the responses to each cycle of the temporal alternation are averaged. The studies we will describe later use a specific form of VEP, the sweep-VEP. In this variant, one of the stimulus parameters is “swept” through a series of values to find the value at which some criterion response level is reached; for the spatial CSF, the contrast of each grating is varied, in a series of small steps from high to low, to find the contrast at which the averaged response falls within the EEG signal's noise (Norcia, 1993; Norcia, Clarke, & Tyler, 1985; Tyler, Apkarian, Levi, & Nakayama, 1979). In EMV (Abramov et al., 1984; Hainline, Camenzuli, Abramov, Rawlick, & Lemerise, 1986; Hainline, De Bie, Abramov, & Camenzuli, 1987; Hainline & Abramov, 1997), detailed records of eye movements are made while the infant views a moving pattern of stripes with different levels of visibility. As with FPL, an observer blind to the characteristics of the stimulus “votes” on each trial on which way the stimulus is moving, using the infant’s eye movements. Using standard psychophysical methods, one can determine a CSF from the pattern of correct and incorrect responses of the observer, as a function of stimulus contrast and spatial frequency.

There are several sets of data on the development of the spatial CSF for human infants, including but not limited to: Atkinson et al. (1977), Banks and Salapatek (1978), Hainline et al., 1986; Norcia, Tyler, and Allen (1986), and Norcia, Tyler, and Hamer (1990); some of these are shown in Fig. 1.3. Although these studies differ in the absolute sensitivities that are found, they agree in finding a gradual improvement in both absolute contrast sensitivity and in the range of spatial frequencies to which the system is sensitive; that is, with increasing age over the first year, the CSF (plotted as in Fig. 1.2b) shifts laterally to higher frequencies and upward to higher sensitivities. There is also evidence that sensitivity to low and middle frequencies approaches the adult asymptote at an earlier age than does the sensitivity to high spatial frequencies close to the acuity limit (Norcia et al., 1990). There is general agreement that these changes in spatial contrast sensitivity largely reflect retinal immaturities, with lesser contributions from neural processing contributed by higher centres (Banks & Crowell, 1993; Banks & Shannon, 1993).

Although showing the same general developmental pattern across age, the absolute sensitivities obviously differ as a function of the method used, with FPL producing the lowest estimates of sensitivities, VEP about 10 times higher, and EMV intermediate. It is possible that some of these differences come from differences in the specific stimuli used in the different studies cited; for adults, factors such as luminance and field size influence sensitivities (Kelly, 1961; Robson & Graham, 1981; Van Nes & Bouman, 1967). FPL is, of course, dependent on detecting infants’ relatively gross attentional discriminations between more and less visible stimuli; it has been assumed that these attentionally driven responses are voluntary and thus dependent on “higher” neural centres, although the method may also be tapping reflexive orienting to salient peripheral stimuli. A VEP depicts an average of the composite neural activity to repeated presentations of a stimulus, after processing at many different levels in the visual system. Except for the need for a minimal level of alertness, it does not depend on voluntary behaviours of the infant. VEP is generally believed to reflect the spatial sensitivity of central retina. Our EMV method is based on following eye movements elicited by moving gratings. Such eye movements are a behavioural response to moving stimuli; some kinds of following eye movements have been described as reflexive, although in practice it is not easy to discriminate between reflexive and voluntary following eye movements. Although it is a behavioural method, EMV may be more similar to VEP than FPL: in both cases, stimuli are presented directly in front of the subject and so the thresholds are more likely to represent those of central, rather than peripheral, retina.

Another major difference among methods is how long it takes to get an estimate of the CSF. The methods clearly demand different degrees of cooperation from the infant. FPL is time-consuming. With FPL, derivation of a full CSF extends over several lengthy sessions, with the risk of significant habituation to the stimuli and to the testing situation itself, as well as the possibility of centrally caused reductions in alertness. Compared with FPL, VEP is a rapid method; in some sweep-VEP procedures, a complete spatial CSF can be obtained in as little as 10–15 minutes. The EMV method takes about the same time as the sweep-VEP to derive a CSF.

Thus there are many reasons why one would not expect identical spatial CSF from the different methods – indeed, the similarities in results are all the more striking in the face of these methodological variations. It is likely that postnatal development of infant contrast sensitivity reflects predominantly the development of the fovea and central visual field, although in infant monkeys who have been used as a model of infant human visual development there is also evidence of peripheral development.
development (Kiorpes, 1996). It is important to underscore, however, that the level of spatial vision that young infants possess is normally adequate to make many of the visual discriminations that infants need to make to respond adaptively to events around them.

**Spatiotemporal interactions**

Visual stimuli vary simultaneously in both space and time, and the visual system's sensitivities to these dimensions are inextricably linked. The particular relevance in the case of infants is the possibility that, as a result of poorer eye movement control, infants cannot keep patterns stable on their retinas. The visibility (i.e., contrast at threshold) of a grating of any given spatial frequency depends on the temporal parameters of the stimulus; that is, a csf is best expressed as a particular section through a three-dimensional spatiotemporal solid (Kelly, 1979a,b). Rather than there being one csf, there is an infinite family of spatiotemporal csfs that describe an entire spatiotemporal contrast sensitivity surface. Moreover, the details of any such surface also depend on a host of other stimulus parameters such as luminance, spatial extent, and so on. An example of a spatiotemporal surface we have obtained from adults is shown in Fig. 1.4. The figure shows the smooth surface fitted over the separate csfs obtained for different rates of pattern movement. The exact ordinate values for such a surface will depend, of course, on stimulus details. For example, Fig. 1.4 shows that we did not use gratings of very high spatial frequencies, and hence the front corner of the solid is truncated; the reason was simply that the mean luminance of our display was relatively low and under these conditions, very high frequencies were not visible. Spatiotemporal surfaces have, however, general properties that hold over a range of experimental conditions.

