Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 46 (2006) 556–567
Vision Research
www.elsevier.com/locate/visres

Is the direction of second-order, contrast-defined motion patterns visible to standard motion-energy detectors: A model answer?

Timothy Ledgeway *, Claire V. Hutchinson

School of Psychology, University of Nottingham, University Park NG7 2RD, UK

Received 14 February 2005; received in revised form 23 June 2005

Abstract

Previous psychophysical studies (e.g., Smith & Ledgeway, 1997) have provided evidence that under some conditions, the detection of a particular class of stimuli (contrast-modulated static noise) widely employed to study second-order motion processing may be inadvertently based on encoding local imbalances in luminance motion energy. In particular when static noise composed of relatively large noise elements is used, direction-identification performance at threshold may actually be mediated by the same mechanisms that respond to first-order motion, due to the presence of persistent spatial clusters of noise elements of the same polarity. However, Benton and Johnston (1997) modeled the responses of conventional motion-energy detectors to contrast-modulated static noise patterns and found no evidence of any systematic directional biases in such stimuli when the mean opponent motion energy was used to quantify performance. In the present paper we sought to resolve this discrepancy and show that the precise manner in which computational models are implemented is crucial in determining their response to contrast-modulated, second-order motion patterns. In particular we demonstrate that by considering the information encapsulated by the peak (rather than the mean) opponent motion energy and the predominantly local nature of imbalances in motion energy that can arise in contrast-modulated static noise, it is possible to readily model the patterns of empirical results found.

Keywords: Computational modeling; First-order motion; Second-order motion; Motion-energy detectors

1. Introduction

It is well established that the visual system is adept at encoding movement conveyed by a diverse range of image characteristics including luminance variations (“first-order motion”) and textural differences such as contrast variations (“second-order motion”) (e.g., Cavanagh & Mather, 1989). Chubb and Sperling (1988) applied the term “non-Fourier” to second-order motion to emphasise the point that the direction of such motion is not explicitly conveyed by its Fourier power spectrum unlike first-order motion. That is, the Fourier spectra of many, though not all, second-order motion stimuli contain no net motion energy (any they contain is equal in opposite directions). Indeed Chubb and Sperling (1988, 1991) described many second-order motion stimuli of this type (e.g., a drifting modulation in the contrast of a random noise field), which they termed “drift-balanced” and provided formal mathematical proofs that they were also “micro-balanced”. That is, remain drift-balanced even after any arbitrary spatiotemporal filtering by the

* Corresponding author. Tel.: +44 0 115 8467343; fax: +44 0 115 9515324.
E-mail address: txl@psychology.nottingham.ac.uk (T. Ledgeway).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.07.008

visual system. Consequently these stimuli have proved to be invaluable for studying the underlying basis of second-order motion perception.

Using modeling techniques Johnston and colleagues (e.g., Benton, 2002; Benton & Johnston, 2001; Benton, Johnston, McOwan, & Victor, 2001; Johnston, McOwan, & Benton, 1999; Johnston, McOwan, & Buxton, 1992) have shown that first-order motion and some types of second-order motion could be detected by the same (common) mechanism. However the balance of empirical evidence (e.g., for reviews see Baker, 1999; Smith, 1994; Sperling & Lu, 1998) overwhelmingly suggests that the two varieties of motion are each encoded, at least initially, by distinct (separate) visual mechanisms.

One line of evidence that suggests that first-order motion and second-order motion are detected by distinct visual mechanisms comes from studies that have compared threshold sensitivity for identifying the spatial form (e.g., orientation) and the direction of moving patterns. These studies have measured the minimum stimulus modulation depth (threshold amplitude) needed to identify accurately both the orientation and the drift direction of first-order motion patterns (luminance-defined, sinusoidal gratings) and second-order motion patterns (sinusoidal modulations of the contrast of a 2-d noise carrier). In the case of first-order motion stimuli it is well established that the two thresholds are the same, except at very low temporal frequencies and high spatial frequencies (e.g., Green, 1983; Watson, Thompson, Murphy, & Nachmias, 1980). For second-order motion patterns, however, thresholds for identifying direction are typically consistently higher (by ~50%) than those for orientation over a wide range of drift rates (Smith & Ledgeway, 1997, 1998). This suggests that first-order motion and second-order motion are each encoded separately and that the mechanism that processes second-order stimuli cannot determine direction at the absolute threshold for spatial form.

This difference in thresholds has been considered a characteristic signature of second-order motion-detecting mechanisms (Ledgeway & Hess, 2002; Ledgeway & Hutchinson, 2005; Smith & Ledgeway, 1997, 1998; Smith & Scott-Samuel, 1998) and is most readily seen when dynamic visual noise is used as a carrier. Interestingly the two thresholds tend to converge (i.e., are the same as they are for first-order motion) when static noise carriers composed of relatively large ($\geqslant$~4 arcmin) noise elements are used. This has been taken as evidence that when the contrast of a 2-d, static carrier composed of 'large' noise elements is modulated by a drifting sinusoidal waveform, spatially extensive and persistent clusters of noise elements with the same luminance polarity in the image give rise to systematic local directional biases that are visible to conventional motion sensors (Smith & Ledgeway, 1997).[1] The use of either static noise carriers composed of small elements (with no spatial variation in luminance within each element) or dynamic noise carriers, may minimise the presence of these potential luminance-based motion cues in second-order motion patterns.

Interestingly Benton and Johnston (1997) found no evidence for the existence of a consistent directional imbalance in the output of an (idealised) opponent motion-energy detector (Adelson & Bergen, 1985) that was applied to space–time (x–t) images representing drifting contrast modulations of noise carriers. They applied a standard energy model to 100 instantiations of contrast-modulated static and dynamic noise stimuli. The energy model utilised two quadrature pairs of x–t oriented Gabor filters which were maximally sensitive (tuned) to the same absolute spatial and temporal frequencies (matched to those of the drifting contrast waveform) but responded best to motion in opposite directions. They squared and then added the outputs within each quadrature pair to give a directionally sensitive and phase invariant measure of local motion energy. By subtracting the outputs of the two motion sensors at each point in space and each instant in time they derived a local measure of opponent motion energy. Benton and Johnston (1997) then pooled the output measures over all 100 instantiations (motion sequences) and calculated the overall mean opponent energy and standard deviation of responses. Although the mean response of the model exhibited greater variability as the size of the individual elements comprising the noise carriers increased (for both static and dynamic noise), there was no systematic net directional bias that would be indicative of the presence of a dominant luminance-based motion signal.

