5 Visual Recalibration and the Development of Contrast and Optical Flow Perception

Martin S. Banks
University of California, Berkeley

INTRODUCTION

A central question in perceptual studies is how constant perceptions are attained from varying inputs. Stated another way, how do we identify common objects and determine their location reliably despite variations in their distance, orientation, slant, context, illumination, and other confounding factors? Developmentalists have sought an understanding of what is innate and what is learned in these capabilities. With the recent advent of techniques for examining the visual capacities of young infants, evidence favoring the nativist position has accumulated. Specifically, several sophisticated visual capabilities have been demonstrated at birth or shortly thereafter. These capabilities include recognition of familiar objects presented in various ways (Fagan, 1976), segmentation of one object in the field from another (Kellman & Spelke, 1983), the perception of depth (Fox, Aslin, Shea, & Dumais, 1980; Gordon & Yonas, 1976), and more.

Although these observations lend credence to the nativist view of visual development, they also pose a significant problem for that view. The eye and the visual nervous system are distinctly immature when these competencies emerge: The optical components of the eye are still growing in complicated ways, the photoreceptors of the retina are maturing and migrating, and the dendritic trees and synapses of central visual neurons are forming. The problem is to explain how infants at such young ages are able to maintain these perceptual capabilities in the face of age-related changes in the visual apparatus. Thus, a question similar to the one at the opening of this chapter arises: How does the developing child maintain constant perceptions despite varying

inputs AND a varying visual apparatus? This chapter considers this question and argues that visual experience provides crucial guidance to the maintenance of perceptual stability during development.

Visual recalibration is the process by which perceptual stability is achieved despite changes in the visual apparatus. Recalibration can be demonstrated in adults in several ways. Consider, for example, the experiments of Wallach, Moore, and Davidson (1963). They examined the use of binocular disparity to perceive depth when the effective distance between the eyes was doubled. The relative distance of two points in space can be determined from the resulting binocular disparity (the difference in the projected images to the retinas). The relationship can be expressed in the following way:

$$d_1 - d_2 = \frac{d_2^2 \cdot \Delta\theta}{S - d_2 \Delta\theta} \quad (1)$$

The terms $d1$ and $d2$ refer to the distances of the two points. The term S refers to the separation between the two eyes. $\Delta\theta$ is the binocular disparity. Turning the equation around to determine relative distance from disparity, we obtain:

$$\frac{d_1}{d_2} = 1 + \frac{d_2 \cdot \Delta\theta}{S - d_2 \cdot \Delta\theta} \quad (2)$$

Thus, to make veridical judgments of relative depth, the observer needs to "know" the binocular disparity and the interocular separation. Wallach et al. (1963) found that adults were able to use binocular disparity to judge the three-dimensional form of an object accurately. They also examined the effects of doubling the effective interocular distance by asking adults to view the object through a telestereoscope. As one might expect from the above equations, the object's form was initially misperceived. Indeed, the object seemed to deform nonrigidly when it was rotated. (This occurred because a given distance between two points was perceived to be less when the points were in the frontoparallel plane than when they rotated to positions out of the frontoparallel plane.) After 10 minutes of inspecting the rotating object, however, perceived depth was once again veridical, so the object's form was again perceived accurately. Thus, the subjects were able to take into account the change in interocular distance when judging depth based on binocular disparity. This experiment reveals that disparity-based judgments of depth are affected by alterations of interocular separation. It also demonstrates that the visual system can compensate and reestablish veridical depth judgments when given several minutes of experience with such alterations. A similar recalibration presumably occurs during normal development. Interocular distance increases by roughly 60% from birth to adolescence (Aslin & Jackson, 1979) yet infants and children seem to be able to use binocular disparity rather effectively during most of that time.

Several researchers have recognized the importance of recalibration to visual development (e.g., Bower, 1974; Harris, 1980; Held & Bossom, 1961; Rosinski, 1977), but the great majority of empirical work has been conducted on adults. Thus, before discussing visual recalibration in the developing child, a few comments on adult adaptation experiments and their relevance to developmental investigations are useful.

Practical rather than theoretical considerations motivate the use of adults; with a few notable exceptions (e.g., von Hofsten, 1977), techniques for conducting adaptation experiments in infants and young children simply have not been developed. In the prototypical adult adaptation experiment, a device is attached to the subject that distorts the relationship between the distal properties of the environment and the retinal image. An example is the telestereoscope of Wallach et al. (1963). Another is the wedge prism. A prism placed in front of the eye displaces the entire visual field in one direction and distorts the shape of objects (Rock, 1966). Initially the subject is quite incapable of accurate visually guided behavior. After several minutes or even hours of adaptation, however, visuomotor performance improves significantly provided that active exploration of the environment is allowed (Held & Bossom, 1961). Although adaptation is readily observed, it is frequently not complete (Hay & Pick, 1966; Held & Rekosh, 1963; Pick & Hay, 1964). In the Pick and Hay experiment, for example, the observed adaptation to prismatic deformation was only 11% of total adaptation possible after 3 days and 30% after 42 days.

The common failure to observe complete visual adaptation is puzzling in some ways. For one thing, adaptation seems to be complete in many situations. For instance, the adults in the Wallach et al. (1963) experiment seemed to adapt completely to the change in effective interocular separation. Another example is adaptation to new spectacles. Adults seem to adapt completely to spatial distortions (pincushion or barrel distortion, see Fig. 5.2) induced by the new lenses. When the spectacles are first put on, the distortions are quite noticeable, particularly in the peripheral visual field. The distortions are particularly distressing when the head is moved about; contours appear to bend rendering a nonrigid appearance to quite solid objects. The perceived distortions gradually diminish. Within 2 weeks, no spatial distortions or contradictions between head movement and perceived stability of the visual environment are noticed. This common experience implies that visual adaptation to optical distortions can occur to a significant degree.

Because complete adaptation is observed in some cases and not in others, one wonders what the required conditions are. I propose two such conditions and suggest that many of the previous adult adaptation experiments have not satisfied these conditions. First, experimentally induced distortions should mimic the effects of normal developmental processes. My reasoning is the

following. Recalibration mechanisms, like any other visual mechanism, have been subject to natural selection. They have had to cope with distal-proximal distortions that result from normal ocular and neural growth and not with bizarre distortions like those induced in many of the experiments in the visual adaptation literature. (An example of a bizarre distortion is inverting prisms; Stratton, 1897.) Consequently, the capabilities of existing recalibration mechanisms may be restricted to only those sorts of distortions produced by normal growth. From this view, one predicts that alterations of the distal-proximal relationship that mimic normal developmental processes should lead to greater recalibration than alterations that do not mimic normal development. Second, the experimental subjects must be provided adequate visual and visuomotor experience to determine what sort of distortion has occurred and how to compensate for it. In many experimental reports, adequate visual and visuomotor experience is not provided during the adaptation phase of the experiment. A well-known example is the work of Held and his colleagues, which shows that adaptation is much more complete under conditions of self-initiated movements than in passive conditions. Another example is illustrated in the fourth section of this chapter. That section describes how subjects could determine the sort of spatial distortion that is introduced in an adaptation experiment from the optic flow field. The analysis in that section shows that the flow field must be large and must contain more than one depth plane in order to provide information that specifies the distortion. Flow fields created by impoverished scenes simply do not provide enough information. Thus, rich environmental stimulation including complex visual targets and active motoric exploration may be required for appropriate recalibration to occur.

We do not know yet how well adult experiments inform us about recalibration during normal development. As suggested in the preceding paragraph, one might examine this question by conducting adult experiments in which the experimental distortion (for example, an optical distortion created by spectacles) is either similar or dissimilar to the distortion created by ocular or neural growth. Greater adaptation observed when the distortion is similar would be indirect evidence that the adult phenomenon is homologous to the developmental one. In such cases, adult experiments would be more likely to inform us about developmental recalibration than in cases in which the magnitude of adult adaptation does not match expectations based on developmental theory.

Let us now return to the topic of primary interest here: visual recalibration during development. There are several plausible models of developmental recalibration. The first is a genetic prespecification model which claims that recalibration is specified by a predetermined maturational program. The program modifies the relationship between proximal stimuli and their distal interpretations in a manner that compensates for ocular and neural growth.

Genetic prespecification models are not particularly attractive because the existing evidence suggests that environmental feedback is important to recalibration (some of this evidence is reviewed in the fourth section). Different versions of environmental models are feasible. They differ in their explanations of how the environment provides the required information. To illustrate, let me return to the example involving binocular disparity and perceived depth. The problem to be explained again is how the developing child maintains constant depth percepts in spite of the changes in binocular disparity caused by the increasing interocular separation. One environmental model would claim that the stereoscopic depth system becomes rescaled by referring to information from another system altogether. For example, the infant might note disagreements between the output of the stereo system and the outcome of motoric behavior such as reaches. The infant then uses such disagreements to recalibrate the stereoscopic depth percepts. Another environmental model would claim that the stereo system is rescaled by referring to information from a different, but closely related, system. For example, infants might note discrepancies between the outputs of the stereo system and the optic flow system and use those discrepancies to adjust stereoscopic depth percepts. Finally, another environmental model would claim that the stereo system is recalibrated without referring to another system. The stereo system is rescaled instead by noting internal discrepancies among stereoscopic depth judgments. These three environmental models differ, therefore, in the extent to which they are self-contained. The first requires information from another system altogether, so it is not at all self-contained. The third does not require information from another system, so it is quite self-contained.

Although there are no detailed models of developmental recalibration, the literature contains examples of two of these three environmental models. Held's reafference model (Hardt, Held, & Steinbach, 1971; Held & Hein, 1963) is an example of a non-self-contained model; visual recalibration occurs primarily in reference to self-initiated movements the body, head, and limbs. Bower's (1974) recalibration model is less self-contained and most similar to the second of the above-mentioned environmental models. He proposed that stereo percepts are recalibrated by noting disagreements between depth judgments provided by the stereoscopic and optic flow systems.

The non-self-contained recalibration models (such as Held's and Bower's) have one unattractive feature: They are subject to the criticism that the standard, against which the system being rescaled is judged, may have its own recalibration problems. For instance, consider Bower's model. If optic flow is to provide a good standard for stereoscopic depth judgments, its own depth judgments must be unaffected by physical growth. A later section of this chapter shows that this precondition is implausible: Determinations of depth from optic flow are also very likely to be affected by physical and neural growth. Because non-self-contained environmental models are burdened

with this unattractive feature, this chapter examines self-contained models most closely.

Gibson (1966) also seemed to favor recalibration models that do not require information from another system altogether. In his discussion of spectacle and prism adaptation experiments, for example, he described how the visual system might adjust to maintain veridical perception.

> There must be invariants over time in the flowing array of optical stimulation to specify the rectilinearity, the constancy, and the rigidity of the world. This assumption holds as much for vision without spectacles as for vision with spectacles. When they are first put on, the observer must learn what the new constants are in the stimulus flux. The extraction of invariants by the perceptual system is taken to be the crux of the explanation of the phenomenal adaptation. (p. 302)

The primary goal of this chapter is to identify such invariants and to describe how they could be used in the recalibration of contrast and optical flow perception. Before discussing these invariants, however, I describe the sorts of distal-proximal distortions with which the developing visual system has to cope.

OCULAR AND NEURAL GROWTH

Several ocular and neural changes occur during infancy and childhood that modify the relationship between distal and proximal stimuli. I examine two distal-proximal changes in this chapter: (a) alterations of the spatial mapping between points in the environment and their neural representations and (b) alterations of the relationship between the physical and represented contrasts of features in the environment. I argue that these age-related changes would disrupt object and spatial layout perception unless some sort of recalibration occurred. The recalibration must in some way compensate for the potentially distorting effects of ocular growth and neural development.

First consider the changes in the spatial mapping between environmental points and their neural representations. Two age-related changes are most significant to altering this distal-proximal mapping: (a) changes in the projective geometry of the eye's optics (by projective geometry, I mean the relationship between points in the environment and their projections onto the retinal surface) and (b) changes in the positions of photoreceptors (a topic discussed by Aslin, this volume). The eye grows from birth to adolescence, most of the growth being compressed into the first 2 years (Larsen, 1971a, 1971b, 1971c). The pattern of growth is not simple; that is, it cannot be described as a simple expansion of all ocular components. Figure 5.1 emphasizes this fact by displaying two eyes, one adult and one newborn. Note how

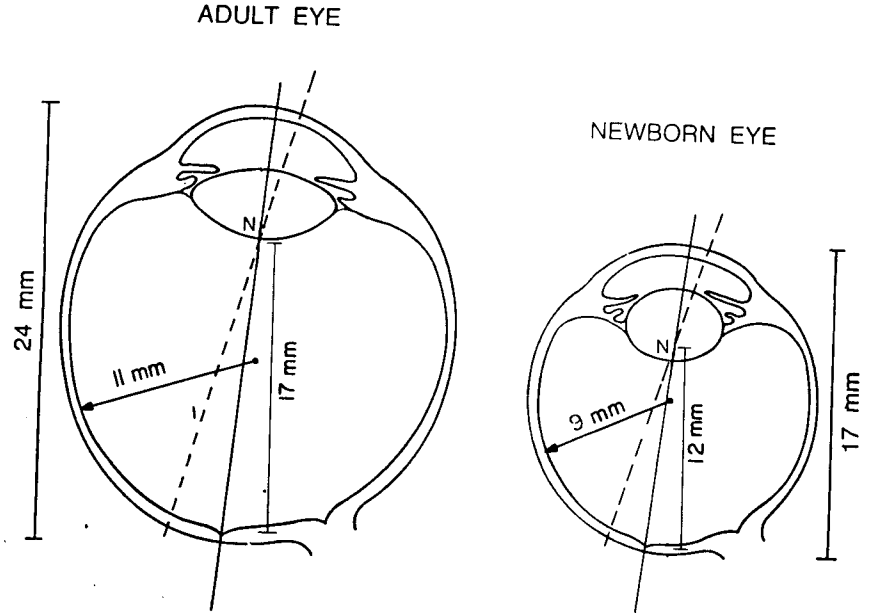


FIG. 5.1 Horizontal cross-sections of newborn and adult eyes. The dimensions shown for each eye are axial length, posterior nodal distance, and radius of curvature of the retina.

different structures grow by widely varying amounts. The thickness of the crystalline lens, for example, changes very little compared to the depth of the anterior chamber. This relatively complex pattern of growth presumably alters the projective geometry of the eye.

Figure 5.2 provides examples of common changes in projective geometry. When first fitted with spectacles, adults experience predictable spatial distortions. When the spectacles correct for myopia, light rays projecting to the retinal periphery are magnified and pincushion distortion results. The opposite occurs with a hyperopic correction, so barrel distortion results. These spatial distortions can be quite disturbing at first, but adults generally overcome them within 1–2 weeks.

I have estimated the sort of optical distortion that occurs between birth and adulthood. My estimates are somewhat crude because there is not a wealth of information on the dimensions of various structures at different ages. In making these calculations, I used Lotmar's (1976) schematic eye for newborns and Gullstrand's simplified schematic eye for adults (Bennett & Francis, 1962). Figure 5.1 displays the critical dimensions for my calculations of projective geometry. The newborn and adult eyes are on the right and left, respectively. The posterior nodal point, which is indicated in both eyes, is the

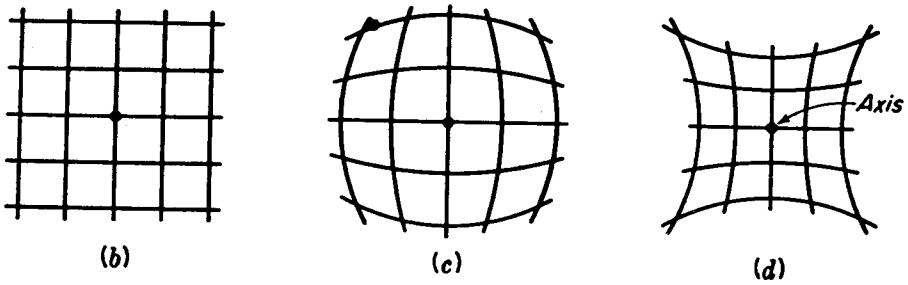


FIG. 5.2 Demonstration of changes in projective geometry of the eye. Images of concentric squares with (b) no distortion, (c) barrel distortion, and (d) pincushion distortion. From Jenkins and White, 1976.

point through which points in the visual scene seem to project on their way to the retina. The posterior nodal distance is the distance from that point to the retina. Also shown is the radius of curvature of the retina. To illustrate age-related changes in projective geometry, I calculated how a pattern of concentric squares would project onto the retinal surface in newborn and adult eyes. I have assumed that the distribution of receptors is the same for the two eyes, so only optical projection effects are shown here.

