# Non-Fourier motion analysis

Charles Chubb*, James McGowan*, George Sperling†, Peter Werkhoven‡

*Department of Psychology, Rutgers University, New Brunswick, NJ 08903, †Department of Cognitive Science, University of California at Irvine, Irvine, CA 92717, USA and ‡TNO Institute for Human Factors, 3769 ZG Soesterberg, The Netherlands*

*Abstract.* It has been realized for some time that the visual system performs at least two general sorts of motion processing. *First-order* motion processing applies some variant of standard motion analysis (i.e. spatiotemporal Fourier energy analysis) directly to stimulus luminance, whereas *second-order* motion processing applies standard motion analysis to one or another grossly non-linear transformation of stimulus luminance. We have developed a method for disentangling the different sorts of mechanisms that may operate in human vision to detect second-order motion. This method hinges on an empirical condition called *transition invariance* that may or may not be satisfied by a family $\Phi$ of textures. Any failure of this condition indicates that more than one mechanism is involved in detecting the motion of stimuli composed of the textures in $\Phi$. We have shown that the family of sinusoidal gratings oriented orthogonally to the direction of motion and varying in contrast and spatial frequency is transition invariant. We modelled the results in terms of a single-channel motion computation. We have new results indicating that a specific class of textures differing in texture element density and texture element contrast decisively fails the test of transition invariance. These findings suggest that in addition to the single second-order motion channel required by our earlier results there exists at least one other second-order motion channel. We argue that the preprocessing transformation used by this channel is a pointwise non-linearity that maps stimulus contrasts of absolute value less than some relatively high threshold $\tau$ onto 0, but increases with $|c| - \tau$ for contrasts $c$ of absolute value greater than $\tau$.

*1994 Higher-order processing in the visual system. Wiley, Chichester (Ciba Foundation Symposium 184) p 193–210*

## Standard motion analysis and covariant motion analysis

Various computations have been suggested to detect motion in a dynamic input function. Some of the prominent models in this class are motion energy analysers (Adelson & Bergen 1985), Reichardt detectors (Reichardt 1961, van Santen & Sperling 1985), motion sensors (Watson & Ahumada 1985) and gradient detectors (Marr & Ullman 1981). Although there are superficial differences among these devices, there are deep similarities among the computational ends

they achieve (Adelson & Bergen 1986, Simoncelli & Adelson 1991). In particular, all of these models respect the principle of motion from Fourier components (Watson et al 1986). This principle asserts that the motion percept elicited by a dynamic visual input function $I$ is consonant with the distribution of the energy of $I$ in the Fourier domain. Thus, for instance, it dictates that if most of the energy in $I$ resides in rightward drifting Fourier components (sinusoidal gratings), then the predominant direction of motion displayed by $I$ should be rightward. We shall refer to any motion computation conforming to the principle of motion from Fourier components as standard motion analysis.

Consider the result of applying standard motion analysis to a four-frame stimulus $F$ of the sort illustrated in Fig. 1. $F$ assigns the values $A$ and $B$ uniformly throughout rectangles marked 'A' and 'B', respectively, and the value 0 throughout all unmarked rectangles. Let us call the diagonal motion path containing only rectangles of value A the homogeneous path and the oppositely inclined motion path containing alternating rectangles of values A and B the heterogeneous path. Although Fig. 1 shows the homogeneous path running rightward, and the heterogeneous path leftward, the stimulus $F$ schematized should be viewed as a realization of a random stimulus that could equally well have reversed the relative directions of these two paths.

A simple argument shows that the principle of motion from Fourier components predicts the following: the strength of homogeneous motion elicited by $F$ (homogeneous strength of $F$) is given by:

$$\text{homogeneous strength of } F = h(f^2(A) - f(A)f(B)) \tag{1}$$

for some strictly increasing function $h:\mathbb{R} \to \mathbb{R}$ and $f:\mathbb{R} \to \mathbb{R}$ satisfying $h(0) = f(0) = 0$. Thus, if $A = B$, *homogeneous strength of* $F = 0$, indicating that the motion elicited

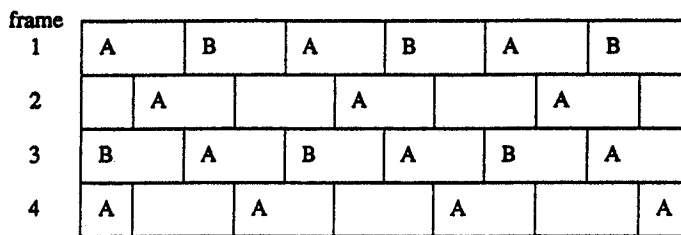| frame | | | | | | |
|---|---|---|---|---|---|---|
| 1 | A | B | A | B | A | B |
| 2 | | A | | A | | A |
| 3 | B | A | B | A | B | A |
| 4 | A | | A | | A | |

