Chapter 4

Properties of early motion processing: Implications for the sensing of egomotion

Ken Nakayama

Smith-Kettlewell Eye Research Foundation
San Francisco, California

4.1 Introduction

J.J. Gibson's seminal insight regarding the pick-up of higher-order invariants in the optic-array was outlined by 1950, one decade before the recent wave of neurophysiological advances which began in the early 1960's. Gibson stressed motion, especially the global field of velocities, as particularly important, both for the delineation of surface properties (Gibson, 1950) and for the perception of self-motion (Gibson, 1958). Electrophysiological recordings from cortical area MT in primate reveal cells which are sensitive to differences in velocity between adjacent portions of the visual field (Allman, 1985). Such mechanisms are likely to encode just those variables that Gibson understood as important for vision: velocity differences rather than absolute values of velocity (see Nakayama & Loomis, 1974; Frost &

Nakayama, 1983). Thus Gibson's insight was perhaps prophetic, yet in ways not personally anticipated, since his own outlook eschewed a concern for the necessary details of neural computation.

In this paper we concentrate more on these biological details of information pick-up, asking what specific range of velocities are available and utilized for the sensing of egomotion. Gibson's theory of ecological optics suggests only how we might determine what is potentially available in the optic array. To understand the biological registration of egomotion, however, we need to ask what is actually available, which subset of potential information is used and which subset is not.

As an analogy, we note that the polarization of visible electromagnetic energy contains reliable information as to the position of the sun in an otherwise overcast sky. This property of the optic array is useful for insects in their navigation (Waterman, 1966). Mammals, such as ourselves, however, have no known mechanisms available to distinguish one plane of polarization from another. Consequently, we are unable to pick-up and use large amounts of information that could be useful for ego-orientation and egomotion.

In a similar vein, I will argue that the limitations imposed by early visual processing, perhaps the properties of the photoreceptors themselves, largely dictate the range of velocities that can be extracted for calculating egomotion. A wider understanding of these specific limitations could have two beneficial effects: (1) It could advance our understanding of the details of egomotion computation and (2) it may suggest some considerations for the design of computer driven visual simulators for research and training.

4.2 Variation of spatial resolution across the visual field

If we compare the eye to man-made imaging devices, we see an important difference. In contrast to a TV or regular camera, the eye (and in particular the retina) is non-uniform with its specialized fovea and periphery. Yet despite the profound difference between these regions, we should be careful not to over-dichotomize the fovea-periphery distinction. The variation between central and peripheral vision is a smoothly continuous difference in quantity rather than any abrupt break in quality. No sharp functional dividing line between fovea and periphery is discernible (Weymouth, 1958). The bar width of the finest sinusoidal grating that can be resolved rises continuously as a linear function of eccentricity and this increase is mirrored by a proportional increase in receptive field size in primate striate cortex (Hubel & Wiesel, 1974). Aside from showing the smooth continuity from fovea to periphery, the neurophysiological data suggest that grating resolution is likely to be mediated by the receptive field size of cortical neurons.

How can these psychophysical-neurophysiological correlations regarding visual acuity have any bearing on the perception of egomotion, especially since this capacity is probably based on specialized motion sensors not directly related to detailed vision? Visual resolution, however, is important because it provides for each retinal eccentricity, a rough and simple index as to the spatial scale of processing.

A more complete description of this spatial scale of processing is the contrast sensitivity function as measured by sinusoidal grating stimuli. Contrast sensitivity refers to the reciprocal of threshold contrast for a given set of spatial frequencies. To measure contrast sensitivity using sinewave gratings at different retinal eccentricities, however, requires some consideration of the retino-cortical projection because increasing the retinal eccentricity of a grating pattern will excite less and less cortical tissue. If one takes this cortical magnification function into account and thus scales the size of the grating stimulus to increase approximately in proportion to the linear cortical magnification function, the similarity of processing for all retinal regions emerges. Virsu and Rovamo (1979) have demonstrated that the contrast sensitivity functions of all regions of retina begin to look very similar, each have the same inverted U function and are simply scaled downward in spatial frequency for the more eccentric positions. Even more remarkable is the fact that, by employing the single scaling factor of cortical magnification, these functions can be collapsed into a single function relating contrast sensitivity to cortical spatial frequency (i.e., sinusoidal cycles per millimeter distance on the cortical surface). See Figure 4.1a and 4.1b, adapted from papers of Virsu and Rovamo (Virsu & Rovamo, 1979; Rovamo & Virsu, 1979). This striking superimposition of functions emphasizes the continuity of processing across the visual field and suggests that visual processing across different regions of cortex are basically the same (again see also Hubel & Wiesel, 1974).