Consider the changes in the shape of the spatial csf that would result from changes in the temporal rate. Such curves can be obtained by slicing the solid parallel to the spatial frequency abscissa. At very low temporal rates, overall sensitivity of the spatial csf is depressed and the peak is shifted towards higher spatial frequencies. As temporal rate is increased, the peak shifts to lower spatial frequencies and sensitivity to those frequencies increases. However, because of the shift of the function towards lower spatial frequencies, there is a loss in sensitivity at the highest frequencies (i.e., the acuity limit is shifted to lower values), and a gain in sensitivity at low frequencies; indeed, the usual drop in sensitivity at low spatial frequencies may no longer be evident at very high temporal rates. This is the reason why the infant vep csfs in Fig. 1.3 fail to “turn down” at lower spatial frequencies: in order to obtain a vep, it is necessary to use a time-

---

**FIG. 1.3.** Examples of infant spatial csfs obtained from the three methods that have been used to derive infant csfs: fpl, vep, and emv. Note the differences in absolute sensitivities, but the general similarity of the shapes of the curves, in most cases. (Adapted from Hainline & Abramov, 1992.)
varying stimulus, which increases sensitivity at lower spatial frequencies compared to a static stimulus.

The eyes of even highly motivated adult observers fixating static targets can drift at rates of the order of 0.5 to 1 deg/sec, unless subjects' heads are rigidly stabilised (Skavenski, Hansen, Steinman, & Winterson, 1979), and this can be demonstrated to produce changes in threshold. Except in the few cases when the effects of eye movements have been nullified (e.g. Kelly, 1979b), unknown temporal variations influence measures of the spatial CSF, even for stimuli that the experimenter intends to be "static". How much this affects the resulting CSF is an empirical question and depends on many factors. It is likely that similar effects occur for infants.

Besides this sensory effect of spatiotemporal interactions, motion of the stimulus has attentional implications for infants. A fact established fairly early about infant preferences is that even for simple stimuli, infants prefer moving and time-varying stimuli over the equivalent static stimulus (e.g. Carpenter, 1974; Volkman & Dobson, 1976), possibly because movement is an important characteristic of the stimuli infants naturally attend to; it is likely that infants find the stimuli in VEP studies and in our EMV studies more interesting than in many PPL versions and this may influence measured thresholds. Gibsonians and neo-Gibsonians would of course argue that motion is not interesting because of its influence on sensory thresholds, but because much information about objects and our relationship to objects is revealed by motion (e.g. Gibson, 1979). Indeed, it can be argued that for very young infants, the components of the world that are perceived as relating to "objects" are those components that move as a coherent unit (Kellman, 1993; Kellman & Spelke, 1983); this motion could either be due to real motion of the object relative to the background or to motion of the infant (Kellman, Gleitman, & Spelke, 1987). The ability to detect both the degree and direction of movement has been studied developmentally (see Braddock, 1993; Braddock, Atkinson & Hood, 1996, and Wattam-Bell, 1996) and appears to change with age during infancy in a way that reflects higher levels of processing, particularly for faster-moving stimuli. However, the current data support the notion that even the young infant has the rudiments to attend to, process, and deal with moderate movement in visual stimuli.

Functional consequences of the infant CSF

A subject's CSF can be regarded as a selective filter for the spatial frequencies represented by the Fourier power spectrum of a given stimulus; this is the central assumption of "linear systems" approaches in vision (see e.g. De Valois & De Valois, 1988; Ginsberg, 1986). Although it is clear that above threshold vision is in many respects nonlinear, the linear systems approach has been successful in describing many aspects of visual performance. Some studies, for example, suggest that object recognition depends heavily on information at lower spatial frequencies (Ginsberg, Evans, Sekuler, & Harp, 1982); low frequencies communicate information about an object's broad features. Discrimination among similar objects may depend, however, much more on high-frequency information; this is clearly seen in letter discrimination tasks like the traditional Snellen test, in which reductions in high spatial frequency vision (i.e. acuity), often due to optical blur, severely impair visual performance (e.g. Thorn & Schwartz, 1990). At this point we must add that it is a serious oversimplification to consider only the power spectrum of a stimulus: the phase spectrum (i.e. the starting point on its cycle of each sinusoidal component of the stimulus) is also very important; if a stimulus is recreated using its original power spectrum but with a random phase spectrum, the object may no longer be recognised (Piotrowski & Campbell, 1982).

The ability of the visual system to recognise and discriminate among
patterns will depend both on the individual viewer’s csf and on the Fourier spectra of the particular stimuli involved. The assumptions of linear systems analysis allow us to simulate what various objects might look like to a young infant: in essence, to substitute their visual system for our own in a limited way. Because faces are such an important part of the infant’s visual world, we have used them to illustrate some of these points. In Fig. 1.5a we start with a sharply focused image of a face, which we will filter so that when it is viewed by an adult with normal vision it will appear as if the adult’s visual system had been replaced by that of a 2-month-old infant. In Fig. 1.5b we show the csf for a normal adult and a typical 6- to 8-week-old infant that will be used in the filtering process. These particular curves are in fact examples of Kelly’s (1979b) model of spatial csf that we have generated for a specific set of viewing conditions, as follows: a moderate photopic luminance level (e.g. equivalent to viewing the figure under a good reading light), and a temporal parameter value approximating a normal, casual, nonlaboratory degree of fixation stability (1deg/sec). The original image of the face was analysed to obtain its two-dimensional Fourier spectrum; the power at each spatial frequency was then rescaled according to the infant’s sensitivity to each frequency to produce an effective (or filtered) power spectrum of the image. (Strictly speaking, the csf, or filter, in Fig. 1.5b is for a one-dimensional stimulus, and a two-dimensional filter had to be derived from it for the filtering operation.) This filtered spectrum could have been used to resynthesise a filtered image of the face. But had we done so, the resulting image would have been quite misleading. The reason is that when adults look at such a filtered image, they are imposing a second stage of filtering — that due to their own visual systems. The problems raised by this are best appreciated by considering some low spatial frequency to which the adult and the infant are approximately equally sensitive; at that frequency the double filtering would be equivalent to squaring the infant’s filter, which would lead to overfiltering at these frequencies. To compensate for this, we have rescaled the infant-filtered power spectrum by the inverse of the adult’s csf before resynthesising the image. Thus, assuming that the viewer’s csf is like the adult csf in Fig. 1.5b, the inverse filtering according to that function is negated by the filtering imposed by the viewer’s own csf, and the viewer should see the face as if the viewer’s visual system had been replaced by that of the infant. Because the spatial frequency content of a face necessarily changes with viewing distance, we have repeated the operations for two viewing distances of 30cm and 150cm. The resulting images are shown in Fig. 1.5c,d. At the near distance, the stimulus is clearly still recognisable as a face, although at the further distance, this becomes a questionable identification. In all of these manipulations, the original phase relationships have been maintained, otherwise the image would have been “scrambled”. In short, at the distances at which infants are called on to recognise faces (arm’s length), even the visual system of a young infant is probably adequate to discern a face’s critical features.