FIG. 1. A diagram of a four-frame dynamic visual display, $F$. $F$ assigns the values A and B uniformly throughout rectangles marked 'A' and 'B' respectively, and the value 0 throughout all unmarked rectangles. Call the motion path containing only rectangles of value A the *homogeneous* path and the path containing alternating rectangles of values A and B the *heterogeneous* path. Although the figure shows the homogeneous path running rightward, and the heterogeneous path leftward, the stimulus $F$ schematized should be viewed as a realization of a random stimulus that could equally well have reversed the relative directions of these two paths.

by $F$ will be ambiguous. If *homogeneous strength of $F<0$*, $F$ should display heterogeneous motion with probability greater than 0.5. If *homogeneous strength of $F>0$*, $F$ should display homogeneous motion with probability greater than 0.5.

A special case of Equation (1) is obtained by supposing that:

$$homogeneous \ strength \ of \ F = f^2(A) - f(A)f(B) \tag{2}$$

for some increasing function $f: \mathbb{R} \rightarrow \mathbb{R}$ satisfying $f(0) = 0$. We shall refer to such a computation as covariant motion analysis.

### First- and second-order mechanisms for detection of motion

It has been realized for some time now that the human visual system performs at least two general sorts of motion processing. First-order motion processing applies standard motion analysis directly to stimulus luminance, whereas second-order motion processing applies standard motion analysis to one or another grossly non-linear transformation of stimulus luminance (e.g. Ramachandran et al 1973, Sperling 1976, Lelkins & Koenderink 1984, Derrington & Badcock 1985, Green 1986, Pantle & Turano 1986, Cavanagh & Mather 1989, Cavanagh et al 1989, Turano & Pantle 1989, Victor & Conte 1990).

Chubb & Sperling (1988) showed how to construct stimuli that could selectively engage second-order mechanisms without systematically stimulating first-order ones. They went on to describe several different stimuli whose motion was obvious to human subjects, yet which would be invisible to first-order mechanisms. For each of these stimuli, $I$, although its motion could not be detected directly by standard motion analysis, its motion could be revealed to standard motion analysis by application of a non-linear transformation.

The motion of some of these stimuli could be disclosed to standard motion analysis by simple rectification of stimulus contrast. Consider, for instance, a vertical bar stepping from left to right across a grey visual field. If the bar retains a single intensity (e.g. black or white) as it steps along, then its motion will be available directly to standard motion analysis. However, if the bar alternates randomly between white and black with every step its takes, then its motion is no longer available directly to standard motion analysis. In particular, it can be shown that such a bar, $C$, is *drift balanced*: that is, the expected energy in $C$ of any given Fourier component is equal to the expected energy in $C$ of the component of the same spatial frequency drifting at the same rate but in the opposite direction. Although $C$'s motion cannot be detected by a direct application of standard motion analysis, we can disclose $C$'s motion to standard motion analysis simply by rectifying $C$'s contrast at each point in space–time. This changes all instances of both black and white to the same high value, yielding a bar of uniform, high value stepping across a field of uniform low value. We can easily use standard motion analysis to detect the motion of this transformed bar.

Simple pointwise rectification suffices to expose the motion of a bar of randomly reversing contrast to standard motion analysis, but other stimuli require more complicated preprocessing for their motion to be exposed to standard motion analysis. Imagine, for instance, a stimulus $G$ comprising a field of static black and white visual noise across which we sweep an imaginary vertical bar that causes the pixels within it to flicker rapidly between black and white. This stimulus presents obvious and compelling motion. $G$, again, is drift balanced. Moreover, simple rectification (or any other pointwise transformation) of stimulus contrast applied to $G$ yields a spatiotemporal function that is still drift balanced. Perhaps the easiest way to disclose $G$'s motion to standard motion analysis is to apply a temporal bandpass filter to $G$ and rectify the output. Such a transformation yields high values throughout the flickering regions within the moving bar and low values throughout the background static noise field, and thus succeeds in revealing the motion of the flicker-defined bar to standard motion analysis.

