PERGAMON

Vision Research 40 (2000) 1135–1142

Vision Research

www.elsevier.com/locate/visres

Computational modelling of interleaved first- and second-order motion sequences and translating $3f + 4f$ beat patterns

Christopher P. Benton [a,*], Alan Johnston [a], Peter W. McOwan [a,b]

[a] *Department of Psychology, University College London, Gower Street, London WC1E 6BT, UK*
[b] *Department of Mathematical and Computational Sciences, Goldsmiths College, New Cross, London SE14 6NW, UK*

Received 30 June 1999; received in revised form 28 September 1999

Abstract

Despite detailed psychophysical, neurophysiological and electrophysiological investigation, the number and nature of independent and parallel motion processing mechanisms in the visual cortex remains controversial. Here we use computational modelling to evaluate evidence from two psychophysical studies collectively thought to demonstrate the existence of three separate and independent motion processing channels. We show that the pattern of psychophysical results can largely be accounted for by a single mechanism. The results demonstrate that a low-level luminance based approach can potentially provide a wider account of human motion processing than generally thought possible. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Motion perception; Computational modelling; Second-order; Feature tracking

1. Introduction

The notion of a low-level motion detector that operates directly upon image luminance has gained a wide acceptance and is perceived to be the basic building block for the biological detection of motion (van Santen & Sperling, 1984, 1985; Adelson & Bergen, 1985, 1986; Watson & Ahumada, 1985). Such a mechanism cannot readily account for motion that is not defined by luminance (Chubb & Sperling, 1988; Benton & Johnston, 1997). Fig. 1a shows a sinusoidal contrast modulation (the envelope) translating over a textured background (the carrier). This motion may be referred to as contrast-defined and is part of the wider class of texture-defined motion. Examples of the latter are translating modulations of the spatial frequency, temporal frequency or orientation of texture elements. These can be described as 'second-order' because they involve the modulation of second-order image statistics (Cavanagh & Mather, 1989). The term 'first-order' is used to describe luminance-defined motion.

In order to account for the perception of texture-defined motion, a dedicated texture motion mechanism has been proposed (Chubb & Sperling, 1988; Clifford, Freedman & Vaina, 1998; Clifford & Vaina, 1999). In this approach, motion analysis is preceded by a texture grabber, a mechanism that spatio-temporally filters the image and then applies some gross non-linearity such as squaring or full-wave rectification about mean luminance (Sperling, 1989). This process makes texture-defined motion detectable by standard low-level motion analysis.

One can also create types of motion that are not luminance-defined and can also not be extracted through the operation of the texture channel outlined above. Examples of these are motion defined by complex features (Cavanagh, 1992; Lu & Sperling, 1995a,b), motion generated by motion boundaries (also termed theta motion (Zanker, 1993)), and motion defined by binocular disparity (Patterson, Donnelly, Phinney, Nawrot, Whiting & Eyle, 1997; Smith & Scott-Samuel, 1998). The fact that we can detect motion in these stimuli suggests that some additional process or processes may operate. One possibility is a feature tracking mechanism, a concept that has a long history within the field of human motion perception. Although the existence of a feature tracking process is

* Corresponding author. Tel.: +44-20-75045301; fax: +44-20-74364276.
E-mail address: c.benton@ucl.ac.uk (C.P. Benton)

PII: S0042-6989(00)00026-2

generally recognized, there is little consensus regarding the underlying mechanisms. Additionally, there is evidence that feature tracking may underlie the perception of texture-defined motion (Seiffert & Cavanagh, 1998; Derrington & Ukkonen, 1999). If this is the case, one may reasonably question the utility of dedicated texture motion mechanism.

In a recent study examining patients with neurological damage, a double dissociation for first- and second-order motion was reported (Vaina, Cowey & Kennedy, 1999). This finding is not supported by other studies which have sought to detect and identify different processing areas for first- and second-order motion (Victor & Conte, 1992; Greenlee & Smith, 1997; Smith, Greenlee, Singh, Kraemer, & Hennig, 1998; Somers, Seiffert, Dale, & Tootell, 1998, 1999). These have reported small or no differences in the cortical areas activated by the two classes of stimuli. The lack of concordance may well imply that where differences are found, these reflect the response of a single mechanism to different stimuli, rather than a functional subdivision in the cortical motion analysis system.

A number of studies have described neurons in the visual cortex of cat and macaque that respond to both luminance and texture-defined motion (Albright, 1992; Zhou & Baker, 1993; Mareschal & Baker, 1998; O'Keefe & Movshon, 1998). Although these studies provide evidence for the neural processing of second-order motion, they provide no direct evidence for a dedicated texture motion channel. Such a mechanism would predict the existence of cells that respond mainly to texture-defined motion whilst showing little or no response to luminance-defined motion. As yet, there is no evidence to support this prediction.

In the absence of firm neurophysiological and electrophysiological evidence, support for the existence of a separate motion processing channels has come mainly from the domain of psychophysics. Here we examine two psychophysical paradigms: one that has been taken to demonstrate the existence of separate luminance and texture motion mechanisms, and one that has been taken to demonstrate the existence of separate luminance and feature based mechanisms. The first utilises an interleaved sequence of luminance- and contrast-defined motion (Ledgeway & Smith, 1994), the second, a beat pattern created from the addition of $3f$ and $4f$ components (Hammett, Ledgeway & Smith, 1993).