If the eye simply expanded uniformly from birth to adulthood, the pattern of stimulated retinal points would be identical at the two ages, except for a change in scale. As illustrated in Fig. 5.1, however, the posterior nodal distance increases more with age than does the radius of retinal curvature. In other words, as the eye grows, the posterior nodal point moves farther from the center of the eye. In consequence, the pattern of retinal points stimulated by the concentric squares changes with age. Figure 5.3 illustrates this change. On the left are the concentric squares. Assuming that the squares look like squares to a young infant, one can calculate their appearance to an adult, if no recalibration of perceived visual directions occurred. The distorted squares on the right are my estimate of the appearance of the concentric squares to the adult eye, based on the assumptions just mentioned. The distortions shown increase with eccentricity. The figure shows the effects out to 56 deg of visual angle. Larger distortions would be observed at greater eccentricities. The larger size of the adult eye has not been taken into account in this figure. If this factor were included, it would simply cause the squares on the right to appear larger.

Figure 5.3 demonstrates graphically how ocular growth should change the relationship between points in the environment and their optical projections onto the retinal surface. Another important factor has not been considered yet: the distribution of photoreceptors at different ages. Perhaps the maturational program for retinal development has cleverly built in a plan for

changes in receptor distribution that compensates for age changes in projective geometry. Let us examine this possibility.

Aslin discusses retinal development at some length in his chapter, so I only describe the properties that are directly relevant to the current discussion. Abramov et al. (1982), Hendrickson and Yuodelis (1984), and Yuodelis and Hendrickson (1986) found that the density of cones is much lower in the newborn's fovea than in the adult's. Indeed, cone density in the center of the newborn's fovea is not much different from cone density in the neonatal parafovea. By the end of the first year, foveal cones are packed much more tightly, but still not as tightly as in the adult fovea. One might hypothesize that the increasing cone density results from the birth of new cones in the fovea. This hypothesis can be ruled out, however, because a full complement of cones is present at birth. An alternative hypothesis is that cones in the central retina migrate centripedally for several months after birth. This hypothesis is suggested by the morphology of axons from cones to bipolar and retinal ganglion cells (the processes are called the fibers of Henle). The

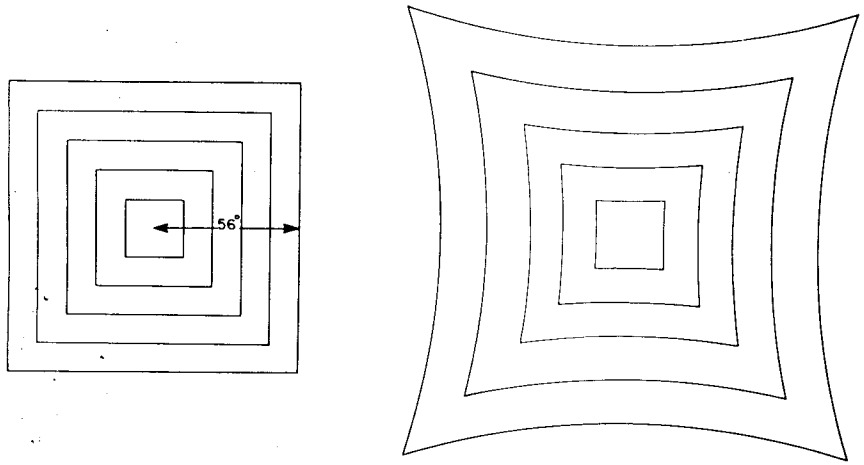


FIG. 5.3 Demonstration of age changes in projective geometry of the eye. In the demonstration I assume that a newborn fixates the middle of the concentric square pattern on the left. The angular subtense from the middle to the edge of the pattern is set to 56 deg. I also assume that the newborn perceives the pattern veridically; that is, the straight lines are perceived as straight. If no recalibration of visual directions occurred, a change in projective geometry of the eye would change the percept. To demonstrate this perceptual change, I calculated the retinal points stimulated by the squares in newborn and adult eyes, assuming central fixation. The distorted pattern of squares on the right is my estimate of the adult's percept if no recalibration occurred from birth to adulthood. It is the pattern that would have to be presented to the newborn in order to stimulate the retinal points the concentric squares would stimulate in the adult eye. Pincushion distortion is predicted for the reasons given in the text.

connections are established before birth and do not appear to be broken later on. In year-old and adult eyes, the fibers fan out centrifugally forming the walls of the foveal pit. In neonatal eyes, however, the fibers run roughly perpendicular to the retinal surface. The change from perpendicular to tangential projections suggests that the foveal cones move relative to the bipolar and retinal ganglion cells to which they are connected. The movement stretches and shears the fibers of Henle until they lie almost tangential to the retinal surface. Thus, changes in the geometry of retinal connections are consistent with the idea that foveal cones migrate centripedally for many months after birth. We do not know, unfortunately, when the distribution of cones becomes adultlike, but Yuodelis and Hendrickson's (1986) observations suggest that the migration persists for at least the first year and a half of life.

The postnatal migration of central cones, therefore, is another factor that alters the relationship between points in the environment and their neural representations. Aslin (this volume) points out that there is insufficient information about cone density as a function of retinal eccentricity in newborns to calculate the pattern with any precision. Nonetheless, two qualitative points can be made. First, centripedal migration with age implies barrel distortion, the opposite of the distortion induced by eye growth. Second, centripedal migration probably occurs near the fovea only (the connection from photoreceptors to bipolar and retinal ganglion cells are perpendicular to the retinal surface in the retinal periphery of adults, suggesting no postnatal receptor migration). Thus, receptor migration probably causes a tendency toward barrel distortion, but only near the fovea. So if Figure 5.3 were modified to incorporate the effect of receptor migration, barrel distortion (or at least less pincushion distortion) would be observed for squares near the center, but the pincushion distortion observed in the near and far retinal periphery would be unaltered.

These two growth effects—one concerning optical projection and the other photoreceptor displacement—could profoundly disrupt the use of spatial information to infer properties of the environment unless some sort of recalibration occurred. In a later section, I consider how these changes in distal-proximal spatial mapping could confound the use of optical flow information in deducing the spatial layout of the environment and one's motion through it.

Now let us turn to the second disrupting effect of ocular and neural growth: changes in the relationship between physical and perceived contrasts of environmental features. I argue that several optical and neural factors are likely to affect this relationship across age. I also argue in the next section that changes in the relationship between distal and proximal contrast are likely to confound object recognition and identification, unless some sort of recalibration occurs.

What is the evidence that distal-proximal contrast relationships are altered with age? To answer this question, one must first delineate the optical and neural factors that are known to influence the perceived contrast of visual stimuli in adults. In discussing these factors, it is useful to employ Fourier's theorem, sinewave gratings, and the contrast sensitivity function. Fourier's theorem implies that any two-dimensional, achromatic stimulus can be represented by the combination of sinewave gratings of different spatial frequencies, contrasts, orientations, and phases (if these terms are unfamiliar, refer to Banks & Salapatek, 1981, or Cornsweet, 1970, for details). The visibility of sinewave gratings is not affected significantly by changes in orientation or phase. Contrast, however, is an important determinant of visibility: As intuition implies, low-contrast gratings are less visible than high-contrast gratings. Another, less intuitive, determinant is spatial frequency. The relationship between detectability and spatial frequency is indexed by the contrast sensitivity function (CSF). The CSF represents the amount of contrast needed to detect sinewave gratings of various spatial frequencies. An adult with good vision exhibits a CSF like the one shown in the lower portion of Fig. 5.4. Note that sensitivity is greatest for intermediate spatial frequencies (2–5 c/deg) and lower for low and high frequencies. To illustrate, a grating, varying in spatial frequency and contrast, is displayed in the lower portion of the figure. The grating increases in spatial frequency from left to right and increases in contrast from top to bottom. The physical contrast of the grating is constant along any horizontal line in the photograph, but its perceived contrast is not. Clearly, perceived contrast is greatest at intermediate frequencies and lower for low and high frequencies.

The shape of the CSF reflects the operation of several basic visual mechanisms. Banks, Geisler, and Bennett (1987) recently examined the steady loss in adults' contrast sensitivity for spatial frequencies above 5 c/deg. They found that they could precisely predict relative sensitivities from 5 to 40 c/deg from three factors alone: (a) the optical quality of the eye, (b) the size of individual foveal cones, and (c) quantal fluctuations inherent in visual stimuli (that is, stimulus noise due to the quantal nature of light). Only the first two factors are discussed here. The third is not discussed because its contribution is dependent on neural pooling and we know very little about the development of pooling.

First consider the contribution of optical quality. The optics of the eye are imperfect. Optical transfer measurements in the adult eye reveal that less than 100% of the contrast of a stimulus is transmitted to the retina and that this percentage depends heavily on the spatial frequency of the stimulus. For example, the retinal image contrast of a 5 c/deg grating is about 70% of the stimulus contrast under normal, perfectly focused viewing conditions, whereas the retinal contrast of a 20 c/deg grating is merely 15% of stimulus

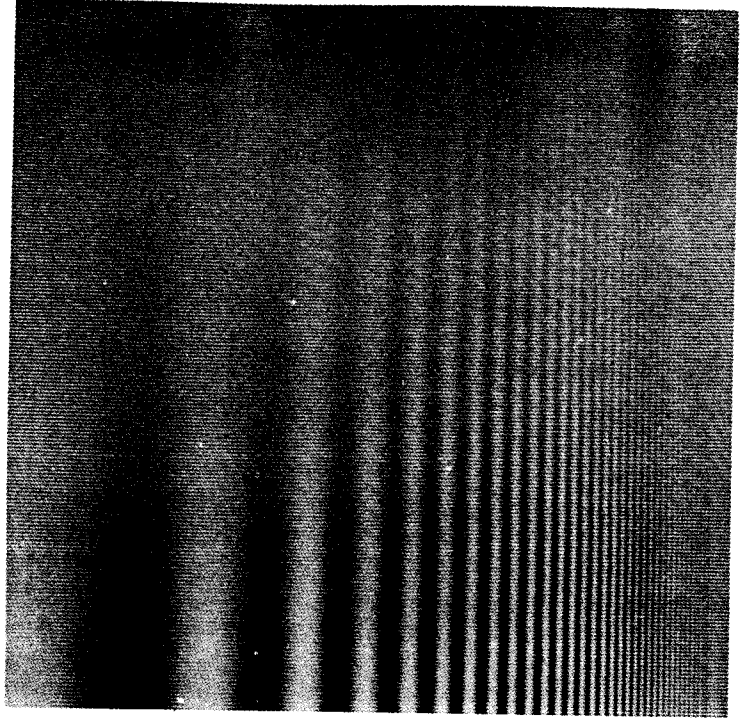


FIG. 5.4

contrast (Campbell & Gubisch, 1966). In other words, the contrast of very fine pattern information is reduced much more by the eye's optics than is the contrast of coarse pattern information. This optical effect accounts for most of the high-frequency loss in adults' contrast sensitivity illustrated in Fig. 5.4.

Additional losses occur due to the spatial summating properties of photoreceptors. Foveal cones subtend a small, but finite, visual angle. As Banks et al. (in press) and Miller and Bernard (1983) point out, a finite receptor aperture implies that high spatial frequencies cannot be signaled by the receptor mosaic as well as low frequencies can be. Thus, the foveal cone aperture also attenuates high-frequency gratings. The effect is small, however, because foveal cones subtend only about 0.5 minutes of arc in adults.

Even greater attenuation of high frequencies relative to lows can occur from neural summation among retinal and central visual neurons. Banks et al. concluded that attenuation due to neural summation was insignificant in the adult fovea, but it is quite significant in the peripheral visual field (Green, 1970).

The relevance of this discussion to the developing visual system is the following. All of these factors—optical quality, receptor aperture, and neural summation—are likely to change with age.

Although no measurements of optical quality as a function of age exist for humans, the dramatic changes that occur in the dimensions and shapes of various ocular structures lead one to suspect that optical quality changes. Moreover, optical quality improves significantly with age in kittens (Bonds & Freeman, 1978) and infant monkeys (Williams & Boothe, 1981), so improvement is likely in human infants, too.

The aperture of foveal cones changes dramatically with age. According to Yuodelis and Hendrickson (1986), the width of the cones' inner segments (the dimension that is relevant to calculating attenuation) shrinks from 5–7.5 microns at birth to 1.8–2.2 microns in adults. Thus, an individual foveal cone subtends 1.5–2.2 minutes of arc in neonatal eyes and about 0.5 minutes in adult eyes (taking into account the shorter posterior nodal distance of the newborn compared to the adult eye). The neonatal aperture is large enough to have rather pronounced effects on the encoding of high spatial frequencies. By my calculations, the newborn's receptor aperture would attenuate a grating of 5 c/deg to about 94% of its original contrast and a grating of 20

FIG. 5.4 *(Opposite page)* A sinewave grating and a typical adult contrast sensitivity function (CSF). The upper part of the figure displays a sinewave grating in which spatial frequency increases from left to right and contrast increases from top to bottom. The lower part of the figure shows a typical adult CSF. Contrast sensitivity, the reciprocal of contrast at threshold, is plotted as a function of spatial frequency. Scales relating spatial frequency to Snellen equivalents and stripe width in minutes of arc are shown for comparison. If the figure is viewed from a distance of 60 cm, the scales at the bottom indicate the actual frequency values of the grating in the upper part of the figure.

c/deg to only 31% of its original contrast. The corresponding adult values are 99.7% and 95% for 5 and 20 c/deg, respectively.

There are no existing measures of neural summation as a function of age, so its potential effects on contrast perception cannot be pinned down. We can, however, compare the shapes of adults' and young infants' CSFs as an index of the overall attenuation due to quantal fluctuations in the stimulus, optics, receptor aperture, and neural summation. The CSFs of young infants have been measured by four research groups. Atkinson, Braddick, and colleagues have used behavioral and evoked-potential techniques to measure CSFs in infants from a few days to 6 months of age (Atkinson, Braddick, & Moar, 1977; Harris, Atkinson, Braddick, 1976). Banks and Salapatek (1978) used a behavioral technique to measure CSFs in 1- to 3-month-olds. Pirchio and colleagues measured these functions in 2- to 10-month-olds using evoked potentials (Pirchio, Spinelli, Fiorentini, & Maffei, 1978). These reports agree remarkably well in light of the differences in technique and stimuli. More recently, Norcia, Tyler, and Allen (1986) have measured infant CSFs using a novel VEP technique. Although they reported higher contrast sensitivities than the other research groups, the shapes of their infant CSFs were similar to those of previous reports. Thus, all of the observed infant CSFs are shifted to lower spatial frequencies and lower contrast sensitivities than the adult CSFs. Figure 5.5 shows some of the data from Pirchio et al. (1978). These CSF data imply that the attenuating characteristics of the visual system, which are presumably the result of optics, receptor aperture, and neural summation (as well as stimulus noise), change rather substantially from early infancy to adulthood. This point can be illustrated by considering the relative sensitivities of infants and adults at particular spatial frequencies. According to the data of Fig. 5.5, adults require about 50% more contrast at 1.5 c/deg than at 4 c/deg to respond to a sinewave grating. The situation is reversed for 2.5-month olds. They require 200% more contrast at 4 c/deg than at 1.5 c/deg to respond equally well. Thus, the visibility of one spatial frequency relative to another appears to change substantially with age. The consequence would be alterations of the relationship between physical (distal) contrast and perceived (proximal) contrast, unless some sort of recalibration occurred.

MATURE CONTRAST PERCEPTION

Clearly, the relationship between distal and proximal contrast varies with spatial frequency, even in adults. Moreover, the distal-proximal relationship seems to change rather substantially with age. Before considering the prospect of developmental recalibration of this relationship, I discuss the implications of the mature visual system's differential sensitivity to spatial frequency

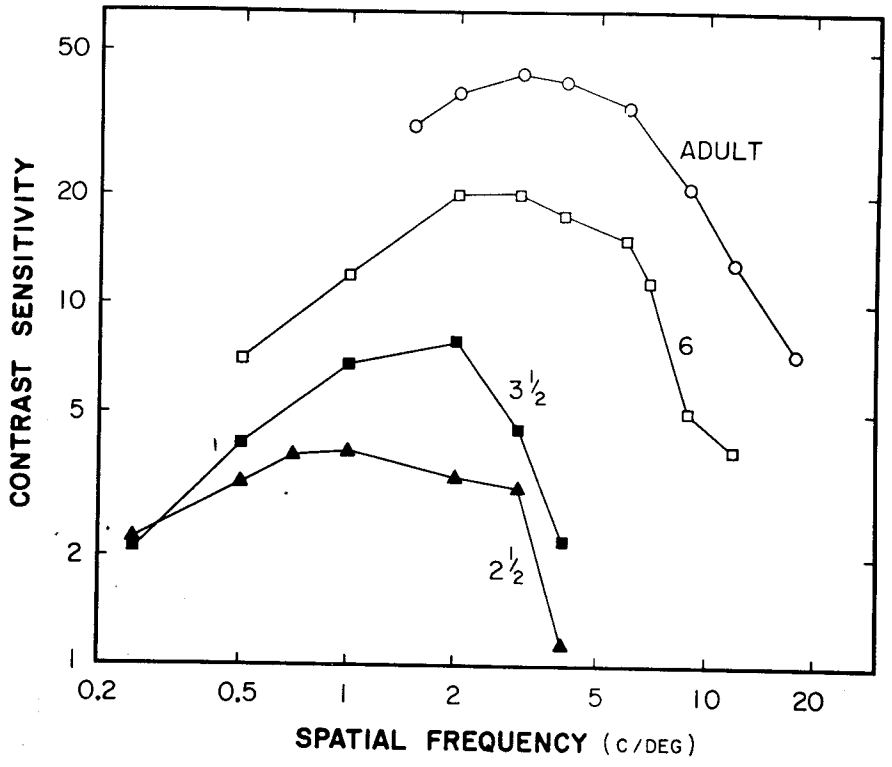


FIG. 5.5 Contrast sensitivity functions at different ages as reported by Pirchio et al. (1978). The axes are the same as in Fig. 5.4. The functions were measured using the VEP in one infant at 2.5, 3.5, and 6 months of age. The adult data are from another observer. Adapted from Pirchio et al., 1978.

to object recognition and identification in adults. Spatial-frequency-dependent sensitivity poses a problem for recognition and identification of an object at various distances. If a common object, such as a hand, is viewed at close range, the spatial frequencies corresponding to its gross and fine structure are quite low, so they are represented rather faithfully in the retinal image. However, when the distance to the object increases, the frequencies associated with those same structures increase. As those frequencies encroach upon the high-frequency falloff of the CSF, their relative amplitudes in the retinal image will change rather dramatically. It should be difficult to recognize that the same object produced both images, unless the visual system compensated for the differences in relative retinal image contrasts. In fact, the mature visual system does appear to compensate in a useful way.