The stimulus $G$ results from modulating the temporal frequency of spatially white noise as a function of space and time. Other sorts of drift-balanced stimuli, represented by a stimulus $Q$, can be produced by quilting together patches of various sorts of static texture so that motion is carried purely by spatial textural properties, but not systematically by luminance (Chubb & Sperling 1991).

A straightforward approach to exposing the motion of such a stimulus $Q$ to standard motion analysis is to apply a spatial filter to $Q$ that is differentially sensitive to the textures of which $Q$ is composed, and to rectify the filtered output.

### How many different second-order mechanisms for detection of motion are there and what preprocessing transformations do they use?

For each of these example stimuli, one can imagine a simple, non-linear transformation that suffices to expose the motion of that stimulus to standard motion analysis. However, we can also imagine a single transformation that suffices to expose the motion of all three different stimuli. Specifically, the motion of each of $C$, $G$ and $Q$ could be disclosed to standard motion analysis by (a) an appropriate space–time separable linear filter followed by (b) a rectifier (e.g. a pointwise square or absolute value transformation). The question thus arises as to whether the motion of all three stimuli, $C$, $G$ and $Q$, is detected by a single mechanism that uses a transformation of this sort, or whether there exist multiple second-order motion mechanisms, using different preprocessing transformations, which are collectively responsible for detecting the motion of $C$, $G$, $Q$ and other non-Fourier stimuli.

### A method for investigating the dimensionality of second-order motion preprocessing

We can use Fig. 1 as a template for producing second-order motion stimuli. Throughout this section we write $\mu$ for the 'blank texture' consisting of a uniform

field of mean luminance. For any textures $\alpha$ and $\beta$, let I($\alpha$, $\beta$) be the stimulus obtained by filling rectangles marked 'A' and 'B' in Fig. 1 with patches of $\alpha$ and $\beta$, respectively.

All the stimuli that we shall consider in this section are drift balanced and hence strictly second order. We shall assume that, for some integer $N$, the direction of motion elicited by $I(\alpha, \beta)$ depends on the responses of $N$ second-order motion channels, each of which applies covariant motion analysis (see Equation (2)) to some non-linear, preprocessing transformation of the visual input. $T_1$, $T_2$, . . . $T_N$ are the preprocessing transformations used by our $N$ motion channels. We assume that the space-average response of $T_i$ to $\mu$ is 0 for $i = 1$, 2, . . . $N$. For any texture $\gamma$, we write $E_{\gamma,i}$ for the space-average response of $T_i$ to $\gamma$.

We shall assume that for each channel, $i = 1$, 2, . . . $N$, the strength of homogeneous motion given by that channel alone in response to $I(\alpha, \beta)$ is given by

$$homogeneous\ strength_i\ of\ I(\alpha,\ \beta) = f_i^2(E_{\alpha,i}) - f_i(E_{\beta,i})f_i(E_{\alpha,i}) \qquad (3)$$

for some strictly increasing function $f_i : \mathbb{R} \to \mathbb{R}$ satisfying $f_i(0) = 0$. (Thus, a negative value of $homogeneous\ strength_i\ of\ I(\alpha, \beta)$ indicates that the motion elicited in the $i^{\text{th}}$ channel by $I(\alpha, \beta)$ will tend to be in the heterogenous direction.)

We further assume that the signals of individual channels combine additively to determine the percept of global motion elicited by $I(\alpha, \beta)$. Specfically, for some strictly increasing function $H : \mathbb{R} \to \mathbb{R}$ satisfying $H(0) = 0$, we assume that

$$
\begin{aligned}
& homogeneous\ strength\ of\ I(\alpha,\ \beta) \\
& = H\left[ \sum_{i=1}^{N} homogeneous\ strength_i\ of\ I(\alpha,\ \beta) \right] \\
& = H\left[ \sum_{i=1}^{N} f_i^2(E_{\alpha,i}) - f_i(E_{\beta,i})f_i(E_{\alpha,i}) \right]
\end{aligned}
\qquad (4)
$$

## Transition invariance

We call any family $\Phi$ of textures transition invariant (Werkhoven et al 1993) if the following condition holds for all textures $\alpha$, $\beta \in \Phi$: if $I(\alpha, \mu)$ elicits homogeneous motion and $I(\alpha, \beta)$ is ambiguous, then $I(\beta, \alpha)$ is also ambiguous.

