# Maximum displacement (dmax) in compound bandpass-filtered images: A simple averaging rule

Walter F. Bischof & Vincent Di Lollo

University of Alberta

An image displayed successively in adjacent screen locations is seen as moving in the direction of the displacement. The maximum displacement at which directional motion can be seen is known as *dmax*. Models based on individual motion sensors can account for *dmax* in one-dimensional sinewave gratings but not in two-dimensional bandpass filtered images. In earlier work we found that *dmax* in bandpass filtered images varies according to a simple averaging rule. To wit, *dmax* is proportional to the

average horizontal frequency of all frequency components of the image, weighted by the observer's contrast sensitivity function. The generality of this rule was tested in two experiments with bandpass filtered images that differed in centre frequency and in orientation bandwidth. There was excellent agreement between the empirical results and quantitative predictions based on the averaging rule. These outcomes place definite constraints on models of motion integration.

An image displayed successively in adjacent screen locations is perceived as moving in the direction of the displacement. Such apparent motion has been studied extensively with random-dot images (Baker & Braddick, 1985; Braddick, 1974; Dawson & Di Lollo, 1990) and with bandpass-filtered images (Bischof & Di Lollo, 1990; Chang & Julesz, 1985; Cleary & Braddick, 1990).

Quality of perceived motion depends on the magnitude of displacement. At small displacements, apparent motion is indistinguishable from real motion. As displacement is increased, more parts of the image are seen to move in various directions until, at even greater displacements, coherent motion is no longer seen. The maximum displacement at which coherent directional motion can be seen is known as dmax.

Theoretical accounts of apparent motion have postulated discrete motion signals produced by populations of directionally-selective, frequency-tuned sensors (Adelson & Bergen, 1985; Marr & Ullman, 1981; Reichardt, 1961; van Santen & Sperling, 1985; Watson & Ahu-

This work was supported by Grants No. OGP38521 to WFB and No. OGP0006592 to VDL from the Natural Sciences and Engineering Research Council of Canada. Requests for reprints should be sent to either author, Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada.

mada, 1985). However, such accounts are unambiguous only if the stimuli are sinewave gratings: Additional considerations arise with two-dimensional images. If the stimulus is a sinewave grating, a simple relation can be inferred between activity of motion sensors and perception of motion in the whole image. This is so because the only sensors stimulated by a sinewave grating are those tuned to an orientation orthogonal to that of the grating. Since sensors preferentially tuned to other frequencies or orientations are activated weakly or not at all, motion in the whole image is signalled unambiguously by the most active sensors.

In images other than sinewave gratings, perception of motion is related less simply to sensor activity. Consider an image of isotropically bandpass-filtered noise. By definition, the image contains components of all orientations. When translated, the image will stimulate sensors tuned not only to the direction of motion but to other directions as well. The motion signals generated by sensors stimulated in "spurious" directions will vary in complex ways with the magnitude of displacement (Bischof & Di Lollo, 1990; Cleary & Braddick, 1990). In this case, an account based solely on the outputs of individual sensors is no longer sufficient. Instead of just one, there are now several motion signals, issued from groups of sensors tuned to different orientations. Clearly, the activity of only one such group cannot provide reliable information as to direction of motion in the whole image. What needs to be known is how the outputs from the various groups of sensors combine to yield the perception of overall motion.

In a recent paper, we proposed one way in which the outputs from many motion sensors may be integrated (Bischof & Di Lollo, 1991). In brief, we suggested that the outputs of all motion sensors, weighted by the observer's contrast sensitivity function (CSF), can be averaged to yield an expected value of dmax. In an initial series of studies, we found that this simple averaging rule could account successfully for directional motion detection in bandpass-filtered images (Bischof & Di Lollo, 1991). The present work was designed to extend the earlier studies. The main objective was to determine whether the simple averaging rule could account for detection of motion in more complex compound images. Each compound image was constructed from component images passed by filters of different frequency and orientation bandwidths. In the first of two experiments, the simple component images were used separately to provide a basis for comparison with the spectrally more complex images. In the second experiment, the stimuli were compound images constructed by summing pairs of images used in Experiment 1. The simple averaging rule provided a remarkably good fit to the outcomes of both experiments.

## Experiment 1

## Methods

Observers. One of the authors and two paid undergraduate students, naive as to the purpose of the research, served as observers. All had normal or corrected-to-normal vision.