The linear systems approach to visual discrimination has already been applied to normal infant preference data with some success: simply by using an age-appropriate csf to filter stimuli, it is possible to predict infants’ visual preferences (Atkinson, 1977; Banks & Salapatek, 1981; Gayl, Roberts, & Werner, 1983; Slater, Earle, Morison, & Rose, 1985). A lot of the work on stimulus recognition and discrimination has dealt with faces, probably because of the high “ecological validity” of such stimuli. There have recently been some attempts to examine the relative importance of phase and power spectra for recognition of faces by infants. As already noted, for adults the phase spectrum of a stimulus is a much more important determinant of recognition than is the power spectrum. Thus, in the recent developmental studies, “face-like” stimuli were analysed into their Fourier components and then recombined using different phase relations, but preserving the original power spectra (Dannemiller & Stephens, 1988; Kleiner, 1987; Kleiner & Banks, 1987). The results have been interpreted as showing that infants younger than 2 months are responding in their preferences on the basis of total power, irrespective of the phase relations among the components. That is, they do not always prefer the stimuli seen as face-like by adults. However, these interpretations must be treated cautiously: for example, Kleiner (1987) compared newborns’ preferences for a schematic face versus a lattice made up of squares on a blank background; comparisons also pitted preferences for stimuli created from the phase spectrum of one and the power spectrum of the other. Kleiner’s conclusions focus on the fact that newborns, unlike 2-month-olds (Kleiner & Banks, 1987), do not prefer the lattice stimulus with the face’s phase relations (which looks like a slightly distorted face to adults). However, the data also show that newborns, like 2-month-olds, prefer the original schematic face to the stimulus with the face’s power spectrum but the lattice’s phase spectrum, which shows that they are not phase-insensitive and that the face’s phase structure may be inherently attractive, even to newborns (Morton, Johnson, & Maurer, 1990).

Beyond the problems described, there is a more general question of whether such studies are indeed examining the perception of “faceness”. The designation of these schematic and simplified stimuli as “face-like” is based on adults’ categorisations. It is not clear that infants categorise such stimuli equivalently (Cohen, 1988). Before concluding that young infants respond to faces primarily on the basis of some low-level sensory signal (such as the total effective power in a stimulus as derived from a linear systems analysis), the studies need to be repeated using real faces.

Although powerful quantitatively, linear systems analysis may not tell
1. THE DEVELOPMENT OF BASIC VISUAL ABILITIES

us all that we need to know about stimulus appearance. Strictly, linear systems approaches to stimulus discriminations can predict performance only when responses vary linearly with intensity. When contrasts are close to threshold, it is probably safe to assume that the visual system is behaving linearly. However, when contrasts are well above threshold, as in normal viewing conditions, the system may become nonlinear. Furthermore, the nonlinearities may be associated more strongly with certain regions of the spatial frequency spectrum. Finally, different tasks may be biased towards particular frequency ranges; for example, reading may depend more on higher spatial frequencies, but categorising something as this or that object may be more affected by sensitivity to lower frequencies. These factors make it difficult to predict exactly how a given CSF will affect visibility of real objects above threshold.

Most of the studies of the effects of the CSF on perception have dealt with static “snapshots” of the stimulus. Gibsonian theory reminds us, however, that stimulus recognition may be more crucially dependent on invariant features of dynamic displays extended in time (e.g. Arterberry, Craon, & Yonas, 1993; Bertenthal, Proffitt, Speet, & Thomas, 1985; Kellman, 1993; Kellman & Spelke, 1983; Ruff, 1982). Until recently, motion of the stimulus or of the subject with respect to the stimulus have been regarded as nuisance variables rather than as a necessary dimension of study. As we saw in the discussion of spatiotemporal interactions, it is far from trivial to deal with both spatial and temporal dimensions in a controlled fashion, and the visibility of a target in motion may be different from that of an object at rest. These problems notwithstanding, linear systems analysis of the capacities of the developing visual system offers a powerful means to understand the implications of neural immaturity, to predict what visual information is available to the infant for other developmental functions, and to evaluate early visual system dysfunction. Comparisons of results from different methods can serve as the source of converging information to give a perspective on the development of spatial vision not accessible from any single method of data collection.

DEVELOPMENT OF COLOUR VISION

The ability to see light with different physical wavelengths as having different colours necessarily requires an interaction of mechanisms at the retina. Normal humans have two classes of photoreceptors, rods, which are maximally sensitive in low light levels, and cones, which are sensitive at higher light levels. There are three different cone types in the normal human (most forms of “colour blindness” result from genetic anomalies leading to abnormalities or absence of one of these cone types). The rods and cones of the retina contain different photosensitive chemicals, the