Let us explore the implications of this property. First, some general notation: for any texture $\gamma$, set $v_{\gamma,i} = f_i(E_{\gamma,i})$ for $i = 1$, 2, . . . $N$, and compose the vector

$$v_\gamma = (v_{\gamma,1},\ v_{\gamma,2},\ .\ .\ .\ v_{\gamma,N}) \qquad (5)$$

For any textures $\gamma$ and $\theta$, we can now write simply

*homogeneous strength of* $I(\gamma,\theta) = H(v_\gamma \cdot v_\gamma - v_\theta \cdot v_\gamma) = H(|v_\gamma|^2 - v_\theta \cdot v_\gamma)$   (6)

By assumption, $E_{\mu,i} = 0$ and $f_i(0) = 0$ for $i = 1, 2, \ldots N$. Thus, for all $i$, $v_{\mu,i} = f_i(E_{\mu,i}) = 0$, yielding the fact that $v_\mu = (0, 0, \ldots 0)$.

Note then that for any texture $\gamma$,

$$\text{\textit{homogeneous strength of} } I(\gamma, \mu) = H(|v_\gamma|^2 - v_\mu \cdot v_\gamma)$$
$$= H(|v_\gamma|^2) \qquad (7)$$

By assumption, $H(0) = 0$ and $H(x) > 0$ for any $x > 0$. Thus, $I(\gamma, \mu)$ elicits ambiguous motion if $v_\gamma = (0, 0, \ldots 0)$; otherwise, its motion is homogeneous.

Now, let $\alpha$, $\beta$ be textures such that $I(\alpha, \mu)$ displays homogeneous motion (implying that $v_\alpha \neq (0, 0, \ldots 0)$), whereas $I(\alpha, \beta)$ is ambiguous in motion content. From Equation (6) we note that the strength of homogenous motion elicited by $I(\alpha, \beta)$ is

$$H(|v_\alpha|^2 - v_\beta \cdot v_\alpha) \qquad (8)$$

Hence $I(\alpha, \beta)$ is ambiguous precisely if

$$v_\alpha \cdot v_\beta = |v_\alpha|^2 \qquad (9)$$

or equivalently if

$$v_\alpha \cdot (v_\beta - v_\alpha) = 0 \qquad (10)$$

Next note, by virtue of Equation (10), that

$$
\begin{aligned}
|v_\beta|^2 &= |(v_\alpha + (v_\beta - v_\alpha))|^2 \\
&= |v_\alpha|^2 + 2v_\alpha \cdot (v_\beta - v_\alpha) + |v_\beta - v_\alpha|^2 \\
&= |v_\alpha|^2 + |v_\beta - v_\alpha|^2
\end{aligned} \qquad (11)
$$

We note from Equation (11) that

$$|v_\beta|^2 \geqslant |v_\alpha|^2 = v_\beta \cdot v_\alpha \qquad (12)$$

from which it follows that

$$H(|v_\beta|^2 - v_\beta \cdot v_\alpha) \geqslant 0 \qquad (13)$$

Note that if $v_\beta = v_\alpha$, then the left and right sides of Equation (13) are equal, indicating that the motion of $I(\beta, \alpha)$ is ambiguous. On the other hand, if $v_\alpha \neq v_\beta$, then we see that

$$|v_\beta|^2 = |v_\alpha|^2 + |v_\beta - v_\alpha|^2 > |v_\alpha|^2 = v_\alpha \cdot v_\beta \tag{14}$$

implying that

$$H(|v_\beta|^2 - v_\beta \cdot v_\alpha) > 0 \tag{15}$$

and hence that $I(\beta, \alpha)$ displays predominantly homogeneous motion.

In summary, then, we have shown the following: given that $I(\alpha, \mu)$ elicits homogeneous motion and $I(\alpha, \beta)$ is ambiguous, then $I(\beta, \alpha)$ will be ambiguous if and only if $v_\alpha = v_\beta$; otherwise, $I(\beta, \alpha)$ will elicit homogeneous motion.