We apply a recent model of cortical motion processing (Johnston, McOwan & Benton, 1999a) to the stimuli to see if the effects can be accounted for within a unified architecture. The model is an elaboration of a simple low-level motion detection strategy — the spatio-temporal gradient model (Fennema & Thompson, 1979; Horn & Schunck, 1981). The simulations show that the pattern of psychophysical results from the two tasks can be accounted for by a single luminance based mechanism. It is clear that the model can behave in a manner that is generally thought indicative of multiple motion mechanisms. Our results offer the possibility that luminance based mechanisms can provide a wider account of human motion processing than is generally perceived to be the case.

2. A computational model of biological motion processing

This section contains a general description of the model employed in the study. A full description is given by Johnston et al. (1999a). We initially describe a version that operates in only two dimensions (i.e. one spatial dimension, x, and one temporal dimension, t).

The model is derived from the gradient approach in which velocity is calculated by dividing the temporal derivative of image brightness by the spatial derivative of image brightness. Because partial differentiation of an image followed by filtering is equivalent to applying a partially differentiated filter to the image (Bracewell, 1965), a gradient model may be realised by applying pairs of filters to an image, one of which is the temporal derivative and one the spatial derivative of the same filter kernel. Local velocity can then be calculated by taking the ratio of the outputs of the two filters.

One potential problem is that, when the output of the spatial filter is zero, velocity is undefined. In order to condition the velocity ratio calculation, one can include additional measures. Rather than thinking of a single filter kernel, one can think of a vector of filter kernels. By taking the spatial derivatives of each of the filter kernels we derive one vector of filters, and by taking the temporal derivatives we derive another. A least squares estimate of velocity can then be calculated as the ratio of dot products:

$$\text{velocity} = \frac{\boldsymbol{X} \cdot \boldsymbol{T}}{\boldsymbol{X} \cdot \boldsymbol{X}} \tag{1}$$

where $\boldsymbol{X}$ is the vector of filter outputs from the spatially differentiated filters, and $\boldsymbol{T}$ is a vector of filter outputs from the temporally differentiated filters. In this case, for the denominator to equal zero, all of the measures in $\boldsymbol{X}$ must equal zero. If we increase the number of measures then the probability of this occurring is reduced.

Here, however, one is faced with the problem of how to construct the vector of filter kernels prior to differentiation. What filters should be used, and how should they be weighted relative to one another? One way to solve this problem is to utilise a proposal put forward by Koenderink and van Doorn (1987). In this framework, local image structure is represented as a truncated Taylor series expansion. In other words, image structure in a local region is represented by the

weighted outputs of a series of filters applied at a point in the image. These filters are generated from a single filter (the *blur kernel*) by progressively increasing the order of the spatial and temporal differential operators applied to the kernel.

From the Taylor series expansion, one can estimate image brightness at a distance from the point at which the measures are taken. The weights that are attached to the various derivatives are dependent upon the direction and length of the vector joining the point at which the measures are taken and the point at which we wish to estimate image brightness. Therefore, for any point in the image we can construct a vector of weighted filter functions which will serve as the initial vector of filter kernels. The calculation of velocity uses this property of the Taylor series to integrate over a region surrounding the point at which the bank of filters is applied.

The 2D space–time version of the model summarised so far, has been described by Johnston, McOwan and Buxton (1992) and Johnston and Clifford (1995a). In the current study we use a version that has been extended to accommodate three dimensions (i.e. two spatial dimensions, x and y, and one temporal dimension, t). To generate our vector of filter kernels from a single blur kernel, we differentiate the kernel in increasing orders of x, y and t. As in the 2D version of the model, the blur kernel is a Gaussian in space and log time.

In the 3D version of the model, the reference frame (i.e. the x and y axes defining the local direction of spatial differentiation) is rotated through a number of orientations with respect to the input image. A total of 24 orientations (evenly spaced over 360°) are employed. For each orientation, three vectors of filters are created by differentiating the vector of filter kernels with respect to x, y and t. Coupled with the rotation of the reference frame, this produces a population of filters that are tuned to different orientations and spatial frequencies and may show either transient or sustained temporal properties. From the measures derived by applying the three vectors of filter functions to the image, we calculate the following four speed related measures: speed, orthogonal speed, inverse speed and orthogonal inverse speed:

$$\text{speed} = \frac{\boldsymbol{X}\cdot\boldsymbol{T}}{\boldsymbol{X}\cdot\boldsymbol{X}}\cos^2\phi = \frac{\boldsymbol{X}\cdot\boldsymbol{T}}{\boldsymbol{X}\cdot\boldsymbol{X}}\left(1+\left(\frac{\boldsymbol{X}\cdot\boldsymbol{Y}}{\boldsymbol{X}\cdot\boldsymbol{X}}\right)^2\right)^{-1} \tag{2}$$

