Watson, A. B. & Eckert, M. P. (1994). Motion-Contrast Sensitivity: Visibility of Motion Gradients of Various Spatial Frequencies. *Journal of the Optical Society of America A 11*(2), 496-505.

Motion-Contrast Sensitivity: Visibility of Motion Gradients of Various Spatial Frequencies

Andrew B. Watson

Michael P. Eckert

NASA Ames Research Center

Moffett Field, CA 94035-1000

Abstract

The purpose of these experiments was to estimate basic sensitivity to motion gradients, and to evaluate the evidence for second-order integration and differentiation of motion signals. We measured sensitivity to spatially sinusoidal contrast modulation between two oppositely-moving bandpass-filtered noise images. The *motion-contrast sensitivity function*, defined as the inverse of threshold modulation amplitude as a function of modulation spatial frequency, was band-pass in shape with declines at both highest and lowest frequencies. The functions for three noise spatial frequencies were approximately the same shape when modulation frequency was expressed as a fraction of the noise frequency. We compared the data to a model in which linear motion filters, whose outputs are squared or rectified, are followed by a second stage of excitatory and/or inhibitory pooling. The data are consistent with a model in which 1) all excitatory pooling occurs at the linear stage, and 2) the second stage contains a large inhibitory pooling area, with a radius about 8 times that of the linear receptive field.

Introduction

There now exists a plausible and rigorous model for the sensing of motion at the earliest levels of the human visual system. This *filter model* consists of linear, direction selective filters that are tuned for spatial frequency [1,2]. In one variant, the outputs of two such filters in quadrature phase are squared and added to compute a "motion energy"[3,4]. In another variant, the energies for opposite directions are subtracted, to form an opponent motion signal[5,6]. The linear filter in each of these models has a corresponding receptive field, with a size that is related to the filter spatial bandwidth. As in other "multiple channel" models, the receptive fields are thought to come in a range of sizes, with a corresponding range of spatial frequency optima. It is assumed that these receptive fields cover the visual field, yielding for each size an array of responses distributed over space and time.

While this model can account for many aspects of the detection of motion and the sensing of direction of motion [7,8,9,10,11,12,13], it does not deal directly with the sensing of spatial motion gradients. Sensitivity to motion gradients is of interest for many reasons. First, they represent the next higher order elaboration of the motion field beyond a simple uniform translation, and are a ubiquitous component of our visual experience. Second, motion gradients and discontinuities are important cues in defining the boundaries of objects in motion. Third, motion gradients are fundamental for sensing three-dimensional self-motion as well as the three-dimensional motion and depth of objects and surfaces [14,15,16,17]. These important roles make the study of motion gradients of interest in its own right, but they also suggest that there might exist special mechanisms for detection and estimation of spatial motion gradients.

By analogy to the processing of luminance, we might expect such mechanisms to employ basic operations such as spatial integration and differentiation of the outputs of the first order filters. And also by analogy to luminance, the nature of these operations might be revealed by measurement of a *motion-contrast sensitivity function*. Consider an abstract stimulus as depicted in Fig. 1.

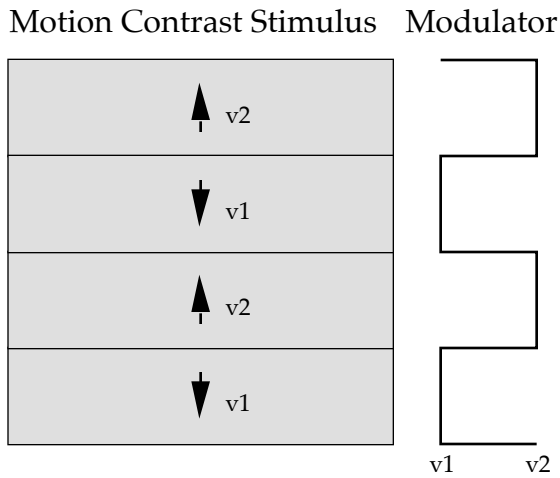

Figure 1. A generic motion-contrast grating consisting of alternating strips moving with different velocities $\mathbf{v}_1$ and $\mathbf{v}_2$. The modulator shown on the right controls the velocity at each point in the stimulus.

It shows an image divided into alternating stripes of two particular velocities. The velocities are carried by some spatial contrast pattern, perhaps noise, which we will call the carrier. At each point in space, the velocity of the carrier is defined by a *modulator*, the function depicted to the right of the image. If there is a mechanism that pools filter outputs over a large area, then it will become insensitive as the stripes become narrow, because it will receive equal input from both velocities. Thus the decline in sensitivity with increasing modulator frequency is a measure of the pooling or integrating behavior of higher-level mechanisms. This is a close analog to the use of luminance contrast sensitivity functions to define the size of luminance-contrast receptive fields. For this reason, we call such generic functions motion-contrast sensitivity functions. To pursue the analogy further, if the higher-level mechanism spatially differentiates the outputs of first order filters, we would expect sensitivity to decline at the broadest stripe widths (lowest modulating frequencies).

An early use of a periodic motion-contrast stimulus was that of van Doorn and Koenderink [18]. They used alternating stripes of spatial white noise moving with respective velocities $\mathbf{v}_1$ and $\mathbf{v}_2$, veiled by some amount of uncorrelated dynamic white noise. When $\mathbf{v}_1$ and $\mathbf{v}_2$ were opposite to each other and orthogonal to the border (compression), sensitivity declined systematically at higher modulating frequencies, typically falling from its peak by a factor of four at around 3 cycles/deg.

The van Doorn and Koenderink stimulus may be thought of as one in which there are two noise images, with velocities $\mathbf{v}_1$ and $\mathbf{v}_2$, whose

contrasts are determined by the modulator. An alternative approach taken by Nakayama and Tyler[19] is to use a modulator that determines the *velocity* of individual random dots. Using a sinusoidal modulator, they measured modulation amplitude detection thresholds for motion parallel to the border (shear) at various modulation frequencies. Here also, sensitivity declined with increasing modulation frequency, again falling from its peak by a factor of four at about 3 cycles/deg. Later experiments with compressive motion showed a much slower decline[20], a point to which we will return in the discussion.

While quite different in detail, both these experiments may tell us something about pooling of motion signals. In particular, both appear to suggest pooling over an area of roughly 1/3 degree. But how can we tell whether this is pooling at a higher level, or pooling within the linear early filter mechanisms? Fig. 2 illustrates two possible sites of spatial pooling. Early linear mechanisms, with some spatial extent, feed a later stage of nonlinear pooling, with a larger extent. The later pooling must be nonlinear, or it could not be distinguished from the early linear pooling, and must be of larger extent, since it incorporates the pooling from the earlier level.

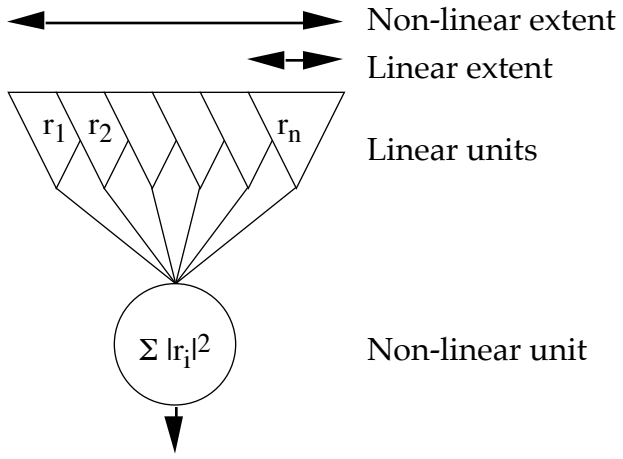

Figure 2. Two types of pooling of motion signals. Early linear units pool over a small extent, later non-linear units pool over a larger extent.

The experiments described above cannot distinguish between these two types of pooling. The theory of early linear filters assumes a bank of filters differing in preferred spatial frequency, and with receptive field sizes inversely proportional to frequency. Because the previous experiments used broadband white noise or random dots, they stimulated the full range of frequency-tuned mechanisms, and hence we cannot say

which linear mechanism, with which receptive field size, was responsible for the psychophysical judgment. Here we propose a method that will allow us to select which spatially tuned filter is operative. This will allow us to compare the spatial extent of pooling inferred from motion-contrast sensitivity measurements to the spatial extent of the putative linear filter. This in turn will reveal the extent of any subsequent non-linear pooling.

The key to our technique is the use of bandpass filtered spatial noise as the carrier. By setting the peak frequency of this noise to a particular value, we target a particular size of motion filter. The first step in the creation of our stimuli is to generate two such bandpass noise images, which we will call c_1 and c_2. The second step is to set both images in motion, with respective velocities $\mathbf{v}_1$ and $\mathbf{v}_2$. In all of the experiments reported here the two noise fields move with equal speed in opposite directions, ($\mathbf{v}_2$=-$\mathbf{v}_1$.). The third step is to create a pair of modulators, m_1 and m_2. These are one-dimensional functions that will multiply the contrasts of c_1 and c_2, respectively, along their vertical dimensions. It is important to note that while the noise fields c_1 and c_2 move, the modulators m_1 and m_2 are stationary (with occasional noted exceptions). Figure 3A shows examples of sinusoidal modulators at motion-contrasts of 1 and 0.5. The sinusoidal modulator has the virtue that, unlike the square wave used by Koenderink and van Doorn, it will not distort the spatial frequency of the noise pattern until relatively high modulation frequencies are reached. After multiplication by the modulator, c_1 and c_2 are added together, their sum is multiplied by a Gaussian temporal window $w(t)$ to ensure gradual onset and offset, and the resulting contrast image itself modulates the luminance of the display. The typical appearance at high motion-contrast is of alternating horizontal stripes of oppositely moving spatial noise. At zero motion-contrast, one sees either a form of dynamic spatial noise, or perhaps transparent sliding of two otherwise unstructured superimposed noise images. The task of the observer is to distinguish between a stimulus with zero motion-contrast and one with non-zero motion-contrast.