Note $v_\alpha$ and $v_\beta$ have the same dimensionality; it is only possible to have $v_\alpha \neq v_\beta$ if their mutual dimensionality is greater than 1. To see this, note that if $v_\alpha$ and $v_\beta$ are just real numbers, the fact that $v_\alpha \neq 0$ plus the fact that $v_\alpha^2 = v_\beta v_\alpha$ requires that $v_\alpha$ is equal to $v_\beta$. Here we encounter the

> *principle of transition invariance: a failure of transition invariance for textures $\alpha$ and $\beta$ reveals that more than one second-order motion mechanism is involved in determining the global motion percepts elicited by the stimuli $I(\alpha, \beta)$ and $I(\beta, \alpha)$.*

## Empirical results using the principle of transition variance

*Texture-defined motion is one dimensional for the class of sinusoidal gratings oriented perpendicular to the direction of motion*

Werkhoven et al (1993) showed that the family $\Xi$ of sinusoidal gratings oriented orthogonal to the direction of motion, but varying freely in contrast and spatial frequency is transition invariant. The results supported a single-channel model of texture-defined motion perception for the class $\Xi$. The non-linear preprocessing transformation used by this mechanism consists of a broadly tuned bandpass spatial filter, with a preference for low spatial frequencies, followed by a rectifier.

*Texture-defined motion is multidimensional for the class of white noise textures differing in noise particle density and contrast*

Define texture $X_{c,d}$ as follows. Partition the visual field into small squares (in the experiment described below, squares were 6.25'). Every square $s$ will be assigned uniform intensity, $X_{c,d}[s]$, in the texture. These intensities are jointly independent, identically distributed random variables. The distribution of $X_{c,d}[s]$ can be described most easily as follows. With each square $s$ associate a random variable, $\phi_s$ distributed as follows:

$$\phi_s = \left\{ \begin{array}{ll} 1 & \text{with probability } d/2 \\ 0 & \text{with probability } (1-d) \\ -1 & \text{with probability } d/2 \end{array} \right. \tag{16}$$

Then we set

$$X_{c,d}[s] = m(1 + c\phi_s) \tag{17}$$

for $m$ the mean luminance of the display. In other words, with probability $d$, $s$ becomes a *noise square*, in which case $X_{c,d}[s]$ is either $m(1+c)$ or $m(1-c)$ with equal probability; with probability $(1-d)$, $s$ becomes a blank square, in which case $X_{c,d}[s]$ takes the value $m$ (mean luminance).

Thus, for instance, $X_{1.0,1.0}$ presents a texture with maximal noise square density (i.e. all squares are noise squares) at maximal contrast. At the other extreme, for any q, $0 \leqslant q \leqslant 1$, the three textures $X_{q,0}$, $X_{0,q}$ and $X_{0,0}$ are identical, each being a uniform field of mean luminance. More interesting examples are textures such as $X_{0.42,1.0}$ which presents a dense field of medium-contrast noise squares, and $X_{1.0,0.18}$, which presents a sparse field of high-contrast noise squares. These two textures are presented side by side in Fig. 2. Obviously, the textures $X_{1.0,0.18}$ and $X_{0.42,1.0}$ are perceptually distinct. What is interesting about them is that their expected energy spectra are identical. (Both are flat and of equal energy over all low frequencies, diverging from flatness identically over frequencies higher than the fundamental components of the texture sampling grid.) Indeed, the expected energy spectra of all textures in $\mathcal{E}$ have the same shape. They differ, if at all, only in overall energy. This means that only a preprocessing transformation that preserves the information in the original image histogram will be able to discriminate between $X_{1.0,0.18}$ and $X_{0.42,1.0}$. The simplest sort of transformation we might use to discriminate between them is a pointwise transformation of some sort applied directly to stimulus contrast. (Note, however, that a squaring rectifier will not do the job.) The transformations that do not work very well are those that begin with a low-pass spatial filter, or indeed any spatial filter that pools over multiple texture elements to produce its output at any given point. It is an implication of the central limit theorem that any such initial linear transformation will obliterate the differences between these textures.