4.3 Temporal factors

We've mentioned the spatial grain of the system and how it varies with retinal eccentricity. What can be said for the temporal sensitivity of the visual system as we move across different portions of the visual field? Temporal frequency sensitivity can also be measured using sinusoids by varying the temporal frequency of a sinusoidally varying light source and measuring the

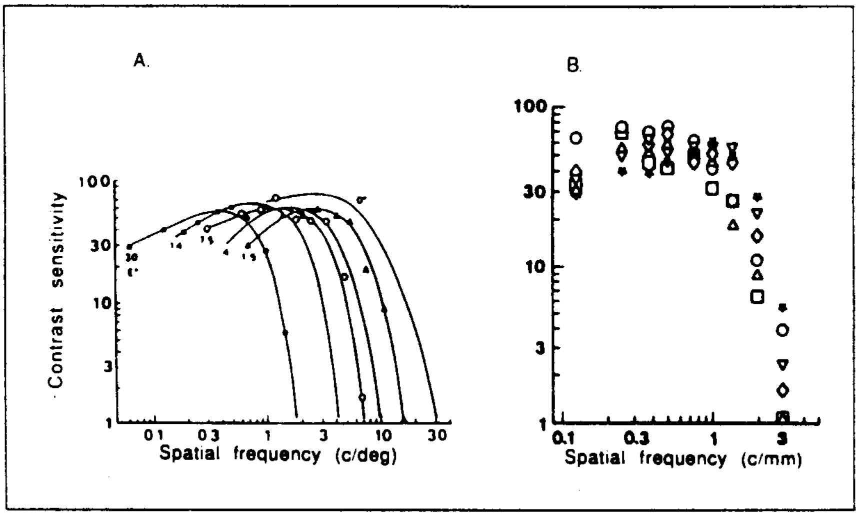

Figure 4.1: A: Contrast sensitivity functions for different retinal eccentricities in the inferior visual field. Observer's task was to discriminate whether the gratings moved to the left or to the right for a .5 s presentation. Size of the target was scaled in accord with the cortical magnification function. Note that with the exception of the translation along the spatial frequency axis, the contrast sensitivity functions for the different regions are very similar (From Figure 3 of Virsu & Rovamo, 1979). B: Same data except the abscissa has been scaled to the cortical magnification function such that spatial frequency refers to c/mm of cortical tissue (From Figure 4 of Rovamo & Virsu, 1979).

threshold amplitude to just perceive flicker (de Lange, 1958). Using this technique, Tyler (1985) demonstrated that temporal frequency sensitivity does change significantly as one moves from central to peripheral regions. He scaled the spatial extent of flickering targets so as to conform to the cortical magnification function and found that there was a 2:1 difference in critical flicker fusion between more central regions and the far periphery with the periphery having the highest temporal sensitivity. Correspondingly, however, the periphery is poorly suited to encode slower temporal fluctuations and its modulation sensitivity to low temporal frequencies was reduced. Tyler hypothesised that this difference might be on the receptors themselves, perhaps related to the well known anatomical differences between foveal and peripheral cones.

So, significant differences do exist between the fovea and the periphery in terms of temporal frequency sensitivity and we will return to this later in our discussions regarding receptor limitations in motion encoding. In comparison to the nearly 35-fold difference in spatial resolution across the retina, this seemingly small 2:1 ratio may appear insignificant. Yet it does contribute to some prominent characteristics in the processing of slow and fast velocities and we shall also argue that it offers a clue when looking for the critical link in the chain of visual processing that limits the pickup of different velocities.