---

**FIG. 1.5.** The effect of the infant CSF on visibility of an “ecologically valid” stimulus. (a) A photograph of a prominent British subject. (b) Hypothetical CSFs for a normal 6-to-8-week-old infant and an adult under conditions of good photopic luminance and moderate fixational stability. These are used to filter the image in (a), as described in the text. (c) The same face as it might be seen by a 2-month-old viewing from a distance of 30cm. More correctly, when the figure is viewed by an adult with a normal CSF, it will be seen as if the adult’s visual system had been replaced by one with the CSF of the infant in (b). (d) A similarly filtered image of our subject, but at a distance of 150cm. Image processing was done by Shuai Chen. This figure is published as a tribute to Princess Diana.
photopigments, which differ from one another in what wavelengths they react to (absorb) maximally. To tell that a light is of a different hue than another (rather than just more or less intense) requires a comparison of the outputs of at least two receptors with different photopigments. The photoreceptors whose output is being compared could be two (or more) types of cones, or a rod and a cone. How sensitive to light a given photoreceptor is depends on how much opportunity the incoming light has to interact with the photopigment in a particular photoreceptor; this is directly related to the shape of the photoreceptor outer segment. Long, slender outer segments will have a better chance of responding to light, on average, than short, stumpy outer segments. This is probably why absolute sensitivity to light increases over the first months of infancy, as the receptors mature, more for the rods than the cones (Brown, 1993; Fulton, Hansen, Dorn & Hendrickson, 1996; Teller & Lindsey, 1993a), which is consistent with what we know about the structural development of photoreceptors in the infant retina. The difficulties in studying colour vision in young infants have made definitive conclusions about the nature of infant colour vision difficult to obtain. The data justify the straightforward conclusion that colour vision develops rapidly in the first year. One-month-olds have often been reported to fail to make some colour discriminations (e.g. green vs. yellow) that one would expect they should be able to make if they had all 4 functional photoreceptors (rods plus 3 cone types), but by 2 months, evidence suggests that normal infants can make most discriminations, particularly if stimuli are large and differences between stimuli are marked (Clavadetscher, Brown, Ankrum, & Teller, 1988). One hypothesis (e.g. Brown, 1990) is that in the earliest months, at least one cone mechanism is not contributing, but rods are playing a role important to chromatic discriminations among stimuli. Another interpretation of the data is that there is an overall loss of sensitivity of all chromatic and luminance mechanisms, possibly due to the developmental differences in the morphology of the photoreceptors, with the least sensitive cone type undetectable because of both weaker neural signals and greater noise in the neural signals, i.e. a lower signal-to-noise ratio (Banks & Bennett, 1988; Banks & Shannon, 1993; Brown, 1993). VEP methods, which in tests of spatial vision appear to have greater sensitivity than behavioural methods like FPL, yield data pointing to the existence of the normal complement of colour mechanisms, albeit with weaker outputs, during early infancy (e.g. Allen, Banks, Norcia & Shannon, 1993; Burr, Marrone, & Fiorentini, 1996; Knoblauch, Bieber, & Werner, 1996; Volbrecht & Werner, 1987).

Even if young infants have all three cone types, and all are functioning, the data do not unequivocally tell us about the salience of colour to infants. The problem is that the outputs of the different types of cones are combined in two fundamentally different ways, one of which communicates information about hue and the other brightness. In one neural pathway, the responses of the different cones are subtracted from each other to yield what is called "spectrally opponent" or "chromatic" mechanisms that communicate hue. In the other pathway, cone outputs are combined additively to yield "achromatic" luminance-detecting visual mechanisms. The colour (or strictly, the hue) of a stimulus is determined by the spectrally opponent mechanism, but the strength or salience of the colour (strictly "saturation") depends on a comparison of the relative strengths of the chromatic to achromatic responses (see Teller & Lindsey, 1993b). There is evidence that young infants are less sensitive to chromatic than luminance differences. Stimulus size and chromatic differences must be large for infants to display the chromatic discriminations that their spectrally opponent mechanisms seem capable of communicating, so, as we have seen before, aspects of the stimulus other than its wavelength will affect how an infant responds to it.

It is difficult to know for certain, but it is probably safe to say that even young infants have a form of colour vision and it is probably largely like that of adults. As a result of the lower responsiveness of their cones and related chromatic mechanisms, colours probably appear less intense than the same colours would to an older infant or an adult. If required to make discriminations among colours, infants would not be able to distinguish as many colour differences, although outside of the laboratory this is not likely to be a task often posed to young infants. Their colour vision should be enough, however, to be able to use commonality of surface colour to distinguish the boundaries of many objects as those objects move across the infant's field of view, which is an important use of colour in the viewing of natural scenes. [None of the data allow us to know whether infants (or dogs or cats) dream in colour – a question that for some reason I am frequently asked when people learn that I study infant vision. Sorry!]

DEVELOPMENT OF ACCOMMODATION AND VERGE

In our discussions of spatial resolution and the CSR, we assumed implicitly that changes in the eyes' optical properties had been controlled or optimised. However, in order to produce a good image on the retina, one that preserves high spatial frequencies, the eye must be appropriately focused; also, to see a single image of an object, both eyes must rotate correctly to point at the object. The abilities to do these things change with age and must be considered when evaluating infants' visual performance.
Acmodation

Acmodation refers to the process of changing the curvature of the crystalline lens to vary the eye's refractive power. Refraction is a measure of the bending of light rays by an optical system, expressed in diopters (D), a unit defined as the reciprocal of the focal length of the lens in metres (and also of viewing distance, given correct accommodation). For example, 1D corresponds to a 1 metre distance, 2D to 1/2 metre, 4D to 1/4 metre, etc. Less accommodation is needed to focus far objects, more to focus near objects. Emmetropia refers to the condition when an individual completely relaxes accommodation and the resulting plane of focus is very far (optical infinity, effectively about 6 metres); by increasing accommodation, it is possible to focus nearer objects. In near-sightedness or myopia, when accommodation is relaxed, the eye is focused at a relatively close distance and therefore distant objects will appear blurred. In contrast, in far-sightedness or hyperopia, the plane of focus when accommodation is relaxed is beyond optical infinity and all objects appear blurred; hyperopes must accommodate to see clearly targets at any distance. If they start with completely relaxed accommodation, all three of the described types of individuals can, in principle, focus nearer objects. Hyperopes must accommodate to some degree for all distances, including very distant ones. Near-sighted or myopic individuals must accommodate for objects closer than their relaxed plane of focus. Note, however, that myopes cannot use accommodation to focus on objects beyond their plane of relaxed accommodation, which is why for myopes without optical correction, far objects appear blurred. Also, because the ability to accommodate declines steadily with age, the range of focussable distances ultimately declines for all individuals, particularly after age 40 (e.g. Millodot, 1982).