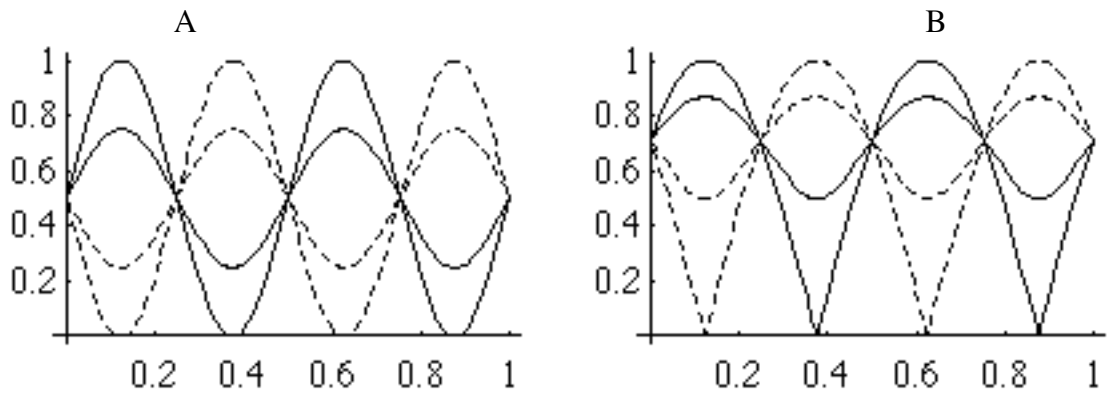

Figure 3. Modulating functions m_1 (solid lines) and m_2 (dashed lines). Motion-contrasts of 1 and 0.5 are shown. Modulation frequency is 2 cycles/image. A) constant peak contrast; B) constant variance.

These two modulators in Fig. 3A are designed to keep the peak luminance contrast constant ($m_1 + m_2 = 1$) everywhere in the image, regardless of the motion-contrast. This is intended to force the observer to use motion gradients, rather than gradients in the luminance contrast, which are zero everywhere. However, if we recall that each pixel in c_1 and c_2 is drawn from a probability distribution with some variance σ^2, then we realize that a pixel in the composite, with a value $m_1 c_1 + m_2 c_2$, will have a variance $\sigma^2(m_1^2 + m_2^2)$. If perceived contrast depends more nearly upon local contrast variance than upon local peak contrast, then the modulators in 3A will produce perceptible stripes in static frames of the stimulus. In fact, such stripes are visible. For this reason, we used the modulators shown in Fig. 3B, which are simply the square roots of those shown in 3A. It is evident that these will keep the local contrast variance constant, and indeed no stripes were visible in individual static frames of the stimuli. Note that when the motion-contrast is below about 0.5, sine and square root of sine are quite similar, so for most purposes the modulators may be thought of as sinusoidal, and we will describe them as such below.

Stimuli constructed with the modulators in Fig. 3 have four parameters of particular interest: the carrier frequency f_c, equal to the peak radial spatial frequency of the isotropic bandpass filtered noise; $\mathbf{v}_1$, the velocity of one noise image (the other was always $v_2 = -v_1$); the modulation frequency, f_m; and the modulation amplitude, or motion-contrast m. Note that the modulators were always vertical (horizontal stripes). As noted above, the task of the observer is to distinguish between zero and non-zero motion-contrast stimuli. With all other parameters fixed, measuring such thresholds as a function of the modulation frequency will define a motion-contrast sensitivity function. In the following experiments,

we collected such functions for a range of velocities and carrier frequencies.

Stimuli

To create samples of isotropic spatial noise with a particular radial spatial frequency bandwidth we defined a "Gaussian ring" filter, given by the convolution in the frequency domain of a Gaussian of a particular spatial scale and a ring impulse function of a certain diameter. In the space domain, the kernel of this filter is the product of a Gaussian and a Bessel function:

$$g(\mathbf{x}) = \mathrm{J}_0\left[2\pi f_c|\mathbf{x}|\right]\exp\left[-\pi\left(|\mathbf{x}|/s\right)^2\right] \tag{1}$$

where $\mathbf{x}$ (spatial coordinate), and s (scale) are in degrees, and f_c is in cycles/degree. We used a scale sufficient to yield a one-dimensional half-amplitude full bandwidth of one octave, as given by the formula

$$s = f_c^{-1}\frac{2^b+1}{2^b-1}\sqrt{Log(2)/\pi} \tag{2}$$

where b is the bandwidth in octaves. We used a value of b =1 octave.

A Discrete Fourier Transform (DFT) of the kernel was used to filter the noise samples. The noise samples were drawn from a uniform distribution with a range of {-1,1}.

The overall luminance contrast of the stimuli was controlled by a Gaussian temporal window:

$$w(t) = c\,g\exp\left(-\pi\left(t/d\right)^2\right) \tag{3}$$

where c is the peak luminance contrast of the stimulus, and g is a constant equal to the inverse of half the largest magnitude in either of the noise images. Unless otherwise noted, the time scale was d= 267 msec (16 frames). The complete stimulus was 533 msec (32 frames).

The spatial modulation of motion was accomplished by a pair of modulating functions:

$$m_1(y) = \left\{\frac{1}{2}\left[1 + m\sin\left(2\pi\,\mathrm{f}_m y\right)\right]\right\}^{1/2} \tag{4}$$

$$m_2(y) = \left\{\frac{1}{2}\left[1 - m\sin\left(2\pi\,\mathrm{f}_m y\right)\right]\right\}^{1/2} = \left\{1 - m_1^2(y)\right\}^{1/2} \tag{5}$$

where f_m is the modulation frequency in cycles/deg, m is the motion-contrast in the range {0,1}, and y is vertical position within the image in degrees. For the special case of $f_m = 0$, the sine was replaced with a constant value of 1. If we write c_1 and c_2 for the two noise images, then the complete stimulus can be written

$$L(\mathbf{x},t) = L_0\left\{1 + w_t(t)\left[m_1(y)c_1(\mathbf{x} - t\mathbf{v}) + m_2(y)c_2(\mathbf{x} + t\mathbf{v})\right]\right\} \tag{6}$$

where L_0 is the mean luminance, $\mathbf{v}$ is the velocity of image motion, and where the noise images wrap-around (the coordinates in $\mathbf{x}$ are interpreted modulo the height and width of the image).

Two successive frames from one stimulus are presented as a stereo pair in Fig. 4. For those able to free-fuse, this should appear as a sinusoidal corrugation in depth.

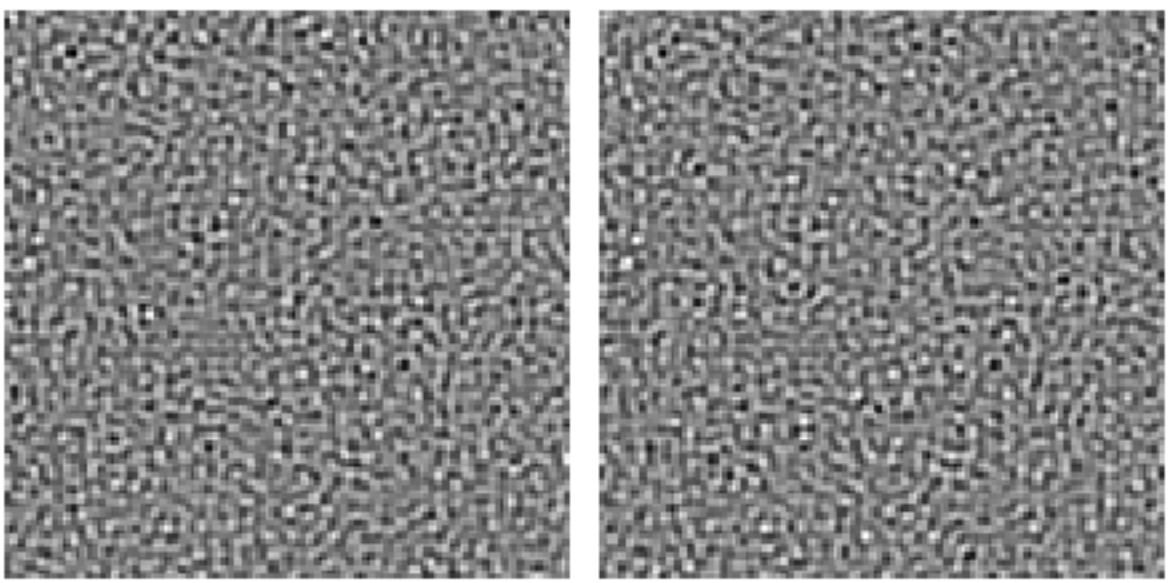

Figure 4. Two successive frames from a stimulus sequence. The modulation frequency f_m was 2 cycles/image, and the motion was horizontal (shear) at $|\mathbf{v}_1| = 1$ pixel/frame. The carrier frequency f_c was 32 cycles/image, the image width is 128 pixels.