Before describing this compensation, I point out that the problem of recognizing objects at various distances is a general one. It arises, for exam-

ple, in satellite photography. Pictures of objects on earth are significantly degraded by the atmosphere and the camera's imperfect optics. These effects can be minimized by "deblurring" the photograph using computer-enhanced imaging techniques. One begins by measuring the defocusing effects of the atmosphere and the camera's optics. These defocusing effects are represented as a blur function that describes how much the contrasts of various spatial frequencies are attenuated. One then computes the Fourier transform of the original photograph and multiplies the amplitudes of the resulting spatial frequency components by the inverse of the blur function.[1] The result is a much improved image that represents the earthbound objects more veridically than the original photograph did. A schematic of the process is presented in Fig. 5.6. Panel A displays the intensity distribution of a simple object (a bright bar on a dark background). The next panel shows the amplitudes of the sinewave components of this object. These values, called the *amplitude spectrum,* were obtained by Fourier transformation of the object (Gaskill, 1978). Panel C displays the blur function of a common camera. Note that higher spatial frequencies are attenuated much more than lows. The next panel shows the result of viewing the object through the camera. The values plotted are the amplitudes of the various sinewave components after they have been attenuated by the camera's optics. The values were obtained by simply multiplying the input amplitude spectrum (panel B) by the blur function (panel C). Finally, panel E displays the intensity distribution of the object viewed through the camera. As predicted by intuition, the image is a smeared version of the original object. This intensity distribution was obtained by inverse Fourier transformation of panel D (Gaskill, 1978). The application of deblurring techniques to this blurred image renders it more similar to the original object. To deblur, one multiplies the values of the amplitude spectrum in panel D by the inverse of the blur function. In other words, amplitudes at low spatial frequencies are multiplied by 1.0 whereas amplitudes at high frequencies are multiplied by values greater than 1.0. By choosing the appropriate values one could restore the blurred image to a rather faithful rendition of the original object.

An example of deblurring is displayed in Fig. 5.7. The left half of the figure shows a severely blurred photograph. The amplitude spectrum of this photograph was computed and multiplied by the inverse of the blur function.

[1]The amplitude of the input is sometimes zero or nearly zero in a particular spatial frequency band. In such a case, multiplying by the inverse of the gain function is inappropriate because it leads to amplification of noisy data. Furthermore, when the amplitude of the blur function is zero or nearly zero, multiplying by the inverse is impossible. In either case, the best strategy is simply to set the output amplitudes at the affected spatial frequencies to zero. The mature visual system appears to amplify only in those cases in which the input amplitude and the blur function are well above zero. This behavior is illustrated most clearly in Fig. 5.8. Veridical contrast matching is not observed until the input amplitude is about 1 log unit above contrast threshold.

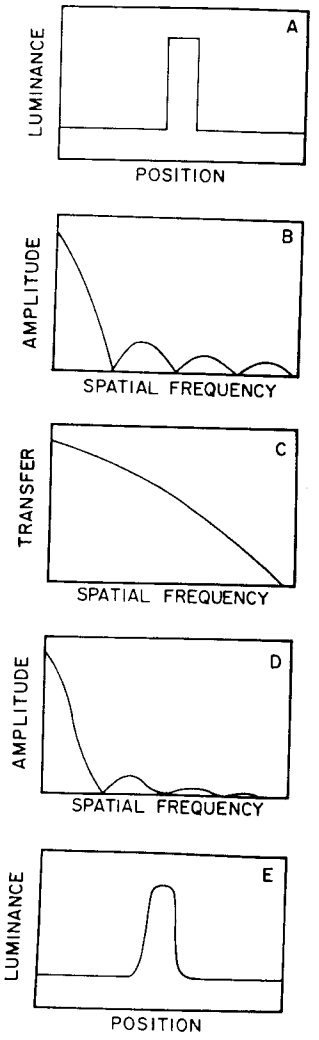


FIG. 5.6 Graphical illustration of blurring. The object is a white bar on a dark background. Assume that it is photographed by a camera with mediocre optics. (A) The intensity distribution of the original object. The object is dark on the left and right and bright in the middle. (B) The amplitudes of the spatial frequency components in the object. The amplitude of each component is plotted as a function of spatial frequency. This graph is the Fourier transform of the one above it. (C) The optical transfer of the camera. Transfer, which is the ratio of output amplitude to input amplitude, is plotted as a function of spatial frequency. The graph shows that low frequencies are transferred without loss from the object to the camera's film but high frequencies are attenuated rather severely. (D) The amplitudes of the object's frequency components after they have passed through the camera's optics. This graph was obtained by multiplying the two above it. Notice that low spatial frequencies from the object are represented quite well in the output but high frequencies are not. (E) The predicted intensity distribution on the camera's film. This graph is the inverse Fourier transform of the one above it. Thus, it represents the image that results when high frequencies are attenuated in the manner depicted in C.

Fourier synthesis was then used to reconstruct the image. The result is illustrated in the right half of the figure. The deblurred photograph is a much more veridical representation of the original object.

The mature visual system also appears to deblur inputs, at least under suprathreshold conditions (e.g., Blakemore, Muncey, & Ridley, 1973; Georgeson & Sullivan, 1975; Kulikowski, 1976). Georgeson and Sullivan demonstrated this quite clearly. They asked adults to adjust the contrast of a sinewave grating at one spatial frequency (the "comparison" grating) until it appeared to match the contrast of a grating at a different frequency (the "standard" grating). In their main experiment, the standard was 5 c/deg, a

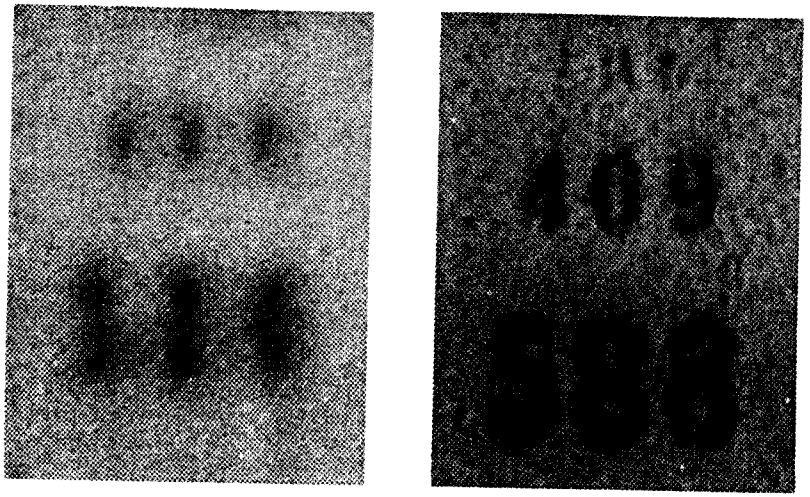


FIG. 5.7 Demonstration of deblurring using Fourier techniques. On the left is the blurred original. On the right is the deblurred version produced by compensating for the attenuation of the original blurring. From Gennery, 1973.

value near the peak of the adult CSF. The results are illustrated in Fig. 5.8. The contrast of the standard grating is plotted on the ordinate and the matched contrast of the comparison grating on the abscissa. When the contrast of the standard was low (that is, when the grating was just visible), adults set the contrast of the comparison gratings to higher values. Those values were predictable from the CSF. For example, adults set the contrast of a 20 c/deg comparison grating to a value eight times higher than the contrast of the 5 c/deg standard. This ratio of 8:1 was equal to the ratio of contrast thresholds for those spatial frequencies. The most interesting result occurred when the contrast of the standard was set to a value well above threshold. Adults in this situation adjusted the contrast of the comparison to the same physical value as the contrast of the standard. This result is surprising because, as mentioned above, two gratings of equal contrast but different spatial frequencies produce different retinal image contrasts. In other words, when the adults set 5 and 20 c/deg gratings to equal physical contrasts, they were accepting as equal in apparent contrast two gratings whose retinal image contrasts differed substantially. This implies that the nature visual system compensates at suprathreshold contrasts for the defocusing effects of the eye's optics and for spatial summation by peripheral neural elements.

Georgeson and Sullivan dubbed this phenomenon "contrast constancy" and noted an important perceptual consequence. As an object moves away from an observer, its spatial frequency content shifts to progressively higher values. The presence of contrast constancy means that the apparent contrast of medium- and high-contrast features will not change as the object moves

away. Thus, contrast constancy confers a useful property on the perception of real objects: So long as the contrast that defines an object and its features is above threshold, apparent contrast remains roughly invariant across a wide range of distances.

How is contrast constancy achieved? Georgeson and Sullivan noted that the central visual system must in some way undo the optical attenuation that occurs in the formation of the retinal image and the neural attenuation caused by peripheral spatial summation. They proposed a simple model based on multiple spatial-frequency channels with narrow tuning.[2] According to Georgeson and Sullivan's model, high-frequency channels have steeper contrast-response functions than medium-frequency channels. So as the contrast of a target is raised, the activity of high-frequency channels increases more rapidly than the activity of medium-frequency channels. Consider how this model could be used to explain apparent contrast matches between a medium- and a high-frequency grating at low and high contrasts. At near-threshold contrasts, apparent contrast matches would reflect differences in the contrast thresholds of the two gratings; observers would require more contrast in the high-frequency grating before they appeared equal in contrast because high-frequency channels have higher contrast thresholds. As one raised the contrast of the targets, however, apparent contrast matches would become veridical because the activity of high-frequency channels increases more rapidly than that of medium-frequency channels. Experimental evidence supports this model and suggests that the differential response to contrast is mediated by the visual cortex (Albrecht & Hamilton, 1982; Georgeson & Sullivan, 1975; Hess, Bradley, & Piotrowski, 1983).

If Georgeson and Sullivan's hypothesis is correct, contrast constancy should not be observed in a visual system without multiple narrowband channels. Banks, Stephens, and Hartmann (1985) demonstrated that 12-week-olds, but not 6-week-olds, possess multiple spatial-frequency channels with narrow tuning. Therefore, if Georgeson and Sullivan are correct, contrast constancy should not be observed at 6 weeks but might be observed at 12 weeks. To test this possibility, Stephens and Banks (1985) used a visual preference procedure to measure an analog to apparent contrast matches in 6- and 12-week-olds. Two sinewave gratings, differing in spatial frequency by a factor of three, were presented side by side. The lower-frequency grating was referred to as F and the higher-frequency grating as $3F$. From previous experiments, Stephens and Banks knew that the contrast threshold for $3F$ was about four times greater than that for F. For a given experimental session, the contrast of $3F$ was fixed at either a near-threshold level or one of two

[2]Spatial-frequency channels are mechanisms that respond to a narrow range of spatial frequencies only. Their existence has been demonstrated psychophysically and physiologically (Braddick, Campbell, & Atkinson, 1978).

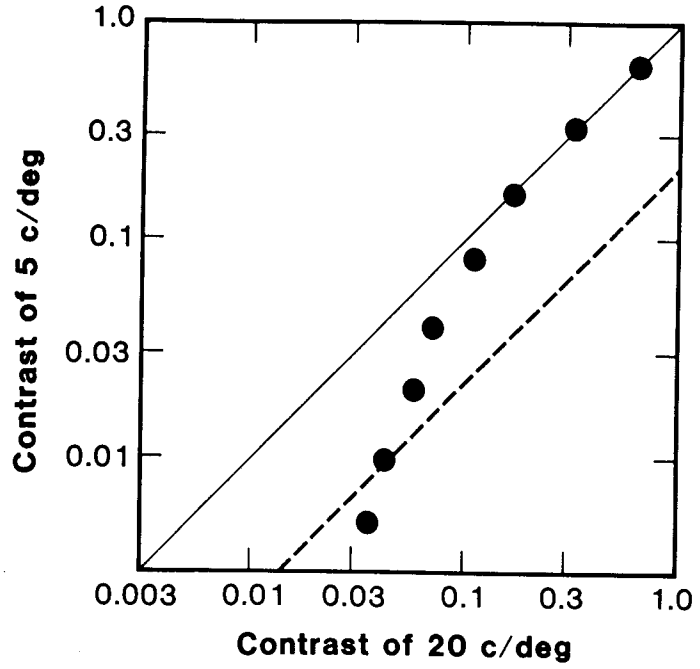


FIG. 5.8 Contrast matching plot showing contrast matching data and retinal image contrasts. The data points represent adult contrast matching data from Georgeson and Sullivan (1975). The contrast of a 5 c/deg standard grating is plotted on the ordinate and the contrast of a 20 c/deg comparison grating on the abscissa. The solid diagonal line indicates where the data would lie if contrast matches occurred at equal physical contrasts. The broken diagonal line indicates stimulus pairs for which the retinal image contrasts are the same. The data would lie on this line if contrast matches occurred at equal retinal contrasts. Adapted from Georgeson and Sullivan, 1975.

suprathreshold levels. Then the contrast of F was varied in order to estimate the contrast at which preference for the two gratings was equal. Stephens and Banks assumed that the equal preference point corresponded to an apparent contrast match. The equal preference points for 6-week-olds were always predictable from their contrast thresholds. That is to say, the younger infants required four times more contrast in $3F$ than in F, at all three contrast levels of $3F$, to distribute their looking time equally to the two stimuli. Some of the 12-week-olds' equal preference points were also predictable from their contrast thresholds. For near-threshold stimuli, they required four times more contrast in $3F$ than in F for equal preference. At suprathreshold levels, however, the equal preference points were not predictable from threshold data. Indeed, equal preference occurred when the two stimuli had the same physical contrast. If one accepts the assumption that the equal preference point in infants is analogous to an apparent contrast match in adults (Stephens & Banks, 1985, evaluate the validity of this assumption in some detail), these data imply that contrast constancy is present at 12, but not 6, weeks of age. This finding is consistent with Georgeson and Sullivan's hypothesis that narrowband channels are required for contrast compensation.

RECALIBRATION OF CONTRAST PERCEPTION

It is important to note that veridical contrast matching observed in adults (and apparently in 12-week-olds) implies that high frequencies are amplified just the right amount to overcome optical and early neural attenuation. This point is illustrated in Fig. 5.8. The dotted line represents the contrast matches one would expect for 5 and 20 c/deg if amplification did not compensate for the effects of optical defocus. For all contrasts, the 20 c/deg grating should require five times more contrast to achieve a perceived match. Obviously, the data do not conform to this prediction, but rather they fall on the veridicality line at suprathreshold contrasts. This implies that the amplification of retinal image contrast at 20 c/deg relative to retinal contrast at 5 c/deg is fivefold. Notice that one of the data points actually falls below the dotted line. This is because the difference in contrast thresholds at 5 and 20 c/deg is actually eightfold rather than fivefold. The discrepancy is probably caused by neural summation in the peripheral visual system. If this explanation is correct, the overall amplification required is eightfold. The developmental question of interest is, How does this relative amplification adopt the appropriate value? In discussing this question, I refer to the optical and neural attenuation of high frequencies as the "blur function." This function simply represents the loss of contrast as a function of spatial frequency. I refer to the postretinal amplification of high frequencies relative to lows as the "gain function." This function represents the relative enhancement of perceived contrast as a function of spatial frequency. (Recall that the mechanism of postretinal amplification is the increased slope of the contrast-response functions. Therefore, amplification occurs at suprathreshold contrasts only. The gain function describes this amplification at contrasts reasonably well above threshold.) Veridical contrast matching implies that the gain function is the inverse of the blur function; any other form would lead to under- or overcompensation for attentuation due to the blur function.