Vergence

Functional vision depends not only on the clarity of the image for each eye but also on how the eyes are jointly positioned. Vergence refers to the rotation of the eyes so that their viewing axes intersect, ideally on a plane for which the eyes are accommodated. The angle of rotation is measured in prism diopters (a prism of one diopter shifts the image by 1cm at a distance of 1m). The primary function of vergence is to reduce the disparity of the two retinal images and to bring the target into single binocular registration. In most adults, the accommodative and vergence systems are synergistically linked so that each system influences the other's performance (see Schor & Ciuffreda, 1983). Thus, there is another component of vergence caused by a change in accommodation (accommodative vergence); similarly, an important component of the accommodative response is accommodation caused by a change in vergence (vergence accommodation). As each eye looks at objects from slightly different angles, there is a disparity between the images in the two eyes that is inversely related to viewing distance. The images of an object on the plane to which the eyes are converged fall on corresponding retinal regions and are normally fused and seen as a single object. There is also a region, known as Panum's fusion area, which extends in front and back of the binocular fixation plane, within which all points are fused and seen as single objects, even though they do not fall on strictly corresponding points; these targets also are seen at difference distances in front of or behind the convergence plane. For adults, vergence driven by disparity is much more precise than accommodative vergence, and so fusional vergence is more likely to be seen under normal conditions (Schor, 1979). As fixation disparity increases, a point is reached at which Panum's region is exceeded and the fusion breaks down, resulting in double vision or diplopia. Stereoscopic information about an object's depth also depends on disparity, although it appears to be a higher function than fusion; one can have single vision, that is, fusion, without achieving stereoscopic depth perception. It is likely that these skills depend on the development of different cortical centres.

For both accommodation and vergence, the spatial frequency content of the stimulus and the observer's spatial contrast sensitivity (or, more correctly, spatiotemporal contrast sensitivity) interact with response. From a developmental standpoint, we would expect to see some correspondence between developments of the spatial csf and the accommodation/vergence system. With cooperative, instructable subjects, the most common method to measure refractive status is retinoscopy, in which a practitioner evaluates the refractive error by determining the power of lenses that need to be substituted externally to null the amount of refraction of a test light by the various optical elements of the eye. Retinoscopy with infants requires considerable practice. In order to facilitate the evaluation of infant refractive status, several photographic methods have also been devised; among these are orthogonal, isotropic, and paraxial photorefraction (Atkinson et al., 1981; Bobier & Braddock, 1985; Howland & Howland, 1974; Norcia, Zadnik, & Day, 1986). The methods differ in the exact optical principles employed, but all provide an instantaneous photographic sample of refractive state from an uninstructed subject. Vergence is measured by monitoring changes in the eyes' angle of rotation. A common method is to measure changes in the position within each pupil of the image of a fixed reference light source reflected from the eye's cornea. The position of this corneal reflection and its shift with a given degree of eye rotation depend on a number of physical features of the eye, including corneal curvature, the location of the fovea and the separation of the eyes, all of which are known to change developmentally.
One of the widely used methods for infant refraction is paraxial photorefraction, also called eccentric photorefraction, photoretinoscopy, and static photographic skiascopy (Abramov, Hainline, & Duckman, 1990; Bobier & Bradick, 1985; Howland, 1985). First developed by Kaakinen (1979), the method takes advantage of a problem encountered by amateur photographers: snapshots of people taken with inexpensive cameras often show an effect known as “red-eye”, caused by the reflection of light from the flash back from the retina into the lens of the camera. With appropriate optics one can measure the degree of misaccommodation by measuring the size and position of a crescent of light. A system we use in our laboratory, with small lighted dolls at different distances as targets, is illustrated in Fig. 1.6a.

Figure 1.6b presents a series of pictures from an adult and an infant, both with normal refractive status. As the subject accommodates on the progressively closer targets, the crescents that are photographed through the eyes’ pupils grow in size. The crescents are all across the top of the pupil, which indicates myopia in our system, because the targets are all closer than the camera distance: a subject who accommodates correctly for any of these targets will appear myopic with respect to the camera. To relate the size of a crescent to accommodation, in diptors, when calibrated, yields a sensitive estimate of plane of focus on each picture. Vergence is measurable from evaluation of the reflections of the flash source in the two pupils simultaneously. The method works effectively both to delineate the normal developmental progress of accommodation and to screen for infants with problems of vergence and/or accommodation.

Developmental studies of accommodation and vergence

**Accommodation** Young infants have been reported to accommodate poorly, if at all. The earliest study of infant accommodation by Haynes et al. (1965) used dynamic retinoscopy. They reported that infants under 1 month had essentially frozen accommodation; however, refractive power changed when the subjects fell asleep, and thus the findings from the subjects when awake could not have been caused by factors such as inherent inelasticity of the lens. Not until 4 months did infants accommodate competently to targets at different distances. Subsequent research (e.g. Banks, 1980; Bradick, Atkinson, French, & Howland, 1979; Brookman, 1983) using different methods and stimuli, lowered the age of acceptable accommodation closer to 2 months, but still concluded that before 2 months, infants were poor accommodators. Assuming that accommodation was driven primarily by the need to keep high spatial frequencies

---

*Fig. 1.6. Paraxial photorefraction of infants (from Abramov et al., 1990). (a) Schematic view of photorefractor and stimuli, showing the relationship between the camera, the flash unit and the series of accommodation targets to the subject. The infrared television equipment is used to position the subject correctly for each photograph. (b) Paraxial photorefractive pictures of a visually normal adult and infant viewing targets at different distances (demands to accommodation). The critical aspects of the photograph are the size and the position of the bright “crescent” in each eye.*
in focus (i.e. to reduce retinal blur), Banks (1980) proposed that young infants did not accommodate better because the small sizes of their pupils together with their low acuity created an optical system with a large depth of focus, making accommodation unnecessary and irrelevant. With small viewing apertures, a wide range of focal planes is in focus simultaneously, reducing the need to accommodate. This argument overlooks another critical factor in determining depth of focus of an optical system. Decreasing pupil diameter, like reducing the aperture in a camera, increases depth of focus, but only if the focal length of the optical system is constant. That is not the case in infants’ eyes, which not only have smaller pupils, but also are shorter in axial length. The small size of the infant eye is compensated for by the greater optical power of the infant lens and cornea (Bennett & Francis, 1962; Lotmar, 1976). Relative depth of focus can be measured by the ratio of an optical system’s focal length to the diameter of its pupil. When both factors are taken into account, there is no great difference in depth of focus between infant and adult eyes (Banks & Bennett, 1988), so this cannot be an explanation to account for developmental differences in accommodation.