To preserve the spectral purity of the stimuli, the modulator frequency must be appreciably lower than the carrier frequency. With our one octave spatial bandwidth, the distortion products (at $f_c - f_m$ and $f_c + f_m$) are negligible provided that $f_m / f_c < 1/2$, which we ensured.

The stimuli were computed in advance as brief movies of 16 frames duration with each frame 128^2 pixels in size. Each frame could be displayed a number of times (dt) on our 60 Hz display. For most of our data, we used dt=2 (30 Hz). This update rate was selected as the best compromise between memory requirements and temporal aliasing. With the spatial and temporal parameters used, little temporal aliasing was expected or observed.

The stimuli were stored in the image memory of a PIXAR II display system. The 10 bit color look-up tables were used to linearize the display[21] (software effectively reduced the size of the look-up table to 512 entries). The stimuli subtended 4 degrees in the center of an otherwise dark screen, and were viewed binocularly with natural pupils from a distance of 48.6 cm (display resolution was 37.6 pixels/cm). Between trials the screen was kept at the mean luminance of 40 cd/m^2. A small dark central fixation point was present at all times.

Procedures

Each 2AFC trial consisted of two temporal intervals each containing a stimulus presentation; in one, the motion-contrast was 0, in the other it was non-zero. We call these null and signal, respectively. The observer attempted to select the interval containing the signal. A QUEST staircase [22] of 64 trials, and subsequent fitting by a Weibull function [23] were used to find the motion-contrast yielding 82% correct. Sensitivity is defined as the inverse of this motion-contrast threshold. Typically three replications were obtained for each threshold. Two observers (the authors) took part. The direction of motion was reversed on a random half of the trials. This means that for a compression stimulus at a given border (antinode of the modulator) the two noise fields might be moving toward (occlusion) or away (dis-occlusion) from each other. The modulators were always arranged to be in sine phase (at an antinode) at the center of the display, and the observers were aware of this.

Results

Preliminary data at a luminance contrast of 1.0 showed erratic performance. Eventually this was attributed to the generation of strong motion aftereffects localized to the stripes in the motion-contrast stimulus. This lead to the appearance of "signal" in both signal and blank intervals. To reduce this problem, the remaining data were collected with the carrier at 8 times its detection threshold. For this purpose, we began by collecting luminance contrast thresholds.

Luminance Contrast Thresholds

Luminance contrast detection thresholds were collected with a 2AFC QUEST staircase. The observer selected between a blank (zero luminance contrast) and a stimulus with zero motion-contrast but non-zero luminance contrast. Results are shown in Fig. 5.

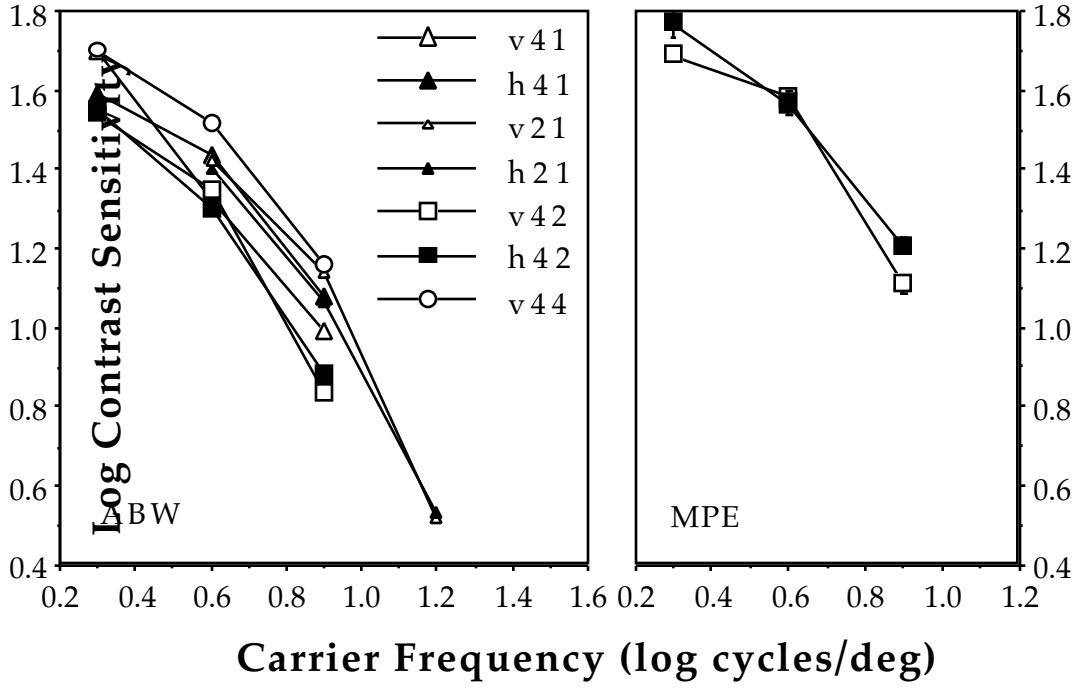

Figure 5. Luminance contrast sensitivity for noise carriers. Motion was horizontal (filled symbols) or vertical (open symbols). Viewing distance was 48.6 or 97.2 cm, resulting in image sizes of 4 (large symbols) or 2 deg (small symbols). Update interval (dt) was 1, 2, or 4 frames, resulting in speeds of 1.875 (circles), 0.938 (squares), and 0.469 deg/sec (triangles). The legend notation is (direction size dt).

Detection thresholds were collected at two viewing distances (48.6 and 97.2 cm), with consequent image sizes of 4 and 2 deg. The number of times each movie frame was exposed (dt) was 1, 2, or 4. Combined with a fixed image displacement of 1 pixel/frame, this resulted in image velocities of 1.88, 0.94, and 0.47 deg/sec, and window time constants of 133, 267, and 533 msec, respectively. The data show the typical effect of spatial

frequency upon sensitivity. Other effects, of image size, velocity, and duration, are more modest.

Motion-contrast Thresholds

Most of our motion-contrast thresholds were collected at $|\mathbf{v}_1| = 0.94$ deg/sec (dt=2), with $\mathbf{v}_2=-\mathbf{v}_1$. The modulator was always vertical (horizontal stripes), and motion was either vertical (compression) or horizontal (shear). Image size was usually 4 deg (viewing distance = 48.6 cm). Carrier frequency was 2, 4, or 8 cycles/deg. For each carrier, luminance contrast was set to eight times the detection threshold.

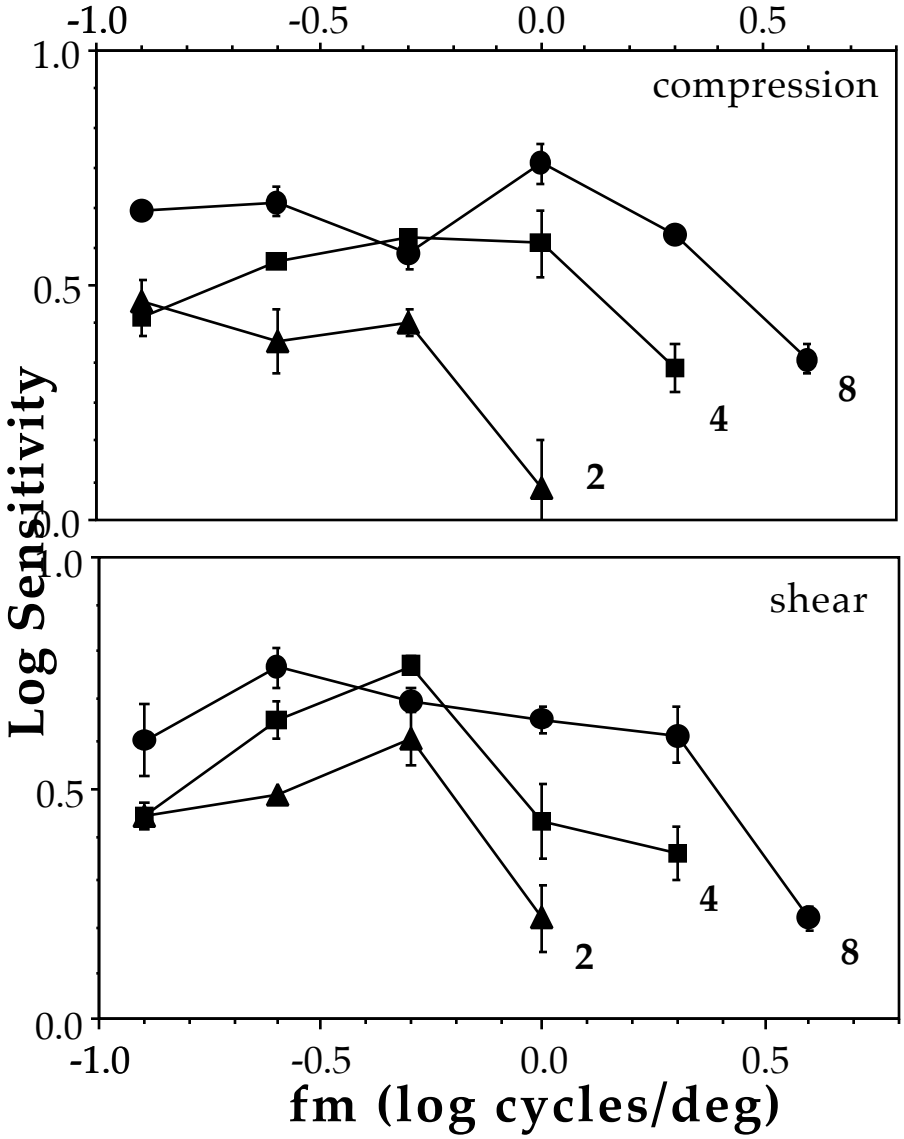

Figure 6 Motion-contrast sensitivity functions for compression and shear for observer ABW. The number near each curve indicates the carrier frequency. Data for f_m =0 are plotted at -0.9. Error bars are ± SE.

