Pergamon

PII: S0042-6989(97)00007-2

*Vision Res.*, Vol. 37, No. 14, pp. 1873–1883, 1997

Printed in Great Britain
0042-6989/97 $17.00 + 0.00

# Independent Coding Across Spatial Scales in Moving Fractal Images

NUALA BRADY,*† PETER J. BEX,* R. ERIC FREDERICKSEN*

*Received 16 August 1995; in revised form 9 January 1996; in final form 6 December 1996*

**We compared observers' ability to discriminate the direction of apparent motion using images which varied in their spatial characteristics; white or flat spectrum noise, and 1/f noise which has an amplitude spectrum characteristic of natural scenes. The upper spatial limit for discrimination ($d_{max}$) was measured using a two-flash random dot kinematogram (RDK), which consisted either of a pair of bandpass filtered images or of a bandpass filtered image and its broadband counterpart. Six bandpass central frequencies were used, ranging from 0.25 to 5.66 cyc/deg. Subjects could perform the direction discrimination task for all six central frequencies in both the bandpass–bandpass and bandpass–broadband sequences for the 1/f images, and $d_{max}$ values were found to be approximately equal in these two conditions at all spatial scales. However, for the white noise images, direction discrimination was not possible at the lowest central frequencies in the bandpass–broadband task. These data show that information from a wide range of spatial scales is equally salient to the human motion system in images whose amplitude spectra fall as 1/f. However, for white noise images, information at the higher spatial frequencies is more salient and dominates performance in the direction discrimination task. These results are consistent with a model in which spatial frequency filters in the input lines of motion detectors have octave constant spatial frequency bandwidths and equal peak sensitivity. In line with a number of recent studies, this suggests that the spatial properties of motion sensitive cells are matched to the statistical properties of natural scenes. © 1997 Elsevier Science Ltd.**

Motion  Direction perception  Spatial filtering  Natural scenes  Displacement

## INTRODUCTION

Random dot kinematograms (RDKs) have been used extensively in the study of low-level motion detection mechanisms. In these displays a spatial noise pattern is presented twice or more in rapid sequence, while undergoing a fixed spatial displacement between presentations. Under appropriate spatial and temporal conditions, observers perceive smooth, continuous motion. When the displacements are below an upper limit known as $d_{max}$ observers can readily discriminate the direction of motion, but beyond this limit they report that motion is incoherent and they can no longer identify its direction.

It was originally proposed that $d_{max}$ reflected a fixed spatial limit on our ability to establish correspondences between features in the image across successive presentations (Braddick, 1974). However, later research has led to a revision of this proposal. The maximum spatial displacement for apparent motion in RDKs increases with increased stimulus area or retinal eccentricity (Baker & Braddick, 1982, 1985; van de Grind *et al.*, 1986), and $d_{max}$ also depends on the spatial frequency content of the stimulus (Chang & Julesz, 1983; Cleary & Braddick, 1990a,b; Bischof & DiLollo, 1990). Using RDKs with bandpass filtered noise, Chang & Julesz (1983) found that $d_{max}$ varied inversely with the central frequency of the bandpass. These findings suggest that $d_{max}$ is a limit set by the properties of motion sensitive neurones rather than an absolute limit to a matching process. Physiological data are consistent with this idea. Baker & Cynader (1986) measured the response of direction selective cells to sequentially flashed bars and found that the maximum spatial displacement between the bars was inversely related to the cells' preferred spatial frequency. The variation in $d_{max}$ with spatial frequency is also predicted by a number of computational models of motion sensing (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Watson & Ahumada, 1985). For example, in the elaborated Reichardt detector models a directionally selective subunit consists of a pair of filters separated by a distance equal to one-quarter of a cycle of the filters' preferred spatial frequency. This variation of spatial offset with spatial frequency is purpose-built to avoid aliasing and

*Vision Research Centre, McGill University, 687 Pine Avenue West, Montreal H3A 1A1, Quebec, Canada.
†To whom all correspondence should be addressed at present address: Department of Psychology, University of Manchester, Manchester M13 9PL, U.K.

effectively sets the upper spatial limit to discrimination just below one half cycle.

Assuming that the visual system has access to information at a broad range of spatial scales, we might expect the maximum spatial displacement for apparent motion in a broadband RDK to be determined by the lowest spatial frequencies in the pattern. This prediction fails when tested. Values of $d_{max}$ for broadband binary noise patterns are considerably less than predicted from the presence of low spatial frequencies in these images. Furthermore, when such images are lowpass filtered, $d_{max}$ increases as the filter cut-off frequency is successively decreased below about 3.5 cyc/deg (Cleary & Braddick, 1990b). How then does information from different spatial scales interact in a broadband pattern?

One suggestion, associated with energy-based models of motion detection, is that at large spatial displacements of a broadband pattern, incoherent motion signals from high frequency channels mask or inhibit coherent motion signals from low frequency channels (Cleary & Braddick, 1990b). A quite different proposal is that $d_{max}$ is limited by the mean separation between features in an image (Morgan, 1992). According to this view, information from different spatial frequency channels is combined to localize features in an image before motion analysis, which is assumed to involve a matching of these features across space and time. Using a novel version of the conventional RDK, Morgan & Mather (1993) asked subjects to identify the direction of motion in a two-flash kinematogram which consisted of a lowpass binary noise pattern and the broadband pattern from which it was filtered. They found that motion discrimination was impossible when the upper cut-off frequency of the lowpass filtered image fell below some critical value. The authors interpreted these findings as evidence for feature-based motion analysis; when the mean separation between features in the lowpass and broadband images reaches a critical difference, discrimination is impossible despite the fact that the image sequence contains motion energy at common spatial scales. An explanation based on inhibition of low frequencies by high seems less likely in this case, as there is no motion signal at high spatial frequencies. In a recent replication of this study, but using noise patterns which had been filtered to have a 1/f amplitude spectrum, Bex *et al.* (1995) found that motion discrimination was possible in the lowpass–broadband sequences for a series of lowpass cut-offs ranging from 16 to 0.125 cyc/deg. $d_{max}$ was roughly constant across this range, suggesting that for these patterns the upper limit to displacement is determined by the low spatial frequencies.