On the basis of their modeling results Benton and Johnston (1997) proposed an alternative explanation of why direction thresholds are typically higher than orientation thresholds for second-order motion stimuli based on the proportion of energy in the Fourier spectra of the images that contains motion direction information. They showed that dynamic noise carriers generate approximately twice as much "motion direction noise" as static noise carriers. The increased motion direction noise associated with a dynamic, compared with a static,

[1] These stimuli are still strictly micro-balanced in that for any given instantiation the expected motion energy in opposing directions is equal when averaged across the *entire* spatiotemporal extent of the stimulus. However stochastic clusters of same polarity static noise elements can lead to spatially *localised* imbalances in motion energy in the same direction as the contrast modulation. Although previous studies have referred to such local directional biases as "artifacts" (e.g., Smith & Ledgeway, 1997), they are an intrinsic property of contrast-modulated static noise patterns and so that terminology will not be adopted in the present manuscript (we thank an anonymous reviewer for this suggestion).

noise carrier could cause increased thresholds on a direction-identification task, but may not necessarily affect performance on an orientation-identification task. Thus they claimed that differences in performance between contrast modulations of static noise and dynamic noise may reflect an interaction (a selective motion masking effect) between the nature of the threshold tasks used and the nature of the stimuli, rather than the operation of two distinct motion-detecting systems (i.e., a system responding to local, luminance motion energy and another separate motion system sensitive to drifting, second-order, contrast modulations).

However this theory fails to explain why direction-identification thresholds can be higher than thresholds for identifying orientation even when static (as opposed to dynamic) noise carriers are used, provided that the noise elements are small relative to the spatial wavelength of the contrast modulation (Smith & Ledgeway, 1997, 1998). We have recently confirmed this important result under a range of conditions (Ledgeway & Hutchinson, 2005) and some key data from that study are replotted in Fig. 1 to illustrate this phenomenon. Hence the reason for the discrepancy between these empirical findings and the modeling results of Benton and Johnston (1997) is presently unclear, and warrants further investigation.

There are three key aspects of Benton and Johnston's (1997) model implementation, however, that may have led to the inability to detect the direction of stimulus motion when static noise carriers composed of large elements were used. First, the response metric was collapsed over both the area of the x–t input images and instantiations, rather than allowing a decision to be made as regards to motion direction for each individual motion sequence. A more realistic model would need to incorporate an explicit decision stage that would determine the overall direction of each individual motion sequence. Second, the assumption that the overall mean opponent motion energy is the most appropriate measure for characterising local directional bias may not necessarily be valid (although it does establish whether the stimulus is micro-balanced or not). In a direction-identification threshold task the most visible motion components will presumably be those that reach threshold first, and thus it may be more judicious to assess the magnitude of the peak opponent motion energy in each image and use this to derive a motion direction judgement. Third, imbalances in motion energy that potentially arise due to spatial clumping of noise elements in contrast-modulated static noise carriers are by definition inherently local and may be most readily detected when motion-energy filters with a range of sizes are used. As the frequency bandwidth of any filter is inversely related to its spatiotemporal extent (the region of space and time it samples), it may be pertinent to vary the Gabor bandwidth because motion sensors with larger bandwidths may be crucially more sensitive to the presence of local, luminance-based, directional biases. Benton and Johnston (1997) did not specify the bandwidths of the spatial and temporal filters used to construct their motion-energy detectors.

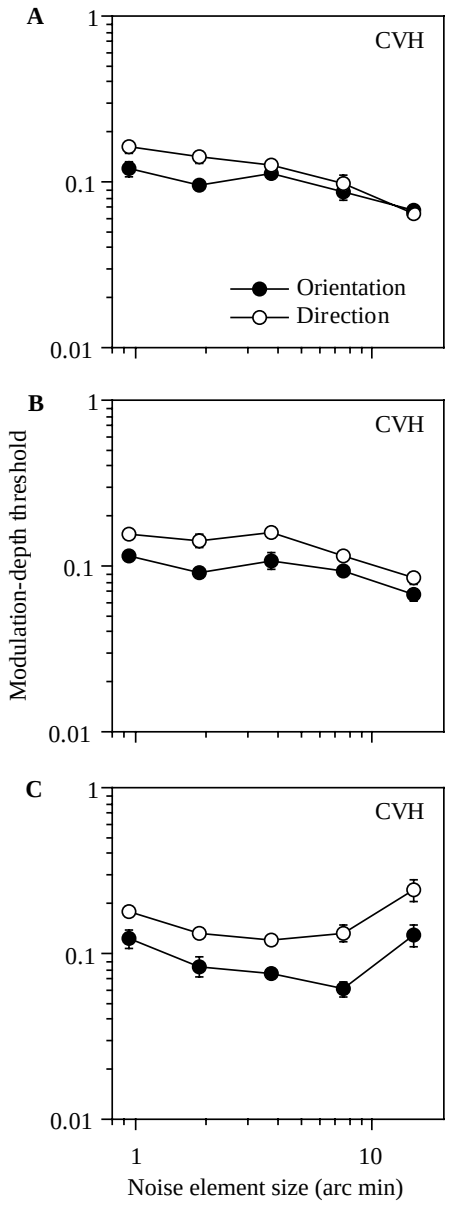

Fig. 1. Modulation-depth thresholds (corresponding to 75% correct performance) for one observer (CVH one of the authors) for identifying the spatial orientation (filled symbols) and drift direction (unfilled symbols) of contrast-modulated static noise when (A) luminance was allowed to vary within each element of the noise carrier (cf. Smith & Ledgeway, 1997) and (B) luminance did not vary within each noise element. (C) Modulation-depth thresholds are plotted for identifying the orientation and drift direction of contrast-modulated dynamic noise. Testing was carried out over a range of carrier noise element sizes (0.94–15 arcmin). The spatial frequency and temporal frequency of the drifting modulation were 1 c/deg and 1 Hz, respectively. The mean Michelson contrast of the 2-d noise carrier was 0.15. Error bars above and below each datum represent ± 1 SEM.

To address these issues contrast-modulated noise patterns were subjected to conventional motion-energy analysis but (1) a simple decision rule for assessing motion direction was implemented after the presentation of each motion sequence and this was based on (2) either the mean opponent motion energy in the stimulus (cf. Benton & Johnston, 1997) or the peak opponent energy (a "winner-takes-all strategy"). (3) Furthermore the model's direction-identification performance was measured for a range of bandwidths of its underlying spatial and temporal frequency filters.

2. Methods

2.1. Apparatus

Modeling was performed on an *Apple Macintosh G4* computer running custom software written in the C programming language.

2.2. Input images

x–t plots representing drifting modulations of the contrast of either static or dynamic visual noise served as input images to the model (see Fig. 2 for examples). Each *x–t* plot was composed of a 512×512 square pixel array so that resolution was comparable in both dimensions. The spatial and temporal wavelengths of the contrast modulation were directly analogous to those used to obtain the empirical results shown in Figs. 1B and C and were set to 64 pixels/cycle (nominally 1 c/deg) and 75 pixels/cycle (nominally 1 Hz), respectively. The initial spatial phase of the contrast waveform was randomised for each instantiation, the modulation depth was always unity and drift direction was leftwards. The carrier was either static or dynamic noise and was composed of elements with a mean Michelson contrast of 0.15 that were either 1, 2, 4, 8 or 16 pixels in spatial extent. There was no luminance variation within each noise element and the noise was generated anew at the beginning of each motion sequence. In the case of dynamic noise a different stochastic noise sample was generated for each frame (horizontal row of pixels) of the *x–t* plot. For every condition to which the model was applied 100 instantiations of each stimulus were generated.