Four models of how the gain function might achieve the appropriate form are listed in Table 5.1. There are two general classes of models, which for want of better terms I have called *genetic* and *experiential* models. These terms are, of course, too simple because all four models require both genetic and environmental influence. I use the terms simply to convey the nature and relative importance of environmental factors for each of the models.

TABLE 5.1

Genetic Models:
Model I: Prespecified gain function that is inverse of blur function
Experiential Models:
Model II: Distribution of environmental information is known. Blur function deduced by comparing distribution of output information to distribution of environmental information.
Model IIA: One type of environmental feature is known. Blur function deduced by comparing output distribution for that feature to known input distribution.
Model III: Distribution of environmental information is not known nor is a particular environmental feature. Blur function deduced by comparing output distribution for an object undergoing a smooth change in distance.

In the genetic model, the gain function is assumed to be prespecified genetically. Visual experience is required to sustain visual function, but is not crucial to setting how much high frequencies are amplified relative to lows. According to this viewpoint, contrast constancy develops once multiple spatial-frequency channels emerge.

The genetic model is not very attractive for two reasons. First, it has difficulty explaining how contrast constancy emerges at 12 weeks of age. The infant's blur function is assuredly quite different from the adult's, a point that is illustrated by the differences between infant and adult CSFs. Twelve-week-olds, for instance, are most sensitive to 0.5 c/deg and quite insensitive to spatial frequencies above 3 c/deg (Atkinson et al., 1977; Banks & Salapatek, 1978). In contrast, adults are actually more sensitive to 3 c/deg than to 0.5 c/deg, and they are able to detect spatial frequencies as high as 40–50 c/deg. Changes in the optics of the eye probably cause some differences in high-frequency sensitivity between infants and adults, but there are no definitive data yet. Additional neural differences are implied by the striking immaturity of the young infant's retina (Abramov et al., 1982; Hendrickson & Yuodelis, 1984; Yuodelis & Hendrickson, 1986). Whatever the cause, the blur functions of mature and immature eyes are assuredly quite dissimilar. Consequently, the gain functions in 12-week-old infants and adults would have to be rather different to achieve veridical contrast matching at both ages. The genetic code could conceivably carry information to reset the gain function at various ages, but to accomplish this resetting, the code and the resulting maturational plan would have to be rather complex. This required complexity renders the genetic model somewhat unattractive.

The second, and more persuasive, argument against a simple genetic model comes from adults with amblyopia. Amblyopia is a condition of reduced visual acuity in one eye, usually caused by anisometropia (different refractive

errors in the two eyes) or strabismus (crossed eyes). Hess and Bradley (1980) measured the CSFs and contrast matching of amblyopic adults.

Their results are illustrated in Fig. 5.9. The CSFs of the observers' amblyopic and normal eyes were rather dissimilar, the amblyopic eye exhibiting reduced sensitivity to high spatial frequencies. Hess and Bradley asked their observers to match a high-frequency grating presented to the normal eye with a grating of the same spatial frequency presented to the amblyopic eye. When the grating presented to the normal eye was just visible, the observers required more contrast in the target presented to the amblyopic eye. When the grating presented to the normal eye was suprathreshold, however, observers reported identical apparent contrasts when the two gratings were in fact equal in contrast. Clearly these amblyopic observers were using different gain functions for the two eyes in order to compensate for the poorer contrast

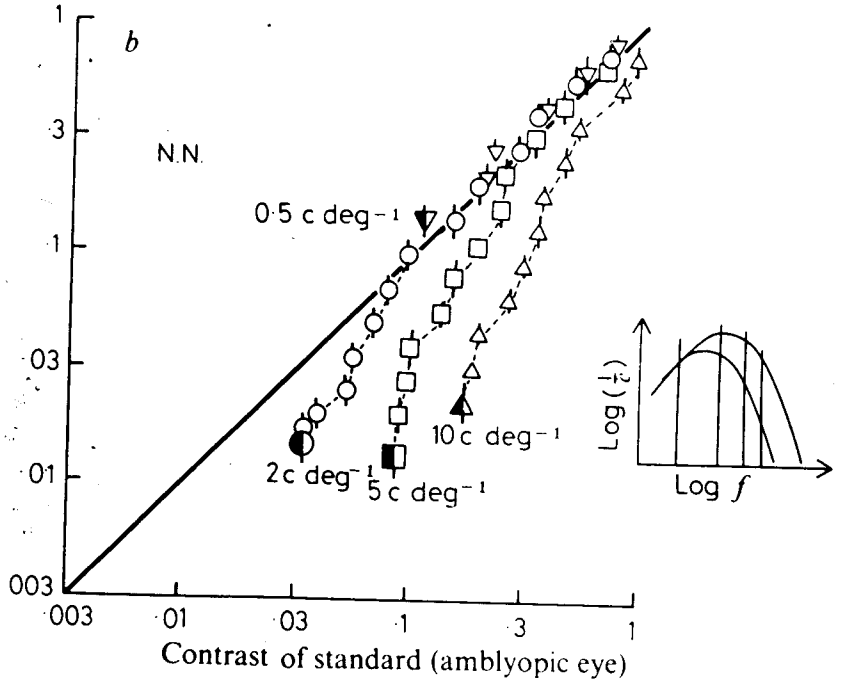


FIG. 5.9 Contrast matching data from amblyopic observer in the Hess and Bradley (1980) report. The CSFs for the ambylopic and nonamblyopic eyes are shown in the inset to the right. The large graph plots the contrast of the grating presented to the nonamblyopic eye, when it appeared to match the contrast of the grating presented to the amblyopic eye, on the ordinate. The contrast of the grating presented to the amblyopic eye is plotted on the abscissa. The solid line indicates where the data would lie if matches occurred at equal physical contrasts. The half-filled symbols represent the matches at threshold. The open symbols represent the matches at suprathreshold contrasts. From Hess and Bradley, 1980.

sensitivity of the amblyopic eye. Genetic models would have difficulty explaining these data because the genetic code would not only have to carry information about how the blur function changes with age in normal children, but also how it changes between eyes in amblyopic children and adults. This seems quite implausible and leads one to suspect that visual experience is involved fundamentally in the acquisition of the proper gain function.

Before describing the three experiential models in Table 5.1, let me characterize clearly what they have to explain. Allow me, for the moment, to use a homunculus (see Fig. 5.10) to represent the information gathering and decision making that must be performed in the central visual system. The homunculus is used only as an explanatory aid. By using him, I am not necessarily claiming that the information gathering and decision making are performed by a high-level, complicated mechanism.

A job description for the homunculus would read: Gather information from stream of visual inputs in order to adjust parameters of gain function. He must adjust the parameters of the gain function to render it in the inverse of the blur function. In the experiential models, the blur function is assumed to be unknown initially, so the homunculus needs information that will allow him to measure it. How should he proceed? The difficulty in answering this question is illustrated by the equation:

$$O(u,v) = I(u,v) \cdot B(u,v) \tag{3}$$

where $O(u,v)$ represents the output of the peripheral visual system as a function of spatial frequency in two dimensions, $I(u,v)$ represents the input to the visual system as a function of frequency, and $B(u,v)$ represents the blur function, that is, how much the contrast of various spatial frequencies is reduced during peripheral processing. The experiential models assume that the homunculus has direct access to $O(u,v)$ only, so he has to solve one equation with two unknowns. This is, of course, impossible unless some additional constraints can be brought to bear. The three experiential models postulate different contraints.

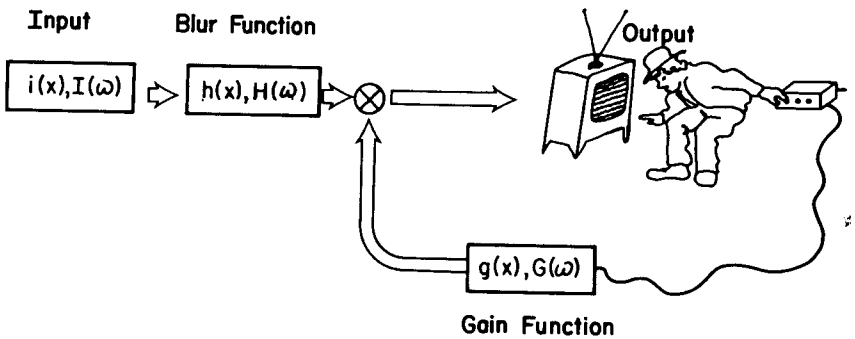


FIG. 5.10. A schematic of the feedback mechanism that is presumably involved in adjusting the contrast compensation (or gain) function. The input to the visual system is represented by the box on the left. The blurring caused by optical imperfections and peripheral neural summation is represented by the next box. The output of the blur process is depicted by the image on the TV screen. The evaluation of that output is represented by the homunculus who adjusts the parameters of the gain function in order to compensate for the attenuation due to the blur function.

Model 2 assumes that the input $I(u,v)$ is known in a statistical sense so the equation reduces to:

$$B(u,v) = \frac{O(u,v)}{I(u,v)} \tag{4}$$

where $O(u,v)$ and $I(u,v)$ are known. Model 2 does not assume that every input presented to the visual system is known; such an assumption would be absurd because the purpose of visual perception itself is to determine what specific inputs are. Instead, it is assumed that the average distribution of spatial frequency information in the environment is known. For instance, the homunculus might assume that the average contrast of 20 c/deg information in the environment is 50% of the average contrast at 5 c/deg. We can rewrite Equation 4 to express this model more clearly:

$$B(u,v) = \frac{O_a(u,v)}{I_a(u,v)} \tag{5}$$

where $I_a(u,v)$ represents the expected average input amplitudes as a function of spatial frequency and $O_a(u,v)$ represents the output amplitudes the homunculus observes on the TV screen. The blur function is deduced by comparing the expected average input distribution to the observed average output distribution. The homunculus can deduce the correct blur function in this manner, but only if the actual input information corresponds on average with his expectations. Is this plausible? If the distribution of spatial frequency information in the environment were simple (that is, could be described by a few parameters), the model could be implemented readily. Unfortunately, the distribution is not simple. Switkes, Mayer, and Sloan (1978) have shown that the distribution of amplitudes as a function of spatial frequency varies from one sort of environment to another. Urban environments, for example, appear to have relatively greater amplitudes at high spatial frequencies and at vertical and horizontal orientations than rural environments do. Thus, Model 2 requires in practice a detailed expectation of the average distribution of amplitudes across spatial frequencies in the environment and this expected distribution would have to vary depending on the kind of environment. The combined requirements for detail and flexibility seem unfeasible and decrease the attractiveness of this model.

Another version of Model 2 is more plausible, however. This model, which I call Model 2A, assumes that the homunculus knows the spatial frequency

distribution of a particular feature in the environment and can identify that feature unambiguously for the range of blur functions he has to deal with. Equation 4 applies for this model, but now the expected input is a particular feature rather than some time average for all imputs. The known feature would have to contain a broad range of spatial frequencies in order to provide information at all the spatial frequencies the homunculus needs to do his job. An example illustrates how this model might work. Suppose that the known feature is a sharp light-dark transition (an edge). In the interest of simplicity, I just consider the edge in one dimension. The one-dimensional Fourier transform of an edge of amplitude k is:

$$I(v) = \frac{k}{\pi jv} \tag{6}$$

where j is the square root of -1 and v is spatial frequency. From previous Equation 3 we have:

$$O(v) = B(v) \cdot \frac{k}{\pi jv}$$

which implies that:

$$B(v) = O(v) \cdot \frac{\pi jv}{k} \tag{7}$$

Because π and j are constants and $O(v)$ can be measured, the homunculus could determine $B(v)$ to within a scale factor $1/k$. In other words, he could in principle deduce the blur function by simply monitoring the output due to a known feature. This model seems feasible. Perhaps the greatest uncertainty is whether the homunculus could reliably identify a particular feature in the environment from the output information provided. For example, without knowing the blur function, the homunculus might be unable to discriminate a diffuse edge, such as a shadow, from a sharp edge. If he mistakenly chose the diffuse edge as the known feature, he would err in his computation of the blur function and adjust the gain function inappropriately. Whether reliable identification of a particular feature is possible could be examined by using a computer implementation of Model 2A. The fact that a computer implementation is possible enhances the utility of these computational models; the performance of the implementation would be an extremely useful index of the model's feasibility. If the implementation could not successfully identify a particular feature embedded in everyday settings, the model should not be pursued further. If it could identify the feature reliably for a reasonably broad range of blur functions, we would know that the model can work in principle.

Model 3 takes a different tack altogether. It assumes that the homunculus deduces the blur function by noting how the output changes as an object undergoes a smooth transition in distance. Figure 5.11 illustrates how the model works. Consider a particular input stimulus $i(x,y)$ and its amplitude spectrum $I(u,v)$. When the stimulus is viewed at 1 meter, its amplitude spectrum in one dimension is the solid line. When the stimulus is viewed at 2 meters, all of the spatial frequency components are shifted upward in frequency by a factor of two. This shifted amplitude spectrum is illustrated by the dashed line in the figure. The homunculus could deduce the blur function if he could determine that the stimuli at 1 and 2 meters were actually produced by the same object and that the change in distance was a factor of two. The equations below show how. In deriving the equations, I assumed that the visual system processes images using separate spatial-frequency-selective mechanisms, an assumption for which there is overwhelming psychophysical and physiological evidence (Braddick, Campbell, & Atkinson, 1978). The mechanisms are roughly one octave wide (that is, they respond to a twofold range of spatial frequencies). The outputs of the various mechanisms are represented by $W(u,v)$, $X(u,v)$, $Y(u,v)$, and so forth. The W mechanism responds to a one-octave band centered at a spatial frequency of $v1$. The X mechanism also

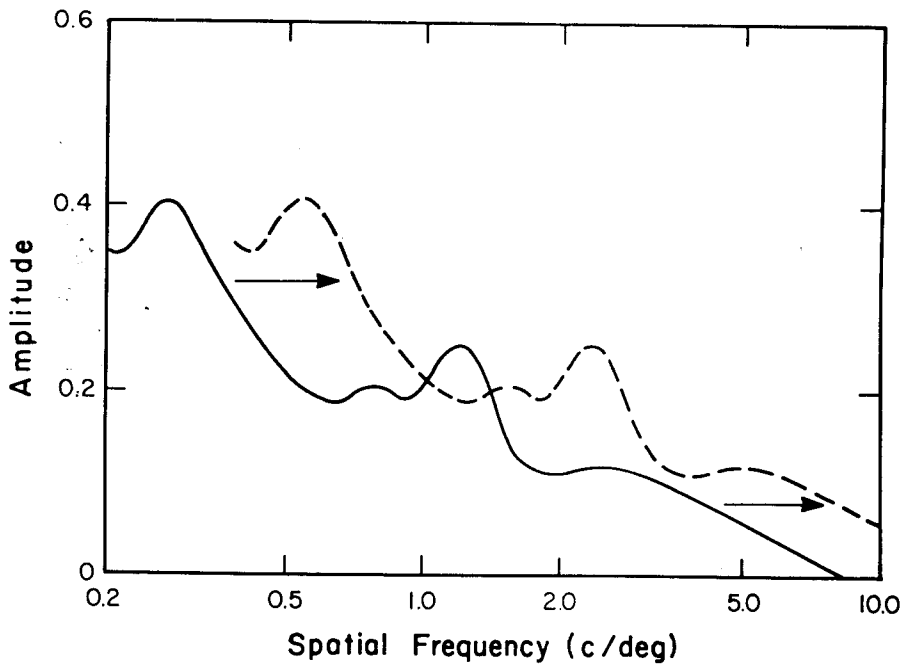


FIG. 5.11 The amplitude spectra of an object when viewed at two different distances. The solid line represents the amplitude spectrum (the amplitudes of the various sinewave grating components) when the viewing distance is one meter, and the broken line the spectrum when the viewing distance is two meters. Notice that all of the sinewave components simply shift upward in spatial frequency by a factor of two.

responds to a one-octave band, but the center of the band is $v2$, a spatial frequency two times higher than $v1$. The center of the Y mechanism's band is $v4$, which is four times higher than $v1$, and so on. The outputs of the three mechanisms are shown in Table 5.2 for the object presented at two distances, $d1$ and $d2$. The second distance in this example is twice the first, but the algorithm will work for other distances, too.

Once the homunculus determines the distance ratio, he can deduce the blur function by comparing one mechanism's output at one distance to another's output at the second distance. In the example, he should compare W's output at $d1$ to X's at $d2$ and X's output at $d1$ to Y's at $d2$. The following equations illustrate how he estimates the blur function by making these comparisons:

$$\frac{W(v_1,d_1)}{X(v_2,d_2)} = \frac{H(V_1) \cdot (I(v_1,d_1)}{H(v_2) \cdot I(v_2,d_2)} \quad \text{and} \quad \frac{X(v_2,d_1)}{Y(v_4,d_2)} = \frac{H(v_2) \cdot I(v_2,d_1)}{H(v_4) \cdot I(v_4,d_2)}$$

But we know that:

$$\frac{I(v_1,d_1)}{I(v_2,d_2)} = \frac{I(v_2,d_1)}{I(v_4,d_2)} = 1$$

therefore,

$$\frac{H(v_1)}{H(v_2)} = \frac{W(v_1,d_1)}{X(v_2,d_2)} \quad \text{and} \quad \frac{H(v_2)}{H(v_4)} = \frac{X(v_2,d_1)}{Y(v_4,d_2)} \tag{8}$$

These ratios are the relative loss of contrast from one band of spatial frequencies to the next. This information is all the homunculus needs to adjust the gain function appropriately.