Stimuli. The stimuli were bandpass-filtered images with a resolution of 128 x 128 pixels. At the viewing distance of 57 cm the square images subtended an angle of 4 deg. Each image contained up to 256 grey levels, and was normalized to a mean luminance of 54 cd/m² (including screen luminance of 10 cd/m²) and peak Michelson contrast of 0.56, yielding a range of 24–84 cd/m². Each image was con-

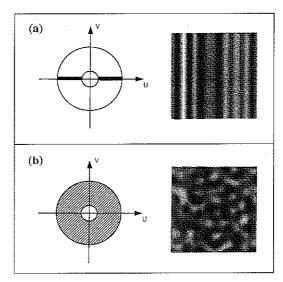


Figure 1: In the left portion of each panel are the Fourier domains of the filters used to produce the stimuli. The filter in panel (a) has an orientation bandwidth equal to zero. The filter in panels (b) has an orientation bandwidth of 180 deg. In the right portion of each panel are sample images passed by the corresponding filter with lower cut-off frequency = 0.5 c/deg, and frequency bandwidth 2 octaves.

structed from a 1282 random-dot matrix in which each dot could be either black or white with a probability of 0.5. The images were filtered with one of the two ideal bandpass filters illustrated in Figure 1. The filters in Figure 1a had an orientation bandwidth equal to zero, and passed only components with an orientation of 90 deg, which produced vertical compound gratings. The filter in Figure 1b had an orientation bandwidth equal to 180 deg; that is, the filter was isotropic and therefore passed components of all orientations. Also shown in Figure 1 are sample images passed by the corresponding filter with frequency bandwidth 2 octaves and lower cut-off frequency 1 c/deg. For each filter, frequency bandwidth was fixed at 2 octaves, but there were five lower cut-off frequencies: 0.5, 0.75, 1, 1.5 and 2 c/deg. In the experimental design, the two orientation bandwidths of the filter (0 or 180 deg) were completely crossed with the five cut-off frequencies to yield a total of 10 experimental conditions. To avoid effects peculiar to the random structure of any individual image, 20 different images were constructed for each experimental condition. On any one trial, the image to be displayed was chosen randomly from the pool of 20 images.

Apparatus. All stimuli were displayed on a Tektronix 608 oscilloscope equipped with fast P15 phosphor. The screen was front-illuminated by a 500-w Sylvania CBA tungsten-halogen projector lamp. A variable neutral-density filter, mounted in the path of the light-source, was adjusted to yield an average screen luminance of 10 cd/m², as measured by a Minolta LS110 luminance meter. The X, Y, and Z (intensity) coordinates of the filtered images were displayed from a fast plotting buffer at the rate of one dot per microsecond (Finley, 1985). To improve brightness and contrast, each image was plotted four times in succession for a total exposure duration of just under 65 ms.

Procedure. On any given trial, the sequence of events was as follows: the observer sat in a dimly-illuminated chamber and fixated on a cross shown in the centre of the screen. Upon a button-press by the observer, the fixation cross disappeared and the first image (F1) was displayed for 65 ms. Immediately upon termination of F1, the second image (F2) was displayed for 65 ms. The two images were identical except that F2 was displaced horizontally with respect to F1 so as to produce the appearance of motion of the left or to the right, randomly. The parts of F2 that were displaced out of the viewing area by the horizontal shift were "wrapped around" to appear at the opposite side of the image. The observer indicated the direction of motion by pressing one of two handheld buttons.

In the present experiment, *dmax* was defined as the F1–F2 displacement that produced 80% correct responses. For this purpose, the range of F1–F2 displacements was adjusted for each observer and condition so as to bracket a level of 80% correct responses. In any one session, observers made 50 observations for every F1–F2 displacement (presented randomly) in one of the 10 experimental conditions.

Contrast sensitivity functions (CSF) were obtained separately for each observer under display conditions matching those of the main experiment. Upon a button-press by the observer, two vertical gratings were displayed sequentially for 65 ms each. The two gratings were identical except that the second was displaced by 0.25 cycles to the left or to the right, randomly. The observer pressed one of two buttons to indicate the direction of motion. Contrast of the gratings was adjusted by a dynamic tracking procedure (PEST, Taylor & Creelman, 1967) to a level that yielded approximately 80% correct responses, which was taken as the threshold for directional motion detection for that spatial frequency. Threshold estimates were taken at spatial frequencies ranging from 0.5 to 9 c/deg. Figure 2 shows the CSFs for the three observers. It should be pointed out that CSF values for the lowest spatial frequency (0.5 c/deg) were probably biased. Namely, owing to equipment limitations, the gratings used to estimate thresholds at the lowest frequency had only two cycles per image. Depending on initial phase angle, these images often produced a strong impression of expanding or contracting horizon-