$$\text{orthogonal speed} = \frac{\boldsymbol{Y}\cdot\boldsymbol{T}}{\boldsymbol{Y}\cdot\boldsymbol{Y}}\sin^2\phi = \frac{\boldsymbol{Y}\cdot\boldsymbol{T}}{\boldsymbol{Y}\cdot\boldsymbol{Y}}\left(1+\left(\frac{\boldsymbol{X}\cdot\boldsymbol{Y}}{\boldsymbol{Y}\cdot\boldsymbol{Y}}\right)^2\right)^{-1} \tag{3}$$

$$\text{inverse speed} = \frac{\boldsymbol{X}\cdot\boldsymbol{T}}{\boldsymbol{T}\cdot\boldsymbol{T}} \tag{4}$$

$$\text{orthogonal inverse speed} = \frac{\boldsymbol{Y}\cdot\boldsymbol{T}}{\boldsymbol{T}\cdot\boldsymbol{T}} \tag{5}$$

where $\boldsymbol{X}$ is the vector of filter outputs from the x differentiated filters, $\boldsymbol{Y}$ is the vector of filter outputs from the y differentiated filters and $\boldsymbol{T}$ is the vector of filter outputs from the temporally differentiated filters. Speed ($\boldsymbol{X}\cdot\boldsymbol{T}/\boldsymbol{X}\cdot\boldsymbol{X}$) and orthogonal speed ($\boldsymbol{Y}\cdot\boldsymbol{T}/\boldsymbol{Y}\cdot\boldsymbol{Y}$) measures are ill conditioned when there is no variation over x and y, respectively. To prevent this from degrading the final velocity estimate, speed and orthogonal speed are both conditioned by measures of the angle of image structure relative to the reference frame (ϕ). From Eq. (2), it can be seen that as $\boldsymbol{X}\cdot\boldsymbol{X}$ decreases, $(\boldsymbol{X}\cdot\boldsymbol{Y}/\boldsymbol{X}\cdot\boldsymbol{X})^2$ increases more rapidly than the speed term, thereby forcing the product to zero as ($\boldsymbol{X}\cdot\boldsymbol{T}/\boldsymbol{X}\cdot\boldsymbol{X}$) approaches infinity. The inverse speed measures however become infinite when the temporal derivative is close to zero, which is appropriate.

As the four speed related measures are calculated from different combinations of different filter outputs, they may all thought of as different estimates that are related to local speed. The inclusion of additional image measures to increase robustness is one of the core design criteria that runs through the model. By taking each of the speed related measures indexed by the orientations through which the reference frame is rotated, we can form the vectors, $\hat{\boldsymbol{s}}_{\parallel}$ (speed), $\hat{\boldsymbol{s}}_{\perp}$ (orthogonal speed), $\check{\boldsymbol{s}}_{\parallel}$ (inverse speed) and $\check{\boldsymbol{s}}_{\perp}$ (orthogonal inverse speed). The final speed estimate is calculated as the square root of the ratio of determinants:

$$\text{speed}^2 = \frac{\begin{vmatrix} \hat{\boldsymbol{s}}_{\parallel}\cdot\boldsymbol{F}_{\parallel} & \hat{\boldsymbol{s}}_{\parallel}\cdot\boldsymbol{F}_{\perp} \\ \hat{\boldsymbol{s}}_{\perp}\cdot\boldsymbol{F}_{\parallel} & \hat{\boldsymbol{s}}_{\perp}\cdot\boldsymbol{F}_{\perp} \end{vmatrix}}{\begin{vmatrix} \hat{\boldsymbol{s}}_{\parallel}\cdot\check{\boldsymbol{s}}_{\parallel} & \hat{\boldsymbol{s}}_{\parallel}\cdot\check{\boldsymbol{s}}_{\perp} \\ \hat{\boldsymbol{s}}_{\perp}\cdot\check{\boldsymbol{s}}_{\parallel} & \hat{\boldsymbol{s}}_{\perp}\cdot\check{\boldsymbol{s}}_{\perp} \end{vmatrix}} \tag{6}$$

where $\boldsymbol{F}_{\parallel}$ and $\boldsymbol{F}_{\perp}$ are vectors containing the cosines and sines of the angles through which the reference frame has been rotated.

In the case of the motion of a simple pattern such as, for example, a translating sine wave grating, the denominator of Eq. (6) takes a value of 1. Measures of speed and orthogonal speed vary sinusoidally with angle of reference frame, with orthogonal speed lagging behind speed by a quarter of a cycle. The numerator shown in Eq. (6) works by taking a measure of the amplitude of the distribution of speed measures that is combined across both speed and orthogonal speed. The denominator (and its measures of inverse speed) are included to stabilise the final speed estimation. This can be shown to be particularly important in the case where motion occurs in the presence of static pattern (see Johnston et al., 1999 for a full analysis).