This model is more attractive than Models 2 and 2A because it does not require too much knowledge about the input stimulus. However, the homunculus does need to know when and where to apply the algorithm, and this decision may not be easy. I propose the following strategy. The homunculus monitors the output of the TV screen for a particular type of event: a smooth symmetric change in the angular subtense of a closed figure (note that he does not actually have to compute the change in distance, but rather only the change in angular subtense). He applies the algorithm of Equation 8 to only those events because they are generally caused by an object moving smoothly in distance and not tilting or slanting toward or away from the eye as it does

TABLE 5.2

Distance 1			*Distance 2*		
$W(\omega_1,d_1)$	=	$H(\omega_1)\cdot I(\omega_1,d_1)$	$W(\omega_1,d_2)$	=	$H(\omega_1)\cdot I(\omega_1,d_2)$
$X(\omega_2,d_1)$	=	$H(\omega_2)\cdot I(\omega_2,d_1)$	$X(\omega_2,d_2)$	=	$H(\omega_2)\cdot I(\omega_2,d_2)$
$Y(\omega_4,d_1)$	=	$H(\omega_4)\cdot I(\omega_4,d_1)$	$Y(\omega_4,d_2)$	=	$H(\omega_2)\cdot I(\omega_4,d_2)$

so. If he could reliably identify such cases, the algorithm could be used to compute the blur function and thereby adjust the gain function. When he identifies a proper event, the change in distance would be determined by the ratio of angular subtenses at the two distances.

The next step in the development of these computational theories will be computer implementation using visual scenes as input. The successes or failures of the implementations should indicate which models are worth pursuing further. Finally, the computational theories should be tested empirically in restricted rearing experiments. For example, Model 3 would predict that a kitten raised in a situation that did not allow the viewing of objects undergoing smooth transitions in distance would not exhibit contrast constancy.

OPTICAL FLOW AND THE PERCEPTION OF DEPTH AND EGOMOTION

The visual system uses several depth cues to deduce the distance of objects in the environment, but most cues have serious limitations. Linear perspective, for example, requires rectilinear objects and, therefore, would be useless in a forest. Familiar size requires familiar objects whose objective sizes do not vary widely. Binocular disparity is useful in a variety of situations because disparity can be computed for any object with horizontal variation in luminance or hue and such objects are very common. But binocular disparity computations are not precise enough to specify the relative distances of objects more than 50 meters away (Ogle, 1950; Schor & Flom, 1969). Furthermore, the just-detectable disparity in the peripheral visual field is so large that the stereoscopic system provides little depth information for objects far from the line of sight (Ogle, 1950).

When an observer moves through the environment, a moving pattern of light falls on the retina. The resulting "optical flow field" provides useful information about the observer's movement and about the three-dimensional layout of the scene (Gibson, 1950, 1966). In principle, optical flow provides useful depth information for a wide range of distances. If one walks briskly through a scene, the relative distances of objects much farther than 50 meters away can be determined from differential motion across the retina. Optical flow also provides useful depth information in the peripheral visual field because velocity discrimination in the periphery is quite good (McKee & Nakayama, 1984), especially in comparison to stereoacuity.

Although few studies have been conducted, the infant literature suggests that sensitivity to optical flow develops quite early in life. Evidently infants can use optical flow to discern their own movements (Bertenthal, Dunn, & Bai, 1986; Butterworth & Hicks, 1977; Lee & Aronson, 1974) and to deter-

mine the depth of objects (Carroll & Gibson, 1981; Kellman, von Hofsten, & Soares, 1985; Yonas et al., 1977; Yonas, Pettersen, & Lockman, 1979).

It was pointed out in the second section that the growth of the eye and retina during infancy should disrupt the computations of egomotion and layout from optical flow. I show in greater detail why this is so and then consider how visual experience might be used to recalibrate the optical flow system and thereby reestablish reasonably accurate computations of egomotion and layout. A specific recalibration mechanism is proposed.

As in the contrast perception example, it is necessary first to provide background information on how optical flow is used by the visual system before discussing the developmental problem of interest. Gibson (1950) and Gibson, Olum, and Rosenblatt (1955) examined the pattern of motion across the retina as an observer moves through a static environment. Figure 5.12 displays the flow of the optic array for a bird in flight. The arrows represent the motion across the retina of texture elements on the earth below. Flow is centrifugal ahead and centripedal behind, traveling in great circles in between. There is a focus of expansion directly ahead and a focus of contraction behind. Gibson argued that this flow field contains two important pieces of information. First, the focus of expansion indicates the direction in which the observer is moving. Second, the magnitude and position of flow vectors indicate the relative distances of points in the scene.

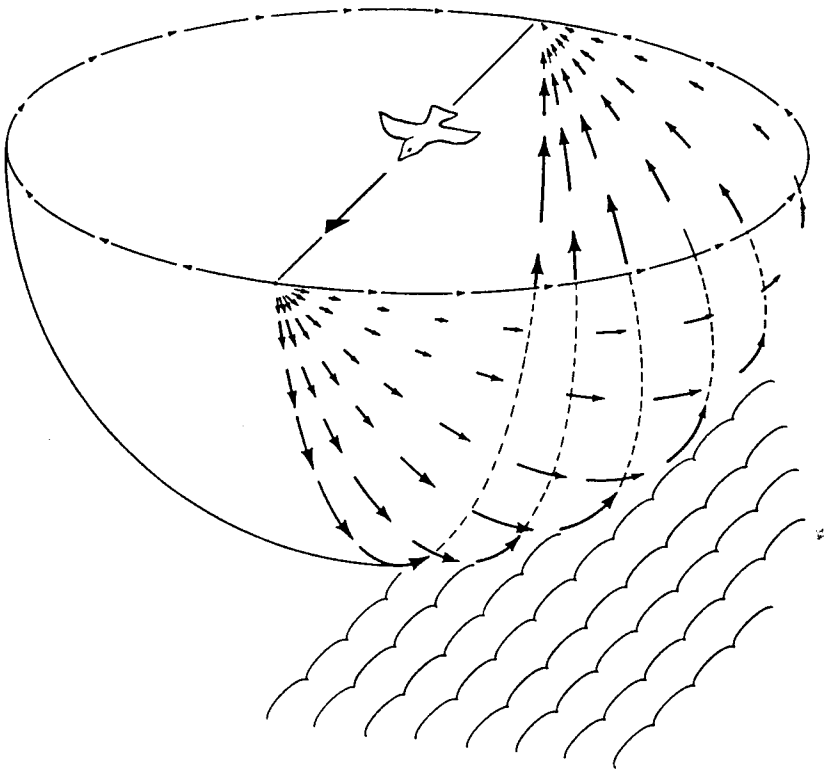


FIG. 5.12 The optical flow field during locomotion. When a bird flies parallel to the ground, texture elements below flow under its eyes in the manner shown. From Gibson, 1966.

Recent psychophysical experiments have illuminated adults' use of optical flow. Consider, for instance, Rogers and Graham's (1982) comparison of adults' sensitivity to binocular disparity and optical flow. They compared the smallest detectable depth interval specified by disparity to the smallest interval specified by motion parallax. The stimuli were random dot patterns on a CRT. In the binocular disparity condition, the two eyes received slightly different patterns. The disparity between the dot patterns specified a surface with sinusoidal corrugations in depth. Rogers and Graham varied the disparity between the dots in order to obtain steresocopic thresholds. The thresholds indicated the smallest detectable modulation in depth. In the motion parallax condition, the same pattern of dots was presented but was viewed monocularly. The observers were instructed to move their heads back and forth while viewing the display. The dot pattern was systematically transformed with each movement of the observer's head to simulate the patterns of relative motion that would have occurred if the observer were viewing a real corrugated surface. This arrangement produced a clear three-dimensional percept whenever the head was moved. When the head was still, the surface appeared flat. Thresholds were measured again by varying the depth interval specified (in this case by varying the relative displacement of the dots). Rogers and Graham found that the binocular disparity and motion parallax thresholds followed similar laws; in both cases, the minimum detectable depth interval occurred at corrugation spatial frequencies of 0.2–0.4 c/deg. They also compared the relative sensitivities of the binocular disparity and motion parallax systems. To do so, they noted that the stereoscopic system provides depth information based on two surface views separated by 6 cm (the interocular distance). Thus, they compared the just-detectable corrugation in the disparity condition to the just-detectable depth corrgation in the motion parallax condition when the head was moved 6 cm. The thresholds were quite similar. For one observer they were virtually identical. For the other two observers, the parallax thresholds were only two times greater than the disparity thresholds. Hence, adults can in fact use optical flow effectively to discern the three-dimensionality of the visual scene.

Warren (1976) examined adults' ability to determine the direction of self motion from the optical flow field. He presented a display of moving texture that simulated locomotion across a flat surface. The observers were able to indicate the direction of simulated egomotion to within 5 deg, suggesting that adults can in fact determine their own motion from optical flow. By the way,

other investigators (Johnston, White, & Cumming, 1973; Llewellyn, 1971; Regan & Beverley, 1982) have reported that adults are not particularly adept at detecting the direction of simulated motion. Their experiments differ in critical aspects from Warren's, a point that I expand later.

The experiments of Rogers and Graham (1982) and Warren (1976) demonstrate that adults are quite sensitive to information in the optical flow field. They do not tell us, however, what computations are required to extract this information in everyday situations. Formal theories of information extraction from realistic flow fields can enhance our understanding. Several formalizations have appeared, but until recently all of them applied to quite limited situations only (Clocksin, 1980; Gibson et al., 1955; Gordon, 1965; Lee, 1980; Nakayama & Loomis, 1974). Generally, they assumed that the observer was moving in a straight line and maintaining fixation in the direction of motion (e.g., looking at an object toward which he is walking). This assumption is overly restrictive because humans frequently move in curved paths while fixating an object that does not lie in the direction of motion. An example is a person walking along a trail and looking at a tree as she walks past it. Such a situation violates the assumptions of earlier formalizations. More recently, Longuet-Higgins and Prazdny (1980) developed a computational theory for determining egomotion and three-dimensional layout for any arbitrary motion of the observer and his eyes. I have modified their theory somewhat and implemented it in a computer simulation.

The simulation computes 3D layout and the direction of egomotion for any arbitrary movement of the observer and his eyes. Figure 5.13 displays the

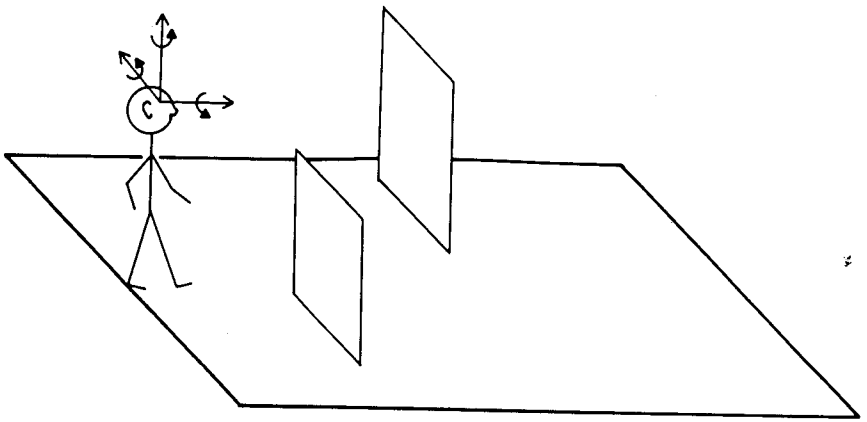
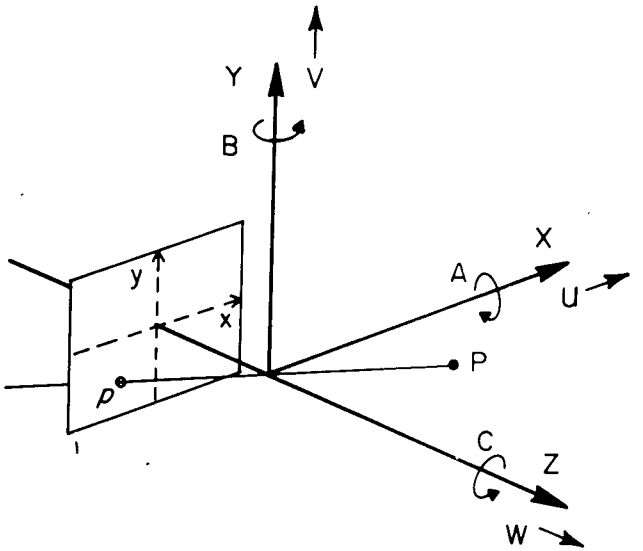


FIG. 5.13 The observer and scene used in the computer simulation. The observer views the scene monocularly. The viewing eye can translate in any direction, which is illustrated by the three orthogonal straight arrows. The eye can also rotate in any direction, which is illustrated by the three circular arrows. The scene consists of a flat surface upon which the observer is standing and flat walls that are positioned at different locations along the surface.



XYZ coordinate system moving with the eye
U,V,W: translation along X,Y, and Z
A,B,C: rotation about X,Y, and Z
x,y: retinal coordinate system
P: position of point in environment
p: retinal position of point P

FIG. 5.14 The three-dimensional coordinate system used in the computer simulation. The center of the coordinate system moves with the eye. The *U*, *V*, and *W* vectors correspond to translations in the *X*, *Y*, and *Z* directions. The *A*, *B*, and *C* vectors correspond to rotations about those directions. The retina is represented by a two-dimensional surface behind the origin of the *XYZ* coordinate system. Point *P* in the scene is projected to point *p* on the retina with coordinates (x,y).

observer and the scene used in the computer simulation. The observer views the scene monocularly while walking through it. He can move in straight or curved paths and rotate his eyes and head in any direction. The objects in the scene are stationary. Walls of different sizes can be placed in a variety of locations. The observer is not provided information about the class of possible scenes. That is to say, he does not know that the supporting surface is flat and that the walls are planar. Discrete texture elements from objects in the scene are provided so the observer does not have to solve the motion correspondence problem (Marr, 1982).

Figure 5.14 shows the coordinate system used. A point *P* in the visual scene has the coordinates (X,Y,Z). The retina is represented by the two-dimen-

sional surface behind the origin. Perspective projection is employed, so the projection of P has retinal coordinates (x,y), where $x = X/Z$ and $y = Y/Z$. From kinematics we know that any instantaneous motion can be described as the addition of a translation and a rotation. Moreover, any translation can be described as the addition of translations in three perpendicular directions, and any rotation can be described as the addition of rotations about those three directions. The simulation represented the observer's motion by the three translation vectors U, V, and W along X, Y, and Z, respectively, and by the three rotations A, B, and C about X, Y, and Z. As the observer moves through the scene, the retinal movement due to P is represented by the vectors u and v where u is velocity in the x direction and v is velocity in the y direction.

In the simulation the observer moves with various combinations of translations and rotations and thereby causes projected points to flow across the retina. Figure 5.15 is a photograph of a scene in which the observer (in this case a camera) moves in a straight line while fixating an object (the woman's face) along that path of motion. The observer's movement is purely translatory, which is the situation assumed by the earlier formalizations. Note that flow is

FIG. 5.15 Photographic simulation of the optical flow field induced by a pure translation. The camera has been translated toward the scene in the direction of the woman's face. The corresponding flow of texture elements in the scene is represented by the smeared dots. From Regan and Beverley, 1982.

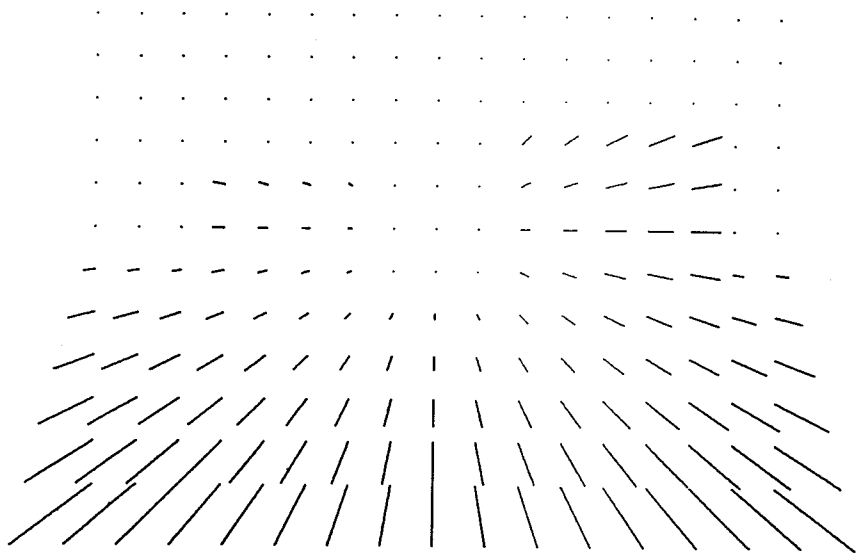


FIG. 5.16 An optical flow field in the computer simulation. In this case, the observer's motion is a pure translation toward the middle of the scene depicted in Fig. 5.13. The flow of texture elements in the scene is represented by the vectors. Long vectors correspond to rapid flow and short ones to slow. The distant wall on the left and the nearer wall on the right can be distinguished from the background on the basis of the flow in those regions.

directed away from a point in the middle of the woman's face. This point is the expansion point.