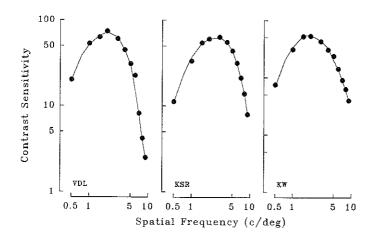


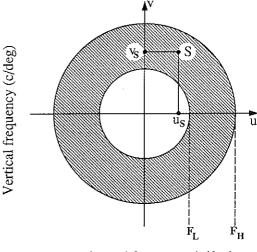
Figure 2: Contrast sensitivity functions for the three observers.

tal motion that interfered with perception of directional motion and decreased response accuracy (see also Hoekstra, van den Goot, van den Brink & Bilsen, 1974). In turn, this may have led to a biasing of the estimate.

#### Results

In the present work, we followed convention by defining *dmax* as the displacement at which direction of motion can be identified correctly with a probability of 80%. Obtained values of *dmax* for all conditions are shown by the unconnected symbols in Figure 3, separately for each observer. Also shown in Figure 3 (by the continuous lines) are the predicted individual values of *dmax* derived from the averaging rule described below.

The averaging rule. The averaging rule encompasses two central assumptions: First, that dmax for horizontal motion is proportional to the average horizontal period of all frequency components of the image and, second, that the frequency components of an image contribute to perceived motion in weighted measure, according to the observer's CSF. The first assumption is best illustrated by reference to Figure 4 which shows the two-dimensional Fourier domain of an image passed by a filter whose frequency band is represented by the shaded area. In Figure 4, S denotes an oblique frequency component having vertical frequency  $v_s$ (with corresponding vertical period pvs) and horizontal frequency us (with corresponding horizontal period pus). Consider a display in

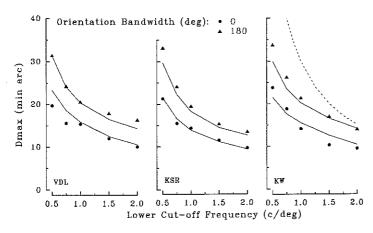


Horizontal frequency (c/deg)

Figure 4: Two-dimensional Fourier domain of an image, showing an ideal bandpass filter with cut-off frequencies  $F_{\rm L}$  and  $F_{\rm H}$ . Also shown is an oblique frequency component S with vertical frequency  $v_{\rm s}$  and horizontal frequency  $u_{\rm s}$ .

which the image is translated to the right. A motion signal  $m = (m_x, m_y)$  generated by an arbitrary motion sensor is said to be correct if  $m_x > 0$ . Now consider an arbitrary frequency component f = (u, v) where u and v denote horizontal and vertical frequency, respectively, as in Figure 4. A correct motion signal is generated for f if the horizontal displacement  $d_{uv} < (2u)^{-1}$ , that is, if the displacement is just short of 1/2 cycle of f with respect to the horizontal axis in Figure 4.

In deriving the predictions shown in Figure 3, the value of *dmax* for an arbitrary image was



Figures 3: Values of dmax in all experimental conditions, separately for the three observers. Empirical values are indicated by the unconnected symbols. The continuous lines show the corresponding results obtained with the averaging rule. The segmented line in the right panel is explained in the text.

taken as the weighted average of all  $d_{uv}$ , weighted in the manner described below. In the present implementation, we used  $d_{uv} = .475u^{-1}$ . This is the value of dmax (in cycles) obtained by the same observers when viewing translating sinewave gratings over a broad range of spatial frequencies (see Di Lollo & Bischof, 1991). For each observer, the predicted values of dmax were calculated as follows. Let F(u, v),  $0 \le u$ , v < n denote the discrete Fourier transform of the filter. Then dmax is defined as

$$d_{max} = 0.475Z^{-1} \sum_{u=1}^{n/2} \sum_{v=0}^{n/2} u^{-1}w(u, v) F(u, v) (1)$$

where Z is a normalizing constant

$$Z = \sum_{u=1}^{n/2} \sum_{v=0}^{n/2} w(u, v) F(u, v)$$
 (2)

and w (u, v) is a weighting factor

$$w(u, v) = [\log_2(r+1) - \log_2(r)]c^2(r)$$
  
 $r^2 = u^2 + v^2$  (3)

The purpose of the term w(u, v) is to weight all frequency components F(u, v) in terms of the observer's contrast sensitivity function. It consists of two factors, both measured in terms of radial spatial frequency (r). The first factor,  $[\log^2(r+1) - \log^2(r)]$ , ensures that frequency components are weighted equally on a logarithmic frequency scale. The second factor, c(r), denotes the observer's contrast sensitivity function obtained under the same spatiotemporal conditions as the motion experiment (see Fig. 2).