The primary goal of the present study is to explore how information at different spatial scales is utilized by the visual system in the class of broadband images it typically encounters. In the experiments reported below we investigate the hypothesis that information at different scales is equally salient to the human motion system for natural scenes and for broadband images with similar power spectra. This is expected from an assumption that the spatial properties of motion sensitive cells are matched to the spatial characteristics of natural images so that such images produce approximately equal activity in cells tuned to different spatial frequencies. We argue that this distribution of activity is a crucial factor in determining $d_{max}$ for broadband images, a perspective which could potentially be used by either energy-based or feature-based models in explaining why $d_{max}$ is smaller than expected in broadband binary RDKs.

Various researchers have suggested that the visual system takes advantage of statistical regularities or redundancy in its input so as to represent information with maximal efficiency. Natural scenes share a number of important statistical properties which distinguish them from other classes of images. For example, contrast energy is roughly independent of viewing distance in such scenes as expected from their fractal or self-similar properties. The power spectra of natural images fall as approximately $1/f^2$ (Field, 1987, 1993; Burton & Moorhead, 1987) so that there is roughly equal energy in octave frequency bands. Field (1987) proposed that the spatial characteristics of mammalian cortical cells are matched to this property of natural scenes. Assuming these cells have spatial frequency bandwidths which are constant in octaves and equal peak gains (i.e., equal contrast sensitivity at their optimal or preferred spatial frequencies), they will respond with roughly equal activity to images with $1/f^2$ power spectra. Psychophysical evidence for this model of sensitivity was provided by Brady & Field (1995), who showed that contrast constancy holds for bandpass images whose spectra scale in a manner similar to that of natural scenes.

The receptive fields of these model cells have an interesting profile.* Because cells tuned to different spatial frequencies have equal response magnitude at their preferred frequencies, the volume under the receptive field is constant for all receptive field sizes. This means that the peak of the point spread function must decrease in proportion to the area under the receptive field. Recent physiological data on the receptive field properties of both M and P retinal ganglion cells show that this arrangement is closely approximated in primate retina. Croner & Kaplan (1995) found that a cell's peak sensitivity (in space) varied inversely with the area of the cell's receptive field such that the "integrated sensitivity" or volume "is (approximately) constant, regardless of the size of the region". As retinal ganglion M cells form the first stage of a pathway which carries information about motion, the assumption that the spatial properties of motion sensitive cells are tuned to the spatial properties of natural scenes is a reasonable one. We consider further evidence for this model in the Discussion section.

This model of sensitivity predicts that cells or filters tuned to different spatial frequencies will respond with

*We are assuming that the model cells are linear, so that the point spread function or receptive field profile can be derived from the frequency tuning function by Fourier transform.

roughly equal activity to images whose power spectra fall as $1/f^2$. By comparison, their response to images with a flat spectrum (such as the binary noise patterns typically used in RDKs) will increase in proportion to their optimal spatial frequency: if cells tuned to different spatial frequencies have equal peak gain, high frequency cells will give a larger response because they integrate over a wider frequency band. So while the Fourier description of 1/f patterns shows them to have increasingly less energy at high spatial frequencies, these patterns are represented neurally by a distribution of activity which is approximately uniformly distributed across scale. This neural transformation is often described as resulting in a distributed frequency response (Field, 1987) or as "whitening" the input (Atick & Redlich, 1992). Similarly, images whose power spectra are flat have a neural representation in which response activity increases in proportion to spatial frequency.

In the three experiments reported below subjects judged the direction of motion in a two-flash RDK which consisted of either a pair of bandpass filtered images or of a bandpass filtered image and the original broadband image. We used bandpass filters to control the scale of information available to observers in the direction discrimination task. Six central frequencies were used (five ranging from 0.25 to 4.0 cyc/deg in octave steps and a sixth at 5.66 cyc/deg) and all filtered images had a 1.6 octave bandwidth. Three types of broadband images were used; 1/f noise which has an amplitude spectrum characteristic of natural scenes and which was rendered with 8 bits of grey level, 8-bit Gaussian distributed white noise, and 1-bit uniformly distributed white noise. These will be referred to as 1/f noise, Gaussian noise and binary noise, respectively. The upper displacement limit ($d_{max}$) was measured in all conditions. Given the hypothesis that information at all spatial scales is equally salient to the motion system in natural scenes, $d_{max}$ is expected to be roughly equal in the bandpass–bandpass and bandpass–broadband conditions for all central frequencies of the 1/f filtered noise patterns. Because the distribution of neural activity is biased toward high spatial frequencies in flat spectrum images, $d_{max}$ is expected to be dissimilar in these two conditions at low central frequencies of the white noise patterns. This is also expected from Morgan and Mather's (1993) results with binary noise patterns.

We used two types of white noise patterns for the following reasons. First, the broadband 1/f noise and the binary noise patterns differ not only in their power spectra but also in the number of grey levels and in the grey level distribution (which was Gaussian for the 1/f noise). The 8-bit Gaussian distributed white noise serves as an experimental control, as it differs from the 1/f noise by its power spectrum alone. Secondly, as all broadband images were rendered at maximum contrast, the binary noise images had greater contrast than the 8-bit Gaussian noise patterns. Comparing performance between these two conditions allowed us to investigate the possible role of image contrast in determining $d_{max}$ in broadband images.

## METHODS

### *Stimuli*

The stimuli were generated on a PC (Gateway 2000 40X2-66E) using our own software. Three types of broadband patterns were used; 1/f noise, binary noise and Gaussian noise images. The 1/f noise patterns were generated by Fourier transforming Gaussian distributed 8-bit noise images using an FFT (Press *et al.*, 1992), filtering their amplitude spectra to 1/f, and inverse transforming. The bandpass images were created by filtering the broadband images with radially symmetric Gaussian bandpass filters. The filters are described in the frequency domain as

$$H(\omega) = \exp(-(\omega - f_0)^2/(2 \times \sigma))$$

where $f_0$ is the central frequency of the filter, $\omega$ is the radial spectral frequency, and $\sigma$ is the standard deviation. The frequency bandwidths of the filters, as measured by $W$ (full width at half height, where $W = 2.35 \times \sigma$) were set equal to the central frequencies, giving a 1.6 octave bandwidth for all of the bandpass filtered images. At the viewing distance used in the experiments the broadband images had a Nyquist frequency of 16 cyc/deg (128 cyc/image with the 256 by 256 pixel image subtending 8 deg of visual angle). The central frequencies of the six bandpass images covered a five and a half octave range; the first five ranged from 0.25 to 4.0 cyc/deg in octave steps, and the sixth bandpass image had a central frequency of 5.66 cyc/deg. Examples of the three types of noise patterns are shown in Fig. 1, along with three bandpass filtered images of low, intermediate and high central frequency.