2.3. Motion-energy detectors

The motion-energy detectors were based on those of Adelson and Bergen (1985) and were constructed from two pairs of quadrature Gabor filters that were oriented in *x–t* as shown in Fig. 3A. The two pairs (L_1, L_2 and R_1, R_2) were identical except that they were maximally

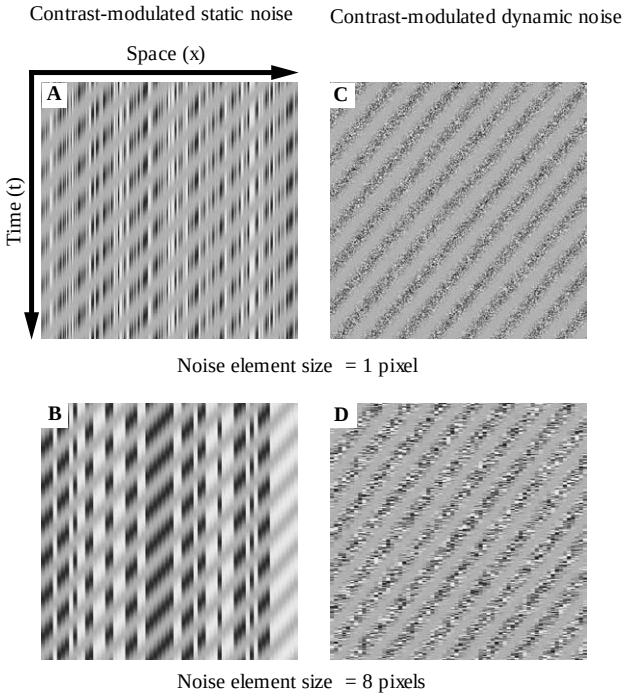

Fig. 2. Examples of space–time (*x–t*) plots representing drifting modulations of the contrast of either (A, B) static or (C, D) dynamic visual noise. Digital versions of these images served as input to the motion-energy detection models. Each *x–t* plot was composed of a 512×512 square pixel array and the spatial and temporal wavelengths of the contrast modulation were 64 pixels/cycle and 75 pixels/cycle, respectively. The carrier was composed of noise elements (mean contrast 0.15) that were either 1, 2, 4, 8 or 16 pixels in spatial extent (only 1 pixel and 8 pixel elements are shown for clarity). See text for further details.

sensitive to motion in opposite directions. The preferred spatial and temporal frequencies of the Gabor filters of the model were identical to those of the sinusoidal contrast modulation in the stimulus, so as to maximise any potential detection of motion energy. In addition the spatial and temporal frequency bandwidths of the Gabor filters were systematically varied by manipulating the standard deviation (spread) of the Gaussian window function and were either 1, 2 or 4 octaves (half-height, full-width).

2.4. Applying the motion-energy model to the input images

Convolution of the input images with the model filters was carried out in the Fourier domain (see Bracewell, 1986) and proceeded as follows: the fast Fourier transform (FFT) of each of the 4 Gabor filters (represented as 235×235 pixel arrays zero padded to the same dimensions as the stimulus input image) was computed together with that of the *x–t* input image as depicted schematically in Fig. 3B. The transform of each filter was then multiplied (in the complex plane) component by component

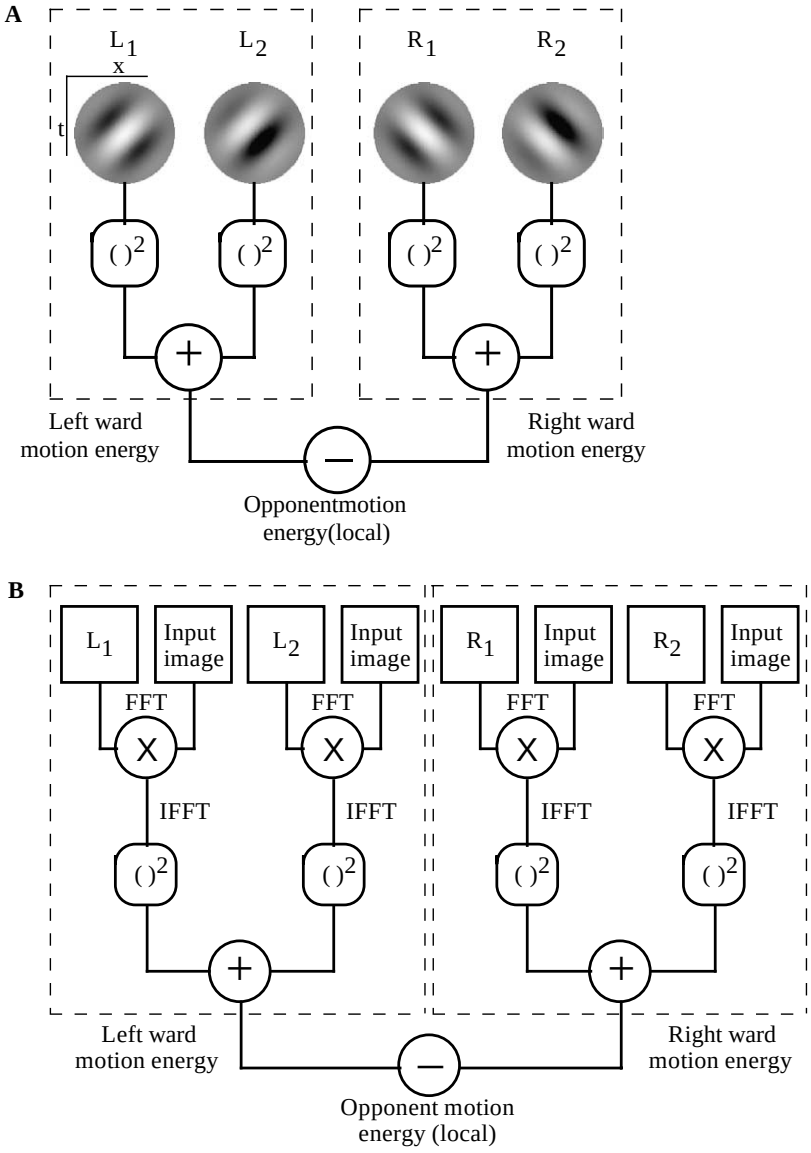

Fig. 3. (A) The motion-energy model of Adelson and Bergen (1985) that was applied, in the present study, to space–time (x–t) images representing drifting modulations of noise contrast. The model utilises two pairs of x–t oriented Gabor filters in spatiotemporal quadrature (L_1, L_2 and R_1, R_2). The two pairs respond to motion in opposing directions and the outputs of each quadrature pair are squared and added to give a directionally sensitive, but phase invariant, measure of local motion energy. The responses for each direction are subtracted to derive the resultant opponent motion energy. (B) Schematic illustration of the actual sequence of operations used to subject the input images to conventional motion-energy analysis. For computational efficiency (and speed) convolution was carried out in the Fourier domain. The fast Fourier transform (FFT) of the input image was multiplied with the FFT of each filter. Inverse fast Fourier transforms (IFFT) were taken and the resulting pair of output images obtained for each direction (e.g., leftwards) were squared and summed to compute the motion energy in that direction. Energies in opposite directions, at each image location, were subtracted (leftwards–rightwards) to give the local opponent motion energy (its sign and magnitude indicate the net direction and 'strength' of motion at each point in space and instant in time).