There may also be problems with the assumption that 1-month-old infants’ poor accommodation is related to their reduced acuity and the role that high spatial frequencies play in accommodation. Acuity develops continuously over the first year (see e.g. Dobson & Teller, 1978), with no sudden increase being reported between 1 and 2 months. Consequently, an explanation of poor accommodation as the result of reduced acuity would seem to require the additional assumption that good accommodation requires some as yet unspecified threshold level of acuity. Alternatively, acuity may not be the most relevant ability here. High-frequency sensitivity develops over a longer time course than sensitivity to low and middle spatial frequencies (see Fig. 1.3). If intermediate spatial frequencies are more important for infants’ accommodation, the time course for the development of acuity would not be immediately relevant in predicting accommodative development. The time course for development of sensitivity to intermediate sensitivities is a closer match to the accommodation data; in the data described earlier for spatial CSF development, sensitivity to intermediate spatial frequencies changes significantly between 1 and 2 months, and then stabilises.

It is also possible that accommodative performance improves for reasons other than a change in visual ability. Studies have found that infants accommodate more accurately to near than to far targets (Banks, 1980; Braddick & Atkinson, 1979). There are also findings that an infant’s attention to stationary objects is inversely related to their distance from the infant, even when retinal size is kept invariant (de Schonen, McKenzie, Maury, & Bresson, 1978; McKenzie & Day 1972). This preferential atten-

tion to near objects is not seen if the objects move, flash, or rotate, presumptively because these manipulations enhance infants’ attention (McKenzie & Day, 1976). The distance of an object will be related to the spatial frequency content of the object; objects farther away have more high frequencies, which young infants may not detect unless middle spatial frequencies are enhanced by movement. Or it may simply be an issue of attention, not detection; if young infants’ accommodative performance were to improve with the addition of attention-enhancing features, we might not need to seek optical factors as an explanation of earlier results.

In research done in our laboratory, we have been measuring the simultaneous accommodation and vergence responses from young infants (Hainline & Riddell, 1995, 1996). What we find is that most young infants (and the occasional older infant) who do not show evidence of varying accommodation to targets at different distances often do not show any evidence of vergence either, suggesting that, for some reason, they are simply not paying attention to the change in target distance. Although we do not have an independent measure of behavioural state, our best guess is that the infants who showed no response in our system were in the open-eyed, trance-like state often seen in young infants; Banks (1980) also reported cases of flat accommodation in drowsy infants. The incidence of attending, correctly accommodating, and verging infants increases with age over the first half-year; but even young infants behave appropriately some of the time.

Such results point up a common problem in work on sensory development; if we use the kinds of stimuli that allow maximal experimental control and that are scientifically justifiable, and indeed, required in adult studies, we may at the same time be stripping the stimulus of many of the dimensions that rouse infants’ interest. Sensory psychophysics typically calls for static, silent stimuli that are hard to see because they are close to threshold. Yet research on infant attentional preferences tells us that infants would rather look at moving, flashing or otherwise changing stimuli, well above threshold contrast and rich in contours. The Gibsonian tradition demands that stimuli have “ecological validity”, usually meaning that a preference is given to real objects or events. When real objects are used as stimuli, however, it easy to make the mistake that the infant is responding to the stimulus as an adult would and end up attributing too much cognitive competence to an infant who is, in actuality, responding to differences across displays on only a sensory level. The obvious answer is to test the same ability with both well-controlled and more complex realistic stimuli, but this is rarely done.

Findings such as these imply that the failure to see accommodative responses in a young infant may not require an explanation in terms of the large depth of focus of the infant eye or low acuity. Under the right
circumstances, it seems that infants can accommodate reasonably well. What we do not know, at this point, is what drives the accommodation of young infants. Is there a relationship within individual infants between peak CSR sensitivity and accommodation, so that younger accommodators would be found to have “precocious” spatial vision? Or are we observing vergence-driven accommodation, executed to maintain fusion? Evidence is beginning to accumulate (Currie & Manny, 1990; Hainline & Riddell, 1996) that binocular fusional mechanisms are important in determining correct accommodation in young infants. More within-subject measures are needed to address these questions.

**Vergence** There is a small number of studies of binocular vergence in infancy. They agree that convergence is relatively poor, particularly for closer targets, until around 2 months, not surprisingly the age at which accommodation is also reported to be markedly improved (Aslin, 1977; Aslin & Jackson, 1979; Slater & Findlay, 1975). Our data show that even at early ages, vergence performance on average is quite good, possibly because of the attractive value of our targets. Unfortunately, the other studies have not simultaneously measured accommodation, so that the direct link between accommodation and vergence could not be checked.

**DEVELOPMENT OF SENSORY BINOCULARITY**

Binocular vision refers to the interaction of vision in the two eyes. “Fusion” is the term used to describe the phenomenon when the brain merges the images from the two eyes (which are slightly different because the eyes are separated by an interpupillary distance of 1–2 inches, a parameter closely correlated with age and growth of the head). The finest cue to depth perception, stereopsis, emerges when specialized cells in the brain’s visual cortex compare the inputs from the two eyes and abstract a signal that conveys very small depth differences for targets that are relatively close – within arm’s length. Binocular fusion depends on having the ability to point the two eyes appropriately when looking at a single object; in our research we (Hainline & Riddell, 1995; 1996) have shown that infants show reasonably good eye alignment from an early age, in most cases. Because of their lower acuities and the probability that Panum’s fusional area is broader in infants (although no one to date has measured it), it is likely that even the youngest infants, if visually normal, demonstrate the behaviors precedent to fusion from an early age. Attention can influence alignment; for example, it is common to see departures from good eye alignment when an infant is very sleepy. But most of the time, infants direct their eyes so that the image of regard falls on corresponding retinal points for both eyes. In other words, consistent eye turns are not normal, even for young infants. Sometimes, the infant may appear to an observer to be “waltz-eyed”, that is, with each eye turned out. This “pseudo-strabismus” comes from differences in the relationship between the size of the eyeball and the position of the fovea for infants compared with adults (Slater & Findlay, 1972). The appearance is enhanced in an infant with a wide bridge of the nose, although measured alignment is usually fine. Our hypothesis is that even though crude, some ability to align the two eyes probably is necessary early on so that as the fovea develops and creates the ability of increasingly fine acuity, the eyes can continue to be kept in alignment with a degree of precision consistent with the level of spatial vision, as a platform for the eventual development of visual functions requiring high binocular acuity such as stereopsis.