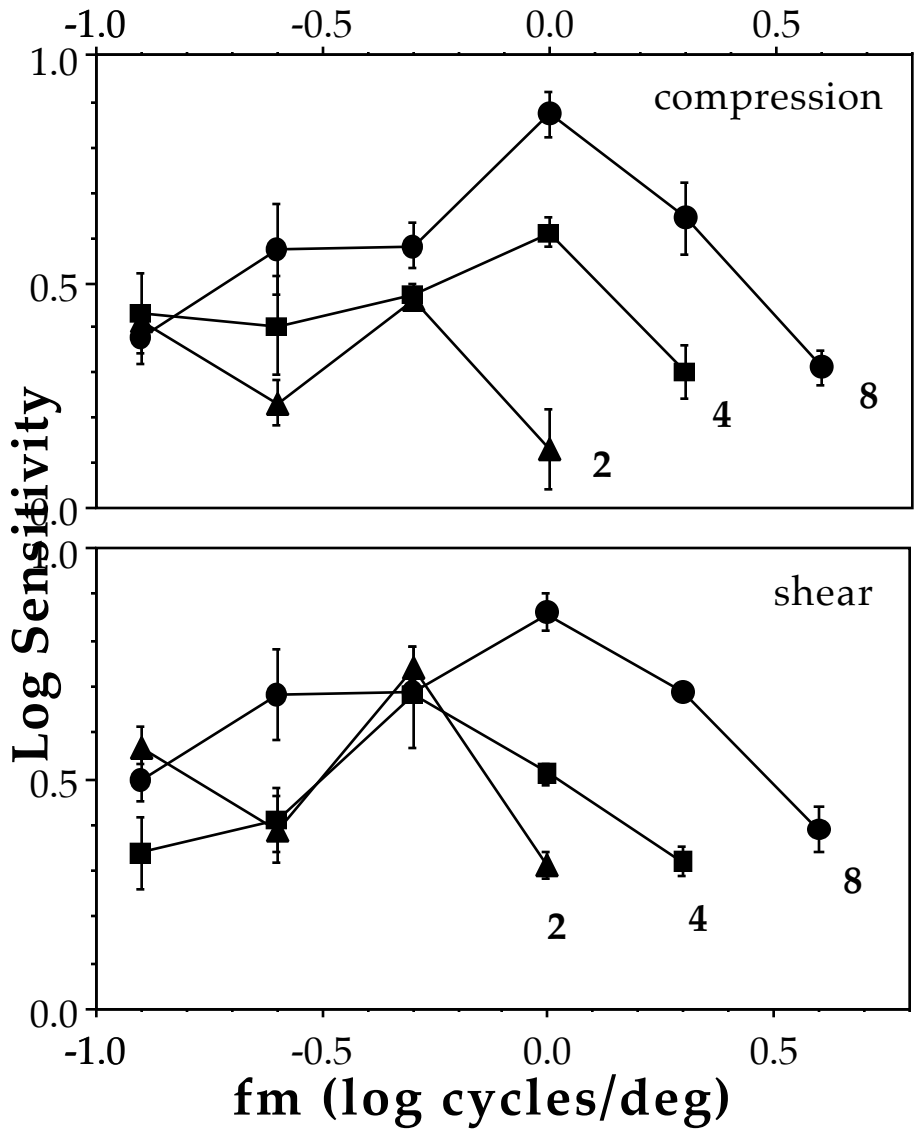

Figure 7 Motion-contrast sensitivity functions for compression and shear for observer MPE. Other details as in Fig. 6.

Figures 6 and 7 show motion-contrast sensitivity functions for compression and shear for both observers. In these and succeeding figures, the data for f_m =0 are plotted at -0.9. The peak sensitivities are rather low, never greater than 10. For each carrier frequency, the data show a decline in sensitivity at the higher modulation frequencies. The location of this decline depends strongly upon the carrier frequency.

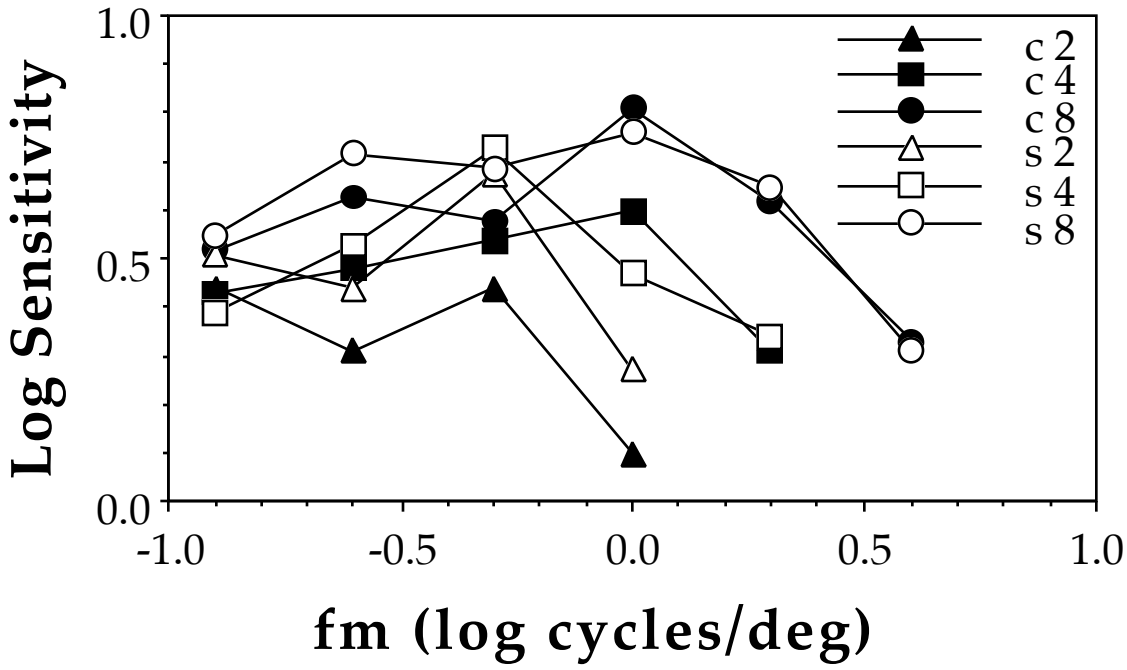

Figure 8. Motion-contrast sensitivity averaged over the two observers. The legend indicates shear (open symbols) or compression (filled symbols) and the carrier frequency.

Since Figs. 6 and 7 show little systematic difference between observers, in Fig. 8 we plot both shear and compression, averaged across observers, to allow easier comparison. Neither is there much systematic difference between shear and compression, except an overall deficit for compression relative to shear at the lowest carrier frequency.

The dependence upon carrier frequency suggests that performance may be "scale invariant", in the sense that the fall-off in sensitivity occurs at a fixed ratio of f_m / f_c . This is confirmed graphically in Fig. 9 which shows the data in Fig. 8, averaged over shear and compression. The data for carrier frequencies of 2 and 4 cycles/deg have been shifted vertically by a small amounts (0.2 and 0.11 respectively), and horizontally by appropriate factors of two so that the horizontal axis may now be labeled f_m / f_c (the data at f_m =0, plotted at -1.8, are of course not shifted). The agreement among the shifted curves, particularly at the higher frequencies, suggests that the second order pooling area, whatever its precise dimensions, is a fixed multiple of the first order linear pooling area.

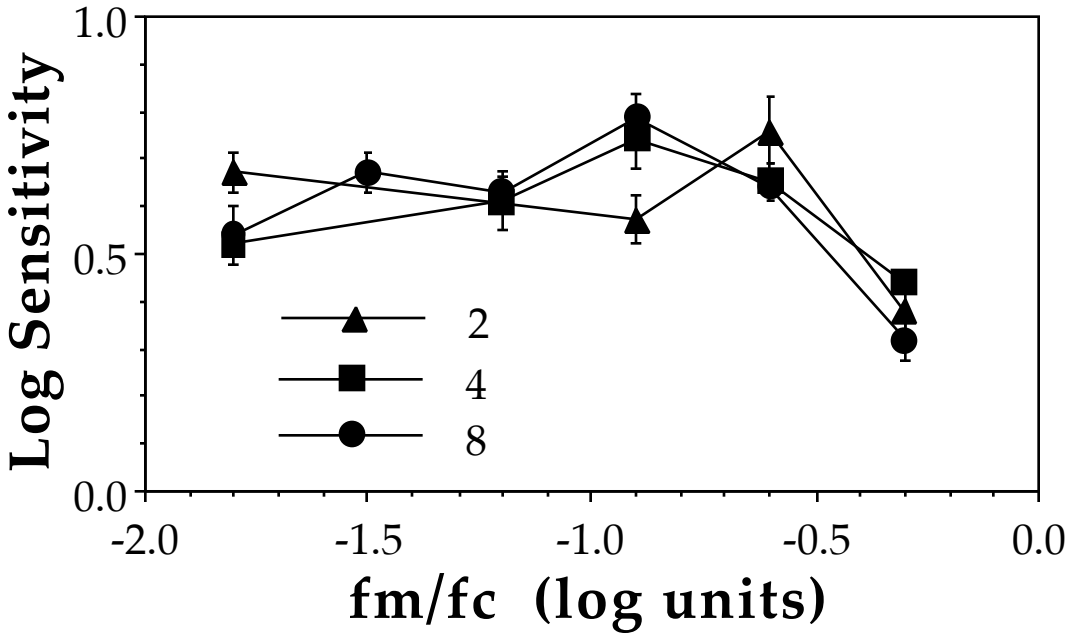

Figure 9. Average motion-contrast sensitivity functions shifted to illustrate scale-invariance. Error bars are ±1 SE.