by the transform of the input image and the inverse FFTs computed. The resulting pair of output images obtained for each direction of motion (e.g., leftwards) were then squared and summed to compute the local motion energy in that direction at each point in space and each instant in time. The responses of motion-energy detectors tuned to opposing directions of motion were then subtracted (leftwards minus rightwards) to give a local estimate of the opponent motion energy (its sign and magnitude giving the net direction and 'strength' of local image motion). Thus, for each x–t input image a final output image was derived which consisted of measures of opponent energy at 278^2 locations ('spoiled' image regions due to edge effects were not included as is conventional in image processing), as was also the case in the study of Benton and Johnston (1997).

2.5. Summary statistics and decision rules for computing the dominant motion direction

Several summary statistics were computed from the opponent energy output image. The mean opponent motion energy (MOME) for each x–t plot was calculated. If the MOME was positive (indicative of leftwards motion—the same direction as the contrast modulation) a counter, which was initially set to zero at the beginning of the modeling procedure, was incremented by one by the decision rule. This procedure was repeated 100 times for each condition that was tested and an overall mean and SEM, collapsed over area and instantiation, were calculated for direct comparison to the results of Benton and Johnston (1997). In addition the percentage of trials on which the decision rule detected a net bias in the same direction as the drifting contrast modulation was also calculated.

For each x–t image to which the model was applied, the peak opponent motion energy (POME) was also calculated for both directions of motion (i.e., the maximum when the sign of the opponent energy was positive and the absolute value of the minimum when the sign was negative). Once again a simple decision rule was applied such that if the absolute magnitude of the positive POME exceeded that of the negative POME (indicating leftwards motion—the same direction as the contrast modulation) a counter, which was set to zero at the beginning of the modeling procedure, was incremented by one. This procedure was repeated 100 times for each condition tested and the overall mean peak and SEM, collapsed over instantiation, were calculated for each direction. In addition the percentage of trials on which the POME decision rule successfully detected a net bias in the direction of the drifting contrast modulation was calculated.

3. Results

3.1. Contrast-modulated static noise

Fig. 4 shows the results of applying the version of the model based on computing the mean opponent motion energy (MOME) to input images composed of contrast-modulated static noise. In Fig. 4A, the overall MOME is plotted as a function of the size of the noise elements (in pixels), separately for each of the 3 frequency bandwidths used for the Gabor filters. The results are in good agreement with those of Benton and Johnston (1997) in that although increasing the noise size leads to a concomitant increase in the variability of the model's output, the responses are directionally balanced such that there is no net directional bias when the mean opponent energy is considered. This was true irrespective of the bandwidth of the motion-energy detectors used. Fig. 4B shows the results of applying the decision rule based on the MOME to each of the 100 motion sequence instantiations tested in each condition. It is clear

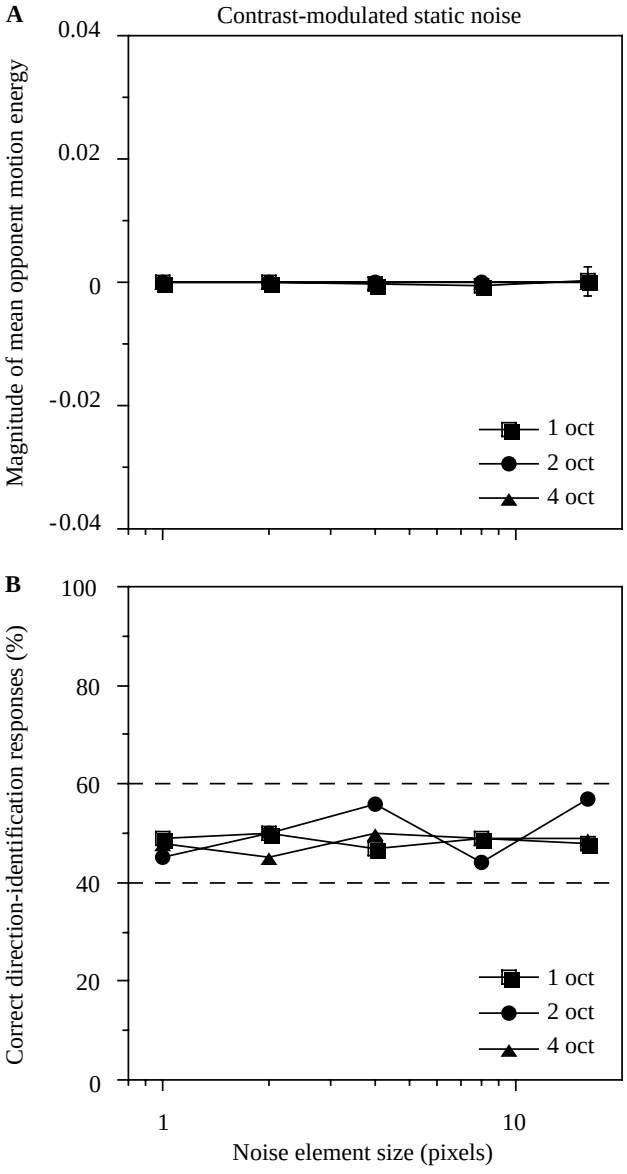

Fig. 4. Results of applying a motion-detection model based on extracting the mean opponent motion energy (MOME) to space–time (x–t) images of contrast-modulated static noise. (A) The overall MOME is plotted as a function of the size of the noise carrier elements (in pixels), separately for each of 3 frequency bandwidths used for the Gabor filters (represented by the different symbols). Positive values indicate net first-order motion energy in the same direction as the contrast modulation (leftwards) and negative values the opposite direction. The vertical bars above and below each datum (where visible) represent ±1 SEM. (B) Shows the results of applying the MOME decision rule to each of the 100 motion sequence instantiations tested in each condition. The percentage of correct direction judgements are plotted as a function of noise element size and different symbols represent different filter bandwidths. The dashed lines indicate the upper and lower 95% confidence intervals corresponding to chance performance (50% correct).

that regardless of the detector bandwidth and noise element size, the percentage of trials on which the model signals motion in the same direction as the contrast waveform was close to 50%, confirming that the average motion energy in the stimulus was equally likely to be rightwards as it was leftwards.