Account of the results. Predictions based on Equation (1), illustrated by the continuous lines in Figure 3, show excellent agreement with the empirical data. The agreement is all the more remarkable in that the simulation contained no free parameters. Both independent variables affected performance in accordance with the simple averaging rule. As predicted, values of dmax were higher at the larger orientation bandwidth, and decreased as the lower cut-off frequency of the filter was increased. Both effects represent systematic changes in the weighted

average of all  $d_{uv}$  in response to changes in the independent variables. In every case, the direction of the empirical and simulated outcomes are in qualitative agreement with the inverse scaling of *dmax* with spatial frequency reported in earlier studies (Bischof & Di Lollo, 1990, 1991; Chang & Julesz, 1985; Cleary & Braddick, 1990).

Also shown in Figure 3 – by the segmented line in the right panel – are the values that dmax would assume were it to remain a constant proportion of the period of the image's lowest frequency throughout the domain. Notably, the segmented curve in Figure 3 is proportional to prediction made by the averaging rule without weighting in terms of the observer's CSF. As can be seen, such a curve is not representative of either the obtained or the simulated values of dmax. Clearly, the pattern of results in Figure 3 cannot be regarded as a trivial consequence of expressing dmax in units of min arc rather than in terms of the period of some frequency component in the image. The soundness of this procedure has been demonstrated by Bischof & Di Lollo (1991).

Next, we check on the generality of the simple averaging rule by using images of substantially more complex spectral composition than the images used thus far. This is done in Experiment 2.

## **Experiment 2**

Orientation bandwidth, which was varied between images in Experiment 1, was found to affect performance in ways that matched predictions based on a simple averaging rule. In Experiment 2, orientation bandwidth was varied within images. That is, as in the previous experiment, the images contained frequency components spanning a range of two octaves. However, for half the images, the orientation bandwidth was zero deg for the lower octave and 180 deg for the upper octave. The reverse was true for the remaining images (see Fig. 5).

Predictions as to how *dmax* should be affected by such *compound* images are virtually impossible to make on a purely intuitive basis but are straight-forward in terms of the averaging rule expressed in Equation (1). The issue

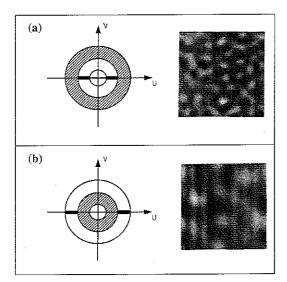


Figure 5: In the left portion of each panel are the Fourier domains of the filters used to produce the stimuli. All filters have a frequency bandwidth of two octaves. The filter in panel (a) has an orientation bandwidth equal to zero for the lower octave, and an orientation bandwidth of 180 deg for the upper octave. The reverse is true for the filter illustrated in panel (b). In the right portion of each panel are sample images passed by the corresponding filter with lower cut-off frequency = 0.5 c/deg.

of just what it is that gets averaged is discussed below. Here, two related events are worth noting: First, predictions based on the averaging rule were as successful with the compound images in Experiment 2 as with the simple images in Experiment 1. Second, each image in Experiment 2 consisted of the *linear summation* of two images from Experiment 1. Taken together, these observations strongly suggest that whatever it is that gets averaged, it does so in a linear fashion.

## Methods

Observers, apparatus, displays and procedures were the same as in Experiment 1, with the exception of the filters used in constructing the images. At each of the five lower cut-off frequencies of the filter (0.5, 0.75, 1, 1.5, and 2 c/deg), there were two types of images, defined in terms of orientation-bandwidth composition. All images had a frequency-bandwidth of 2 octaves. In one set of images, the lower octave had an orientation bandwidth equal to zero,

while the upper octave had an orientation bandwidth of 180 deg. The reverse was true for the other set of images. Each image can be considered as the sum of a compound grating and an isotropic blob pattern. In one set of images, high-frequency blobs were superimposed on a low-frequency compound grating; in the other set, a high-frequency compound grating was superimposed on low-frequency blobs. The filters used for producing the images are shown in Figure 5, as are sample images passed by the corresponding filters.