In particular, the transformation suggested by Werkhoven et al (1993) to account for their results concerning motion carried by sinusoidal grating patches cannot discriminate between $X_{1.0,0.18}$ and $X_{0.42,1.0}$. None the less, viewers clearly discriminate them. This raises the question as to whether there exists a motion mechanism that discriminates them—i.e. whether there exists a motion mechanism that applies a pointwise transformation of some sort directly to stimulus contrast. If there does exist such a 'histogram-sensitive' mechanism,
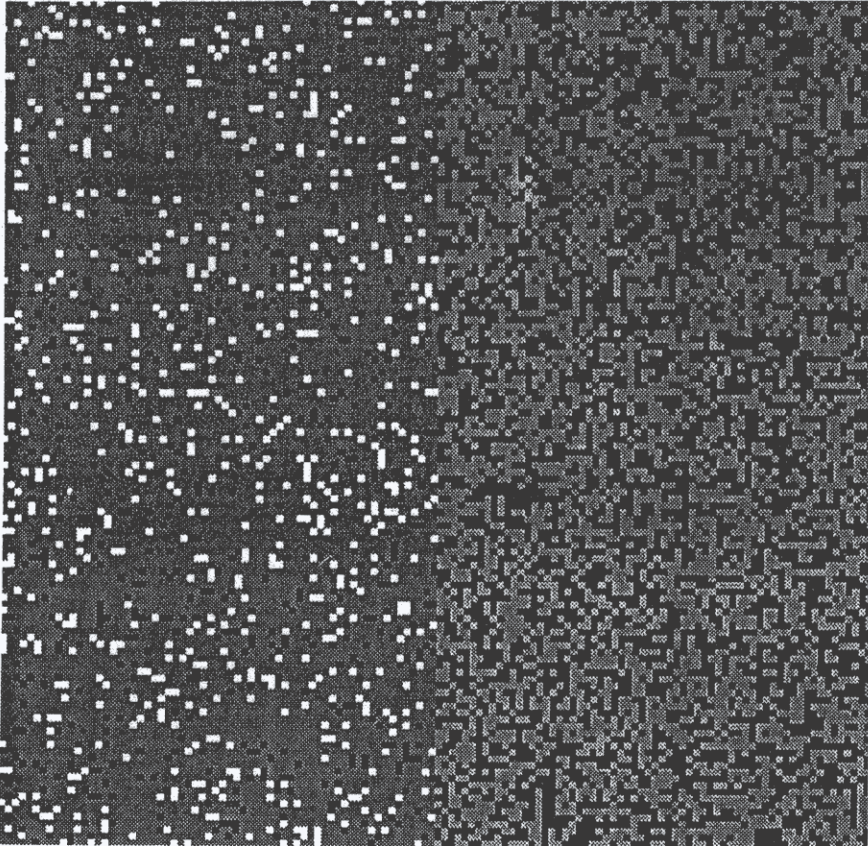
FIG. 2.    Textures $X_{1.0,0.18}$ (on the left) and $X_{0.42,1.0}$ (on the right). $X_{1.0,0.18}$ presents a sparse field of high contrast noise squares, whereas $X_{0.42,1.0}$ presents a dense field of medium contrast noise squares. Although the textures $X_{1.0,0.18}$ and $X_{0.42,1.0}$ are perceptually distinct, their expected energy spectra are identical.

then we should observe a failure of transition invariance across the set $\Xi$, because the textures are also discriminated by the mechanism of Werkhoven et al (1993) purely on the basis of their overall energy content.

*Experimental set up.*    To investigate this question, we used an approach similar to that of Werkhoven et al (1993). Stimuli were presented on a TVM MG-11 monitor under the control of a True Vision AT Vista Graphics Processor. The minimal component of any stimulus was a $4 \times 4$ block of pixels. We shall refer to such pixel blocks as 'texture elements'. A palette of 256 linear grey levels

was available for stimulus construction. Linearization was achieved using a 'by-eye' procedure in which a regular grid of texture elements containing three intensities $L_{lo}$, $L_{hi}$ and $L_{mid}$ (half with intensity $L_{mid}$, 1/4 with $L_{lo}$ and 1/4 with $L_{hi}$) was made to alternate in a coarse vertical square-wave with texture comprising a chequerboard of texture elements alternating between intensities $L_{lo}$ and $L_{hi}$. The screen was then viewed from sufficiently far away that the fine granularity of the texture was barely visible. At this distance, the squarewave modulating between the two types of texture had a spatial frequency of approximately 4 cycles/degree. Since the texture itself cannot be resolved, the squarewave is visible only if the mean luminance of alternating texture bars is different. Thus, the intensity $L_{mid}$ that makes the squarewave vanish is equal to the average of the intensities $L_{lo}$ and $L_{hi}$. We generated a look-up table by reiterating this procedure with different luminances $L_{lo}$ and $L_{hi}$ to determine, in each case, the $L_{mid}$ midway between $L_{lo}$ and $L_{hi}$.

*Stimuli.* We used an annular display $I_{annular}$ analogous to the horizontal motion display $I$ depicted in Fig. 1. The radius of the outer edge of the annulus was 5° of visual angle at a viewing distance of 32.1 cm. The radius of the inner edge was 2.5°. Thus, the entire display subtended 10°. A single texture element (4 × 4 pixel block) subtended 6.25'. The annulus was partitioned into six blocks of equal size. Each block had a radial width of 2.5°.