Fig. 5 shows the results of applying the version of the model based on computing the peak opponent motion energy (POME) to input images composed of contrast-modulated static noise. In Fig. 5A the mean POME (absolute, unsigned, values are shown to aid comparison between signals corresponding to net motion leftwards and rightwards) is plotted as a function of the noise element size, separately for each Gabor filter bandwidth and direction of motion. It is apparent that the POME systematically increases as the noise size increases for all bandwidths tested and that overall its magnitude increases as filter bandwidth decreases. More importantly, for a range of noise sizes the magnitude of the POME signaling motion in the same direction as the contrast modulation (leftwards) is much greater than that for the opposite direction (rightwards).[2] Crucially this effect is strongest when the Gabor filter bandwidth is 4 octaves (note that this corresponds to the most localised filter in both space and time) and when the noise elements are spatially extensive. For the condition producing the largest mean POME difference, the magnitude of the POME signaling leftwards motion is ~1.6 times greater than the POME corresponding to rightwards motion.

Thus a model utilising the peak opponent motion energy in the stimulus can readily detect the direction of contrast-modulated static noise patterns, particularly when the noise size is relatively large and the underlying detection mechanisms are markedly localised in x–t. This is further supported by Fig. 5B which shows the percentage of trials on which a decision rule based on the POME signals the correct direction of second-order motion. For the narrowest filter bandwidth the model's performance is close to chance levels (50% correct) for all noise sizes. However for the 2 larger bandwidths (2 and 4 octaves) the model can correctly identify the direction of motion, but only when the noise element size exceeds about 4 pixels. Indeed under these conditions, 4-octave bandwidth detectors can sense motion in the correct direction on over 80% of the trials. Therefore whether or not contrast-modulated static noise patterns are visible to standard motion analysers depends upon the particular output response measure used (i.e., the motion-energy metric on which direction judgments are based) and the frequency bandwidths (degree of localisation in x–t) of the underlying filters.

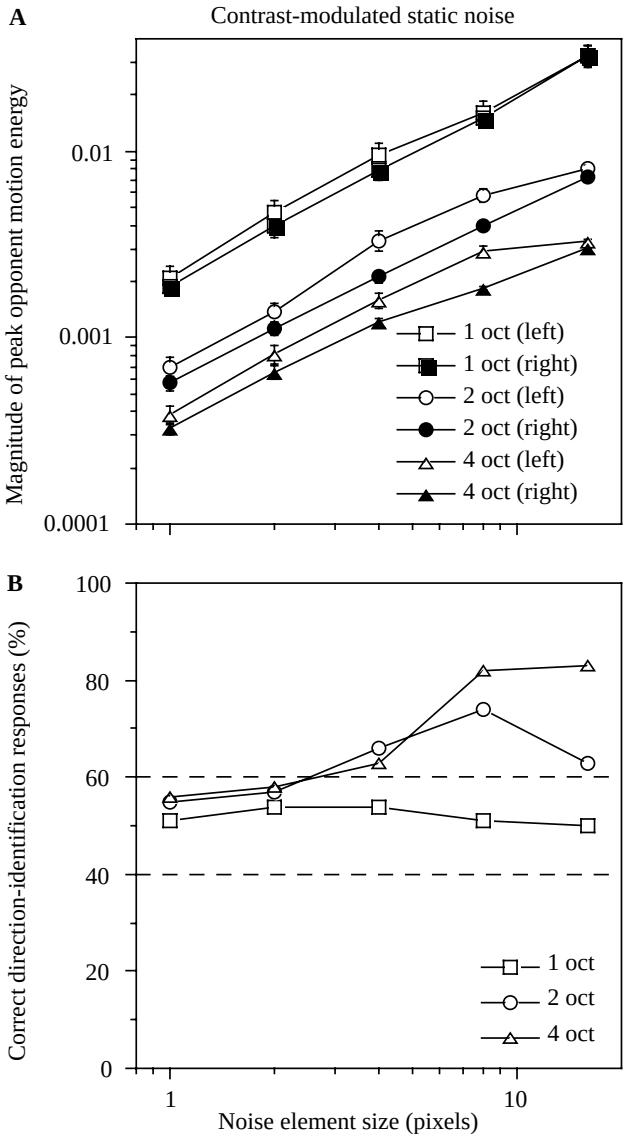

Fig. 5. Results of applying a motion-detection model based on extracting the peak opponent motion energy (POME) to space–time (x–t) images of contrast-modulated static noise. (A) The overall POME is plotted as a function of noise element size for each filter bandwidth used (shown by the different symbols) and direction of motion. The vertical bars above and below each datum (where visible) represent ±1 SEM. (B) The results of applying the *POME* decision rule to the 100 stimulus instantiations in each condition. The percentage of correct direction responses are plotted against the noise element size and different symbols represent different underlying filter bandwidths. The dashed lines show the upper and lower 95% confidence intervals bracketing chance performance.

[2] Additional simulations showed that when the contrast waveform was made to drift rightwards, rather than leftwards, the pattern of model responses was reversed but otherwise identical (i.e., rightwards POME > leftwards POME). This rules out the possibility that the POME differences were due to errors inherent in the limited precision with which digital computers can faithfully represent fractional numbers. Note that double-precision (64-bit), floating point arithmetic was utilised for all non-integer calculations.

3.2. Contrast-modulated dynamic noise

Fig. 6 shows the results of applying the version of the model based on computing the mean opponent motion energy (MOME) to contrast-modulated dynamic noise patterns. Fig. 6A shows the overall MOME plotted as a function of noise element size, separately for each of the 3 frequency bandwidths used for the motion detectors. The results are again in good agreement with those of Benton and Johnston (1997) in that there do not appear to be any systematic biases in the measured mean opponent motion energy, under any of the conditions tested, that could be used to determine motion direction when dynamic carriers are used. This is bolstered by Fig. 6B which shows the results of applying the decision rule based on the MOME to the 100 motion sequences presented in each condition. Correct direction-identification judgments hover around 50% and never exceed chance levels at any noise size and filter bandwidth.

Fig. 7 shows the results of applying the version of the model based on computing the peak opponent motion energy (POME) to input images composed of contrast-modulated dynamic noise. In Fig. 7A the output of the model based on the POME is plotted as a function of the size of the noise elements, for each Gabor filter bandwidth and direction of motion. Although the POME again systematically increases as the noise size increases for all bandwidths (and is greatest for the narrowest bandwidth filters), significantly the magnitude of the POME signaling motion in the same direction as the contrast modulation (leftwards) is always identical to that for the opposite direction (rightwards). Furthermore the percentage of trials on which the POME model encodes the correct direction of second-order motion (Fig. 7B) also shows that performance is close to chance, for all noise sizes and filter bandwidths used. Consequently even for a model based on detecting peak, rather than mean, opponent responses there are no systematic local directional biases present in contrast-modulated dynamic noise stimuli under any of the conditions tested.