Figure 5.16 depicts a similar situation in the computer simulation. Again the observer has moved straight ahead while fixating the expansion point. The scene depicted in Fig. 5.13 was used to derive the flow pattern. Each vector in Fig. 5.16 is simply a line between the projections of a point at the beginning and end of the observer's instantaneous movement. Again the flow vectors are directed away from the expansion point. Note also that the magnitudes of the vectors are dependent on the distance to objects in the scene. The relative distance to any point in the scene can be computed in the following fashion.

$$\begin{aligned} Z/W &= (x-x_o)/u \\ Z/W &= (y-y_o)/v \end{aligned} \qquad (9)$$

where (x_o,y_o) are the retinal coordinates of the expansion point and W is the observer's velocity along Z. Thus, the distance to any point in the scene can be computed to within a scale factor. Of course, the absolute distance to all visible points can be computed if W is known or if the absolute distance to one point is known. These relations allow one to compute the direction of egomotion and the 3D layout of the scene: The 3D coordinates of any point P

can be computed from the calculation of Z and by projecting x and y from the retina to the scene to determine X and Y, and the direction of egomotion is given by the projection of (x_o,y_o) from the retina to the scene.

Figure 5.17 portrays the flow resulting from a similar movement through the same scene. In this case, however, the observer is translating toward the wall on the right while maintaining an eye position several degrees to the left of the direction of motion. This situation is a bit unrealistic in that the observer is not maintaining fixation on a specific object; I present it simply to show the flow resulting from another pure translatory movement. A comparison of Figs. 5.16 and 5.17 shows that the flow patterns are different but, in both cases, the point of expansion is quite distinct and the magnitudes of flow vectors are dependent on the distance to points in the scene. Once again egomotion is specified by the projection of (x_o,y_o) and 3D layout can be determined from Equations 9.

Unfortunately, flow fields are generally not nearly as simple as those shown in Figs. 5.16 and 5.17. Consider, for example, a situation in which the observer moves straight ahead (toward the woman) but fixates the middle of the polka dot wall on the left in Fig. 5.18. As is apparent from a comparison of Figs. 5.15 and 5.18, the flow pattern changes. Despite the fact that the camera is moving in the same direction as it did for Fig. 5.15, the apparent expansion point in Fig. 5.18 is in the middle of the polka dot wall rather than on the woman's face. A somewhat more complicated example for the computer

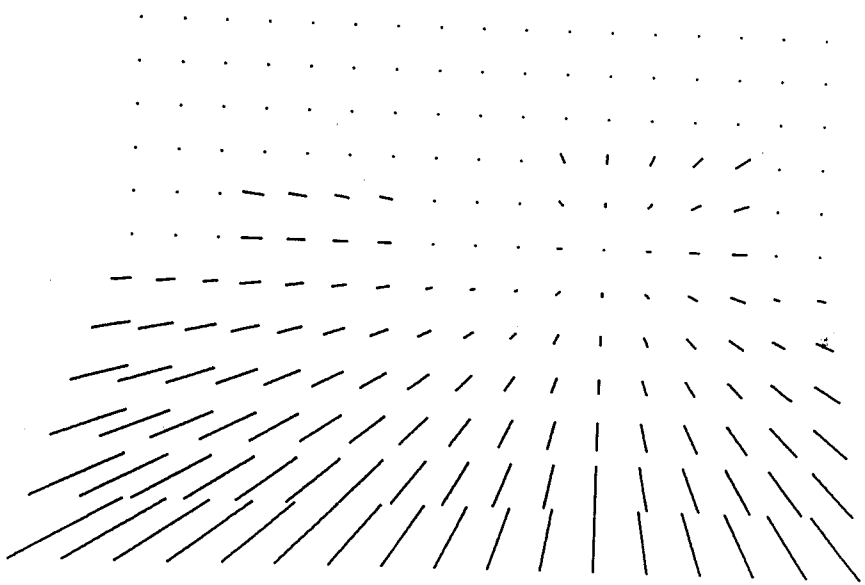


FIG. 5.17 The optical flow field induced by a pure translation toward the right side of the scene. In this case, the observer's motion is toward the wall on the right.

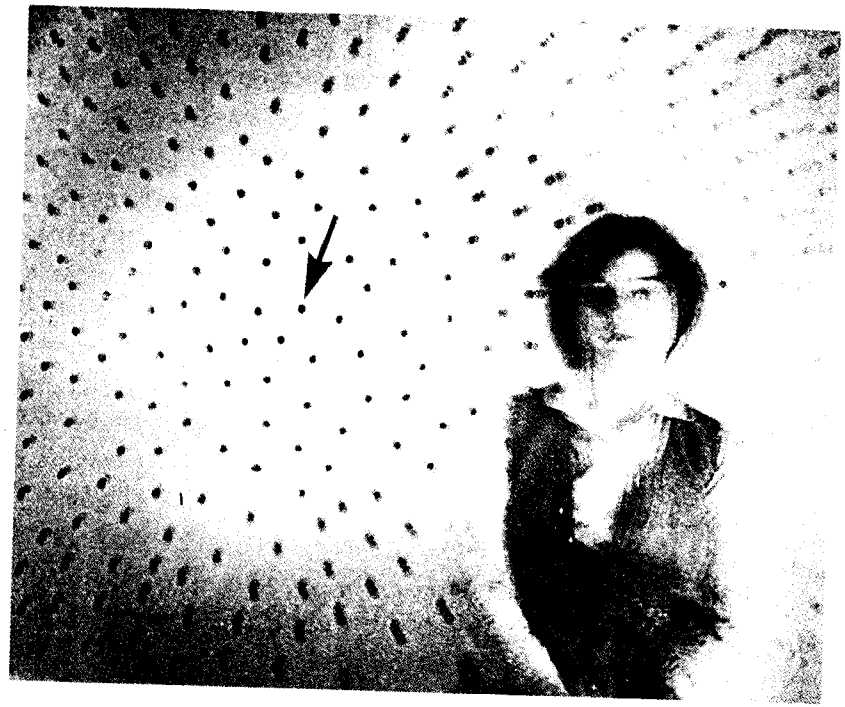


FIG. 5.18 Photographic simulation of the optical flow field induced by a translation plus a rotation. As in Fig. 5.15, the camera has been moved toward the woman's face, but it has also been rotated in order to remain pointed at the texture element indicated by the arrow. From Regan and Beverley, 1982.

simulation is displayed in Fig. 5.19. Here the observer is translating toward a point straight ahead while tilting his head to the side. The combination of translation (due to the observer's linear movement) and rotation (due to the head tilt) produces the complex flow pattern shown in Fig. 5.19. The expansion point is no longer a focus of expansion because many of the flow vectors do not point toward it. Consequently, the direction of egomotion is not as readily discerned from the flow field as before. Notice also that the magnitudes of the flow vectors are no longer proportional to the distance of the points in the scene. The vectors in the upper left corner, for instance, are rather large despite the fact that they represent the motion of very distant points. Hence Equations 9 cannot be applied to compute the relative distance of points in the scene. The flow vectors across the retina are given by the following equations:

$$u = -U/Z - B + Cy + Wx/Z + Axy - Bx^2$$
$$v = -V/Z + A - Cx + Wy/Z - Bxy + Ay^2 \qquad (10)$$

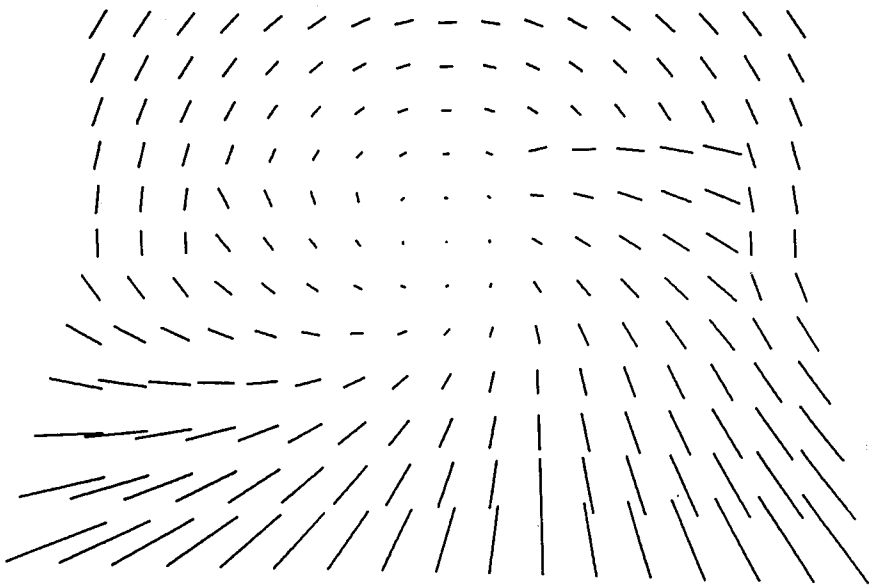


FIG. 5.19 The optical flow field induced by a translation plus a rotation. In this case the observer has moved toward the middle of the scene (the translation) while tilting his head to the side (the rotation). The scene is the one depicted in Fig. 5.13.

Clearly it would be advantageous if one could somehow eliminate the flow due to the rotations *A, B,* and *C*. Gibson (1950) provided the insight that allows one to separate rotation and translation effects. He realized that flow due to rotation is quite different from flow due to translation. Flow due to rotation is independent of the distance to points in the scene. This property is illustrated in Fig. 5.20, which shows the flow field due to the scene of Fig. 5.13 when the observer's head is tilted to the side. The observer has not translated here. The flow field consists of concentric circles of vectors whose magnitudes increase with distance from the middle of the visual field. Most important, the vector magnitudes of all neighboring points are similar even if the points lie on different depth planes in the scene. In other words, flow due to rotations, unlike flow due to translations, is depth-independent. The computer simulation capitalizes on this distinction by comparing the flow vectors in small patches of the visual field. Consider the flow field of Fig. 5.19 in which both translation and rotation effects are present. The simulation goes through the image and subtracts the flow vector of each point from the vector of a neighboring point. Stated another way, the simulation computes difference vectors for neighboring points in the flow field. Most of the difference vectors have magnitudes of zero or nearly zero, but large ones are obtained wherever depth discontinuities occur in the scene. Figure 5.21 shows the difference vectors obtained for the flow field of Fig. 5.19.[3] The difference vectors have been thresholded so they are all zero if they fall below some threshold magnitude. Notice that the nonzero difference vectors occur at the edges of the walls, the regions where depth discontinuities exist. Note also that the difference vectors all point to the expansion point. Thus, the difference vectors indicate the direction of egomotion, even for situations that are seriously corrupted by rotatory flow. The simulation uses this property to compute the true expansion point. Once this point is located, one can determine the flow due to rotation alone. Figure 5.22, which is a duplicate of Fig. 19, illustrates how. The vertical and horizontal lines represent lines in the y and x directions, respectively, that go through the expansion point (x_o,y_o). Recall that flow due to translation is always zero at the expansion point. Thus, any nonzero flow at (x_o,y_o) must be due to rotation. Furthermore, any nonzero flow u (that is, flow in the x direction) at points on the vertical line must also be due to rotation only. Likewise, any nonzero flow v (in the y direction) at points on the horizontal line must be due to rotation. The simulation uses u at (x_o,O) and v at (O,y_o) and u at $(x_o,1)$ to compute the rotations *A, B,* and *C*. The mathematical development is the following. From Equations 10 we note that the rotation effects are:

$$U_r = -B + Cy + Axy + Bx^2 \quad (11)$$

$$V_r = A - Cx - Bxy + Ay^2 \quad (12)$$

[3]Nakayama and Loomis (1974) postulated that neurons would be found that performed an operation similar to the difference vector operation described here. Later von Grunau and Frost (1983) observed neurons with the desired properties in the cortex of the cat.

One of the unattractive features of the computational theory presented here is its complexity. In particular, it is not clear how some of the steps involved in this approach might be implemented biologically. For instance, how would a biological system choose the three points required to compute Equations 15–17? With this criticism in mind, I have recently developed another computational theory of how 3D layout and egomotion might be computed from the optical flow field. The theory is based on direction- and speed-selective differencing operators similar to the ones proposed by Nakayama and Loomis (1974). I have been able to show that computation of these difference vectors allows one in principle to determine the relative depth of all points in the scene and to determine the direction of egomotion. The vector magnitudes are proportional to the relative depth between the two texture elements used to compute a given difference vector. The vector directions intersect at the expansion point. Examination of Fig. 5.21 reveals the need for another assumption in this model. Nonzero difference vectors are only observed at points where depth is changing. Thus, no difference vectors are observed in the middle of the two walls nor in the background behind the walls. To avoid the problem of concluding that any region with no nonzero vectors is very distant, the model assumes a "filling in" process. Whenever a series of nonzero difference vectors is bounded by a region of zero difference vectors, the model assumes that the internal surface is the same distance from the observer as the boundary. Walls and apertures can be distinguished from the directions of the difference vectors at the boundaries. This simple model is not presented formally here because I have not yet developed a simulation of it.

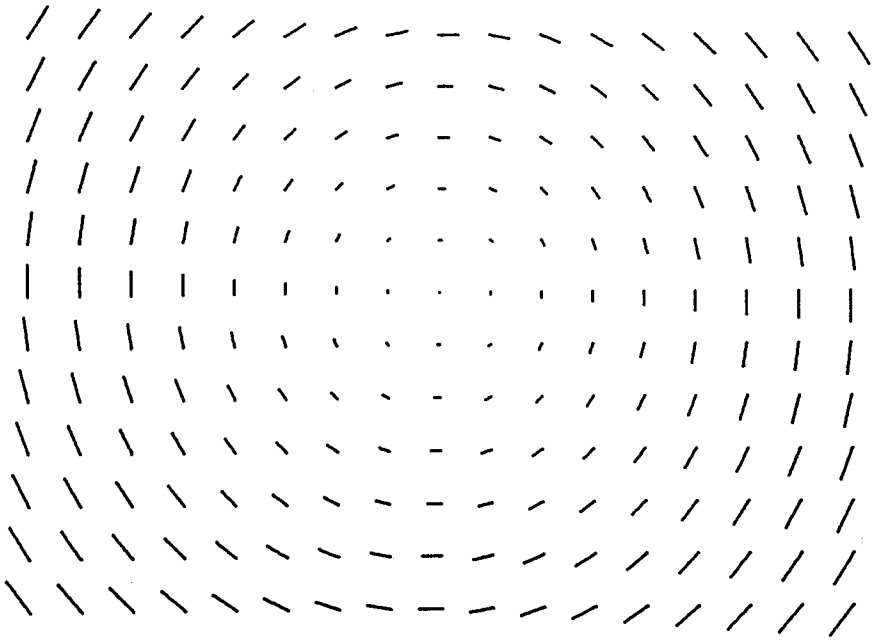


FIG. 5.20 The optical flow field induced by a pure rotation. The observer has simply tilted his head to the side while viewing the scene of Fig. 5.13.

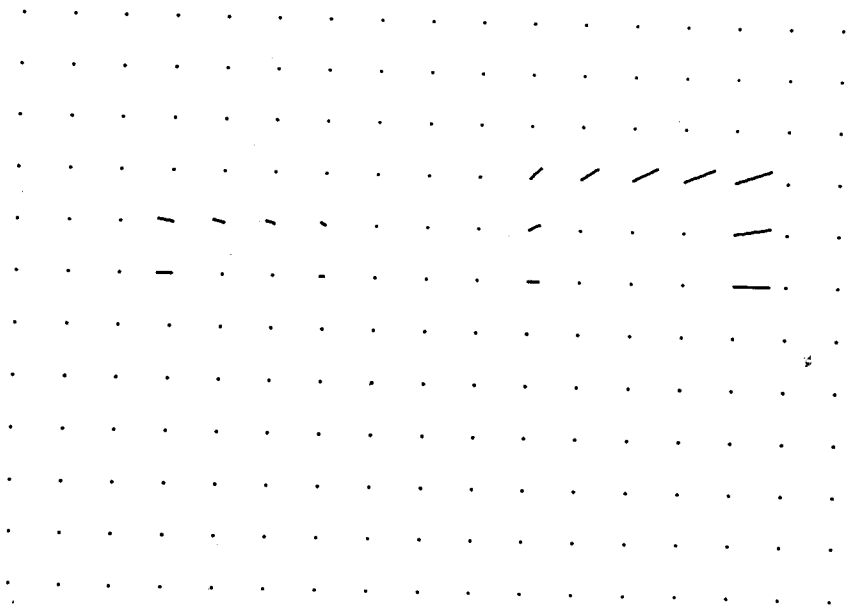


FIG. 5.21 The difference vectors computed by the simulation for the flow field of Fig. 5.19. (See text for details.)