The images were generated in strips which were 2048 pixels wide by 256 pixels high. Before each experimental session, the appropriate images were loaded to the graphics card, a Visual Stimulus Generator 2/3 (Cambridge Research Systems). Six different image strips were used on a "same image" condition (e.g., bandpass–bandpass), and three different image strip pairs were used on a "different image" condition (e.g., bandpass–broadband). On each trial, a stimulus image measuring 256 by 256 pixels was chosen from a big image strip (using a randomly determined horizontal starting co-ordinate and full wrap-around) and was presented to the display. The displaced image was then chosen from either the same image strip (in the "same image" condition) or from the matching image strip (in the "different image" condition), at a starting horizontal co-ordinate offset from that of the first image by the displacement size. This method ensured that novel images were used throughout the experiment.

The broadband images were generated at maximum contrast, which corresponded to approximate root mean square (r.m.s.) values of 100% for the binary noise, 31% for the Gaussian distributed 8-bit noise, and 31% for the 1/f noise. The contrast of the bandpass filtered images was not normalised after filtering. All bandpass filtered 1/f noise patterns were therefore, on average, equal in r.m.s. contrast, whereas in the case of the bandpass

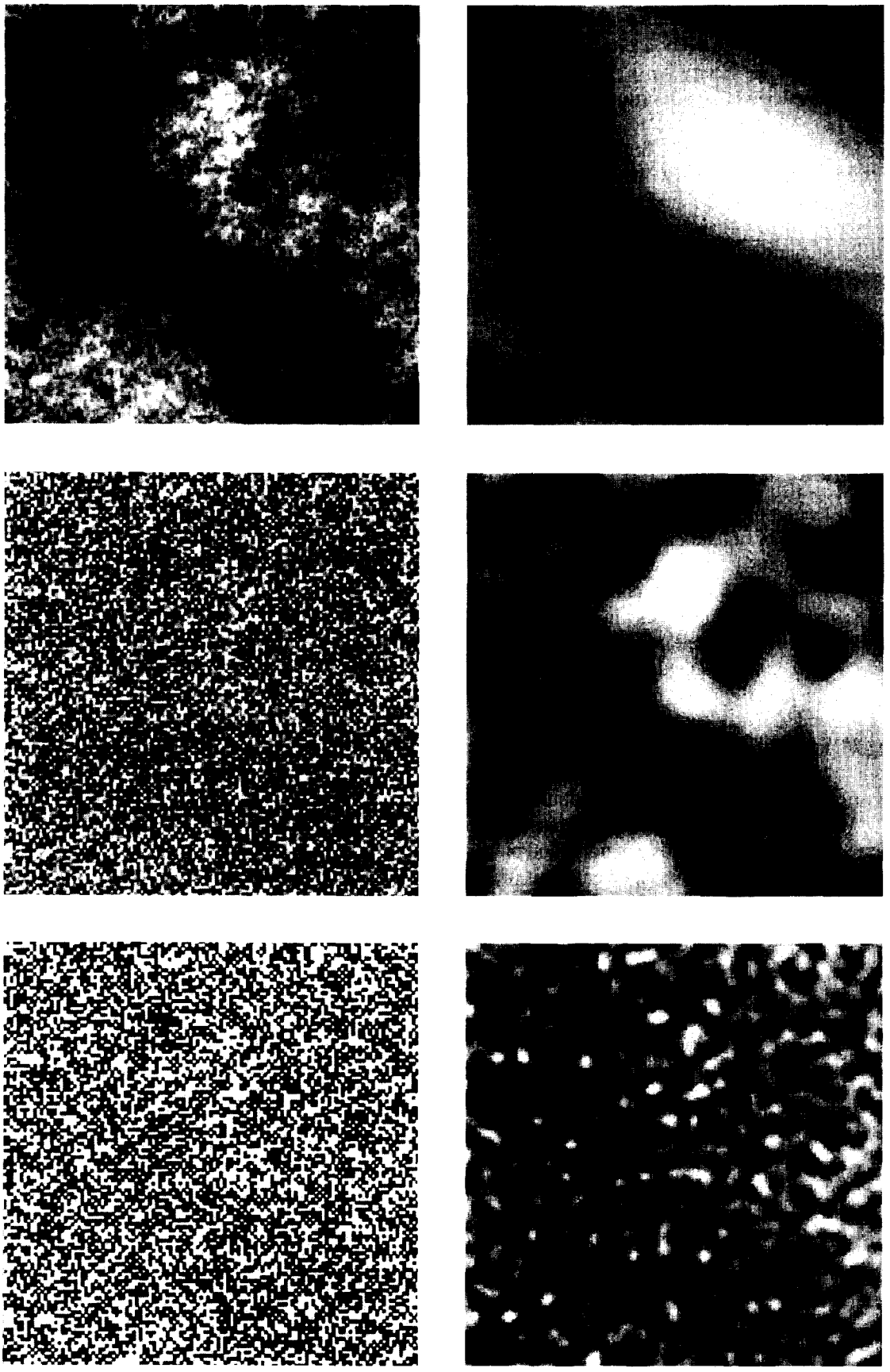

FIGURE 1. Examples of the broadband images used in this study are shown in the left panel; 1/f noise (top) Gaussian noise (middle) and binary noise (bottom). The right panel shows three bandpass filtered images of a low (top), intermediate (middle) and high (bottom) central frequency. All three broadband images in this figure were generated using the same random number sequence. The contrast of the filtered images in this figure has been normalized so they have equal physical contrast as was the case (to a close approximation) for the non-normalized contrasts of the 1/f noise bands in Experiment 1. For the Gaussian and binary noise, both the physical and the apparent contrast of the bandpass filtered images increased with the central frequency of the filter. See text and footnote overleaf for further details.

filtered 8-bit and binary noise patterns, r.m.s. contrast increased in proportion to the central frequency of the filter.*

*Display*

The stimuli were displayed on a Nanao Flexscan 6500 monitor with a frame rate of 100 Hz. The mean luminance of the display was 26 cd/m$^2$. The luminance of the display was calibrated using a UDT Photometer and linearized using a look-up table method. Subjects viewed the screen binocularly in a dimly lit room. The stimulus images were 256 by 256 pixels and subtended 8 deg of visual angle at a viewing distance of 80 cm. They were displayed in the centre of the screen and the remainder of the display was blank and at mean luminance.