4.4 Spatio-temporal contrast sensitivity functions

So far we have dealt with the differences in spatial and temporal frequency sensitivities separately, considering the visibility of static gratings and homogenous fields of flicker. Now we must link these two forms of sensitivity by seeing how the system responds to moving stimuli which vary in both spatial and temporal frequency. To examine the joint sensitivity to both time and space, we discuss experiments which summarize the ability of an observer to detect, to judge the direction, and to estimate the velocity of motion in sinusoidal gratings. By temporal frequency of a moving sinusoidal grating we refer to the number of sinusoidal cycles of luminance variation per second that can be registered at any given point. It should be clear that the temporal frequency of luminance modulation at a point (say for the locus of a given photoreceptor) is a joint function of velocity and spatial frequency.

$$\text{Temporal Frequency} = \text{Spatial Frequency} \times \text{Velocity} \qquad (4.1)$$

To address this issue of spatio-temporal sensitivity, we consider studies where both spatial and temporal frequency are varied and a spatio-temporal modulation sensitivity function is obtained. A reasonably complete examination of this issue has been conducted in several laboratory groups over the years. We first mention the work of Kelly (1979) who made such measurements under conditions of stabilized vision. The results are shown in Figure 4.2 where separate contrast sensitivity functions were obtained at various velocities. Most striking is the fact that for each velocity there is a different range of spatial frequencies which are visible. At very low velocities, say .0016°/s, it is only the very highest range of spatial frequencies that are visible. Most of the range of low spatial frequencies is invisible. Conversely for high velocities, it is only the very lowest spatial frequencies which are visible. Consider the leftmost function which was taken at 32°/s. Here the contrast sensitivity function is shifted to very low spatial frequencies; all spatial frequency components over .5 c/deg are invisible. If we examine the

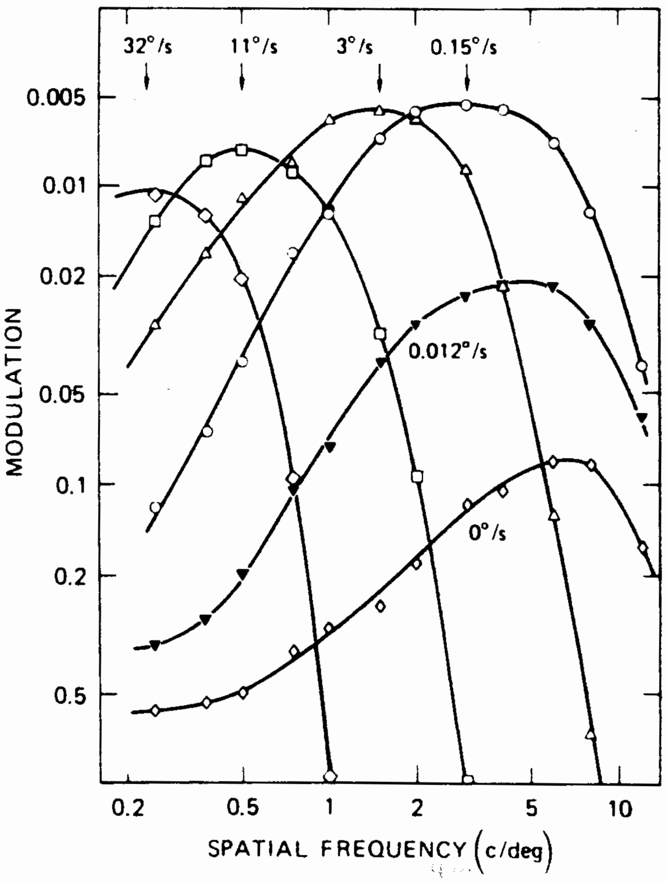

Figure 4.2: Contrast sensitivity for the detection of sinusoidally drifting gratings. Each curve represents the contrast sensitivity (reciprocal of threshold contrast) plotted against spatial frequency for a given velocity. From Kelly (1979). Note, for a range of velocities, that the shift is inversely proportional to the velocity of the drifting grating.

3, 11, and 32°/s curves, which are separated by about a 3:1 ratio, it is clear that their spatial frequency peaks also shift by the same ratio. This implies that what remains constant, at least for this approximately 10:1 range, is temporal frequency tuning.