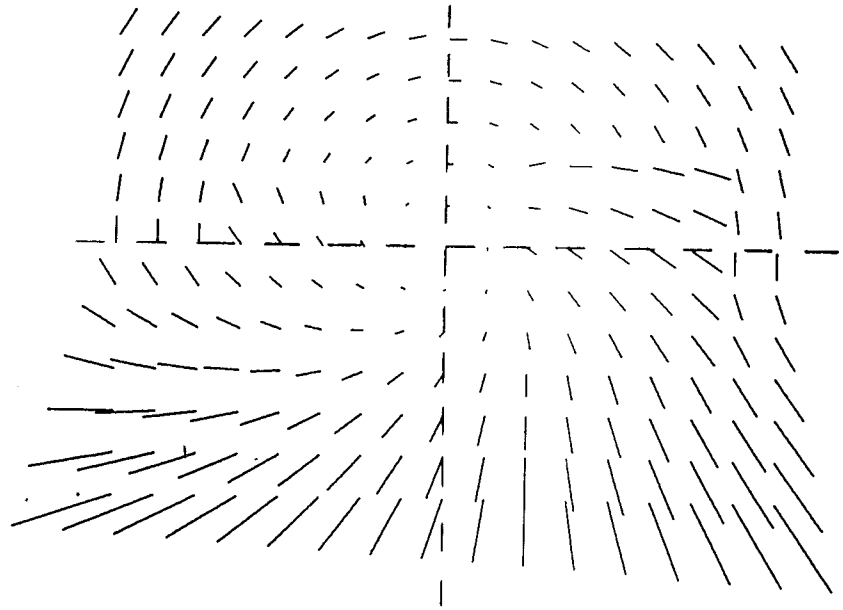


FIG. 5.22 The same flow field depicted in Fig. 5.19. The vertical and horizontal lines go through the computed expansion point in the middle of the field. The simulation uses the flow of points along those two lines to deduce the flow due to rotation. (See text for details.)

At (x_o, O) and (O, y_o) these equations reduce to:

$$u(x_o,O) = -B - Bx_o^2 \quad (13)$$

$$v(O,y_o) = A + Ay_o^2 \quad (14)$$

The subscript r is not required because u at (x_o, O) and v at (O, y_o) are due to rotation only. In other words, u and u_r are equal at (x_o, O), and v and vr are equal at (O, y_o). From Equations 13 and 14 we note that:

$$B = -u(x_o,O)/x_o^2 + 1) \quad (15)$$

$$A = v(O,y_o)/(y_o^2 + 1) \quad (16)$$

The values of all the variables on the right sides of the equations are known, so the rotations A and B can be determined. C is determined first by subtracting $u(x_o, O)$ from $u(x_o, 1)$ and then substituting Equation 16 for A:

$$C = u(x_o,1) - u(x_o, O) - \frac{v(O,y_o) \cdot x_o}{y_o^2 + 1} \quad (17)$$

Once the simulation computes the rotations A, B, and C, it simply subtracts flow due to those rotations from the original flow field (Fig. 5.19). The resulting flow field is shown in Fig. 5.23. Note that this figure is identical to

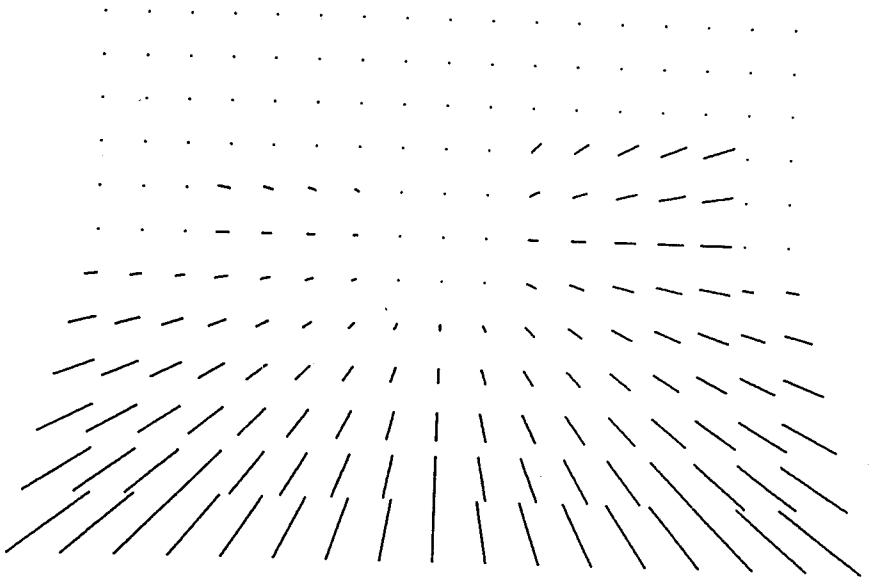


FIG. 5.23 The flow field that results once the rotation effects are subtracted. Notice that this field is the same as the one in Fig. 5.16. This demonstrates the effectiveness of the algorithm to eliminate flow due to rotation. (See text for details.)

Fig. 5.16, which means that only translation effects remain after the rotation effects are subtracted from the field. The simulation now uses Equations 9 to compute Z for all points in the scene. The absolute distance to one point (the distance from the eye to the ground beneath the feet) is given, so the simulation actually computes the absolute distance Z to all points. The direction of egomotion is also determined by simply projecting (x_o,y_o) into the scene.

I have examined the accuracy of depth and expansion point computations for a variety of scenes and movements. The simulation computes both very accurately. Thus, the computer simulation is able to determine the direction of egomotion and the 3D layout of objects in the environment even when the observer moves in complicated ways involving eye and head movements.

Before discussing recalibration of the optical flow system, let me return to the psychophysical experiments conducted on adults' ability to detect the direction of simulated egomotion. Recall that Warren (1976) reported accuracies of 5 deg or less whereas Johnston et al. (1973), Llewellyn (1971), and Regan and Beverley (1982) reported much poorer accuracy. The stimulus appears to be the critical difference between these experiments. Warren simulated motion across a ground plane. Thus, different points were different distances from the observer. The computer simulation presented here, when applied to his stimulus, produces a number of nonzero difference vectors. Those vectors can then be used to locate the true expansion point and

thereby eliminate the effects of rotation. The experiments of Johnston et al., Llewellyn, and Regan and Beverley simulated motion toward a frontoparallel plane. In other words, all points were the same distance from the observer. When the simulation is applied to such stimuli, all difference vectors are zero so the expansion point cannot be located whenever rotations occur. Thus, human adults and the computer program are both able to determine the direction of simulated egomotion accurately in the conditions of the Warren experiment and are unable to do so in the conditions of the other experiments.

Recent work by Rieger and Toet (1985) further establishes the importance of different depth planes to the determination of egomotion. They presented optical flow fields to adults that simulated approach to one or two textured frontoparallel surfaces. The simulated approach involved translation and rotation of the observer. Adults were asked to locate the perceived direction of egomotion. Localization was quite poor when only one surface was presented, but improved dramatically with the addition of the second surface. The computer simulation behaves similarly in those two situations. With only one frontoparallel surface present, all difference vectors are zero. The second surface produces a different flow field and, consequently, several nonzero difference vectors emerge. Those vectors allow the simulation to locate the expansion point and the direction of egomotion. The similarity between the simulation's and real observers' performance suggests that local differences in optical flow are crucial to adults' perception of the direction of egomotion.

OPTICAL FLOW AND RECALIBRATION

With this background, I now consider whether the optical flow system requires recalibration during infancy and early childhood, and if it does, how recalibration might occur. There is a small but consistent, literature on infants' sensitivity to optical flow. Yonas et al. (1977, 1980) found that 3- to 4-month-old infants respond defensively to optical flow specifying approach, rather than withdrawal, of a surface. Carroll and Gibson (1981) reported that 3-month-olds distinguish, on the basis of optical flow, an approaching object from an approaching aperture. Infants responded defensively to the former, but not the latter, situation. Kellman, von Hofsten, and Soares (1985) found that 4-month-olds perceived similarities between an expanding/contracting array of three points and an approaching/withdrawing triangle.

Others have examined infants' use of optical flow to control their posture. Lee and Aronson (1974) placed infants in a room whose walls and ceiling could be moved. The infants were 12 months or older and were able to stand

unsupported. When the room was moved, infants swayed or fell in the direction of the room's motion. Butterworth and Hicks (1977) observed similar behavior in young infants who experienced the room's motion while seated. These experiments show rather dramatically that infants capable of standing or sitting use optical flow information to adjust their posture. Thus, by the time infants are able to locomote under their own control, they can use optical flow to determine egomotion and depth, at least crudely.

As described in the second section, the eye is growing and the retina and central visual system are maturing during this period. These age-related changes alter the mapping between points in the environment and their neural representations. Such modifications of the distal-proximal mapping should affect the computations of depth and egomotion from the optical flow field. Therefore, the optical flow system, in order to maintain accurate judgments of depth and egomotion, must be recalibrated during infancy and early childhood.

Two age-related changes are most significant: (a) changes in the projective geometry of the eye and (b) changes in the positions of photoreceptors. Both of these growth effects are most dramatic during the first 12 months, but they continue in a slower fashion for years afterward. (Eye growth is not complete until adolescence; the photoreceptors migrate for at least the first year and a half.)

These two growth effects—one concerning optical projection and the other concerning photoreceptor displacement—could profoundly disrupt the use of optical flow information unless some sort of recalibration occurred. One can demonstrate such disruptions with the aid of the computer simulation. Figure 5.24 shows two flow fields for the same scene and same observer motion. The one on the right, however, has been modified to mimic the effects of barrel distortion. The simulation can estimate the distance to points in the two fields from Equations 9. The estimates derived from the flow field on the left had an average error of less than 1%. The estimates from the distorted field on the right had a much larger error of 147% on average. Thus, depth computations are quite inaccurate unless distortions in the distal-proximal mapping are compensated for by some sort of recalibration mechanism.

I have developed a recalibration scheme that can in principle compensate for such changes. The difficulty in developing such a scheme lies in the fact that there are three determinants of optical flow: flow due to translation, flow due to rotation, and differences in flow from one region to another due to age-related distortion. A recalibration algorithm has to uncover the effects due to distortion only. The algorithm presented here separates these effects in two ways. First, the algorithm looks for situations in which translation effects are not present. The algorithm computes difference vectors throughout the field and operates only when all of the difference vectors are zero or

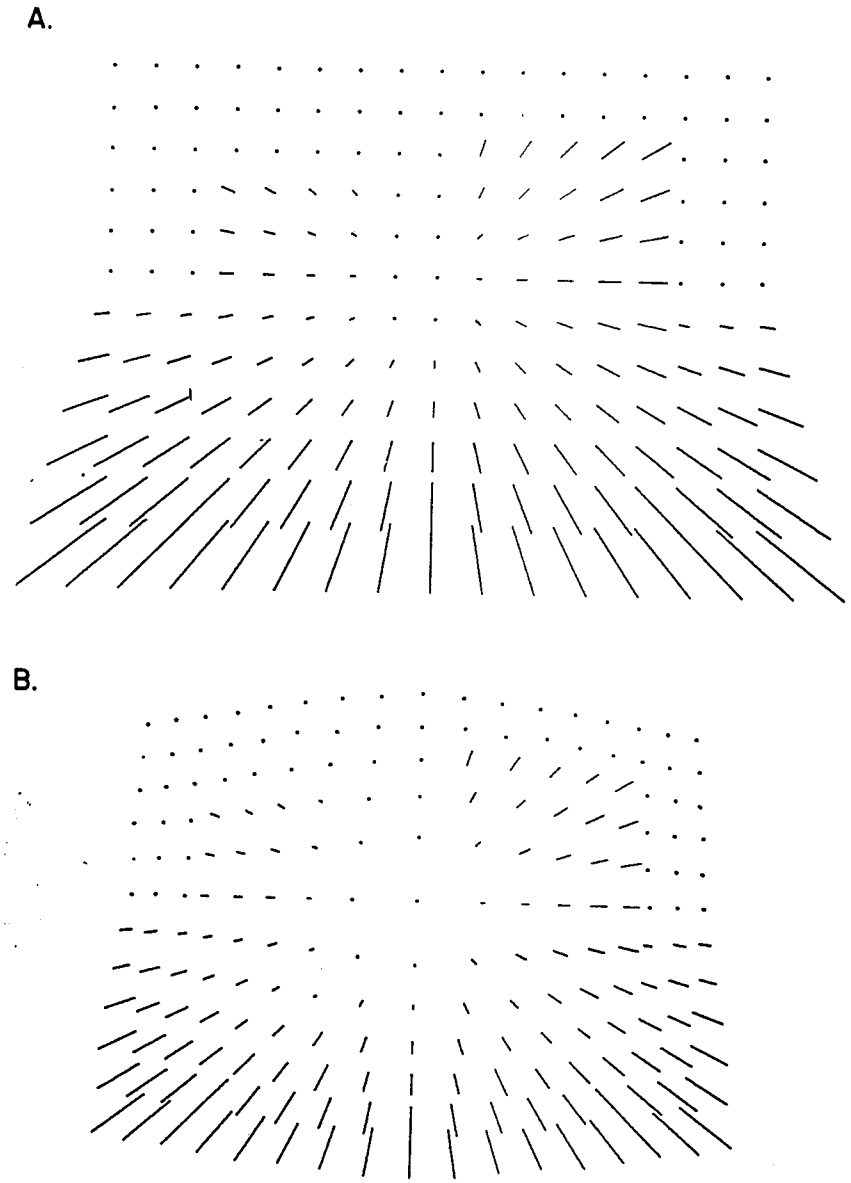


FIG. 5.24 Optical flow fields induced by a pure translation. The field on the top is the same as the one in Fig. 5.16. The one on the bottom is for the same scene and same observer motion, but barrel distortion has been added.

nearly zero. Only two situations can produce flow fields devoid of nonzero difference vectors. One is a pure rotation of the observer. The other is a translation (or a translation plus a rotation) when the scene has only one depth plane. The second situation is exceedingly rare in the natural environment, so it is ignored in the simulation. Limiting the application of the algorithm to pure rotations of the observer reduces the number of flow determinants to two: rotation and distortion. Fortunately, rotation has quite simple effects. As shown in Equations 15–17, the flow due to rotation can be predicted for all points in the visual field from the rotational flow in a small patch. The recalibration algorithm capitalizes on this fact by computing the rotation effects in the central visual field and then predicting what the flow should be in other portions of the field. Any disagreements must be due to distortion in the distal-proximal mapping.

Figure 5.24 illustrates the procedure when barrel distortion has been introduced. The panels on the top and bottom show flow fields due to a pure rotation with and without distortion. The distortion applied to the field on the bottom is isotropic, meaning that the same distortion is applied in all directions from the fovea. The use of isotropic distortions is justified because the

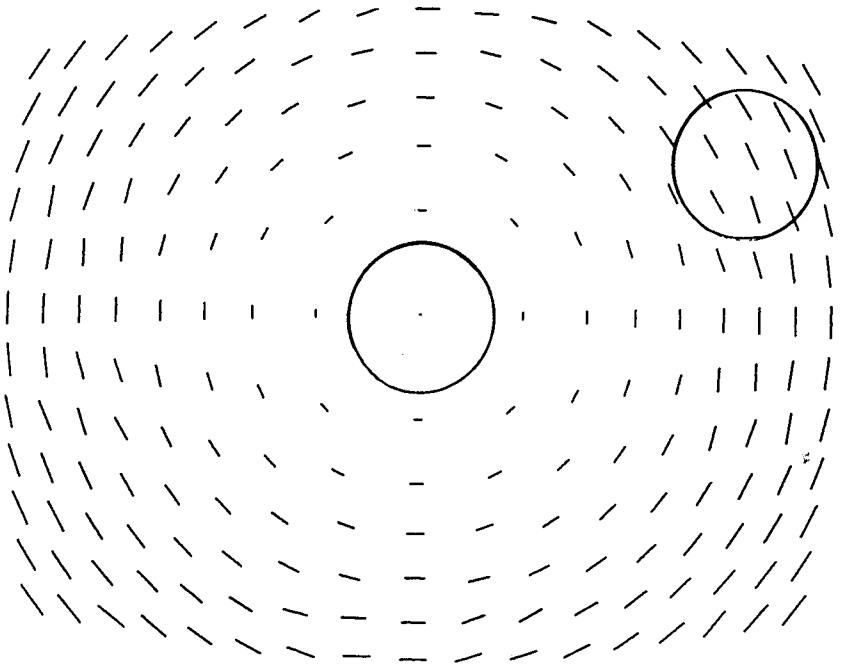


FIG. 5.25 The optical flow field induced by a pure rotation and a distortion that mimics a change in the projective geometry of the eye. The circles represent different retinal regions in which the rotation effects are estimated. (See text for details.)

age-related changes in distal-proximal mappings appear to be isotropic or nearly so. The difference vectors for the distorted flow field of Fig. 5.25 because no translation effects are present. The algorithm computes the rotation effects from flow vectors around the fovea. This patch is indicated by the middle circle. The algorithm then compares the rotation effects from that patch to those in another patch (indicated by the right circle). When distortion has not been introduced (i.e., the top panel in Fig. 5.24), the observed flow agrees with the predicted flow. The flow in Fig. 5.25 does not because it has been corrupted by spatial distortion.[4] The algorithm searches for such disagreements. When they exist, it assumes that they are due to distal-proximal mapping changes and recalibrates the flow field to compensate for them. Specifically, the flow vectors in peripheral patches are adjusted to bring them into accord with the predicted vectors. Because the distortions are isotropic, the algorithm can generalize distortion measured at one peripheral region to other regions with the same eccentricity. The flow vector adjustments are represented by coefficients. Once determined, those coefficients are used to modify vector magnitudes before depth and egomotion are computed from Equations 9.