The need for vertical shifts is a departure from strict scale invariance, except that the image was of constant size, rather than a fixed number of cycles of the carrier frequency, as would be required to test strict scale invariance. We conjecture that enlarging appropriately the size of the lower carrier frequencies would remove the need for vertical shifts.

Model

To assess the relative contributions of first and second-stage pooling we have developed a model of motion-contrast detection. An outline of the model is shown in Fig. 10.

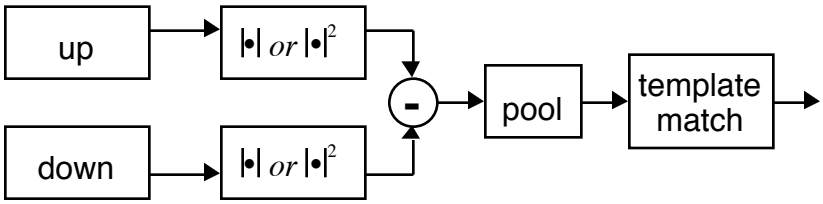

Figure 10. Schematic of the motion-contrast detection model.

The first stage of the model consists of linear direction selective filters for upward and downward motion[1,2]. The spatial filter employed was a Gaussian in frequency centered at the carrier frequency f_c, with an orientation matched to that of carrier motion, and with a bandwidth of 1.4

octaves. The temporal filter employed was a difference of Gamma functions with parameters of $n_1 = 9$, $n_2 = 10$, $t = 0.005$ sec, $k = 1.33$, $z = 1.0$, in the notation of Watson (1986)[24]. These parameters are roughly appropriate for human temporal contrast sensitivity at the experimental mean luminance.

The inputs to the model are actual sequences of digital images, identical to the actual stimuli except that they are one quarter the size (64^2 pixels) and that the noise samples are different. The output of the first stage is two sequences, one for up and one for down. The motion filters we use incorporate both "even" and "odd" phases, and the complex output contains both phase responses as real and imaginary parts. The second stage of the model is a rectifying and "demodulating" nonlinearity. Regarding even and odd filter responses as real and imaginary parts of a complex number, we compute either the magnitude of this number or the magnitude squared. These correspond to "magnitude" and "energy" measures, respectively, and both remove the phase-dependence of the filter response [3,7]. Such operations have been widely conjectured to occur at this stage in the motion pathway, and they correspond roughly with the behavior of many complex cells in primate visual cortex [4,25]. The output of this stage is a pair of image sequences for up and down, respectively.

The next stage subtracts the responses for the two opposite directions. When the squared magnitude (energy) is used, this corresponds to the final opponent stage in the models of van Santen and Sperling[6] and Adelson and Bergen [3]. The output is now a single image sequence.

At the next stage, the opponent signal is pooled spatially over a Gaussian shaped area. The radius of this Gaussian is the parameter of the model that controls the amount of second-stage pooling. Since we assume that an array of pooling units cover the stimulus, this pooling is implemented as a convolution with a Gaussian. The output is again an image sequence; it is, in fact, a Gaussian blurred version of the output at the previous stage. It is convenient to express the radius of the pooling Gaussian in units of the radius of the first-order linear receptive field. We call these "receptive field units" (rfu).

At the final stage we assume a simple template-matching detection process. Our template is a spatial sinusoid of the modulation frequency, approximately equal to the spatial modulator, multiplied by a temporal Gaussian equal to the stimulus time window. This template multiplies the opponent output and the result is integrated over space and time.

Considering the integral operations involved in pooling and matching, it is evident that by changing the order of integration we can, as a matter of implementation only, move the multiplication by the time window and time integration before the spatial pooling and matching, yielding great computational savings.

Multiplying by the time window and integrating the opponent signal over time yields a single image like that in Fig. 11A, in which positive values indicate upward local motion and negative (dark) values, downward. The result of the spatial pooling operation (with a Gaussian radius of 3.84 rfu) is shown in Fig 11B.

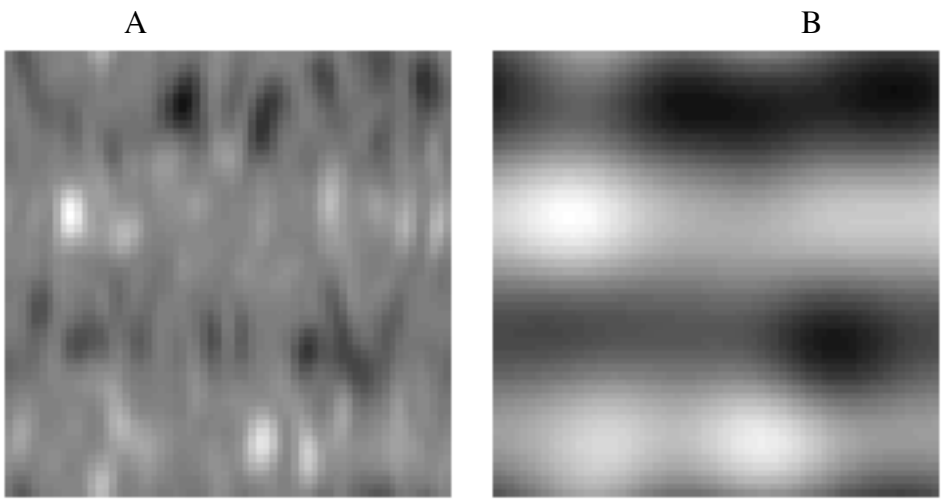

Figure 11. Output of the model after temporal matching and integration. A) before spatial pooling, B) after spatial pooling with radius= 3.84 rfu. The input was f_c = 16 cycles/image, f_m = 2 cycles/image, m=1. Image is 15.3 rfu wide.

Since the spatial template is one-dimensional, spatial matching is implemented by first integrating the image (eg Fig. 11B) over x and then matching the 1D template. Figure 12 shows an example 1D outputs after pooling with radii of 0.24 rfu (essentially no pooling) and 3.84 rfu.

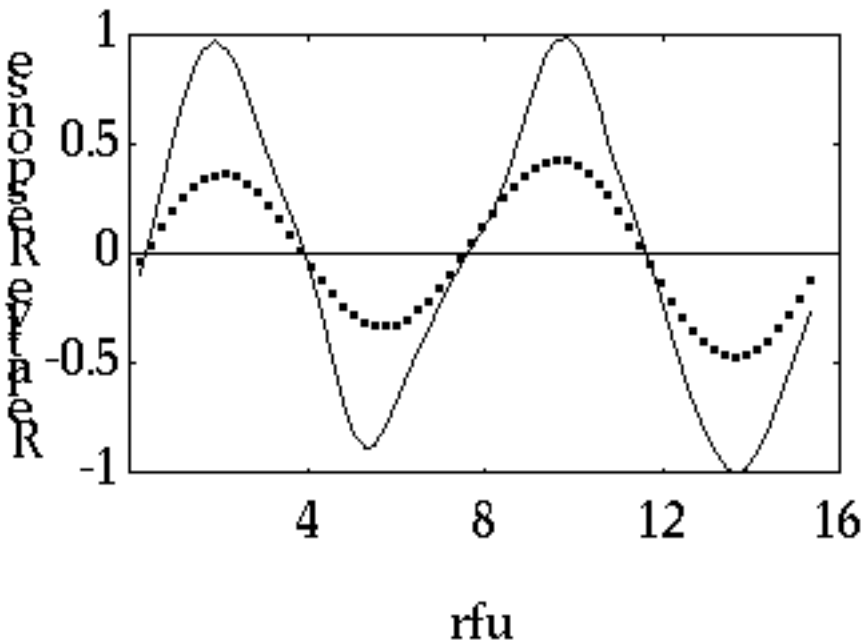

Figure 12 Opponent response (Fig. 11A) integrated over horizontal position after pooling by a Gaussian with a radius of 0.24 rfu (line) and 3.84 rfu (dots).

The 1D pooled opponent response (Fig 12) is then multiplied by the 1D template and integrated. This quantity is computed for both a null stimulus, in which m=0, and the signal stimulus, in which m=0.45. Tests confirmed that the essentials of model performance do not depend much upon the particular value of motion-contrast used. The responses to null and signal were subtracted, and this quantity was assumed to be proportional to sensitivity. Since we are not attempting to predict absolute sensitivity, initially we normalized the results at f_m =0 for all pooling radii. Although the simulations were computed for one particular carrier frequency, we plot the results with respect to f_m / f_c , in keeping with the previous observations regarding scale-invariance. Model predictions are shown in Fig. 13. As in the previous figures, results for f_m =0 are plotted at -1.8. Comparable simulations for the magnitude model and for shear stimuli were very similar.