*Procedure*

The same procedure was employed in all three experiments. Each trial was preceded by a blank screen of mean luminance with a small, bright fixation square in the centre. This remained on until the trial was initiated by the observer. An image was then presented for 100 msec, and the displaced image was presented for a further 100 msec without an intervening ISI. The direction of the displacement (right or left) was randomly determined on each trial, and the observer indicated the perceived direction of motion by pressing one of two keyboard buttons.

Displacement ranges were determined using the method of constant stimuli. The magnitude and direction of displacement was varied from small displacements at which there were few or no errors in direction discrimination, to large displacement at which performance fell to chance. The errors in direction discrimination were recorded at each displacement, and were later fitted with a Weibull function (Weibull, 1951) from which $d_{max}$ was estimated as the displacement at which there were 20% errors. The Weibull function was fit to the declining portion of the psychometric functions (i.e., from zero to 50% errors). Each displacement size was measured at least 40 times. Subjects had substantial practice at the task before the data were collected.

*Observers*

The observers were the first (NB) and second (PB) authors, both of whom participated in all three experiments. Two additional observers who were naïve to the purpose of the experiments also participated; CW participated in all three experiments and MAT participated in Experiments 1 and 2. All observers had normal or corrected to normal visual acuity.

*The 1/f noise patterns have a power spectrum in which power spectral density (p.s.d.) falls in proportion to the square of frequency. Therefore, there is constant power in any octave band. ("Power" or "contrast power" is the integral of the p.s.d. over all spatial frequencies). The flat spectrum images have, on average, equal p.s.d. at all frequencies, so that the power in any octave band increases in proportion to the linear bandwidth of the filters. By Parseval's theorem, the mean square value of an image is equal to the image power (See e.g., Bracewell, 1986). This explains why image variance or r.m.s. contrast is approximately equal for all bandpass filtered 1/f images, but increases as a function of the central frequency (or equivalently as a function of the linear bandwidth) of the bandpass filter in the case of the white noise images.

## RESULTS AND DISCUSSION

Three types of noise pattern were used in this study; 1/f noise (Experiment 1), Gaussian noise (Experiment 2) and binary noise (Experiment 3). There were three conditions in each experiment; subjects were required to identify the direction of motion in a two-flash kinematogram in which both images were bandpass, both were broadband, or in which the first image was bandpass and the second was broadband. Additional data were collected for the reverse presentation (broadband–bandpass) for subject NB. Six central frequencies of the bandpass filtered images were tested.

Figure 2 compares performance between the bandpass–bandpass and bandpass–broadband conditions for all three experiments. The percentage of correct responses is plotted against image displacement, expressed in degrees of visual angle. Data are shown for three of the six central frequencies tested (0.5, 1.0 and 4.0 cyc/deg) for subject NB. (Data from the other subjects were similar.) In agreement with earlier studies (Chang & Julesz, 1983; Bischof & DiLollo, 1990), the range of displacements over which direction discrimination is possible decreased with increasing central frequency of the bandpass. For the 1/f noise patterns (Experiment 1), performance was similar in the two conditions at all spatial frequencies. The range of displacements over which subjects could discriminate the direction of motion when only the first image in the sequence was bandpass filtered was roughly the same as when both images were filtered. The presence of additional high spatial frequencies in the broadband image did not prevent subjects from using the lower frequencies common to both images. This was not the case for the other noise patterns. In Experiment 2 which used Gaussian noise images, performance was similar across conditions only at the highest central frequencies of the bandpass image. At lower central frequencies of the bandpass images, motion discrimination was possible out to higher displacements in the bandpass–bandpass condition. This trend was even more marked for the binary noise patterns (Experiment 3). As shown in the bottom row of Fig. 2, performance in the bandpass–broadband condition at 0.5 cyc/deg did not rise very far above chance for much of the displacement range. In summary, there was a clear difference in performance on the direction discrimination task for images which differed in their spatial characteristics. For noise patterns with a 1/f amplitude spectrum, the upper displacement for motion discrimination was determined by the lowest spatial frequencies common to the image sequence, independently of image bandwidth. However, for noise patterns with a flat spectrum, the presence of high spatial