These experiments used the detection of the grating as the observer's task and so the results only tell us about the visibility of sinusoidal stimulus components moving at given velocities. But they also generalize to the registration of direction.

Burr and Ross (1982) conducted an experiment very much like the one just described by Kelly except they asked the observers to discriminate the direction of motion, not whether the grating was visible. They conducted this study over a very wide range of velocities, from 1°/s to over 800°/s. The main results are seen in Figure 4.3 which extend and mirror Kelly's results almost exactly. They also underline the fundamental importance of temporal frequency since all of the velocity curves (with the exception of the slowest velocities) superimpose on a temporal frequency axis.

The superimposition of these temporal frequency curves provides strong evidence that the limitations of visibility and the limitations of seeing motion direction are largely the result of an underlying limitation in temporal frequency sensitivity. Extrapolating from Tyler's work on the temporal dynamics of central and peripheral retina, we hypothesize that it is the photoreceptors themselves which set this temporal frequency limit.

We suggest that the deviation of the lowest velocity curves in the otherwise superimposed set of functions provides additional circumstantial evidence to support this view. Our argument runs as follows. Points on this lowest velocity curve have the highest spatial frequency (from Eq. 4.1). Since such high spatial frequencies are best processed in the fovea, one would expect that if foveal cones had slower temporal dynamics, the lowest velocity curve would be tuned to lower temporal frequencies, which is indeed the case. The lowest velocity curve (1°/s) deviates significantly from the other curves in terms of temporal frequency sensitivity (see Figure 4.3). Thus we think it reasonable to suppose that these velocity sensitive limitations are dictated very early, probably by the photoreceptors themselves.

4.5 Some examples

Whatever the ultimate biological limit, the spatial and temporal contrast sensitivity functions would appear to go a long way in explaining what aspects of the stimulus will become visible at what velocity and why. We consider two extreme examples to illustrate the point. First, is a large target which is moving slowly across the retina. Ordinarily one might think this is

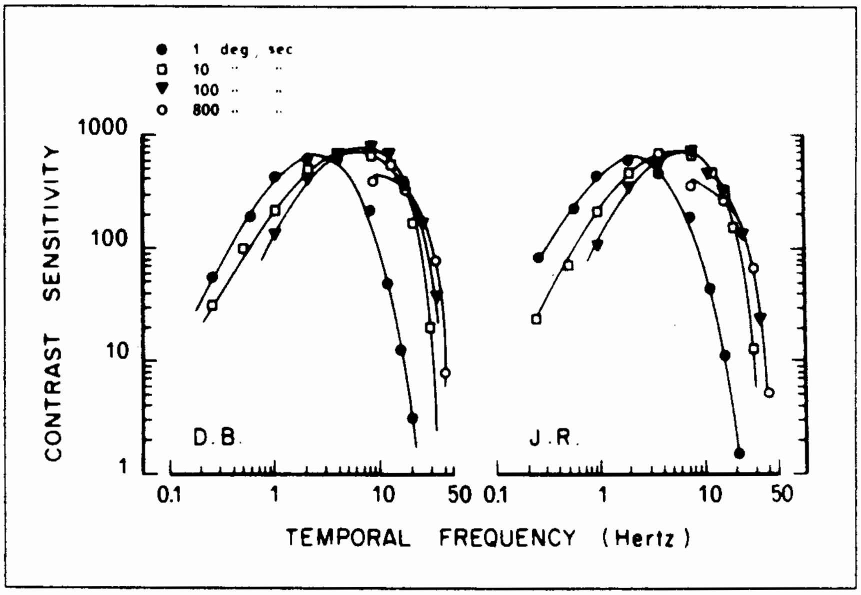

Figure 4.3: Contrast sensitivity for the detection of motion direction of sinusoidally drifting gratings. This data has the same form as that obtained by Kelly (see Figure 4.2) but here the sensitivity is plotted as a function of temporal frequency for different velocity curves. Note that with the exception of the slowest velocity (see text for explanation) all curves essentially superimpose.

a large target so that its motion would be sensed by low spatial frequency mechanisms. Not so. The temporal frequencies corresponding to the very lowest spatial frequency components (via Eq. 4.1) will be too low to elicit a reliable motion signal. Thus to see very slow motion of either big or small targets requires motion mechanisms which have high spatial resolution.