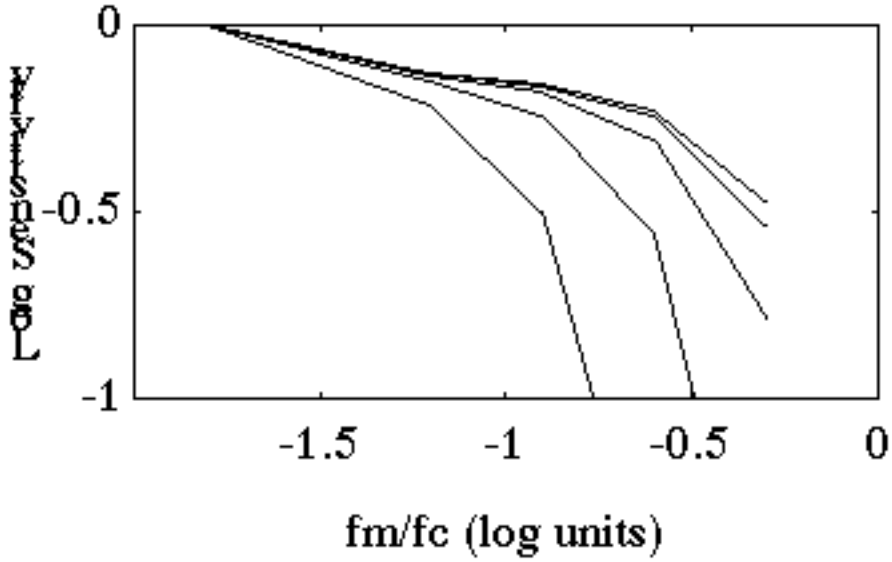

Figure 13. Model predictions for pooling radii of 0.24, 0.48, 0.96, 1.92, 3.84 rfu (from top to bottom). These are energy predictions for a compression stimulus.

At this point the model contains only positive values in the pooling receptive field, hence it produces a strictly lowpass motion-contrast sensitivity function. We do not therefore attempt to simulate the human performance below $\log(f_m / f_c)$=-0.9, and we select the upper portion of each curve in Fig. 13 and shift it vertically to match human sensitivity at this frequency, as shown in Fig 14.

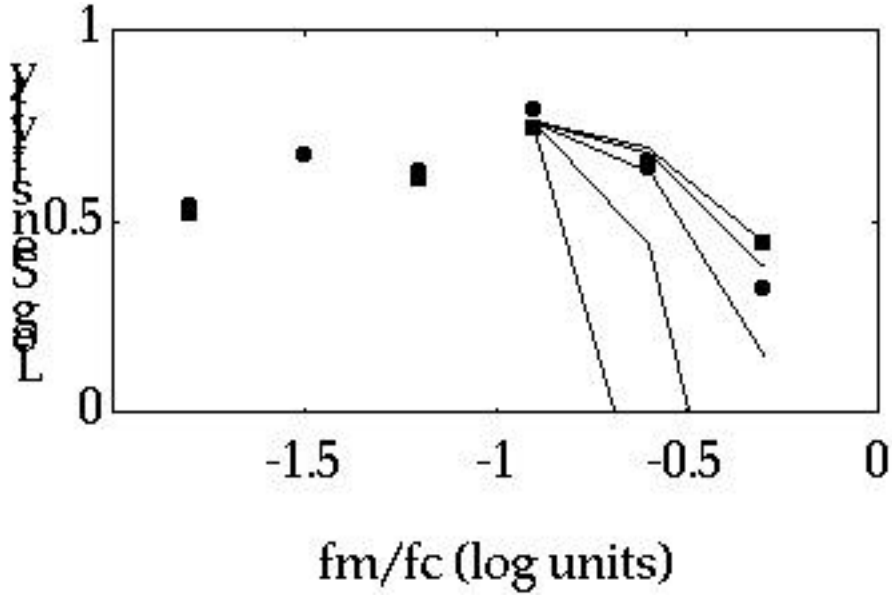

Figure 14. Average data compared to model predictions for pooling radii of 0.24, 0.48, 0.96, 1.92 and 3.84 rfu.

The data are consistent only with the smallest pooling radii, of one half to one quarter the radius of the linear receptive field. The curve at 0.24 rfu is essentially that obtained with no second-order pooling. We conclude that the data show no evidence for any significant second order pooling.

Square-Wave Experiments

Motion borders frequently occur in the natural world as a result of occlusion or dis-occlusion of one surface moving in front of another. The borders are characterized by a relatively sharp discontinuity, rather than the sinusoidal gradient we have used here. To determine the generality of our results, and to ascertain whether sharp motion borders are detected with special sensitivity, we repeated our compression condition at a carrier of 8 cycles/deg with square, rather than sine wave modulators. Results are shown in Fig. 15, along with data shown previously for sine waves.

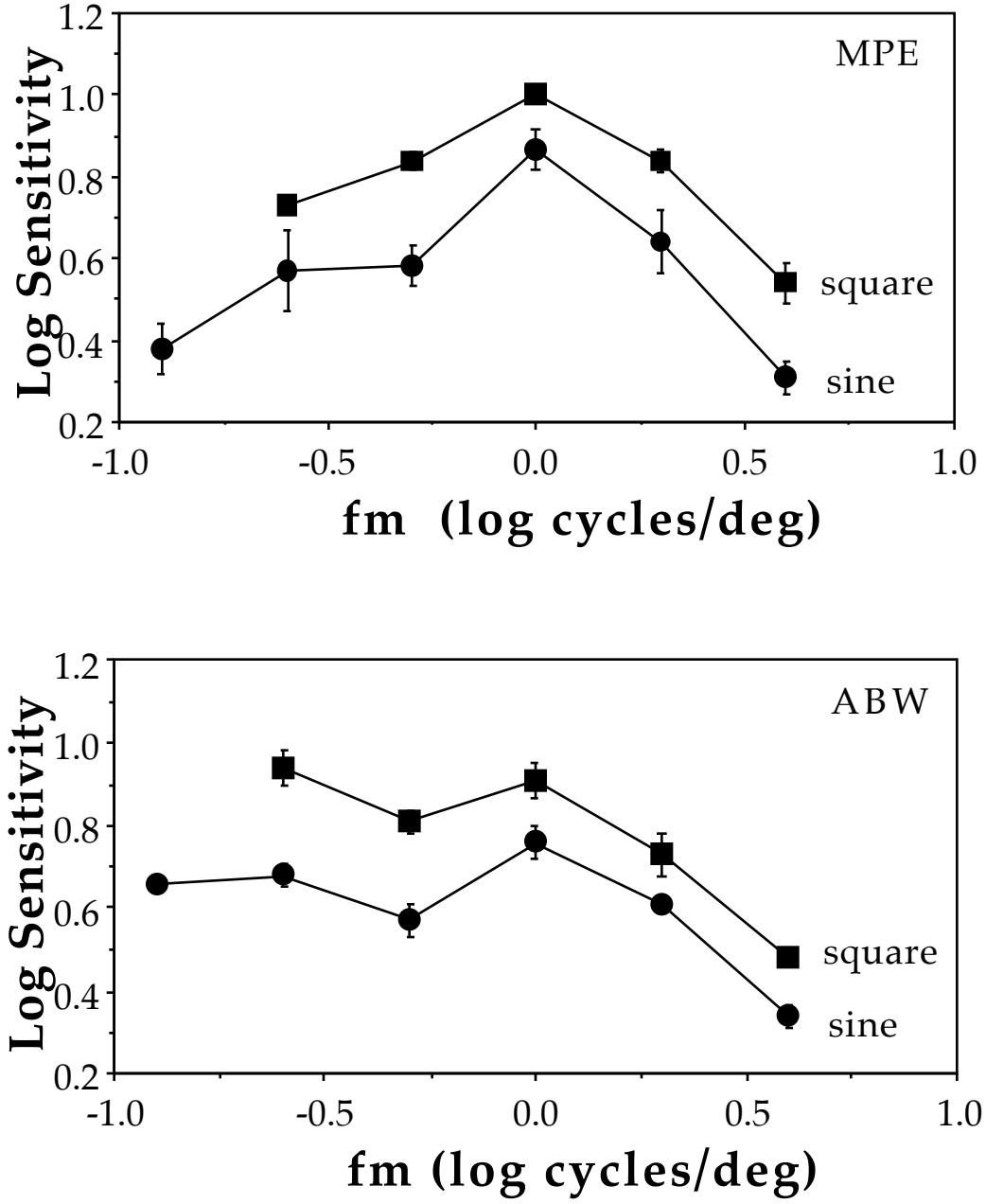

Figure 15. Motion-contrast sensitivity for square-wave and sine-wave modulators.

It is evident that the sensitivity for square waves is higher, but not markedly so. Borrowing an idea from sensitivity to luminance patterns, we may ask whether in fact this slight improvement in sensitivity is consistent with a linear analysis of modulation sensitivity. This would suggest that the square wave should be about $4/\pi$ (~0.1 log unit) greater, since that is the amplitude of its fundamental component [26], and this is

very close to the actual difference. This is consistent with the view that sharp motion borders have no special status, and that our sinusoidal data capture the properties of whatever mechanisms detect these square waves.