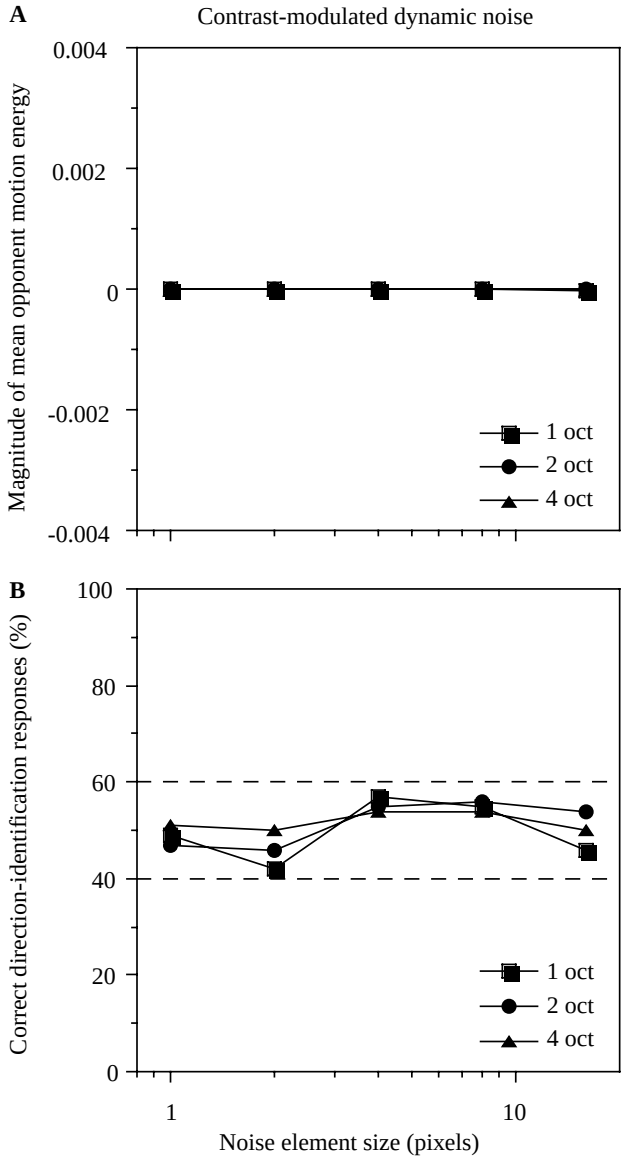

Fig. 6. Same as Fig. 4 with the exception that the *x–t* input images representing the second-order motion sequences, were composed of contrast-modulated dynamic noise. Note the change of scale of the ordinate of the graph shown in this figure (A) compared with that in Fig. 4A.

4. Discussion

The present modeling results are important in that they clearly demonstrate that the precise manner in which computational models of motion are implemented can determine how they respond to the patterns of motion energy present in contrast-modulated noise stimuli. For example the model utilised by Benton and Johnston (1997) employed conventional motion-energy detectors with a fixed (but unspecified) frequency bandwidth and the model's performance was assessed by computing the overall mean opponent motion energy to the entire stimulus set. Additionally no explicit decision rule was applied to judge the motion direction of each individual motion sequence and thus it was not possible to determine the models efficacy at encoding direction on each presentation and compare it to psychophysical data. On the basis of their model results, Benton and Johnston (1997) concluded that Smith and Ledgeway (1997) were incorrect in suggesting that local, unbalanced, luminance-based motion signals present in contrast-modulated static noise patterns could influence direction-identification thresholds for second-order motion (even when the noise elements are very large) as there was no directional bias in the mean opponent energy output of the model.

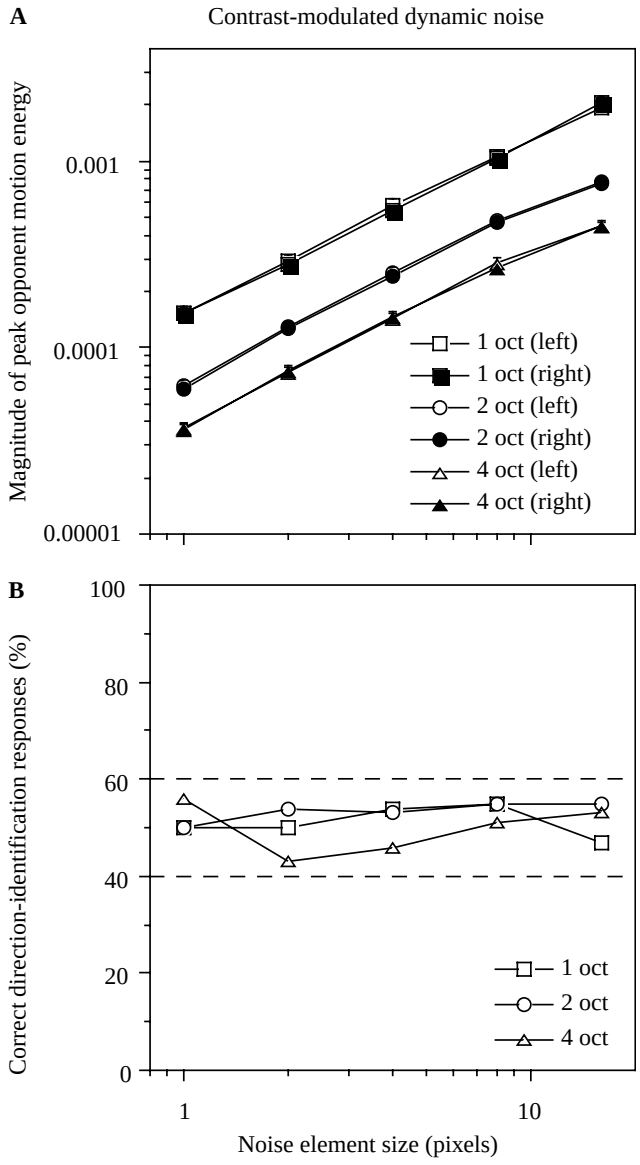

Fig. 7. Same as Fig. 6 with the exception that the *x–t* input images representing the second-order motion sequences, were composed of contrast-modulated dynamic noise. Note the change of scale of the ordinate of the graph shown in this figure (A) compared with that in Fig. 5A.