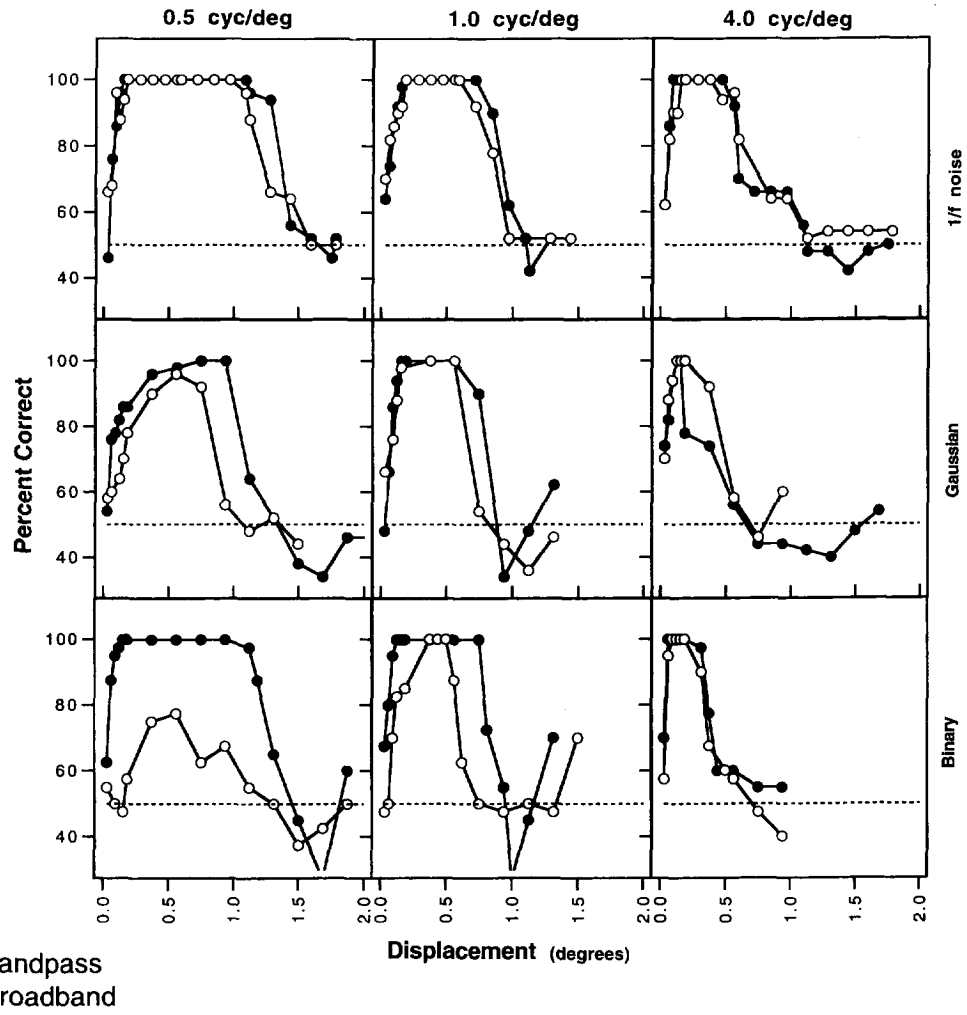


FIGURE 2. Percentage of correct responses plotted against image displacement (expressed in degrees of visual angle). The bandpass–bandpass data are shown by the filled circles, and the bandpass–broadband data by the open circles. Data are shown for three central frequencies of the bandpass filters (0.5, 1 and 4 cyc/deg) from left to right, and for the three noise types (1/f, Gaussian and binary) from top to bottom. The dashed lines indicate 50% correct. All data are from subject NB.

frequencies in one of the images somehow prevented the use of low spatial frequency information in performing the task.

The same conclusion follows when we compare performance between the bandpass–bandpass and broadband–broadband conditions. The data for the first five frequencies in the bandpass–bandpass condition are plotted in Fig. 3. Superimposed on these data are the results from the broadband–broadband condition in which the two-flash kinematogram consisted of a broadband image and a displaced version of itself. In the case of the 1/f noise, the broadband data effectively envelope the bandpass data, suggesting that the motion system has access to information across a wide range of spatial scales in these images. The only exception is at the very lowest frequency (0.25 cyc/deg), where discrimination was possible out to bigger displacements in the bandpass–bandpass condition. For the Gaussian noise patterns, the broadband data encompass the bandpass data only for central frequencies greater than 1 cyc/deg. Note, however, that there are no data for the lowest frequency. At a central frequency of 0.25 cyc/deg, the (non-normalized) contrast of the 1.6 octave bandpass image was too low to allow direction discrimination at all displacements. In the case of the binary noise images, the broadband data envelope the bandpass data only for central frequencies of 2 cyc/deg and higher, suggesting that the lower frequencies are unavailable to the motion system in these broadband images.

The data from all conditions of the three experiments are summarized in Fig. 4 for all four subjects. The upper threshold for direction discrimination is plotted against the central frequency of the bandpass filtered images. The plots also include estimates of $d_{max}$ for the broadband images, shown by the dashed lines. The absence of a data point indicates conditions in which the subject could not reliably perform direction discrimination. For all three noise patterns, $d_{max}$ initially decreased with increasing spatial frequency and then levelled off at frequencies beyond about 4 cyc/deg, consistent with previous findings (Cleary & Braddick, 1990b; Morgan, 1992). When only one of the images in the RDK was filtered, there is

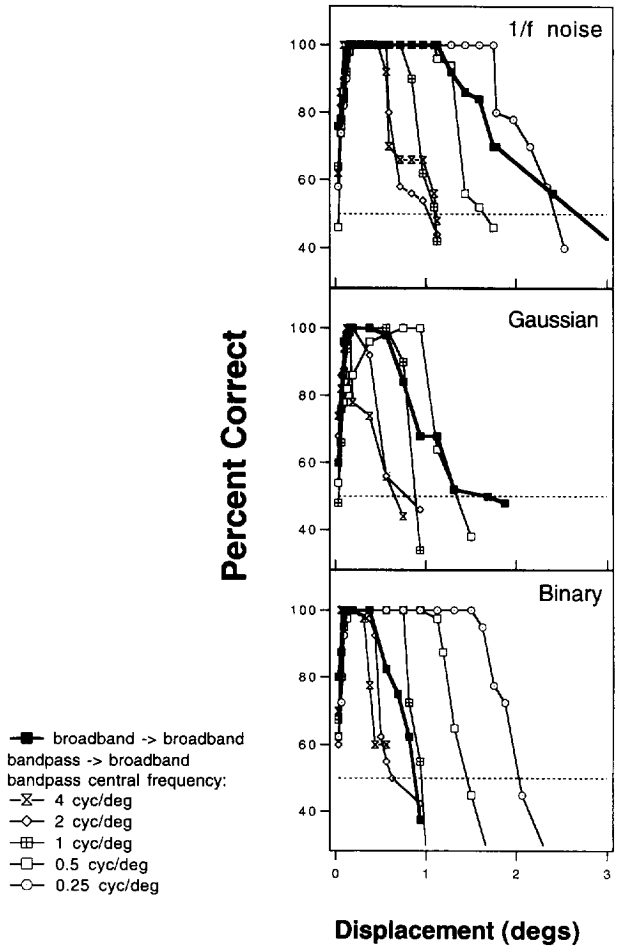


FIGURE 3. Percentage of correct responses plotted against image displacement for all bandpass–bandpass conditions (open symbols) and for the broadband–broadband conditions (filled symbols). Data are shown for the three noise types separately (1/f, Gaussian and binary) from top to bottom. All data are from subject NB.

some suggestion at the low frequencies that $d_{max}$ is higher when the bandpass image was presented first [Fig. 4(a)]. At the low spatial frequencies, this is consistent with Morgan and Mather's (1993) finding for lowpass filtered binary noise images.

There was a clear difference in subjects' ability to discriminate the direction of motion across the three experiments. In the bandpass–broadband condition, subjects could perform the task for all frequencies in the case of the 1/f images, and except for the lowest frequency of 0.25 cyc/deg, $d_{max}$ values were close to those found in the bandpass–bandpass condition. For the Gaussian noise patterns, either $d_{max}$ was larger in the bandpass–bandpass condition at the lowest spatial frequencies or direction discrimination was not possible. The most striking difference in performance across conditions was seen for the binary noise patterns. For subject PB, $d_{max}$ was lower over the entire range of frequencies in the bandpass–broadband condition, and for subject CW values of $d_{max}$ only converged at the highest spatial frequency. Direction discrimination collapsed for subjects NB, PB and CW at the lowest spatial frequencies. As shown by the dashed lines in the plots, direction discrimination was possible out to much larger displacements for the 1/f noise patterns than for either of the flat spectrum noise patterns, and $d_{max}$ was slightly higher for Gaussian noise than for binary noise for all three subjects who participated in all experiments.