Direction of motion is extracted by calculating a measure of phase that is combined across all four speed related measures:

$$\text{direction} = \tan^{-1}\left(\frac{(\check{\boldsymbol{s}}_{\parallel} + \hat{\boldsymbol{s}}_{\parallel}) \cdot \boldsymbol{F}_{\perp} + (\check{\boldsymbol{s}}_{\perp} + \hat{\boldsymbol{s}}_{\perp}) \cdot \boldsymbol{F}_{\parallel}}{(\check{\boldsymbol{s}}_{\parallel} + \hat{\boldsymbol{s}}_{\parallel}) \cdot \boldsymbol{F}_{\parallel} - (\check{\boldsymbol{s}}_{\perp} + \hat{\boldsymbol{s}}_{\perp}) \cdot \boldsymbol{F}_{\perp}}\right) \quad (7)$$

When speed is large (and inverse speed is small) then direction is dominated by the speed measures, however when speed is small (and inverse speed is large) then the measure is dominated by inverse speed. The use of complementary and antagonistic speed and inverse speed measures should prove valuable in any system where small signals are likely to be affected by noise.

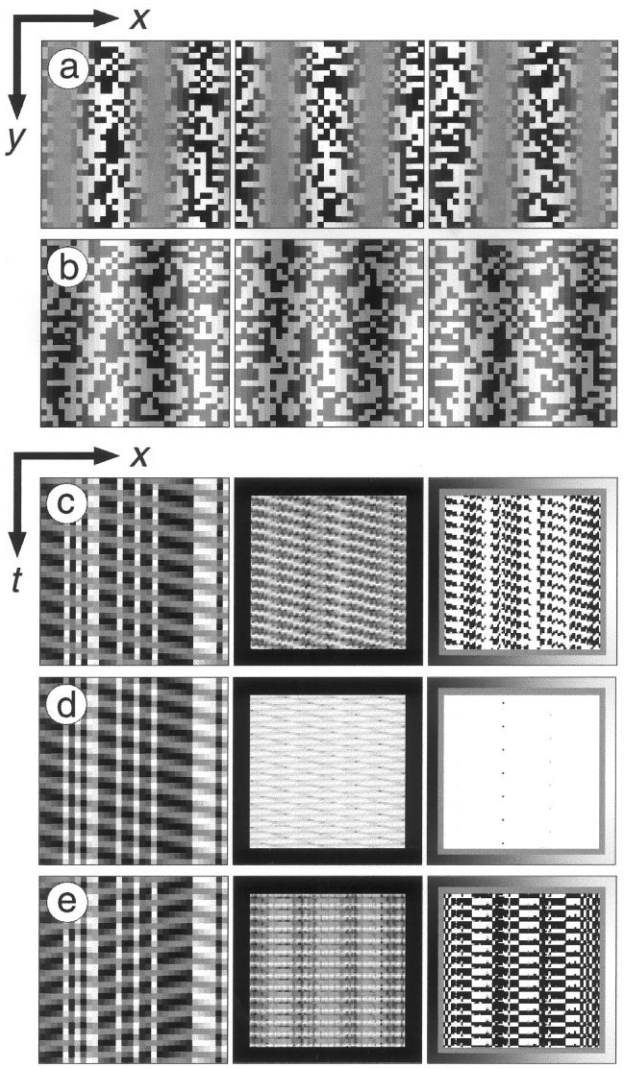

Fig. 1. Model input and output for contrast modulated, luminance modulated and interleaved sequences. All images have been scaled to fill the full luminance range. For clarity, in the examples of luminance-defined patterns, the contrast of the additive sinusoidal component is 0.5. In space–time input/output plots (see Section 3), the leftmost column shows input stimuli, centre column shows speed and rightmost column shows direction. Speed is encoded by brightness, direction is indicated by the grey surround. White indicates forwards motion, black indicates reversed motion. Grey indicates motion perpendicular to that of the envelope. (a) Three frames from a contrast-defined sequence in which the envelope translates from left to right by a quarter of a cycle every frame. (b) Three frames from a luminance-defined sequence with additive static noise in which the envelope translates from left to right by a quarter of a cycle every frame. (c) Space–time model input and output for a contrast-defined sequence. (d) Space–time input/output for a luminance-defined sequence. (e) Space–time input/output for an interleaved sequence.

The model described above incorporates a number of distinct processing stages and integrates the outputs of oriented filters that have a range of spatial and temporal frequency tuning characteristics, consistent with the properties of cortical neurons. In summary, the model can be thought of as a stable and robust extension of the gradient algorithm. The model seeks to resolve problems of mathematical ill-conditioning by making multiple measures on the stimulus rather than by introducing arbitrary constants or thresholds. It should be noted that the model contains no non-linear preprocessing stages that seek to extract texture or features prior to motion analysis; the computation of speed and direction are based directly upon the output of filters applied to the input image.

3. Method and results

3.1. General methodology

All input sequences contained 128 frames of 128 × 128 pixel images. The temporal filters incorporated into the model are calibrated by fitting their temporal frequency tuning curves to psychophysical data. The temporal scaling means that 128 frames represents 1 second (Johnston & Clifford, 1995b). Model input and associated output are shown as space–time plots. The following procedure was used to compress space–space–time model input and output. Within each frame, x refers to the dimension parallel to the direction of motion (horizontal), y refers to the dimension perpendicular to motion (vertical). Space time images of input sequences were created by collating the central horizontal lines from each input frame. Model output gives speed and direction which may be expressed in component form. Space–time model output was derived from space–space–time output by averaging the motion vectors over y. After this procedure, each frame is represented by a horizontal line of averaged motion vectors. These are then expressed as speed and direction, collated over frames and plotted separately as space–time plots.