In order to explain why empirically derived direction thresholds are typically higher than orientation thresholds for second-order motion stimuli, Benton and Johnston (1997) proposed an alternative theory based on the proportion of energy in the Fourier spectra of the images that contains motion direction information. They suggested that the increased motion direction noise associated with a dynamic, compared with a static, noise carrier could selectively impair performance on a direction-identification task but may have little affect on performance for an orientation-identification task. However Benton and Johnston's (1997) proposal clearly cannot account for the psychophysical results depicted in Fig. 1, in that thresholds for identifying the orientation and the drift direction of contrast-modulated noise patterns can be very different even when static noise, albeit composed of relatively small noise elements, is used. Furthermore, if their motion noise masking explanation is correct then one might also expect direction-identification thresholds for first-order motion (sinusoidal luminance gratings) to be higher than those for orientation when a dynamic (but not a static) noise carrier is employed. This prediction is based on the not unreasonable assumption that the presence of a superimposed dynamic noise field should selectively impair direction judgments of a first-order motion stimulus too. We have recently shown that this is not the case (Ledgeway & Hutchinson, 2005).

In the present study conventional motion-energy detectors were used and the results of Benton and Johnston (1997) were replicated when the mean opponent motion energy was utilised as the response metric to quantify performance. However this was not the case when an alternative but equally plausible output measure based on the peak local opponent energy was employed. By considering the direction information encapsulated by the peak, rather than the mean, opponent motion energy and the predominantly local nature of imbalances in motion energy that arise due to spatial clumping of noise elements in contrast-modulated static noise carriers, it was possible to readily model the empirical results. Indeed applying an explicit decision rule to judge motion direction (winner-takes-all) produced a pattern of performance that closely mirrors the psychophysical findings. When the frequency bandwidth of the filters comprising the motion detectors was relatively broad the model correctly predicted the perceived direction of contrast-defined patterns, but only when static noise carriers with relatively large elements were used. While we have not attempted to model the responses of orientation-detecting mechanisms to our input images, it is apparent that spatial clusters of noise elements of the same luminance polarity should not only give rise to local directional biases in motion energy but also analogous biases in the orientation domain when spatially 2-d carriers are used. Empirical results such as those outlined in Fig. 1 clearly suggest that this is indeed the case. When the carrier is static (especially, but not exclusively, when there is a spatial variation in luminance within each noise element) and the noise element size is progressively increased, thresholds for orientation and direction both fall to some extent as they converge. This is precisely the pattern of performance expected if contrast-modulated static noise patterns composed of relatively large elements are encoded using the same mechanisms that respond to first-order motion stimuli. This is because absolute sensitivity to drifting luminance variations is much greater than to comparable contrast modulations (e.g., Smith, Hess, & Baker, 1994) and identification thresholds for direction and spatial form

are convergent when first-order motion stimuli are employed (Ledgeway & Hutchinson, 2005). Although explicit modeling of orientation sensors is beyond the scope of the present study, to account for the convergence of the two thresholds one would simply have to assume that responses to local directional and orientation imbalances arising from stochastic clustering of static noise elements is similar in each case.

The current findings also bear upon models that suggest that the detection of second-order motion has a similar underlying basis to that proposed for first-order motion but that each is encoded, at least initially, by distinct mechanisms. These models typically assume that some form of non-linear processing prior to motion analysis is necessary to encode the direction of second-order motion patterns (e.g., Chubb & Sperling, 1988, 1991). For example Wilson, Ferrera, and Yo (1992) propose that second-order motion is extracted by a specialised processing pathway that utilises the ubiquitous "filter-rectify-filter" (FRF) principle. Specifically the retinal image is convolved with an array of spatial-frequency-selective filters, subjected to a non-linearity (e.g., rectification) and then motion-energy detection at a lower spatial frequency. Interestingly the results of the present study show that with a micro-balanced second-order motion stimulus, a simple energy model can show biases in output at the motion opponent stage if a non-linear decision mechanism (i.e., the POME decision rule) is applied. This adds to the growing body of evidence showing that particular implementations of standard motion models can "see" some (but not all) second-order stimuli without appealing to a FRF front end. Nonetheless the inclusion of a gross non-linearity, implemented in either the decision rules or in pre-decision processing, appears to be the key common characteristic by which sensitivity to second-order motion emerges within these models.

However it is important to emphasise that we are not proposing the POME model as a general alternative to FRF models of second-order motion processing. Indeed one major limitation of the POME model as a candidate generic mechanism for encoding second-order motion stimuli is that, unlike models based upon FRF, it is insensitive to the direction of contrast-modulated dynamic noise patterns. On the other hand FRF theories of motion perception could potentially encode a diverse range of second-order motion stimuli. All that is required is for the first stage filters to respond in a differential manner (i.e., to be selectively sensitive or tuned) to the local texture differences (e.g., in contrast, orientation or flicker rate) that carrying that motion. For example consider an image in which the flicker rate of a random noise field is modulated by a drifting sinusoidal profile. If the image is convolved with an array of linear temporal filters that tile visual space, the resulting activity in the filters will vary across space at the same spatial frequency as the modulation in flicker rate. If the outputs of the filters are then rectified this spatial variation in filter outputs will be converted into a neural signal, with the same periodicity as the original flicker modulation, which could then be analysed by a motion detector. Furthermore the sequence of operations embodied in FRF models is consistent with the properties of some visual neurones in mammalian cortex that are responsive to contrast-defined, second-order motion patterns (e.g., Ledgeway, Zhan, Johnson, Song, & Baker, 2005; Zhou & Baker, 1993, 1994, 1996).

The primary significance of the POME model in the context of the current study is that it serves to illustrate the point that the manner in which the outputs of motion-energy sensors (originally proposed to account for the detection of first-order motion) are combined or compared is crucial in determining the kinds of motion stimuli to which they respond. While the most appropriate manner in which to derive an estimate of the overall image motion from a population of local detectors is still unresolved, it is worth noting that several other authors have also previously considered a winner-takes-all strategy based on selecting the peak (maximum output) of a distribution of responses that has much in common with the POME model (e.g., Heeger, 1987; van Santen & Sperling, 1984; Wilson et al., 1992). Although psychophysical evidence on this issue is equivocal the processes that serve to integrate local direction signals do not necessarily implement a strict averaging strategy (e.g., simple vector summation). The perceived direction of global motion of a display composed of multiple moving dots can be biased, to some extent, towards the strongest directional signal (indicative of a winner-takes-all type strategy based on the mode) and away from the mean when the distribution of dot directions is skewed asymmetrically (Zohary, Scase, & Braddick, 1996). Furthermore it is not necessary to assume that the "winner's" response must propagate throughout the entire stimulus to invoke a global motion percept, since it is well established that human observers can simultaneously encode the overall global direction of image motion whilst being aware of local motion signals that do not necessarily conform to that direction (Braddick, 1997; Watamaniuk & McKee, 1998; Williams & Sekuler, 1984). In terms of physiology Salzman and Newsome (1994) have also provided evidence for a winner-takes-all process in the behaviour of primates who viewed a stimulus moving in one direction whilst simultaneously receiving electrical microstimulation of cells in area V5 that preferred a different direction. The primate's responses corresponded to either the stimulus direction or the electrically stimulated direction rather than a compound (e.g., vector summation) of the two competing neuronal motion signals. Thus the model based on the POME decision rule is not inconsistent with previous model formulations and receives at least some support from the limited empirical evidence available.