Moving and stationary modulators

Another feature of natural motion borders, in contrast to our stimulus, is that the border often moves with its associated object. In our context, this means that the modulator should move with one of the carrier noise fields. To test whether moving modulators were seen with heightened sensitivity we repeated some of our measurements with a square-wave modulator which moved at the same velocity as one of the noise fields. The data are shown in Fig. 16. Rather than enhancing sensitivity, moving the modulator reduces sensitivity substantially at all but the lowest frequencies. This is consistent with the idea of a motion edge detector which is stationary on the visual field (as in the model described below), but is inconsistent with a mechanism tuned to the natural correspondence of velocities of carrier and modulator.

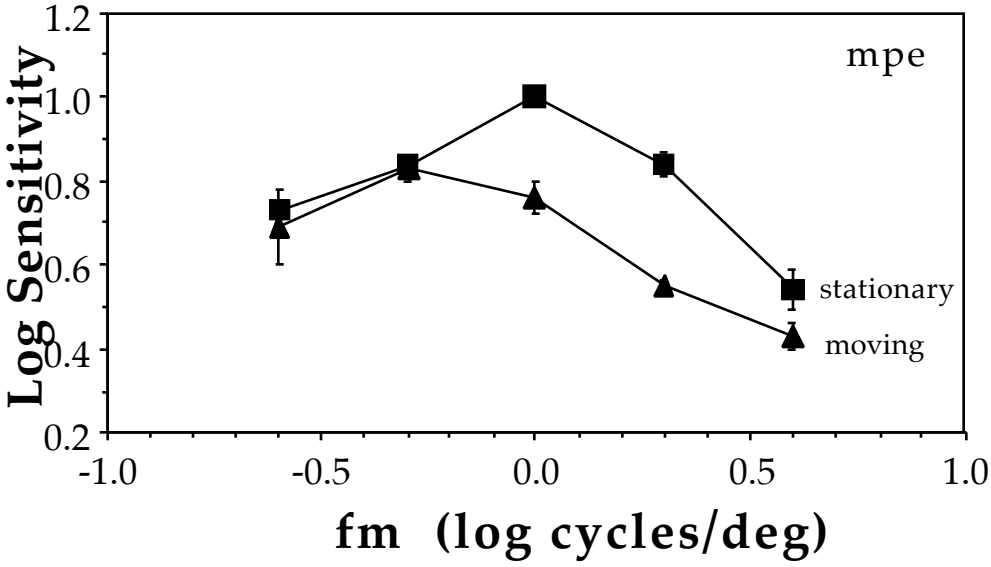

Figure 16. Motion-contrast sensitivity for moving and stationary square-wave modulators.

Inhibitory Pooling

The average data in Figs. 9 and 14 show clear evidence of a decline in sensitivity at low modulation frequencies. Such a decline is consistent with inhibitory pooling, as occurs in the opponent surround of a center-surround luminance receptive field. This sort of higher order motion mechanism, with the center tuned to one direction and an opponent surround tuned to the opposite direction, would be useful in detecting motion discontinuities since it would not respond well to uniform translation. Such units are also found in the visual motion areas of pigeons[27] and primates[28,29], are used in some recent algorithms for estimation of three-dimensional motion estimation [30], and have been conjectured in models of human 3D motion perception [31].

Our model is easily modified to include inhibitory pooling. For the pooling Gaussian, we substitute a difference-of-Gaussians. Keeping the center Gaussian fixed at its smallest value (0.24 rfu), we varied the width and amplitude of the inhibitory Gaussian. The results of the best fitting parameters are shown in Fig. 17, along with the predictions that result when no inhibition is employed (dashed line).

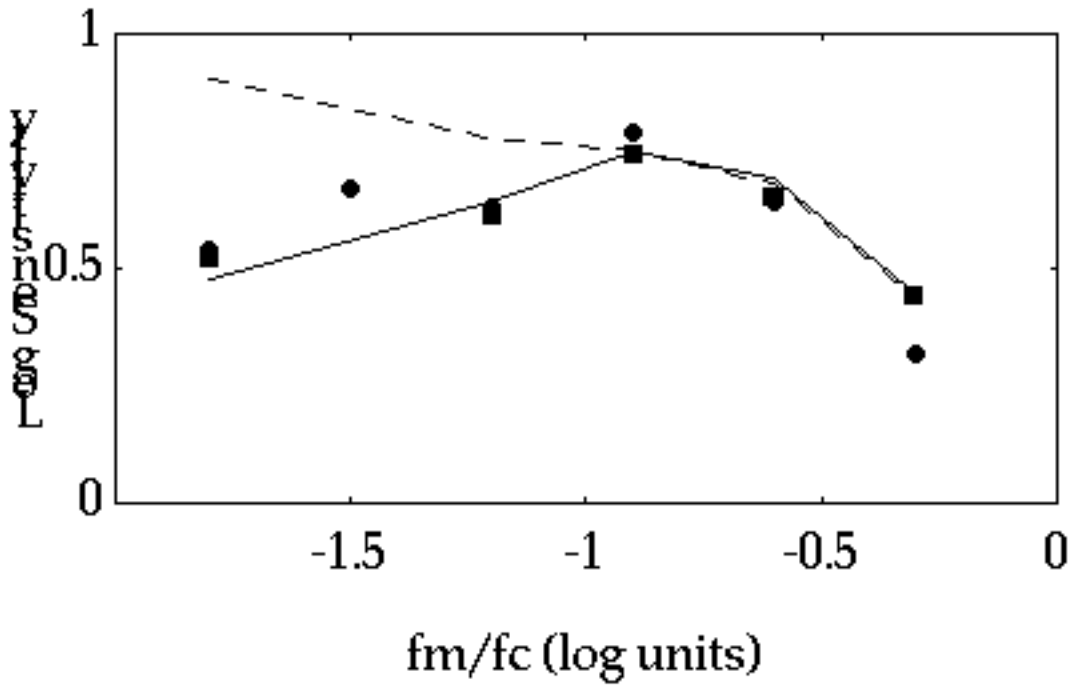

Figure 17. Model predictions for an inhibitory surround in the second order pooling mechanism. The dashed line shows simulations without inhibition.

The predictions have been shifted vertically to agree with the data at $\log(f_m/f_c)$=-0.9. Along with the predictions, we have again reproduced the average data for f_c=4 and 8 cycles/deg. The best prediction is for an inhibitory radius of 7.7 rfu and an inhibitory amplitude of 0.8. This amplitude corresponds to the ratio of volumes of the excitatory and inhibitory Gaussians. Figure 18 provides a picture of the relative size and shape of the first-order linear receptive field and the second order pooling unit. Note that the center of the latter is essentially an impulse, so that there is in effect *only* inhibitory pooling at the second stage. In the following section we discuss a possible physiological basis for this inhibitory pooling.

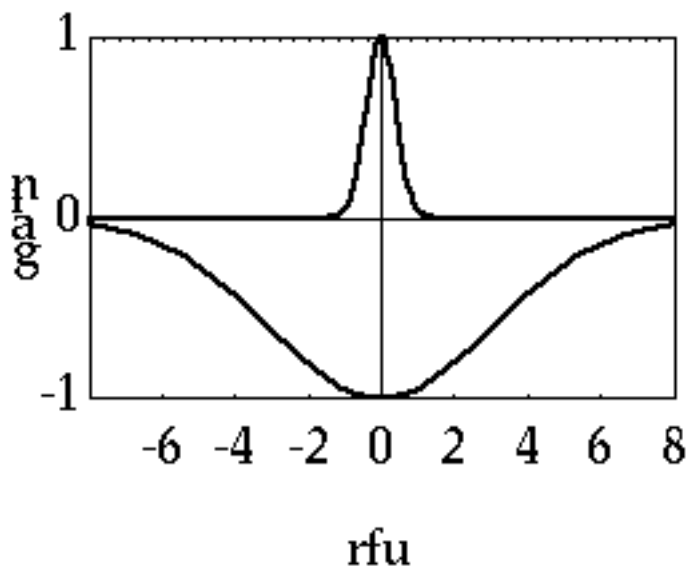

Figure 18. Comparative widths of first-order linear receptive field and inhibitory second-order surround. Curve heights are arbitrary. Horizontal scale is in "receptive field units" (rfu).

Discussion

The purpose of these experiments was to estimate basic sensitivity to motion gradients, and to evaluate the evidence for second-order integration and differentiation of motion signals. On the first question, we found that peak sensitivity was around 0.75 log units (a threshold modulation amplitude of about 18%), for a luminance contrast of eight times threshold. The shape of the motion-contrast sensitivity function is approximately constant when plotted against the ratio of modulator frequency and carrier frequency, indicating that the mechanisms involved

are scale-invariant. Our model displays this sort of scale-invariance for two reasons: 1) the high frequency decline scales because it is due to the first-order filters which themselves scale with the carrier frequency, and 2) the low-frequency decline scales because the radius of inhibitory pooling is assumed to be a constant multiple of the linear receptive field radius.

On the second question, we find no evidence in our data for excitatory spatial pooling beyond the level of the early linear motion filters. It is important to note that this does not mean that such pooling does not occur, only that it does not occur in the pathway used in this task. However, this does create problems for theories of the motion sense which suppose a single serial pathway from V1 to MT and beyond, since neurons in MT are generally reported to have excitatory receptive fields perhaps ten times the diameter of V1 receptive fields from the same visual field location[32].