The data on stereopsis in infancy agree fairly well that initially coarse stereopsis emerges rather suddenly somewhere between 3 and 6 months of age (Birch, 1993; Birch, Gwiazda, & Held; 1982; Shea, Fox, Aslin, & Dumas, 1980; Shimojo, 1993; Shimojo, Bauer, O’Connell, & Held, 1986), followed by a rapid increase in stereocuity (a measure of the smallest binocular disparity detectable) to near adult levels in the second half of the first year. Held (1993) has argued that the appearance of stereopsis signals the development of particular cortical structures that compare the inputs from the two eyes. We have the sense from these and other data that much of the sensory and motor development that has been measured in the first half-year is a kind of preparation for the visual refinements (including stereopsis and the ability to detect very fine stimulus offsets, termed Vernier acuity (Shimojo, Birch, Gwiazda, & Held, 1984) that emerge near the end of that time period.

**OCULOMOTOR DEVELOPMENT**

Human infants are motorically immature compared with other primates. Because poor control of eye movements could adversely influence functional vision, it is interesting to ask how mature their oculomotor systems are. There are many ways to record and measure eye movements, some of which are inappropriate for use with infants. The most common methods in use with infants include simple observation, the electrooculogram (EOG) reflecting changes in electrical potential across two electrodes as the eye moves, and video measurements that focus on landmarks of the eye that change as the eye moves (corneal reflection methods that typically reference the position of a stationary light source to the position of the centre of the eye’s pupil during eye movements). The methods have various advantages and disadvantages, which are discussed in Shupert and Fuchs (1988), Hainline and Abramov (1992), and Hainline (1993), among other sources.
Pointing the fovea: Saccades and fixations

Saccades  Saccadic eye movements, among the most rapid of all muscular responses, serve to point the fovea at targets that must be examined in some detail during a fixation. They are fast presumably to minimise visual “down time” as the eye moves from fixation to fixation. Because we know that the foveae of young infants are immature, we might expect that there should be marked immatures in the saccadic and fixational behaviour of young infants. We have measured the details of saccades and fixations in young infants looking at a variety of targets, using corneal reflection methods. An example of such an eye movement record, and how the various episodes are extracted, is shown in Fig. 1.7.

When we analyse the properties of the saccades, we discover that the quantitative parameters of saccades, for example, the maximum velocity that the eye accelerates to in a saccade, which is related to the amplitude of the saccade, are very similar for infants and adults (Hainline, Turkel, Abramov, Lemerise, & Harris, 1984); this relationship is termed a “main sequence” (Bahill, Clark, & Stark, 1975); see Fig. 1.8.

However, we find that how “adult-like” saccades are depends on whether the stimulus material that the infant is looking at has many contours, associated with infant attention (Fig. 1.8d). With simpler stimuli, infants’ saccades are slower, possibly because of reduced physiological arousal and its effects on the structures in the brain stem reticular formation, implicated in both attention and eye movements.

In terms of amplitude, when freely scanning scenes, the distribution of saccadic amplitudes for infants and adults look similar (Hainline & Abramov, 1985). Some studies report that when infants are asked to look at a new target after the current target disappears, they approach the new target with a series of equal-sized, “step-like” saccades (Aslin & Salapatek, 1975). Not all studies show this effect, however (Hainline & Abramov, 1985), and it may also be the result of low attention, as the frequency of adult step-saccades increases with fatigue and inattentiveness (Bahill & Stark, 1975). An alternative explanation is that infants need to calibrate how much of an eye movement is necessary to reach the target (Aslin, 1993), particularly in an otherwise featureless scene. We do not regularly observe step-saccades when infants look at natural scenes, so they may be an artifact of the laboratory. In general, the saccadic system seems quite mature and ready to function to reorient the fovea at high speed, even early in life.

Fixations  A saccade is usually followed by a fixation, during which the eyes are held more or less stationary and the new target is examined (Fig. 1.8a and b). In contrast to positions that have characterised young infants as “captured” by visual stimuli (Stechler & Latz, 1966), we generally find that when infants are alert and attending, their fixations are actually briefer in many cases than those of un instructed adults. Infant fixation duration appears to be related to how many visual details are available to pull the eye away from the current fixation; with many contours, fixations are briefer than when contours are less prevalent (Hainline & Abramov, 1992; Harris, Hainline, Abramov, Lemerise, & Camenzuli, 1988). Part of the confusion about infant fixation may arise from an unfortunate redefinition of the term “fixation” in the infant attention literature, where the term is used to denote the total amount of attention directed toward a stimulus; we would suggest that this behaviour might be better termed “looking”, leaving the term “fixation” to its traditional meaning in the ocu lmotor literature (psychoanalysis obviously appropriates the term as well, but not in a form likely to cause confusion in discussions of infant vision). It then is not a paradox that when an infant is extremely interested in a visual stimulus, total looking could be high, but the length of individual fixations, as infants actively scan the interesting object, could be brief.

Another issue concerns the question of how stable the eye is during fixation. In nonlaboratory situations, even for adults the eye actually drifts quite a bit (e.g. Skavenski et al., 1979). Infants appear to have more difficulty maintaining stable fixations, but seem to be sensitive to this; higher drift rates are associated with shorter fixations, as if the visual
A final point about fixations concerns the precision of “foveation”, related to the question of whether, when a target is looked at several times (i.e. refixated), how widely scattered about the target are the successive fixations. Even highly trained adults’ refixation of small targets shows some scatter, although small (with a standard deviation of about 0.1 deg according to de Bie, 1986, and Snodderly, 1987). When infants refixated small targets, they showed a higher degree of refixation scatter than for trained adults (0.8 deg ± 5 ± 0 on average), but untrained adults looking at the same targets were not substantially less variable (about 0.4 ± 5; Hainline, Harris, & Krinsky, 1990). These data suggest that infants are using some consistent small retinal area, possibly the fovea, to direct their fixations even when they are fairly young; Hainline et al.'s study did not detect age differences during infancy in the degree of refixational scatter. The present data support the contention that infants possess sufficient oculomotor control to allow deliberate inspection of static visual scenes.