We calculate a directional index from the model output. This is a measure of contrast between forwards and reversed motion. Velocities are expressed in component form, and the sum of the positive values (P) and the sum of the absolute negative values (N) of the velocity components in the direction of motion are calculated. The directional index is defined as $(P-N)/(P+N)$. A positive index indicates forwards/rightwards motion, a negative index indicates reversed/leftwards motion. The measure ranges from 1.0 to -1.0. In this study, each directional index is drawn from 64 frames of model output. The mean indexes described in this

Table 1
Mean directional indexes (see Section 3)

	Contrast-defined	Luminance-defined	Interleaved	Beat *without* inter frame interval	Beat *with* inter frame interval
Mean direction index (+ve = forwards)	0.49	1.00	−0.0025	−0.86	0.18
Standard deviation	4.0×10^{-2}	2.1×10^{-4}	5.1×10^{-2}	7.8×10^{-3}	2.1×10^{-2}

paper are based upon the model's response to 100 instantiations of each input stimulus.

3.2. Interleaved first-order/second-order paradigm

We applied the model to luminance-defined, contrast-defined and interleaved sequences. Fig. 1 shows examples of the stimuli. In contrast-defined sequences (a), the contrast of static binary noise is modulated by a translating sine wave grating. In luminance-defined sequences (b), static binary noise is added to a translating sine wave grating. Noise element size was 4×4 pixels, the spatial frequency of the sine wave was two cycles per image. The sine wave jumped by a quarter of a cycle every four frames (remaining static between jumps). This gives a temporal update rate of 32 Hz. Interleaved stimuli were created by taking alternating four frame segments from luminance- and contrast-defined sequences with the same underlying noise pattern. In contrast modulated sequences the modulation depth was 1.0, mean noise contrast was 0.5. To closely approximate Ledgeway and Smith (1994), the contrast of the sinusoidal component in the luminance-defined motion sequences was 0.1. The contrast of the additive noise was 0.5. Each stimulus instantiation contained a fresh sample of static binary noise.

When either the contrast or the luminance-defined stimuli are viewed, the overall direction of motion is readily apparent. However, when they are interleaved, no coherent motion percept is elicited. It has been argued that this pattern of results demonstrates that luminance- and contrast-defined motion are processed through separate mechanisms (Ledgeway & Smith, 1994). Fig. 1c,d,e shows the results of the model applied to the three sequences, direction indexes are given in Table 1. The results show that motion is signalled in the correct direction for the luminance- and contrast-defined sequences but that no coherent motion is detected in interleaved sequences[1]. This clearly demonstrates that the pattern of psychophysical results is not necessarily indicative of separate motion mechanisms.

3.3. 3f + 4f beat paradigm

Examples of $3f+4f$ beat sequences are shown in Fig. 2. For our simulations, the spatial frequency of the beat pattern (given by the difference between its component frequencies) was 2.0 cycles per image. The amplitudes of the components were identical. The pattern jumped through a quarter beat cycle every eight frames (remaining static between jumps). Inter frame intervals were added by inserting mean luminance grey frames between the eight frame segments. For the results described in this paper, inter frame intervals of zero and eight frames were employed. For each stimulus instantiation, the phases of the components were randomised.

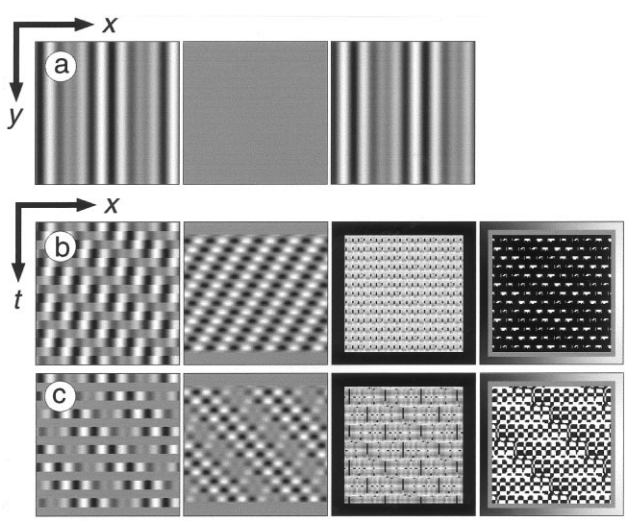

Fig. 2. Model input and output for $3f+4f$ beat patterns. In space–time input/output plots, first column shows input stimuli, second column shows input stimuli filtered with the first derivative of a log Gaussian temporal filter, third and fourth columns show model output. The scale of the temporal filter was determined by fitting to psychophysical data (Johnston & Clifford, 1995b), eight frames represents an inter frame interval of 62.5 ms. (a) Example of three successive frames, including a grey inter frame interval, from a sequence in which the beat patterns translates by a quarter of a cycle from left to right. (b) Space–time input/output for a sequence with no interframe interval. (c) Space–time input/output for a sequence with an eight frame inter frame interval.