In conclusion, some second-order motion patterns can give rise to systematic directional biases in motion energy that are visible to standard models of local motion detection. Benton (2004) recently arrived at a similar conclusion using both a motion-energy model and a gradient model. Following the computation of local direction the outputs of the models were subjected to two additional processing stages. In the first stage the local opponent motion was calculated for each image location and at each point in time (cf. Benton & Johnston, 1997). The output at this stage did not show a net directional bias in response to second-order motion patterns (contrast-modulated static noise) when a global direction index based on the pooled opponent responses was considered. In the second, contrast-normalisation stage the motion-opponent output was then divided by a measure of the static image structure (''static energy'') so that the final output was not contrast dependent. Interestingly at this second stage there was a strong output bias, indicating the correct direction of second-order motion. However this sequence of operations is considerably more complex and speculative than that undertaken in the current study, in that the former requires an additional, intrinsically non-linear, contrast-normalisation process after the local opponent motion signals have been computed. Although Georgeson and Scott-Samuel (1999) provide psychophysical evidence that contrast normalisation may precede decision making regarding the direction of image motion, the adequacy and the parsimony of the current model (based on a straightforward analysis of the peak opponent energy in the stimulus), suggests that it may not be strictly necessary to encode the direction of some second-order motion patterns.

Acknowledgments

This research was supported in part by a University of Nottingham Research Committee grant and a BBSRC Responsive Research Grant (reference BB/C518181/1) to TL.

References

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

Baker, C. L. Jr., (1999). Central neural mechanisms for detecting second-order motion. *Current Opinion in Neurobiology, 9*, 461–466.

Benton, C. P. (2002). Gradient-based analysis of non-Fourier motion. *Vision Research, 42*, 2869–2877.

Benton, C. P. (2004). A role for contrast-normalisation in second-order motion perception. *Vision Research, 44*, 91–98.

Benton, C. P., & Johnston, A. (1997). First-order motion from contrast modulated noise?. *Vision Research 37*, 3073–3078.

Benton, C. P., & Johnston, A. (2001). A new approach to analysing texture-defined motion. *Proceedings of the Royal Society of London B, 268*, 2435–2443.

Benton, C. P., Johnston, A., McOwan, P. W., & Victor, J. D. (2001). Computational modelling of non-Fourier motion: Further evidence for a single luminance based mechanism. *Journal of the Optical Society of America A, 18*, 2204–2208.

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

Braddick, O. J. (1997). Local and global representations of velocity: Transparency, opponency, and global direction perception. *Perception, 26*, 995–1010.

Cavanagh, P., & Mather, G. (1989). Motion: The long and short of it. *Spatial Vision, 4*, 103–129.

Chubb, C., & Sperling, G. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A, 5*, 1986–2007.

Chubb, C., & Sperling, G. (1991). Texture quilts: Basic tools for studying motion-from texture. *Journal of Mathematical Psychology, 35*, 411–442.

Georgeson, M. A., & Scott-Samuel, N. E. (1999). Motion contrast: A new metric for direction discrimination. *Vision Research, 39*, 4393–4402.

Green, M. (1983). Contrast detection and direction discrimination of drifting gratings. *Vision Research, 23*, 281–289.

Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A, 4*, 1455–1471.

Johnston, A., McOwan, P. W., & Benton, C. P. (1999). Robust velocity computation from a biologically motivated model of motion perception. *Proceedings of the Royal Society of London B, 266*, 509–518.

Johnston, A., McOwan, P. W., & Buxton, H. (1992). A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proceedings of the Royal Society of London B, 250*, 297–306.

Ledgeway, T., & Hess, R. F. (2002). Failure of direction-identification for briefly presented second-order motion stimuli: Evidence for weak direction-selectivity of the mechanisms encoding motion. *Vision Research, 42*, 1739–1758.

Ledgeway, T., & Hutchinson, C. V. (2005). The influence of spatial and temporal noise on the detection of first-order and second-order orientation and motion direction. *Vision Research, 45*, 2081–2094.

Ledgeway, T., Zhan, C., Johnson, A., Song, Y., & Baker, C. L. Jr., (2005). The direction-selective contrast response of area18 neurons is different for first- and second-order motion. *Visual Neuroscience, 22*, 87–99.

Salzman, D., & Newsome, W. T. (1994). Neural mechanisms for forming a perceptual decision. *Science, 264*, 231–237.

Smith, A. T. (1994). The detection of second-order motion. In A. T. Smith & R. J. Snowden (Eds.), *Visual detection of motion* (pp. 145–176). London: Academic Press.

Smith, A. T., Hess, R. F., & Baker, C. L. Jr., (1994). Direction identification thresholds for second-order motion in central and peripheral vision. *Journal of the Optical Society of America A, 11*, 506–514.

Smith, A. T., & Ledgeway, T. (1997). Separate detection of moving luminance and contrast modulation: Fact or artifact? *Vision Research, 37*, 45–62.

Smith, A. T., & Ledgeway, T. (1998). Sensitivity to second-order motion as a function of temporal frequency and eccentricity. *Vision Research, 38*, 403–410.

Smith, A. T., & Scott-Samuel, N. E. (1998). Stereoscopic and contrast-defined motion in human vision. *Proceedings of the Royal Society of London B, 265*, 1573–1581.

Sperling, G., & Lu, Z.-L. (1998). A systems analysis of visual motion perception. In T. Watanabe (Ed.), *High-level motion processing* (pp. 153–183). London: The MIT Press.

van Santen, J. P. H., & Sperling, G. (1984). A temporal covariance model of motion perception. *Journal of the Optical Society of America A, 1*, 451–473.

Watamaniuk, S. N. J., & McKee, S. P. (1998). Simultaneous encoding of direction at a local and global scale. *Perception and Psychophysics, 60*, 191–200.

Watson, A. B., Thompson, P. G., Murphy, B. J., & Nachmias, J. (1980). Summation and discrimination of gratings moving in opposite directions. *Vision Research, 20*, 341–347.

Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research, 24*, 55–62.

Wilson, H. R., Ferrera, V. P., & Yo, V. (1992). Psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience, 9*, 79–97.

Zhou, Y.-X., & Baker, C. L. Jr., (1993). A processing stream in mammalian visual cortex neurons for non-Fourier responses. *Science, 261*, 98–101.

Zhou, Y.-X., & Baker, C. L. Jr., (1994). Envelope-responsive neurons in area 17 and 18 of cat. *Journal of Neurophysiology, 72*, 2134–2150.

Zhou, Y.-X., & Baker, C. L. Jr., (1996). Spatial properties of envelope-responsive cells in area 17 and 18 of the cat. *Journal of Neurophysiology, 75*, 1038–1050.

Zohary, E., Scase, M. O., & Braddick, O. J. (1996). Integration across directions in dynamic random dot displays: Vector summation or winner take all? *Vision Research, 36*, 2321–2331.