## Results

Predicted and obtained values of *dmax* are shown in Figure 6, separately for each observer. The predictions show excellent agreement with the empirical data, with the exception of a single point for observers VDL and KW. In both cases, the wayward point was obtained with images that contained extremely low spatial frequencies. We suspect that the poor fit is due to instability of the corresponding point in the observer's CSF, as was noted above.

Irrespective of such minor variations, the analysis fits the data remarkably well. In addition, it offers a coherent account of several aspects of the results that would not be readily interpretable otherwise. For example, on purely intuitive grounds, it is not at all obvious why *dmax* should be higher when the 180-deg orientation bandwidth occupies the lower octave than when it occupies the upper octave (Fig. 6). Yet, this result becomes readily interpretable in terms of the weighted averaging process. It should also be noted that the phenomenal appearance of the displays was vastly different in Experiments 1 and 2. Yet both sets of results are explained equally well in the same terms.

## Discussion

It is generally agreed (Adelson & Movshon, 1982; Bischof & Di Lollo, 1990, 1991), that a comprehensive account of directional motion perception cannot be achieved solely on the basis of individual motion sensors functioning independently one from the other. Individual sen-

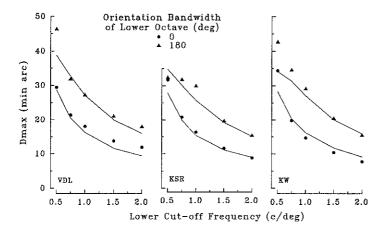


Figure 6: Values of dmax in all experimental conditions, separately for the three observers. Empirical values are indicated by the unconnected symbols. The continuous lines show the corresponding results obtained with the averaging rule.

sors can provide an adequate account of directional motion perception of "one-dimensional" stimuli (simple sinewave gratings) but not of more complex images. The present averaging rule outlines one possible way in which the individual frequency components of an image may be combined to yield perception of directional motion. We hasten to add that, although important as a determinant of dmax, the averaging rule has little to say about direction of motion. This is so because, in the present work, observers reported only the horizontal component of directional motion. However, within these limits, the rule was obviously successful in distinguishing leftward from rightward motion.

What, if anything, can the averaging rule contribute towards an understanding of motion perception? Before answering this question, some basic assumptions must be spelled out. Along with other researchers (e.g., Adelson & Movshon, 1982; Movshon, Adelson, Gizzi & Newsome, 1986) we assume that at least two processing stages are required to account for perception of motion in complex images: motion sensing and motion integration. The initial stage contains a population of orientation-selective, frequency-tuned sensors, each capable of signalling motion only over a very limited area (e.g., Reichardt, 1961; Marr & Ullman, 1981). The outputs of many such sensors must be combined if directional motion is to be seen over the entire visual field. As do Adelson and Movshon (1982), we assume this integrative function to be performed at a separate processing stage, affected by different variables - and bound by different rules – from the stage of "one-dimensional" sensors that supply its input.

The present averaging scheme is largely unconcerned with the motion-sensing stage. For example, no assumption is made as to the precise mechanism of motion detection at the earlier level. In fact, the present scheme is compatible with the major models of motion detection, be they correlational models (Reichardt, 1961), energy models (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Watson & Ahumada, 1985) or gradient models (Marr & Ullman, 1981). All that the present scheme requires is that, at the initial stage, motion perception be mediated by sensors capable of providing input to the second processing stage.

With respect to the second stage, the averaging scheme makes two distinct contributions: First, it places explicit constraints on models of motion integration; and, second, it highlights the importance of the contrast sensitivity function in selecting the frequency components that contribute most to motion perception. The major constraint stems from the finding that, to provide suitable fits to the empirical data (such as illustrated in Figures 3 and 6), all spatial frequency components in the image must be considered. Omission of frequency components from the averaging process results in marked deterioration in the goodness of fit. This places a clear constraint on models of motion integration: It questions the tenability of those that rely on sub-populations of sensors - rather than on the entire population - for providing the motion signals to be integrated. This pertains to "winner-take-all" models such as those proposed by Bischof and Di Lollo (1990), Grzywacz and Yuille (1990), and by Watson and Ahumada (1985). This constraint also questions the tenability of one-dimensional analyses such as that proposed by Cleary and Braddick (1990).