[1] The temporal update rate employed in this study (32 Hz) was approximately double that of Ledgeway et al. (1993). They note that temporal update rate is not a critical factor. When we use an update rate of 16 Hz the pattern of results remains the same (although less clear-cut). Mean directional indexes (and associated standard deviations) are 0.26 (9.66×10^{-3}) for the luminance-defined pattern, 0.26 (2.03×10^{-2}) for the contrast-defined pattern and 0.0032 (2.70×10^{-2}) for the interleaved pattern.

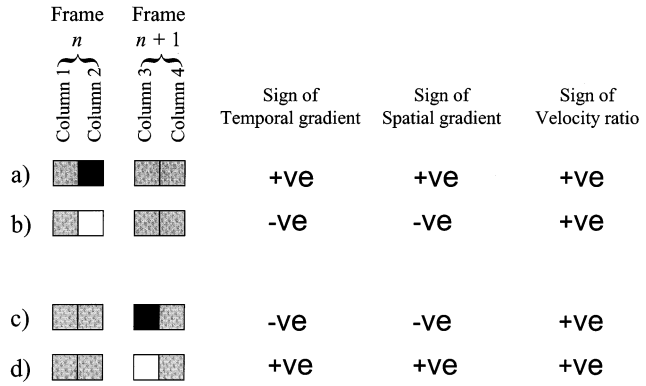

Fig. 3. Schematic diagram showing four small spatially localised areas (2 pixels wide, 1 pixel high) over two successive frames as a square wave contrast envelope, translating from left to right, modulates the contrast of a static binary noise carrier. The first two columns of squares indicate the luminances of two adjacent pixels at frame *n*, the third and fourth columns indicate the luminances of those two pixels on the next frame. In the top two examples, the leading edge of a low-contrast grey region reduces the contrast of dark (a) and light (b) pixels. In the bottom two examples, the contrast of dark (c) and light (d) pixels increases as the trailing edge of the low-contrast region moves to the right. The signs of the spatial and temporal gradients are based on the following differences. Temporal gradient = (column 3 + column 4) − (column 1 + column 2). Spatial gradient = (column 1 + column 3) − (column 2 + column 4). The direction of velocity is given by the sign of the ratio of the temporal gradient over the spatial gradient. The figure demonstrates that at both the leading and trailing edge of the low-contrast regions, velocity is signalled in the same direction whatever the polarity of the underlying noise carrier. Note that if all the noise elements are either dark or light (i.e. a luminance-defined stimulus), the calculated direction of motion is the same as that computed with the texture-defined stimulus.

As the pattern translates by a quarter of a beat cycle, the 4*f* component translates by a whole cycle (and is therefore stationary) whilst the 3*f* component translates forwards by three quarters of a cycle. The motion of the latter is equivalent to the component shifting backwards by a quarter of a cycle. A low-level motion detector should indicate this reversed motion. When the inter frame interval is short, a percept of transparency is obtained with both forwards and reversed motion present. When the inter frame interval is increased, there is a greater tendency to report forwards motion and a concomitant reduction in transparency (Hammett et al., 1993). It is argued that the forwards motion (the direction of pattern motion) represents the influence of a feature tracking mechanism. Fig. 2b,c shows the response of the model to the stimulus. Table 1 shows mean directional indexes. With an inter frame interval of zero, the model's output contains both forwards and reversed motion. When the inter frame interval is increased, there is a greater preponderance of forwards motion. We can conclude from this that the pattern of psychophysical results does not necessarily indicate multiple motion channels. A similar conclusion is reached by Johnston and Clifford (1995b) who, with the 2D version of the model employed in this study, successfully modelled the perceived motion reversal found with a missing fundamental fluted square wave over increases in inter frame interval (Georgeson & Harris, 1990).

4. Discussion

Taken together the results show the following. The model correctly detects the direction of motion in luminance- and contrast-defined patterns but fails to detect motion when these two patterns are interleaved. With the $3f + 4f$ beat sequence, the model's output contains a mixture of reversed and forwards motion with the latter increasingly prevalent as the blank inter frame interval is increased. These simulations indicate that a single channel model can display behaviours normally thought to indicate the presence of multiple motion mechanisms.

The model 'sees' luminance- and contrast-defined motion, yet fails to see motion in interleaved patterns. In the single channel model used for our simulations (Johnston et al., 1999a), there is no initial non-linearity. The detection of contrast-defined motion is based upon a geometrical approach to the analysis of the luminance surface (Johnston & Clifford, 1995a; Johnston, Benton & McOwan, 1999b). A schematic diagram describing how the model detects the direction of motion in translating contrast modulations of static noise is shown in Fig. 3. This shows how the temporal and spatial gradients at the leading and trailing edges of a low-contrast region may be combined to give the correct direction of motion. When the sign of the both the temporal and spatial gradients is changed, the sign of their ratio (i.e. the direction of motion) is conserved (see Johnston et al., 1999b for a more detailed examination). The analysis can also be applied to translating sinusoidal contrast modulations.