For any textures $\alpha$ and $\beta$, $I_{annular}$ ($\alpha$, $\beta$) comprised four frames. We continue to write $\mu$ for the 'null texture', comprising a uniform field of mean luminance. Frame 1 contained alternating patches of textures $\alpha$ and $\beta$. Frame 2 was rotated clockwise 30° relative to frame 1, and patches of texture $\beta$ were replaced by patches of $\mu$. Frame 3 was identical to frame 1 except that the roles of textures $\alpha$ and $\beta$ were reversed: patches that were painted with $\alpha$ in frame 1 were painted with $\beta$ in frame 3 and vice versa. Frame 4 was identical to frame 2, except that the roles of $\alpha$ and $\mu$ were reversed. Each frame had a duration of 67 ms, and frames were presented in immediate succession (with no interframe interval). Thus, stimulus duration was 267 ms.

$I_{annular}$ pits a *homogeneous* motion path comprising successive patches of $\alpha$ against a *heterogeneous* path comprising alternating patches of $\alpha$ and $\beta$. As described in the preceding paragraph, the homogeneous path runs clockwise and the heterogeneous path counterclockwise; however, the direction of the homogeneous path was randomized from trial to trial in the experiments reported here.

Stimuli were viewed in a darkened room. The mean luminance of the display was 5.4 cd/m$^2$. Between trials the subject viewed a screen that was of uniform mean luminance, except for a bright cue spot located in the centre of the display. The subject initiated a trial by pressing a button. There was no interval between the press of the button and stimulus onset. The subject entered his response with another button press, and proceeded with the next trial when ready.

We varied contrast $c$ across the values 0.4, 0.55, 0.7, 0.85 and 1.0, and measured, for each of the stimuli $I_{annular}(X_{0.14,1.0}, X_{c,0.18})$ and $I_{annular} (X_{c,0.18}, X_{0.14,1.0})$, the probability of observing heterogeneous motion as a function of $c$. Each point is the proportion of 36 trials on which the subject registered heterogeneous motion. The data shown were collected in four blocks of trials (3 blocks of 100 and one block of 60). All 10 conditions (corresponding to the 10 results) were mixed in each block.

## Results and discussion

The results, for one subject, are plotted in Fig. 3. These data clearly indicate that for this subject, the class $\Xi$ of textures is *not* transition invariant. The dotted line gives the probability of observing heterogeneous motion from $I(X_{0.14,1.0}, X_{c,0.18})$ as a function of $c$, and the solid line gives the probability of observing heterogeneous motion from $I(X_{c,0.18}, X_{0.14,1.0})$ as a function of $c$. If a unique second-order motion mechanism were responsible for detecting the motion of these two stimuli, then their curves should cross the dashed 'ambiguity line' at the same place. However, this is not the case. On the one hand, we note that $I(X_{0.14,1.0}, X_{0.68,0.18})$ is ambiguous, whereas $I(X_{0.68,0.18}, X_{0.14,1.0})$ yields
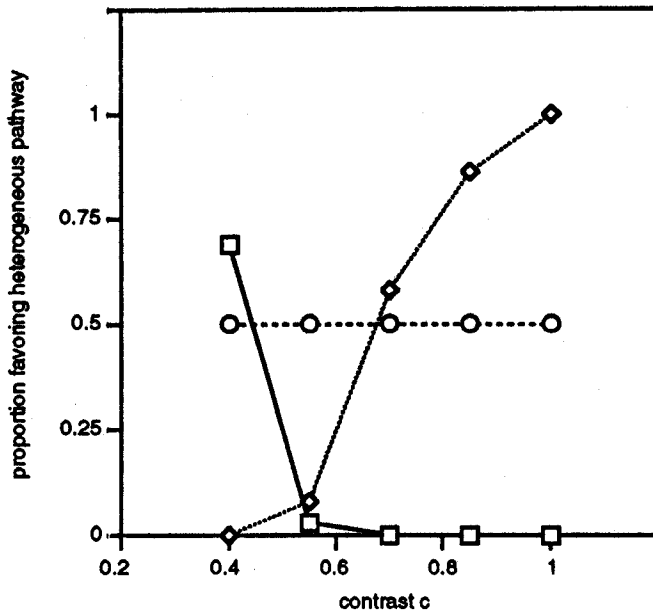


FIG. 3. The probability, as a function of $c$, of heterogeneous motion being elicited by $I(X_{0.14,1.0}, X_{c,0.18})$ (dotted line) and $I(X_{c,0.18}, X_{0.14,1.0})$ (solid line). Dashes indicate the 'ambiguity line'.

homogeneous motion with probability near 1.0. And on the other hand, $I(X_{0.45,0.18}, X_{0.14,1.0})$ is ambiguous, whereas $I(X_{0.14,1.0}, X_{0.45,0.18})$ yields homogeneous motion with probability near 1.0.