*General discussion*

The results of these experiments reveal clear differences in observers' ability to discriminate direction of motion for patterns which differ in their spatial characteristics. Consistent with previous findings (Chang & Julesz, 1983; Cleary & Braddick, 1990a; Morgan, 1992), the upper displacement limit for direction discrimination was found to vary inversely with central spatial frequency when the motion sequence consisted of pairs of bandpass filtered noise patterns. For the mixed motion sequences of bandpass filtered and broadband white noise, our data are consistent with previous data showing that direction discrimination is not possible between broadband binary noise images and lowpass images with a low cut-off frequency (Morgan & Mather, 1993). However, for 1/f noise images, observers could discriminate the direction of motion at all central frequencies of the bandpass patterns in the mixed motion sequences, and except for the very lowest central frequency (0.25 cyc/deg), $d_{max}$ values were similar to those obtained with pairs of bandpass filtered images. This suggests that while low spatial frequency information does not determine $d_{max}$ in images which have flat amplitude spectra, information at a wide range of spatial scales is equally salient to the human motion system in images whose amplitude spectra fall as 1/f.

These findings are compatible with a model of neural sensitivity to contrast in which cells tuned to different spatial frequencies have bandwidths which are constant in octaves and equal peak gain. This model of sensitivity has been proposed previously to explain the visual response to natural scenes (Field, 1987) and to account for contrast constancy (Brady & Field, 1995). Recent physiological data are supportive. For example, the frequency band over which cortical cells integrate information scales with the optimal spatial frequency, so as to produce bandwidths which are roughly constant on a log scale (DeValois *et al.*, 1982; Tolhurst & Thompson, 1981). Although octave bandwidths do decrease somewhat with frequency, it is important to note that this emerges when the mean estimated bandwidth is plotted as a function of optimal frequency. Given the large variability in bandwidth estimates at each optimal frequency (e.g., DeValois *et al.*, 1982), an assumption of equal octave bandwidths seems quite reasonable. Data on the absolute sensitivity of cortical cells tuned to different frequencies are harder to come by because such data are usually normalized. However, in a recent study of the receptive field properties of both M and P primate retinal ganglion cells, Croner & Kaplan (1994) showed that the peak sensitivity (in space) scales inversely with receptive field size so that the volume under the receptive field remains approximately constant,

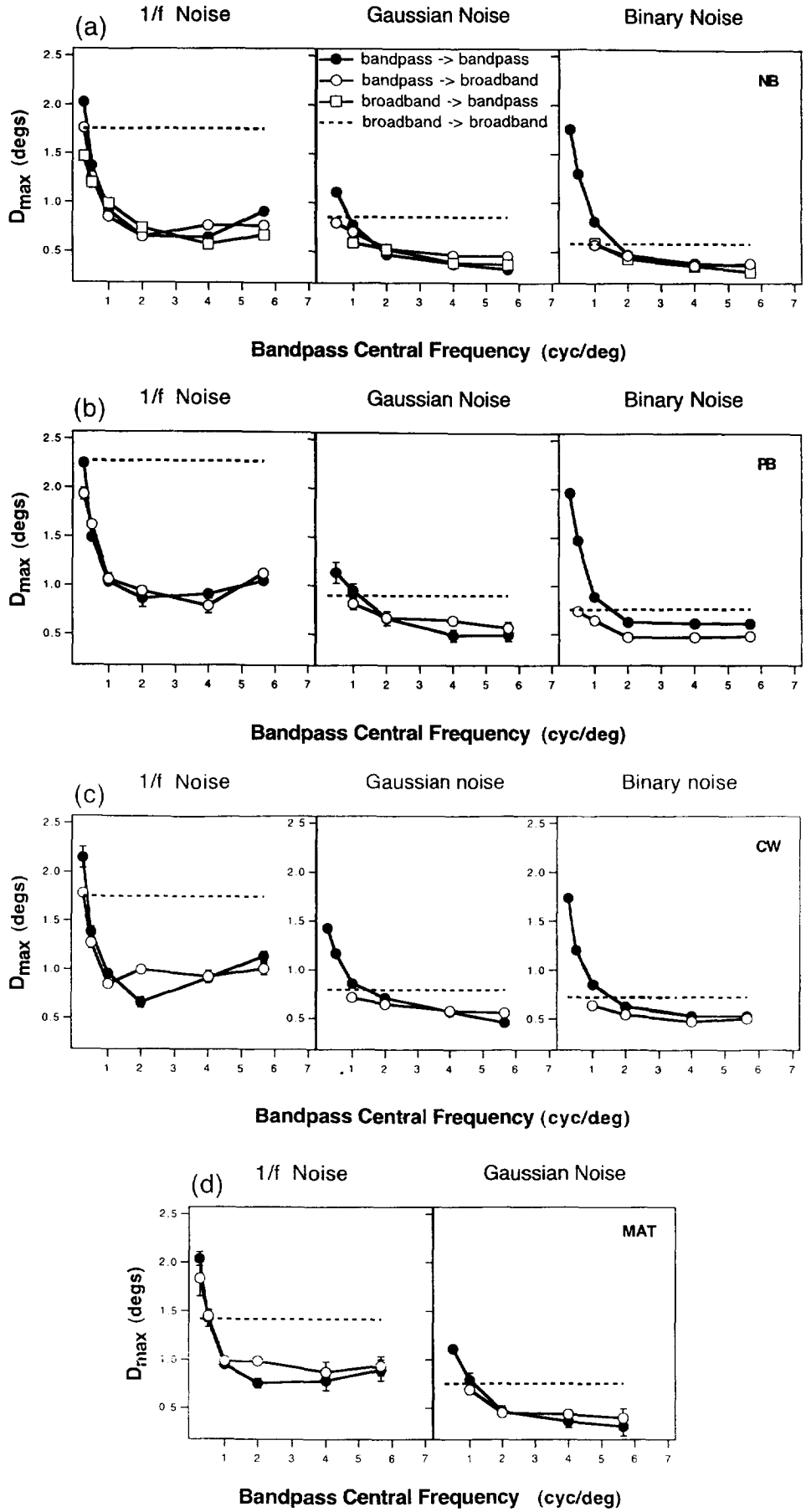


FIGURE 4. $d_{max}$ Plotted against the central spatial frequency of the bandpass images for the 1/f, Gaussian and binary noise (left to right). Bandpass–bandpass data are shown by the filled circles, bandpass–broadband data by the open circles, broadband–bandpass data by the open squares, and broadband–broadband data by the dashed lines. All $d_{max}$ values are the average of four or five estimates. Error bars, where visible, show $\pm 1$ SE. Data are from subjects (a) NB, (b) PB, (c) CW and (d) MAT.

independently of receptive field size. These data provide direct support for the proposal that motion sensitive cells tuned to different spatial frequencies respond with approximately equal activity to their preferred spatial frequencies.