It is clear then that the gradient model can correctly detect the direction of motion in luminance- and contrast-defined stimuli. Why then should it fail to detect motion when the two are interleaved? Fig. 4 shows a representation of three frames of an interleaved sequence. An explanation of why no motion is elicited by this stimulus can be made by appealing to the temporal symmetries inherent in the image.

There is no expected variation in mean luminance across the contrast defined frame shown in Fig. 4a. Any deviations from mean luminance that do occur will be random and are equally likely to signal leftwards or rightwards motion. On a more local level, the relationship between the luminance-defined frames and the noise elements in the high contrast regions of the contrast defined frames is also balanced. Fig. 4b and c show two portions of Fig. 4a which contain the left and

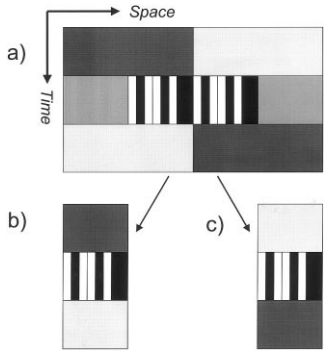

Fig. 4. (a) Schematic representation showing a space time plot of three successive frames of an interleaved motion sequence. The middle frame shows a contrast modulated pattern with a block of high contrast static noise in the centre. (b) A portion of (a) containing the left hand half of the high contrast static noise. (c) A portion of (a) containing the right hand half of the high contrast static noise.

right halves of the high contrast region shown in the centre of the image (Fig. 4a). Based on the combination of spatial and temporal gradients it is possible that motion may be signalled in one particular direction within the left hand side (Fig. 4b). However the right hand side (Fig. 4c) is simply a temporally reversed version of the left. If motion is signalled in one particular direction by the left hand side, the motion signalled by the right hand side must be equal and opposite. Of course the image has been designed to contain this property, it does not necessarily have to do so. However, the configuration of noise seen on the right hand side of the high contrast area is just as probable as that seen on the left hand side. What this must mean is that for every configuration that appears on the left hand side there is an equally probable opposite configuration that may appear on the right hand side. If the underlying noise carrier is truly random, motion must be balanced in this stimulus.

The model also increasingly 'sees' forwards motion in the $3f+4f$ beat pattern as inter frame interval is increased. This forwards motion has been explained by proposing a feature tracking mechanism (Hammett et al., 1993), an assertion that is based upon the immunity of the forwards motion percept to the extended inter frame interval. The ability to integrate over comparatively long temporal and spatial intervals is traditionally associated with a feature tracking mechanism (Braddick, 1980; Georgeson & Harris, 1990).

The results of applying a differential of a log Gaussian temporal filter to the image are shown in the second column of Fig. 2b,c. Input stimuli are shown in the leftmost column. This bandpass or 'transient' filtering operation is incorporated into the model and the filters used provide a good fit to psychophysical measures of human temporal filtering (Hess & Snowden, 1992; Johnston & Clifford, 1995b). With no inter frame interval the static component drops out and the spatio-temporal structure is dominated by the $3f$ component which signals reversed luminance-defined motion. When the inter frame interval is increased, the spatio-temporal structure of the filtered image is very different. There is a clearly oriented contrast modulation indicating motion in the forwards direction. The $3f+4f$ stimulus can therefore be seen to contain both luminance- and contrast-defined motion. As the inter frame interval is increased, the local structural elements within the temporally filtered image that signal forwards motion become more prominent, leading to a shift in the calculated direction of motion. The shift in balance between forwards and reversed motion may simply reflect the operation of a transient temporal filter in human motion perception (see also Shioiri & Cavanagh, 1990; Johnston & Clifford, 1995b), rather than a switch in computational strategy.

5. Conclusions

This paper shows that a low-level luminance based model of human visual perception can exhibit behaviour normally associated with multiple motion processing mechanisms. In particular, the simulations make the following two points. Firstly, the detection of contrast defined motion does not necessarily require a preprocessing non-linearity. Secondly, the inclusion of an interframe interval does not necessarily disable low-level motion processing, although it may well affect its outcome. Our simulations show that two psychophysical tasks, collectively thought to indicate the existence of three separate and parallel motion processing mechanisms, may largely be accounted for by a single coherent motion algorithm. More generally, the results raise the possibility that a low-level mechanism can potentially account for a greater portion of motion phenomena than is normally supposed.

Acknowledgements

This research was supported by a project grant from the EPSRC-BBSRC Mathematical Modelling Initiative. We thank Jason Dale for his assistance with software development.

References

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

Adelson, E. H., & Bergen, J. R. (1986). *The extraction of spatio-temporal energy in human and machine vision*. Paper presented at the Proceedings — Workshop on Motion: Representation and Analysis, Charleston, SC.

Albright, T. (1992). Form-cue invariant motion processing in primate visual cortex. *Science*, *255*, 1141–1143.

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

Bracewell, R. N. (1965). *The Fourier transform and its applications*. New York: McGraw-Hill.