At the other extreme, consider the appearance of a more rapidly moving stimulus, say it is moving across the retina at 32°/s. If we consider the leftmost curve in Figure 4.2 we see that at this velocity, any spatial frequency component above .5 c/deg must be invisible, and the optimal spatial frequency for this velocity is .1 c/deg. Thus a rather small detailed object will either be not seen at high velocities or not identifiable. This conforms to our everyday subjective experience of moving objects. They lack spatial detail.

Insofar as egomotion would seem to require the measurement of velocity to compute key parameters, such as the focus of expansion, we also need to ask how the visual system is able to discriminate the differences

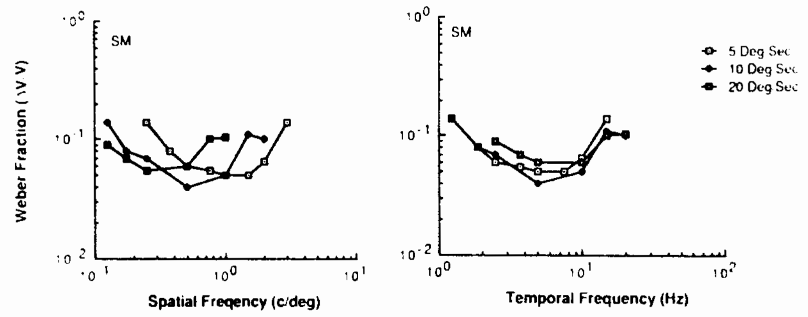

Figure 4.4: A: Proportionate fractions of a base velocity that can be discriminated (velocity Weber fraction) as a function of spatial frequency for 3 different velocities (adapted from Figure 3 of McKee et al., 1986). Note that the best spatial frequency for velocity discrimination shifts downward as velocity is increased. B: Same data plotted as a function of temporal frequency. Note that the curves essentially superimpose indicating that it is the temporal frequency of the stimulus that determines velocity magnitude discriminability (adapted from Figure 4 of McKee et al., 1986).

in velocity of moving images. McKee, Silverman and Nakayama (1986) have made the following observations which are relevant. They measured an observer's ability to see differences in velocity as a function of spatial and temporal frequency of sinusoidal gratings. Figure 4.4a shows such ΔV thresholds as a function of spatial frequency for several velocities. As might be expected from the data already shown by Kelly, Burr and Ross, the best spatial frequency for velocity discrimination shifts downward as velocity is increased. Replotting the velocity discrimination curves as a function of temporal frequency (see Figure 4.4b) also confirms the point. Plotted in these coordinates, the curves essentially superimpose.

The similarity of behavior for pattern visibility (Figure 4.2), discrimination of velocity direction (Figure 4.3) and velocity magnitude (Figure 4.4) again point to our common theme. There exists a set of temporal frequency limits which dictate the encoding of moving visual patterns. If one presents a target outside this temporal limit, information regarding its visibility, shape, and velocity will be highly degraded.

4.6 Considerations for the design of visual simulators

What implications, if any, does the foregoing discussion have for the efficient use of computational resources in the design of visual simulators? Insofar as a major problem in the design of such simulators is the very large amount of information that might be displayed to the observer, it would be an obvious advantage to remove information from the display which is not utilized by the visual system. Here we consider the visibility of various pattern components at different velocities. The results indicate that very different spatial frequency components of the image are visible at very low and very high velocities. If the image is moving very fast, then only the most coarse information is required. Conversely, when the image is moving very slow, only the highest spatial frequency information is required.