The recalibration algorithm has successfully compensated for a variety of pincushion and barrel distortions. Although only isotropic distortions have been used, the algorithm could also be used to compensate for anisotropic distortions. The algorithm would simply have to compute the distortions at all points in the visual field and compensate for them. This procedure would be feasible but much more cumbersome computationally.

CONCLUSION

The existence of early visual competencies suggests that fundamental aspects of visual development are built-in rather than learned. These early competencies, however, pose a significant problem for the nativist view. How are infants and young children able to maintain these capabilities during postnatal ocular and neural growth? The answer, of course, is that some sort of recalibration is required. This chapter examined in detail the recalibration needs involved in two aspects of vision and visual development: contrast perception and optical flow perception. In both cases, computational

[4]The reader can get a feeling for how easy it is to detect such distortions by doing a simple experiment. Find a positive lens, preferably of more than 3 diopters. Place the lens in front of one eye and fixate a long vertical contour. While maintaining fixation on the contour, rotate your head back and forth. This situation best mimics the distortion uncalibrated infants would perceive as they make eye movements. You will see the contour bending in one direction when you rotate to the right and in the other direction when rotating to the left. The bending occurs because patches high and low in the visual field flow at different rates than patches in the middle of the field.

theories were proposed as to how recalibration might occur. The theories illuminate how visual information over time could be used to deduce the distorting effects of optical and neural growth and, once the distortion is assessed, how contrast and optical flow perception could, in principle, be recalibrated. More empirical research is needed, of course, to determine if recalibration actually occurs in the hypothesized fashions, but the predictions of the computational theories provide clear directions for such research.

In closing, I discuss two questions that concern these theories. The first question is, How useful are computational theories of visual recalibration? (For a definition and detailed discussion of computational theories, see Marr, 1982.) In answering this question, I contrast computational theories, like the ones proposed here, with more traditional qualitative theories. In my opinion, computational theories are more likely to aid our understanding of visual recalibration for three reasons. (a) These sorts of quantitative theories require that the theorist be very explicit about several aspects of the problem of interest. When developing a theory that could be implemented as a computer program, the theorist is forced to be very explicit about the relevant aspects of the environment, the computations that need to be performed on the environment, the transformations that occur during processing of the environmental information, and more. The explicitness requirement renders those theories easier to evaluate scientifically compared to more qualitative theories. (b) When a computer implementation does not perform properly, one knows immediately that something is wrong with the theory or with the implementation of the theory. If the theory (and implementation) is sufficiently constrained by environmental restrictions and by empirical observations, the sources of difficulty can usually be pinpointed. Of course, without sufficient constraint, the failure of an implementation is not very informative. Thus, computational theories are most likely to be useful scientifically when the visual skill under consideration is understood well enough for several constraints to be imposed. (c) If the computer implementation does work properly, one can frequently make nonobvious predictions and conduct more experiments to see whether the predictions are accurate. An example is the discovery in the optical flow simulation that egomotion and 3D layout cannot generally be computed if the scene consists of only one frontoparallel surface. Rieger and Toet (1985) found that adults also have great difficulty determining the direction of egomotion when the scene consists of only one frontoparallel surface. When nonobvious predictions are borne out in this way, it enhances one's confidence in the applicability of the computational theory.

The second question is, To what extent can we generalize from models of recalibration in contrast perception and optical flow perception to models of the recalibration of other perceptual capabilities? Unfortunately, the answer

appears to be, not much. My reasoning is the following. The development of the theories presented here involved an analysis of several aspects of the two visual capabilities examined. These included physical facts (e.g., how the optics of the eye attenuates different spatial frequencies, or how optical flow fields are determined by translation and rotation) and the properties of visual mechanisms (e.g., spatial frequency channels, or receptive fields sensitive to flow differences). To the extent that the visual functions considered here are representative of other functions, the analyses suggest that an understanding of the recalibration of other visual capabilities will require a similar in-depth analysis of the physics, psychophysics, and physiology involved. In other words, each visual capability that undergoes recalibration during development may require recalibration mechanisms that are designed quite specifically for that capability. If this speculation proves correct, it means that no general theory of developmental recalibration is likely to emerge. Rather a series of theories, each tailored to the particular capability examined, is more likely.

ACKNOWLEDGMENTS

This research was supported by NIH Research Grant HD-19927. I thank Phil Salapatek and Herb Pick for introducing me to the problem of recalibration in their graduate seminars and for very helpful discussions. I am also indebted to Andy Eisman, who patiently and expertly assisted me with the computer simulations.

REFERENCES

Abramov, I., Gordon, J., Hendrickson, A., Hainline, L., Dobson, V., & LaBossiere E. (1982). The retina of the newborn human infant. *Science, 217,* 265–267.

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology, 48,* 217–237.

Aslin, R. N., & Jackson, R. W. (1979). Accommodative-convergence in young infants: Development of a synergistic sensory-motor system. *Canadian Journal of Psychology, 33,* 222–231.

Atkinson, J., Braddick, O., & Moar K. (1977). Development of contrast sensitivity over the first three months of life in the human infant. *Vision Research, 17,* 1037–1044.

Banks, M. S., Geisler, W. S., & Bennett, P. J. (in press). The physical limits of grating visibility. *Vision Research.*

Banks, M. S., & Salapatek, P. (1978). Acuity and contrast sensitivity in 1-, 2-, and 3-month-old human infants. *Investigate Ophthalmology and Visual Science, 17,* 361–365.

Banks, M. S., & Salapatek, P. (1981). Infant pattern vision: A new approach based on the contrast sensitivity function. *Journal of Experimental Child Psychology, 31,* 1–45.

Banks, M. S., Stephens, B. R., & Hartmann, E. E. (1985). The development of basic mechanisms of pattern vision: Spatial frequency channels. *Journal of Experimental Child Psychology, 40,* 501–527.

Bennett, A. G., & Francis, J. L. (1962). Aberrations of optical images. In H. Davson (Eds.), *The eye* (pp. 87–99). New York: Academic Press.
Bertenthal, B., Dunn, S., & Bai, D. (1986). Infants' sensitivity to optical flow for specifying self-motion. *Infant Behavior and Development (Abstract), 9,* 35.
Blakemore, C., Muncey, J. P. J., Ridley, R. M. (1973). Stimulus specificity in the human visual system. *Vision Research, 13,* 1915–1931.
Bonds, A. B., & Freeman, R. D. (1978). Development of optical quality in the kitten eye. *Vision Research, 18,* 391–398.
Bower, T. G. R. (1974). *Development in infancy.* San Francisco: W. H. Freeman.
Braddick, O., Campbell, F. W., & Atkinson, J. (1978). Channels in vision: Basic aspects. In R. Held, H. W. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology. Perception* (pp. 3–38). New York: Springer-Verlag.
Butterworth, G., & Hicks, L. (1977). Visual proprioception and postural stability in infancy: A developmental study. *Perception, 6,* 255–262.
Campbell, F. W., & Gubisch, R. W. (1966). Optical quality of the human eye. *Journal of Physiology, 186,* 558–578.
Carroll, J. J., & Gibson, E. J. (1981, April). *Infants' differentiation of an aperture and an obstacle.* Paper presented at the meeting of the Society for Research in Child Development, Boston.
Clocksin, W. F. (1980). Perception of surface slant and edge labels from optical flow: A computational approach. *Perception, 9,* 253–259.
Cornsweet, T. N. (1970). *Visual perception.* New York: Academic Press.
Fagan, J. F. (1976). Infants' recognition of invariant features of faces. *Child Development, 47,* 627–638.
Fox, R., Aslin, R. N., Shea, S. L., & Dumais, S. T. (1980). Stereopsis in human infants. *Science, 207,* 323–324.
Gaskill, J. D. (1978). *Linear systems, Fourier transforms, and optics.* New York: Wiley.
Gennery, D. B. (1973). Determination of optical transfer function by inspection of frequency-domain plot. *Journal of the Optical Society of America, 63,* 1571–1577.
Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: Deblurring in human vision by spatial frequency channels. *Journal of Physiology, 252,* 627–656.
Gibson, J. J. (1950). *The perception of the visual world.* Boston: Houghton Mifflin.
Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.
Gibson, J. J., Olum, P., & Rosenblatt F. (1955). Parallax and perspective during aircraft landings. *American Journal of Psychology, 68,* 372–385.
Gordon, D. A. (1965). Static and dynamic fields in human space perception. *Journal of the Optical Society of America, 55,* 1296–1303.
Gordon, F. R., & Yonas, A. (1976). Sensitivity to binocular depth information in infants. *Journal of Experimental Child Psychology, 22,* 413–422.
Green, D. G. (1970). Regional variations in the visual acuity for interference fringes on the retina. *Journal of Physiology, 207,* 351–356.
Hardt, M. E., Held, R., & Steinbach, M. J. (1971). Adaptation to displaced vision: A change in the central control of sensorimotor coordination. *Journal of Experimental Psychology, 89,* 229–239.
Harris, C. S. (1980). Insight or out of sight? Two examples of perceptual plasticity in the human adult. In C. S. Harris (Eds.), *Visual coding and adaptability.* Hillsdale, NJ: Lawrence Erlbaum Associates.
Harris, L., Atkinson, J., & Braddick, O. (1976). Visual contrast sensitivity of a 6-month-old infant measured by the evoked potential. *Nature, 264,* 570–571.
Hay, J. C., & Pick, H. L. (1966). Gaze-contingent adaptation and prism orientation. *Journal of Experimental Psychology, 72,* 640–648.
Held, R., & Bossom, J. (1961). Neonatal deprivation and adult rearrangement: complementary

techniques for analyzing plastic sensory-motor coordinations. *Journal of Comparative and Physiological Psychology, 54,* 33–37.
Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually-guided behavior. *Journal of Comparative and Physiological Psychology, 56,* 872–876.
Held, R., & Rekosh, J. (1963). Motor-sensory feedback and the geometry of visual space. *Science, 141,* 722–723.
Hendrickson, A., & Yuodelis, C. (1984). The morphological development of the human fovea. *Ophthalmology, 91,* 603–612.
Hess, R. F., & Bradley, A. (1980). Contrast perception above threshold is only minimally impaired in human amblyopia. *Nature, 287,* 463–464.
Hess, R. F., Bradley, A., & Piotrowski, L. (1983). Contrast-coding in amblyopia. I. Differences in the neural basis of human amblyopia. *Proceedings of the Royal Society of London, 217,* 309–330.
Jenkins, F. A., & White, H. E. (1976). *Fundamentals of optics.* New York: McGraw-Hill.
Johnston, I. R., White, G. R., & Cumming, R. W. (1973). The role of optical expansion patterns in locomotor control. *American Journal of Psychology, 86,* 311–324.
Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15,* 483–524.
Kellman, P. J., Von Hofsten, C., & Soares, J. (1985). Concurrent motion in infant event perception. *SRCD Abstracts, 5,* 183.
Kulikowski, J. J. (1976). Effective contrast constancy and linearity of contrast sensation. *Vision Research, 16,* 1419–1431.
Larsen, J. S. (1971a). The sagittal growth of the eye. I. Ultrasonic measurement of the depth of the anterior chamber from birth to puberty. *Acta Ophthalmologica, 49,* 239–262.
Larsen, J. S. (1971b). The sagittal growth of th eye. III. Ultrasonic measurement of the posterior segment (axial length of the vitreous) from birth to puberty. *Acta Ophthalmologica, 49,* 441–453.
Larsen, J. S. (1971c). The sagittal growth of the eye. IV. Ultrasonic measurement of the axial length of the eye from birth to puberty. *Acta Ophthalmologica, 49,* 873–886.
Lee, D. N. (1980). The optic flow field: The foundation of vision. *Philosophical Transactions of the Royal Society of London B, 290,* 169–179.
Lee, D. N., & Aronson, E. (1974). Visual proprioceptive control of standing in human infants. *Perception and Psychophysics, 15,* 529–532.
Llewellyn, K. R. (1971). Visual guidance of locomotion. *Journal of Experimental Psychology, 91,* 245–261.
Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of moving retinal images. *Proceedings of the Royal Society of London B, 208,* 385–387.
Lotmar, W. (1976). A theoretical model for the eye of new-born infants. *Albrecht von Graefes Archiv fuer klinische experimentelle Opthalmologie, 198,* 179–185.
Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W. H. Freeman.
McKee, S. P., & Nakayama, K. (1984). The detection of motion in the peripheral visual field. *Vision Research, 24,* 25–32.
Miller, W. H., & Bernard, G. D. (1983). Averaging over the foveal receptor aperture curtails aliasing. *Vision Research, 23,* 1365–1370.
Nakayama, K., & Loomis, J. M. (1974). Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis. *Perception, 3,* 63–80.
Norcia, A. M., Tyler, C. W., & Allen, D. (1986). Electrophysiological assessment of contrast sensitivity in human infants. *American Journal of Optometry and Physiological Optics, 63,* 12–15.
Ogle, K. N. (1950). *Researches in binocular vision.* Philadelphia: Saunders.

Pick, H. L., & Hay, J. C. (1964). Adaptation to prismatic distortion. *Psychonomic Science, 1,* 199–200.

Pirchio, M., Spinelli, D., Fiorentini, A., & Maffei, L. (1978). Infant contrast sensitivity evaluated by evoked potentials. *Brain Research, 141,* 179–184.

Regan, D., & Beverley, K. I. (1982). How do we avoid confounding the direction we are looking and the direction we are moving? *Science, 215,* 194–196.

Rieger, J. H., & Toet, L. (1985). Human visual navigation in the presence of 3-D rotations. *Biological Cybernetics, 52,* 377–381.

Rock, I. (1966). *The nature of perceptual adaptation.* New York: Basic Books.

Rogers, B., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research, 22,* 261–270.

Rosinski, R. R. (1977). *The development of visual perception.* Santa Monica: Goodyear.

Schor, C. M., & Flom, M. C. (1969). The relative value of stereopsis as a function of viewing distance. *American Journal of Optometry and Archives of American Academy of Optometry, 46,* 805–809.

Stephens, B. R., & Banks, M. S. (1985). The development of contrast constancy. *Journal of Experimental Child Psychology, 40,* 528–547.

Stratton, G. M. (1897). Vision without inversion of the retinal image. *Psychological Reviews, 4,* 341–360.

Switkes, E., Mayer, M. I., & Sloan, J. A. (1978). Spatial frequency analysis of the visual environment: anisotropy and the carpentered environment hypothesis. *Vision Research, 18,* 1393–1399.

Von Hofsten, C. (1977). Binocular convergence as a determinant of reaching behavior in infancy. *Perception, 6,* 139–144.

von Grunau, M. W., & Frost, B. J. (1983). Double opponent-process mechanism underlying receptive field structure of directionally specific cells of cat lateral suprasylvian visual area. *Experimental Brain Research, 49,* 84–92.

Wallach, H., Moore, M. E., & Davidson, L. (1963). Modification of stereoscopic depth-perception. *American Journal of Psychology, 76,* 191–204.

Warren, R. (1976). The perception of egomotion. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 448–456.

Williams, R., & Boothe, R. (1981). Development of optical quality in the infant monkey (Macaca nemestrina) eye. *Investigative Ophthalmology and Visual Science, 21,* 728–736.

Yonas, A., Bechtold, A. G., Frankel, D., Gordon, F. R., McRoberts, G., Norcia, A., & Sternfel, S. (1977). Development of sensitivity to information for impending collision. *Perception & Psychophysics, 21,* 97–104.

Yonas, A., Pettersen, L., & Lockman, J. (1979). Young infants' sensitivity to optical information for collision. *Canadian Journal of Psychology, 33,* 268–276.

Yonas, A., Pettersen, L., & Lockman, J. (1980, March). *The perception of impending collision in 3-month-old infants.* Paper presented at the International Conference of Infant Studies New Haven.

Yuodelis, C., & Hendrickson, A. (1986). A qualitative and quantitative analysis of the human fovea during development. *Vision Research, 26,* 847–855.