Stabilising moving images: Pursuit, optokinetic nystagmus, and the vestibulo-ocular reflex

The retina is mounted in a moving eye in a moving head, on a locomoting body. The highest spatial frequencies of moving retinal images are most readily lost due to temporal factors, but as we have seen already, many aspects of visual function may continue uninterrupted based on intermediate spatial frequencies, to which even young infants are relatively sensitive.

Smooth pursuit Smooth pursuit is the ability to rotate the eye smoothly so as to keep the image of a moving target stable on the fovea. To stabilise a moving target the gain of the pursuit (the ratio of eye velocity to target velocity) must be close to unity. Well-trained adults can maintain such gains for velocities as high as 30–40 deg/sec, although this takes extreme attention and the subject fatigues rapidly (Howard, 1982). Early data reported that young infants, whose foveae are quite immature, had poorly developed smooth pursuit (Aslin, 1981; Dayton, Jones, Steele, & Rose, 1964; Shea & Aslin, 1984). Others using targets that may have been more salient to the infant have shown evidence of smooth pursuit at slow speeds and for larger targets in young infants, including newborns (Carchon & Block 1996; Hainline, 1985; Kreminitzer, Vaughan, Kurtzberg & Dowling, 1979; Roucoux, Culee, & Roucoux, 1983; Von Hofsten & Rosander, 1996; Shea & Aslin, 1990). For faster-moving targets, infants and adults both follow the target with a mixture of saccades and smooth pursuit, which is adaptive and requires less concentration, at least for adults (Hainline, 1993).
Optokinetic nystagmus (OKN) and vestibulo-ocular reflex (VOR) OKN and VOR are two oculomotor systems closely linked to the vestibular system and its interactions with the visual system, and are important for compensating for movement of the organism through the world and some of the associated movements of the retinal image (e.g., Cohen, 1974; Cohen, Henn, Raphan, & Dennett, 1981). OKN is a repetitive series of eye movements produced when a large portion of the visual field's retinal image slips across the retina. In response, the two eyes move as a unit with alternating phases of smooth tracking (the "slow" phase of OKN) and return saccades (OKN "fast phase"; see Fig. 1.9). As a result, the slippage of the image on the retina is minimised. OKN is a phylogenetically "old" system that developed in lateral-eyed animals who do not possess a specialised fovea or show smooth pursuit of a small target. Infants show well-developed OKN from birth, although when OKN is stimulated from one eye at a time, the behaviour shows an interesting asymmetry that disappears with age; for young infants, when targets are viewed with one eye only, OKN is elicited when the pattern moves in the direction from the temple towards the nose, but not in the opposite direction. By 4–5 months of age, the monocular response is much more symmetrical (Atkinson & Braddock, 1981; Naegeli & Held, 1982), although infants with various visual problems continue to show the monocular asymmetry (e.g., van Hof-van Duin & Mohn, 1983). The most common interpretation of the "disappearance" of the monocular OKN asymmetry is the emergence of new forms of cortical control in the visual system (e.g., Hoffman, 1979; Preston &Finocchio, 1993; Schor, Narayan, & Westall, 1983), but as yet there are no hard and fast data to substantiate this hypothesis for humans.

**CONCLUSION**

Although OKN in humans may be less important for image stabilisation than the foveational systems, it may still be quite important in its intricate interactions with mechanisms for vestibular stabilisation. The OKN and VOR systems share many neuronal elements, but the VOR is designed to maintain the eye's gaze direction in spite of head rotation. It is possible that, in humans, the major role of these systems under natural conditions is to compensate for head turns; the VOR provides the initial ocular stabilisation, as it is a fast response, and OKN adds any necessary longer-duration components (Preston & Finocchio, 1993). However, both responses can be overridden when voluntarily fixating a target, and may be of less importance developmentally as fixational control and smooth pursuit improve. Some suggestive evidence comes from some recent studies of the development of the VOR. In adults, the gain of the VOR is much less than one, actually about 0.5, so that it functions relatively poorly as an overall image-stabilising system. In contrast, the gain of the VOR in infants and even in young children is close to 1.0, and so compensates better on its own for image slippage caused by head rotations or other bodily movements through space (Finocchio, Preston, & Fuchs, 1990a,b). But if adults' VOR gain is low because other systems are more suitable for keeping the fovea pointed at interesting targets, one might expect to find a close but inverse connection in the developmental rates for smooth pursuit and VOR gain. In fact, there is a fair amount of mismatch between the time courses in VOR and smooth pursuit gains; pursuit is well developed within a year, whereas VOR gain continues to drop to adult levels over at least the first 5–6 years. It is possible that, in this case, the VOR is responding to gross physical development as much as to the parameters of visual stimulation, as the VOR is also important in the maintenance of posture as the body moves. Children's gross motor coordination continues to develop well into the early school years, and so they may need a more "robust" VOR system during that time. The introduction of the physical control of the body raises another level of complexity that we do not have time to develop further here.
sensitive to the characteristics of the organism of interest, the human infant. It is a contrast of some interest how poorly infants often perform in laboratory studies and how well-integrated they appear to be in their natural environments. Part of this effect is probably due to the use of simple, unidimensional stimuli that are not too interesting in their own right in many laboratory studies, which suggests that our research needs to include a wider range of stimuli. We also have not always given sufficient attention to the fact that the various components of vision combine to yield a reasonably competent ensemble of abilities, including not only vision but vestibular reactions. In the main, when exercised in the rich stimulus context of the infant's everyday world, infant vision is a highly functional sense. Although single abilities may be less than fully mature, the ensemble of visual functions allows the infant to respond appropriately to relevant aspects of the environment, including the distinctive features and affordances described by Gibson (1966, 1979). When development proceeds normally, infant vision seems perfectly adequate for the things that infants need to do. There is, however, reason to be worried about the development of infants who start life with some sort of visual deficit; this will be discussed in the next chapter.

REFERENCES


1. The Development of Basic Visual Abilities


1. THE DEVELOPMENT OF BASIC VISUAL ABILITIES