This means that some considerable computational savings might be realized when presenting images at different velocities because much information in the image is invisible to the observer and requires no presentation. This is most evident for high velocity images where only the lowest spatial frequency content of the image need be displayed. A second point regards the use of raster graphics systems to display velocity information of high spatial frequency components. Again, the low upper cutoff in temporal frequency sensitivity mentioned earlier, indicates that motion information provided by high spatial frequency mechanisms is limited to the lowest velocities. This means that high spatial frequency information is essentially wasted when presented to the observer at high velocities for it will have no effect in influencing velocity sensitivity. Worse, however, is the possibility that such information will interact spuriously with the frame rate of the system to create the "wagon wheel effect," the perception of reversed motion in the image. This possibility, known as spatio-temporal aliasing, injects spurious and very visible spatio-temporal frequency components into the display (see also Nakayama, 1985; Watson, Ahumada, & Farrell, 1986). From this discussion it should be clear that low pass spatial frequency filtering of rapidly moving targets could be highly advantageous, both to save on bandwidth resources and to eliminate spurious motion.[1]

---

[1] At this point we should also make a cautionary remark. The application of low pass spatial frequency filtering may only be useful for those portions of the image that are moving rapidly. As we mentioned earlier, it is important to emphasize that low pass spatial frequency filtering of the slowly moving portions of the image will have two deleterious effects. Obviously, it will blur the image and reduce the appreciation of spatial detail. Less obvious, but a consequence of the relations discussed in this paper, it will also degrade the registration of the

4.7 Summary and conclusions

The major point is of this paper is that much of the information imaged on the retina of a moving observer is not registered by the early visual system. Aside from the retino-cortical magnification function which ensures that fine detail is not registered in the peripheral parts of the visual field, high velocities in the image also preclude the pickup of detailed spatial information. We hypothesize that this loss of high spatial frequencies at high velocities is due to an early limit in the temporal dynamics of the visual system, probably in the photoreceptors. Because of these temporal frequency limitations, it is apparent that any system based on the pickup of information in moving images need only sample from a restricted set of early visual mechanisms. Thus, for the pick-up of information regarding one's egomotion, high velocity information need only have very low spatial frequency content. From the perspective of simulator design, it remains to be seen whether hardware and software design can be built to take advantage of these inherent limitations of human vision.

Acknowledgments

This work was supported by grants EY-05408 and EY-01186 from the National Institutes of Health.

4.8 References

Allman, J., Miezin, F., & McGuiness, E. (1985). Direction- and velocity specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception, 14*, 105-126.

Burr, D.C., & Ross, J. (1982). Contrast sensitivity at high velocities. *Vision Research, 22*, 479-484.

DeLange, H. (1958). Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light I. Attenuation characteristics with white and colored light. *Journal of the Optical Society of America*, 48, 777-784.

Frost, B.J., & Nakayama, K. (1983). Single visual neurons code opposing motion independent of direction. *Science, 220*, 744-745.

Gibson, J.J. (1950). *The perception of the visual world.* Boston: Houghton Mifflin.

Gibson, J.J. (1958). Visually controlled locomotion and visual orientation in animals. *British Journal of Psychology, 49*, 182-194.

---

slowest velocities.

Hubel, D.H., & Wiesel, T.N. (1974). Uniformity of monkey striate cortex: A parallel relationship between field size, scatter and magnification factor. *Journal of Comparative Neurology, 158*, 295–305.

Kelly, D.H. (1979). Motion and vision. II. Stabilised spatio-temporal threshold surface. *Journal of the Optical Society of America, 69*, 1340–1349.

McKee, S.P., Silverman, G.H., & Nakayama, K. (1986). Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Research, 26*, 609–619.

Nakayama, K. (1985). Biological image motion processing: A review. *Vision Research, 25*, 625–660.

Nakayama, K., & Loomis, J. (1974). Optical velocity patterns, velocity sensitive neurons, and space perception: a hypothesis. *Perception, 3*, 63–80.

Rovamo, J., & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research, 37*, 479–510.

Tyler, C.W. (1985). Analysis of visual modulation sensitivity. II. Peripheral retina and the role of photoreceptor dimensions. *Journal of the Optical Society of America, A2*, 393–398.

Virsu, V., & Rovamo, J. (1979). Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research, 37*, 475–494.

Waterman, T.H. (1966). Systems analysis and the visual orientation of animals. *American Scientist, 54*, 15–45.

Watson, A.B., Ahumada, A., & Farrell, J.E. (1986). The window of visibility: A psychophysical theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America, A3*, 300–307.

Weymouth, F.A. (1958). Visual sensory units and the minimal angle of resolution. *American Journal of Ophthalmology, 46*, 102–113.