Braddick, O. J. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London*, *B*, *290*, 137–151.

Cavanagh, P. (1992). Attention-based motion perception. *Science*, *257*, 1563–1565.

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

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

Clifford, C. W. G., Freedman, J. N., & Vaina, L. M. (1998). First- and second-order motion perception in Gabor micropattern stimuli: psychophysics and computational modelling. *Cognitive Brain Research*, *6*(4), 263–271.

Clifford, C. W. G., & Vaina, L. M. (1999). A computational model of selective deficits in first and second-order motion processing. *Vision Research*, *39*, 113–130.

Derrington, A. M., & Ukkonen, O. I. (1999). Second-order motion discrimination by feature-tracking. *Vision Research*, *39*(8), 1465–1475.

Fennema, C. L., & Thompson, W. B. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processes*, *9*, 301–315.

Georgeson, M. A., & Harris, M. G. (1990). The temporal range of motion sensing and motion perception. *Vision Research*, *30*, 615–619.

Greenlee, M. W., & Smith, A. T. (1997). Detection and discrimination of first- and second-order motion in patients with unilateral brain damage. *Journal of Neuroscience*, *17*, 804–818.

Hammett, S. T., Ledgeway, T., & Smith, A. T. (1993). Transparent motion from feature- and luminance-based processes. *Vision Research*, *33*(8), 1119–1122.

Hess, R. F., & Snowden, R. J. (1992). Temporal properties of human visual filters: number, shapes and spatial covariation. *Vision Research*, *32*, 47–60.

Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, *17*, 185–203.

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

Johnston, A., Benton, C. P., & McOwan, P. W. (1999b). Induced motion at texture-define motion boundaries. *Proceedings of the Royal Society of London*, *B*, *266*, 2441–2450.

Johnston, A., & Clifford, C. W. G. (1995a). Perceived motion of contrast modulated gratings: predictions of the multi-channel gradient model and the role of full-wave rectification. *Vision Research*, *35*, 1771–1783.

Johnston, A., & Clifford, C. W. G. (1995b). A unified account of three apparent motion illusions. *Vision Research*, *35*, 1109–1123.

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.

Koenderink, J. J., & van Doorn, A. J. (1987). Representation of local geometry in the visual system. *Biological Cybernetics*, *55*, 367–375.

Ledgeway, T., & Smith, A. T. (1994). Evidence for separate mechanisms for first- and second-order motion in human vision. *Vision Research*, *34*, 2727–2740.

Lu, Z.-L., & Sperling, G. (1995a). Attention-generated apparent motion. *Nature*, *377*, 237–239.

Lu, Z.-L., & Sperling, G. (1995b). The functional architecture of human visual motion perception. *Vision Research*, *35*, 2697–2722.

Mareschal, I., & Baker, C. L. (1998). A cortical locus for the processing of contrast-defined contours. *Nature Neuroscience*, *1*, 150–154.

O'Keefe, L. P., & Movshon, A. J. (1998). Processing of first- and second-order motion signals by neurones in area MT of the macaque monkey. *Visual Neuroscience*, *15*, 305–317.

Patterson, R., Donnelly, M., Phinney, R. E., Nawrot, M., Whiting, A., & Eyle, T. (1997). Speed discrimination of stereoscopic (cyclopean) motion. *Vision Research*, *37*(7), 871–878.

Seiffert, A. E., & Cavanagh, P. (1998). Position displacement, not velocity, is the cue to motion detection of second-order stimuli. *Vision Research*, *38*, 3569–3582.

Shioiri, S., & Cavanagh, P. (1990). ISI produces reverse apparent motion. *Vision Research*, *30*, 757–768.

Smith, A. T., Greenlee, M. W., Singh, K. D., Kraemer, F. M., & Hennig, J. (1998). The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging (fMRI). *Journal of Neuroscience*, *18*, 3816–3830.

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.

Somers, D. C., Seiffert, A. E., Dale, A. M., & Tootell, R. B. H. (1998). Second-order motion stimulus-induced activation and attentional modulation of human visual cortical areas MT & V3a. *Investigative Ophthalmology and Visual Science*, *39*, 1129.

Somers, D. C., Seiffert, A. E., Dale, A. M., & Tootell, R. H. (1999). fMRI investigations of motion after-effects with 1st- and 2nd-order stimuli. *Investigative Ophthalmology and Visual Science*, *40*(4), 199.

Sperling, G. (1989). Three stages and two systems of visual processing. *Spatial Vision*, *4*, 183–207.

Vaina, L. M., Cowey, A., & Kennedy, D. (1999). Perception of first- and second-order motion: separable neurological mechanisms? *Human Brain Mapping*, *7*, 67–77.

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

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

Victor, J. D., & Conte, M. M. (1992). Evoked potential and psychophysical analysis of Fourier and non-Fourier motion mechanisms. *Visual Neuroscience*, *9*, 105–123.

Watson, A. B., & Ahumada, A. J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America*, *A*, *2*, 322–341.

Zanker, J. M. (1993). Theta motion: a paradoxical stimulus to explore higher order motion extraction. *Vision Research*, *33*, 553–569.

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