As for using the CSF to weight the frequency components of the image, there is no question that CSF-weighting is crucial to the success of the model's predictions. Without it, there would be poor correspondence between predicted and obtained scores in Figures 3 and 6, as shown by the segmented line in Figure 3. Indeed, goodness of fit would be reduced even if a generalized CSF were used in place of the individual observer's CSF.

On the face of it, use of the CSF as a weighting function hardly needs justifying. It seems reasonable to expect that the magnitude of the signal from a frequency-tuned motion sensor (and hence the input to the second processing stage) should be scaled in terms of the CSF. However, there is evidence that, for stimuli well above threshold contrast, the CSF is essentially flat across the frequency domain (Georgeson & Sullivan, 1975). On the strength of this finding, it could be argued that the shape of the threshold CSF may not apply to the present stimuli whose contrast was decidedly above threshold. On the other hand, there is reason to believe that Georgeson and Sullivan's results may be specific to the contrast-matching procedure used in their experiment. When a magnitude-estimation procedure is used, the characteristic bow-shaped CSF is obtained even at high levels of contrast (Cannon, 1979). Further arguments justifying the use of the threshold CSF with above-threshold stimuli have been presented by De Valois and De Valois (1988, pp. 167-168).

## References

- Adelson, E.H. & Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, 2A, 284–299.
- Adelson, E.H. & Movshon, J.A. (1982). Phenomenal coherence of moving visual patterns. *Nature, London*, 300, 523–525.

- Baker, C.L. & Braddick, O.J. (1985). Eccentricity-dependent scaling of the limits for short-range apparent motion detection. Vision Research, 25, 803–812.
- Bischof, W.F. & Di Lollo, V. (1990). Perception of directional sampled motion in relation to displacement and spatial frequency: Evidence for a unitary motion system. Vision Research, 30, 1341–1362.
- Bischof, W.F. & Di Lollo, V. (1991). On the half-cycle displacement limit of sampled directional motion. Vision Research, 31, 649–660.
- Braddick, O.J. (1974). A short-range process in apparent motion. Vision Research, 14, 519-527.
- Cannon, M.W. (1979). Contrast sensation: A linear function of stimulus contrast. Vision Research, 19, 1045– 1052.
- Chang, J.J. & Julesz, B. (1985). Cooperative and noncooperative processes of apparent movement of random-dot cinematograms. Spatial Vision, 1, 39–45.
- Cleary, R. & Braddick, O.J. (1990). Direction discrimination for band-pass filtered random dot kinematograms. Vision Research, 30, 303–316.
- Dawson, M. & Di Lollo, V. (1990). Effects of adapting luminance and stimulus contrast on the temporal and spatial limits of short-range motion. Vision Research, 30, 415–429.
- De Valois, R.L. & De Valois, K.K. (1988). Spatial Vision. Oxford: Oxford University Press.
- Di Lollo, V. & Bischof, W.F. (1991). Displacement limit (dmax) of sampled directional motion: Direct and indirect estimates. Perception & Psychophysics, 49, 176–186.
- Finley, G. (1985). A high-speed point plotter for vision research. *Vision Research*, 25, 1993–1997.
- Georgeson, M.A. & Sullivan, G.D. (1975). Contrast constancy: Deblurring in human vision by spatial frequency channels. *Journal of Physiology (London)*, 252, 627–656.
- Grzywacz, N.M. & Yuille, A.L. (1990). A model for the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society of London*, 239B, 129–161.
- Hoekstra, J., van der Goot, D.P.A., van den Brink, G. & Bilsen, F.A. (1974). The influence of the number of cycles upon the visual contrast threshold for spatial sine wave patterns. *Vision Research*, 14, 365–368.
- Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London*, 211B, 151–180.
- Movshon, J.A., Adelson, E.H., Gizzi, M.S. & Newsome, W.T. (1986). The analysis of moving visual patterns. Experimental Brain Research, (Suppl.), 11, 117–152.
- Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information. In Rosenblith, W.A. (Ed.) Sensory Communication. Cambridge, MA: MIT Press.
- Taylor, M.M. & Creelman, C.D. (1967). PEST: Efficiency estimates on probability functions. *Journal of the Acoustical Society of America*, 41, 782–787.
- van Santen, J.P.H. & Sperling G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America*, 2A, 300–320.
- Watson, A.B. & Ahumada, A.J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society* of America, 2A, 322–341.