Another set of experiments that have argued for the existence of spatial pooling of motion estimates are those of Williams and Sekuler [33]. They reported that observers perceived “global coherent motion” from fields of random dots whose individual motions were defined statistically. They argued that this required integration over space of many local motions. In a more recent paper, Watamaniuk and Sekuler[34] show that such integration can occur over areas as large as 63 deg^2. However, it is important to note that while observers may report "coherent unidirectional flow", there is no difficulty in seeing in such stimuli a profusion of local motions in many directions. Furthermore, in the latter experiments, the task was to discriminate the mean direction of the random dots, not to detect incoherence. Thus while they may have demonstrated a mechanism that pools motion signals over large areas, it is clear that other mechanisms exist which preserve motion information on finer scales. Furthermore, since we do not know the effective scale or spatial frequency of their stimulus, their results may be due to integration within large first order linear receptive fields.

We find little difference between shear and compression, particularly at the highest carrier frequency, which we consider our "best" data. This is consistent with our model in which 1) the envelope of the linear receptive fields is circularly symmetric, and 2) the inhibitory pooling Gaussian is circularity symmetric. In this sense, it suggests a general radial symmetry of the motion mechanisms at both first and second levels. It is at variance with the results of Nakayama *et al.*[20] who found compression to be more visible than shear, and with the results of van Doorn and Koenderink[18],

who found the opposite. There were, however, considerable differences in methods among these three experiments.

We do, however, find evidence for *inhibitory* pooling. This intriguing result is suggestive of "motion edge detectors" or spatially opponent motion receptive fields. We have incorporated them into our model by supposing a second stage receptive field consisting of an excitatory impulse and a concentric inhibitory Gaussian. We call this a DIG (Difference of Impulse and Gaussian) receptive field. These inhibitory effects were larger for one observer (MPE) than the other (see Fig.s 6 and 7), and were not always evident. We hope to explore further the conditions that promote or inhibit these effects.

There is an intriguing correspondence between our inhibitory pooling areas and the inhibitory surrounds discovered by Allman, Miezen and McGuinness[28] in the owl monkey. For about 3/4 of their cells, motion in the receptive field surround was inhibitory. For about 60% of these cells the inhibition was tuned to the preferred direction of the cell, as in the model proposed here. The size of the inhibitory region was estimated to be about 7 to 10 times the size of the excitatory receptive field, essentially the same as the figure of about 7.7 estimated here. The maximum magnitude of suppression, as deduced by eye from their figures 6, 10, and 14, was about 80%, again essentially the same as the inhibitory amplitude of 0.8 estimated here. However, a big difference between our model and these physiological results is the size of the excitatory receptive field. We estimate excitatory diameters that are several wavelengths of the carrier frequency employed, for example about 1/4 degree for a carrier of 8 cycles/deg. Allman *et al.* in contrast report excitatory diameters of 5 deg and larger, reflecting a discrepancy in scale of a factor of about 20 in this example. This discrepancy could be resolved by assuming that Allman *et al.* were recording from very large (low frequency) neurons, while we, through our choice of carrier, tap into much smaller (higher frequency) cells.

Acknowledgments. We thank Albert Ahumada, Jeffrey Mulligan, and Kathleen Turano for their advice and assistance. Preliminary reports of this work have been provided elsewhere[35,36]. This work was supported by NASA RTOPs 505-64-53 and 506-71-51.

References

1. A. B. Watson and A. J. Ahumada Jr. "A look at motion in the frequency domain," in *Motion: Perception and representation*, J. K. Tsotsos, ed. (Association for Computing Machinery, New York, 1983).

2. A. B. Watson and A. J. Ahumada Jr. "Model of human visual-motion sensing," Journal of the Optical Society of America A. **2**(2), 322-342 (1985).

3. E. H. Adelson and J. R. Bergen. "Spatiotemporal energy models for the perception of motion," Journal of the Optical Society of America A. **2**(2), 284-299 (1985).

4. R. C. Emerson, J. R. Bergen and E. H. Adelson. "Directionally selective complex cells and the computation of motion energy in cat visual cortex," Vision Research. **32**(2), 203-218 (1992).

5. J. P. H. van Santen and G. Sperling. "A Temporal covariance model of human motion perception," Journal of the Optical Society of America A. **1**(5), 451-473 (1984).

6. J. P. H. van Santen and G. Sperling. "Elaborated Reichardt detectors," Journal of the Optical Society of America A. **2**(2), 300-321 (1985).

7. A. B. Watson. "Optimal displacement in apparent motion and quadrature models of motion sensing," Vision Research. **30**(9), 1389-1393 (1990).

8. E. H. Adelson and J. A. Movshon. "Phenomenal coherence of moving visual patterns," Nature. **300**, 523-525 (1982).

9. S. J. Anderson and D. C. Burr. "Spatial and temporal selectivity of the human motion detection system," Vision Res. **25**, 1147-1154 (1985).

10. G. J. Anderson. "Perception of Self-Motion: Psychophysical and Computational Approaches," Psycho Bulletin. **99**(1), 52-65 (1986).

11. S. J. Anderson and D. C. Burr. "Receptive field properties of human motion detection units inferred from spatial frequency masking," Vision Research. **29**, 1343-1358 (1989).

12. S. J. Anderson and D. C. Burr. "Spatial summation properties of directionally selective mechanisms in human vision," Journal of the Optical Society of America. **8**(8), 1330-1339 (1991).

13. S. J. Anderson, D. C. Burr and M. C. Morrone. "Two-dimensional spatial and spatial frequency selectivity of motion-sensitive mechanisms in human vision," Journal of the Optical Society of America. **8**(8), 1340-1351 (1991).

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

15. M. L. Braunstein and G. J. Anderson. "Velocity gradients and relative depth perception," Perception and Psychophysics. **29**(2), 145-155 (1981).

16. A. J. van Doorn and J. J. Koenderink. "Visibility of movement gradients," Biological Cybernetics. **44**, 167-175 (1982).

17. B. Rogers and M. Graham. "Motion parallax as an independent cue for depth perception," Perception. **8**, 125-134 (1979).

18. A. J. van Doorn and J. J. Koenderink. "Spatial properties of the visual detectability of moving spatial white noise," Experimental Brain Research. **45**, 189-195 (1982).

19. K. Nakayama and C. W. Tyler. "Psychophysical isolation of movement sensitivity by removal of familiar position cues," Vision Reasearch. **21**(4), 427-433 (1981).

20. K. Nakayama, G. H. Silverman, D. I. A. Macleod and J. Mulligan. "Sensitivity to shearing and compressive motion in random dots," Perception. **14**, 225-238 (1985).

21. A. B. Watson, K. R. K. Nielsen, A. Poirson, A. Fitzhugh, A. Bilson, K. Nguyen and J. A. J. Ahumada. "Use of a raster framebuffer in vision research," Behavioral Research Methods, Instruments, & Computers. **18**(6), 587-594 (1986).

22. A. B. Watson and D. G. Pelli. "QUEST: A Bayesian adaptive psychometric method," Perception and Psychophysics. **33**(2), 113-120 (1983).

23. A. B. Watson. "Probability summation over time," Vision Research. **19**, 515-522 (1979).

24. A. B. Watson. "Temporal Sensitivity," in *Handbook of Perception and Human Performance*, K. Boff, L. Kaufman and J. Thomas, ed. (Wiley, New York, 1986).

25. D. J. Heeger and E. P. Simoncelli. "Model of visual motion sensing," in *Spatial vision in humans and robots*, L. Harris and M. Jenkin, ed. (Cambridge University Press, New York, 1992).

26. F. W. Campbell and J. G. Robson. "Application of fourier analysis to the visibility of gratings," Journal of Physiology, Lond. **197**, 551-566 (1968).

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

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

29. J. R. Maunsell and D. C. Van Essen. "Functional properties of neurons in middle temporal visual area of the macaque monkey I. Selectivity for stimulus direction, speed, and orientation," Journal of Neurophysiology. **49**(5), 1127-1147 (1983).

30. D. J. Heeger, A. D. Jepson and E. P. Simoncelli. "Recovering observer translation with center-surround operators," in *Proceedings of IEEE Workshop on Visual Motion, Princeton NJ*, T. S. Huang and P. J. Burt, ed. (IEEE Computer Society Press, Los Alamitos, CA, 1991).

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

32. T. D. Albright and R. Desimone. "Local precision of visuotopic organization in the middle temporal area (MT) of the macaque," Experimental Brain Research. **65**, 582-592 (1987).

33. D. W. Williams and R. Sekuler. "Coherent global motion percepts from stochastic local motions," Vision Research. **24**(1), 55-62 (1984).

34. S. N. J. Watamaniuk and R. Sekuler. "Temporal and spatial integration in dynamic random-dot stimuli," Vision Research. **32** (12), 2341-2347 (1992).

35. A. B. Watson. "Motion pooling areas estimated from motion contrast sensitivity functions," Investigative Ophthalmology and Visual Science. **33** (4), 974 (1992).

36. A. B. Watson. "Sensitivity to spatial variations in image motion," Optical Society of America Technical Digest Series. **17**, 217 (1991).

Notation

b	spatial bandwidth of noise filter in octaves
c_1	carrier 1
c_2	carrier 2
d	scale of temporal window
f_c	carrier frequency
f_m	modulation frequency
m	motion-contrast
m_1	modulator 1
m_2	modulator 2
n_1	motion filter parameter
n_2	motion filter parameter
s	spatial scale of noise filter
$\mathbf{v}$	velocity
$\mathbf{v}_1$	velocity of carrier 2
$\mathbf{v}_2$	velocity of carrier 1
$w(t)$	temporal window
$\mathbf{x}$	spatial coordinate
t	motion filter parameter
k	motion filter parameter
z	motion filter parameter