In introducing the ecological reasons for this model of sensitivity, our discussion focused exclusively on the spatial properties of static images. This ecological approach has been recently applied to the study of motion perception, providing evidence that various aspects of the visual code are matched to the spatio-temporal properties of natural scenes (Eckert & Buchsbaum, 1993; van Hateren, 1993). Also relevant are Kelly's (1979) studies of threshold sensitivity to moving gratings of constant velocity (i.e., constant temporal to spatial frequency ratio). His estimates of the line spread functions of channels tuned to different spatial frequencies (derived by Fourier transforming the constant-velocity tuning curves) are particularly interesting. Their spatial characteristics are very similar to those of the retinal ganglion cells described by Croner & Kaplan (1995). Kelly (1979) describes the line (or point) spread functions as having a peak sensitivity which decreases with increasing width (or area) so that the area (or volume) under the functions remains roughly constant for different sizes. For linear units, this implies that the peak responses of channels tuned to different spatial frequencies are roughly constant. Psychophysical estimates of channel bandwidths from motion detection studies (Anderson & Burr, 1985) indicate an increase in linear bandwidth with spatial frequency, but fall somewhat short of octave constancy over the entire frequency range, becoming narrower at high spatial frequencies. What are the implications of findings of departures from octave constancy for the model discussed above? It is perhaps useful to reiterate that the way in which the visual response to image contrast scales as a function of spatial frequency depends on both the peak response of the neural filters and their bandwidths. If the peak response of spatial frequency filters is approximately equal across scale, and frequency bandwidths increase linearly (although falling short of octave constancy) then the response to a flat spectrum input will increase as a function of frequency, albeit not in direct proportion to spatial frequency as suggested above.

As noted in the Introduction, there have previously been two different explanations as to why the upper displacement limit for direction discrimination in binary noise images is smaller than predicted from the presence of low frequencies in these images. Either low frequency information is masked or inhibited by high frequency information (Cleary & Braddick, 1990b), or alternatively $d_{max}$ is limited by the mean separation between features in an image (Morgan & Mather, 1993). The difference between these models has been summarized by Morgan and Mather. According to the first view, motion analysis takes places in independent frequency-tuned channels, the responses of which are later combined to form a decision about the direction of motion. Alternatively, the outputs of channels tuned to different spatial frequencies are combined prior to motion analysis to localize features in an image, and motion analysis then proceeds by a feature matching process. Bex *et al.* (1995) recently argued in favour of the former interpretation. Using kinematograms which consisted of a 1/f noise pattern and a lowpass filtered version of that pattern, they showed that subjects could identify the direction of motion for a series of upper cut-off frequencies of the lowpass pattern ranging from 0.125 to 16 cyc/deg. Furthermore, $d_{max}$ remained approximately constant over this range, suggesting that the upper spatial limit for direction discrimination is determined by the lowest spatial frequencies in these patterns. The results of the present study strongly support a multi-channel model interpretation, in which subjects have independent access to information at different scales in natural images, and can use whatever information is pertinent to the task at hand. They also support the arguments of Bex *et al.* (1995). Our observers could discriminate the direction of apparent motion in the bandpass–broadband task at all central frequencies of the bandpass images for the 1/f noise. As can be seen from Fig. 1, these include cases in which the phenomenal separation of features in the bandpass image is much smaller than the phenomenal separation in the broadband image.

Assuming that the spatial properties of motion sensitive cells approximate those described in the model, the proposed masking effects might be expected simply because the high frequency response to flat spectrum images is much stronger than the corresponding low frequency response. A question still remains as to the nature of the high frequency response in these displays. As Morgan & Mather (1993) point out, the bandpass–broadband kinematograms do not contain motion signals at high spatial frequencies when these are present in only one image. However, the high spatial frequencies are temporally modulated by their abrupt presentation, and it is known that direction selective cells in V1 respond to flicker (Hubel & Wiesel, 1962). The activation of high frequency cells by these transients may be sufficient for masking; a motion signal at high spatial frequencies may not be required.

The central hypothesis of this paper, that the spatial characteristics of motion sensitive cells are matched to the properties of natural scenes, may also be of relevance to feature-based models of motion detection. Feature-based models, such as that discussed by Morgan & Mather (1993), involve an initial stage of processing in which image features are "localized" or "extracted" from the outputs of channels tuned to different spatial scales. If the relative sensitivities of these channels approximate those in the model, then 1/f and white noise patterns will produce different distributions of activity in these channels as described above. Depending on the rules by which channel outputs are combined, "feature-based" models could predict a relatively smaller $d_{max}$ for white noise images, given that these images produce a distribution of response activity which is biased toward

the high spatial frequencies. For example, the comparatively low response from low frequency channels might be ignored in computing image features.

Whatever the final explanation, the distribution of activity in cells or filters tuned to different spatial frequencies remains a crucial factor in determining how information from different scales is utilized by the motion system, and this distribution depends crucially not just on the image statistics *per se* (e.g., Morgan, 1995) but on how the visual system integrates information at different scales. All evidence suggests that these filters integrate information across a band of spatial frequencies which increases with the filter's central frequency so as to produce a distributed response to images whose power spectra fall as $1/f^2$. Information at the high spatial frequencies will be more salient in images whose spectra are considerably shallower, as shown in our experiments with white noise patterns. Similarly, information at low spatial frequencies would be expected to dominate performance in patterns whose power spectra are considerably steeper than $1/f^2$. Interestingly, an analogous phenomenon occurs in form perception. For example, both the Gaussian and binary noise images in Fig. 1 appear to be dominated by high frequency structure, whereas the 1/f image appears to have structure at a variety of spatial scales.

Our data do deviate somewhat from our predictions in the case of the 1/f noise patterns, in that all three subjects show a $d_{max}$ value which is somewhat smaller in the bandpass–broadband than in the bandpass–bandpass condition at the lowest central frequency tested (0.25 cyc/deg). Sensitivity to suprathreshold contrast must decline at both very high and very low spatial frequencies, and a possible explanation for this result is that we have reached the lower limits. For example, it may be the case that the bandpass noise pattern centred at 0.25 cyc/deg maximally excites a channel which is tuned to a slightly higher frequency, but that the response activity in this channel is much less than the response in the next highest channel when a broadband pattern is used. That is, $d_{max}$ may be determined by different channels in these two conditions simply because our lowest frequency stimulus has a peak frequency which is lower than the optimal frequency of the lowest channel.

A final point of interest is the difference in subjects' performance for the broadband 8-bit Gaussian noise and the binary noise images. The broadband images were rendered at maximum contrast, so that the binary noise images had greater contrast by virtue of their greater variance. This difference in image contrast clearly results in a difference in performance at the lowest spatial frequency (0.25 cyc/deg) in the bandpass–bandpass condition. While contrast at this bandpass frequency was too low to allow direction discrimination for three of the four subjects in the case of Gaussian noise, this was clearly not the case for the binary noise patterns where discrimination was possible out to large spatial displacements at the lowest spatial scale. In the case of Gaussian noise, it could be argued that performance breaks down at the lowest spatial frequencies largely because a 1.6 octave band of Gaussian noise (rendered as it was in these experiments) is so low in physical contrast as to be close to threshold. Additional factors must be at play in the case of the binary noise patterns, and are necessary to explain why subjects showed a higher $d_{max}$ for the broadband Gaussian noise patterns than for the binary noise patterns.

## CONCLUSIONS

We have shown that information at a broad range of spatial scales is equally salient to the human motion system in images whose power spectra are similar to those of natural scenes. However, in the case of white noise patterns, information at low spatial frequencies is much less salient to the motion system. These findings are consistent with a model of sensitivity in which filters tuned to different spatial frequencies have octave constant bandwidths and equal peak gain. We propose that the relation between this sensitivity profile and the power spectra of broadband patterns is a crucial factor in determining the upper threshold for direction discrimination in apparent motion.

It has previously been argued that the scaling of neural sensitivity to contrast in the mammalian visual system makes ecological sense, given the spatial characteristics of natural scenes. Such images will, on average, produce equal activity in cells tuned to different frequencies (Field, 1987), and this sensitivity profile provides for contrast constancy in natural images (Brady & Field, 1995). From the results of the present experiments, we can also conclude that the spatial properties of motion sensitive cells appear to be well suited to coding motion in natural scenes. The mechanism which allows for a constant response to information at different spatial scales in natural images may, therefore, be a general feature of early visual processing.

## REFERENCES

Adelson, E. H. & Bergen, J. R. (1985). Spatio-temporal energy models for the perception of motion. *Journal of the Optical Society of America A, 2*, 284–299.

Anderson, S. J. & Burr, D. C. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Research, 25*, 1147–1154.

Atick, J. J. & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation, 4*, 196–210.

Baker, C. L. & Braddick, O. J. (1982). The basis of area and dot number effects in random dot motion perception. *Vision Research, 22*, 1253–1259.

Baker, C. L. & Braddick, O. J. (1985). Eccentricity-dependent scaling of the limit for short-range apparent motion perception. *Vision Research, 25*, 803–812.

Baker, C. L. & Cynader, M. S. (1986). Spatial receptive-field properties of direction selective neurons on cat striate cortex. *Journal of Neurophysiology, 55*, 1136–1152.

Bex, P. J., Brady, N., Fredericksen, R. E. & Hess, R. F. (1995). Energetic motion detection. *Nature, 378*, 670–671.

Bischof, W. F. & DiLollo, 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.

Bracewell, R. N. (1986). *The Fourier transform and its applications* (2nd edition). McGraw-Hill.

Braddick, O. J. (1974). A short-range process in apparent motion. *Vision Research, 14*, 519–527.

Brady, N. & Field, D. J. (1995). What's constant in contrast constancy? The effects of scaling on the perceived contrast of band-pass patterns. *Vision Research, 35*, 739–756.

Burton, G. J. & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics, 26*, 157–170.

Chang, J. J. & Julesz, B. (1983). Displacement limits for spatial frequency filtered random-dot cinematograms in apparent motion. *Vision Research, 23*, 1379–1385.

Cleary, R. & Braddick, O. J. (1990a) Direction discrimination for band-pass filtered random dot kinematograms. *Vision Research, 23*, 303–316.

Cleary, R. & Braddick, O. J. (1990b) Masking of low frequency information in short-range apparent motion. *Vision Research, 30*, 317–327.

Croner, L. J. & Kaplan, E. (1995). Receptive fields of P and M ganglion cells across the primate retina. *Vision Research, 35*, 7–24.

DeValois, R. L., Albrecht, D. C. & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22*, 545–559.

Eckert, M. P. & Buchsbaum, G. (1993). Efficient coding of natural time varying images in the early visual system. *Philosophical Transactions of the Royal Society of London B, 339*, 385–395.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, 4*, 2379–2394.

Field, D. J. (1993). Scale-invariance and self-similar "wavelet" transforms: an analysis of natural scenes and mammalian visual systems. In Frage, M. Hunt, J. & Vassilicos (Eds), *Wavelets, fractals and Fourier transforms*. Oxford: Oxford University Press.

van de Grind, W. A., Koenderink, J. J. & van Doorn, A. J. (1986). The distribution of human motion detector properties in the monocular visual field. *Vision Research, 26*, 797–810.

van Hateren, J. H. (1993). Spatiotemporal contrast sensitivity of early vision. *Vision Research, 33*, 257–267.

Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat visual cortex. *Journal of Physiology (London), 160*, 106–154.

Kelly, D. H. (1979). Motion and vision. 2. Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America, 6910*, 1340–1349.

Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature, 355*, 344–346.

Morgan, M. J. (1995). *Nature, 378*, 671–672.

Morgan, M. J. & Mather, G. (1993). Motion discrimination in two-frame sequences with differing spatial frequency content. *Vision Research, 34*, 197–208.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical recipes in C* (2nd edition). Cambridge, U.K.: Cambridge University Press.

vanSanten, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A, 2*, 300–321.

Tolhurst, D. J. & Thompson, I. D. (1981). On the variety of spatial frequency selectivities shown by neurones in area 17 of the cat. *Proceedings of the Royal Society of London B, 213*, 183–199.

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

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics, 18*, 292–297.

*Acknowledgements*—Supported by Canadian NSERC grant OGP0001978 to C. L. Baker and Canadian NSERC grant OGP0046528 to R. F. Hess. We would like to thank Curtis L. Baker for useful discussions of this work, and two anonymous reviewers for comments on an earlier manuscript. We also thank Cristyn Williams and Mary Ann Thomas for participating in the experiments.