Thus we observe a dramatic failure of transition invariance. We conclude that more than one mechanism is involved in computing motion defined by the textures in $\Xi$. Certainly the mechanism suggested by Werkhoven et al (1993) discriminates between textures in $\Xi$ (purely on the basis of their overall energy). Our reasoning above, however, suggests that the other mechanism (or mechanisms) involved in detecting $\Xi$-defined motion uses a preprocessing transformation whose output at a particular location depends primarily on the input intensity at that location—i.e. a transformation that is approximately pointwise. The textures we have investigated thus far suggest that this pointwise non-linearity $r$ is of the following form: for some threshold, $\tau$, and some strictly increasing function $f$,

$$r(c) = \max\{0, f(|c| - \tau)\}. \tag{18}$$

Further experiments will be required to determine the exact form of $r$.

## Summary

We have reviewed a method for disentangling the different sorts of mechanisms that may operate in human vision to detect second-order motion. This method is used to determine whether or not a given family $\Phi$ of textures is transition invariant. Any failure of this condition indicates conclusively that more than one mechanism is involved in detecting the motion of stimuli composed of the textures in $\Phi$. Werkhoven et al (1993) showed that the family of sinusoidal gratings oriented orthogonally to the direction of motion and varying in contrast and spatial frequency was transition invariant. They modelled their results in terms of a single-channel motion computation. Here we have presented new results indicating that the class $\Xi$ of textures differing in texture element density and texture element contrast fails decisively to be transition invariant. These findings suggest that, in addition to the single second-order motion channel required by the results of Werkhoven et al (1993) there exists at least one other second-order motion channel. We have argued that the preprocessing transformation used by this channel is a pointwise non-linearity. Finally, our current data suggest that this pointwise transformation maps stimulus contrasts of absolute value less than some relatively high threshold $\tau$ onto 0, but increases with $|c| - \tau$ for contrasts $c$ of absolute value greater than $\tau$.

*Acknowledgements*

## References

Adelson EH, Bergen JR 1985 Spatio-temporal energy models for the perception of apparent motion. J Opt Soc Am A Opt Image Sci 2:284–299

Adelson EH, Bergen JR 1986 The extraction of spatio-temporal energy in human and machine vision. In: Proceedings of an IEEE workshop on motion: representation and analysis. IEEE, Computer Society Press, Washington DC, p 151–155

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

Cavanagh P, Arguin M, von Grunau M 1989 Interattribute apparent motion. Vision Res 29:1197–1204

Chubb C, Sperling G 1988 Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. J Opt Soc Am A Opt Image Sci 5:1986–2007

Chubb C, Sperling G 1991 Texture quilts: basic tools for studying motion from texture. J Math Psychol 35:411–442

Derrington AM, Badcock DR 1985 Separate detectors for simple and complex grating patterns? Vision Res 25:1869–1878

Green M 1986 What determines correspondence strength in apparent motion? Vision Res 26:599–607

Lelkins AMM, Koenderink JJ 1984 Illusory motion in visual displays. Vision Res 24:1083–1090

Marr D, Ullman S 1981 Directional selectivity and its use in early visual processing. Proc R Soc Lond Ser B Biol Sci 211:151–180

Pantle A, Turano K 1986 Direct comparisons of apparent motions produced with luminance, contrast-modulated (CM), and texture gratings. Invest Ophthalmol & Visual Sci 27:141(abstr)

Ramachandran VS, Rao VM, Vidyasagar TR 1973 Apparent movement with subjective contours. Vision Res 13:1399–1401

Reichardt W 1961 Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In: Rosenblith WA (ed) Sensory communication. Wiley, New York, p 303–317

Simoncelli EP, Adelson EH 1991 Relationship between gradient, spatiotemporal energy, and regression models for motion perception. Invest Ophthalmol & Visual Sci 32:893 (abstr)

Sperling G 1976 Movement perception in computer-driven visual displays. Behav Res Methods Instrum 8:144–151

Turano K, Pantle A 1989 On the mechanism that encodes the movement of contrast variations: velocity discrimination. Vision Res 29:207–221

van Santen JPH, Sperling G 1985 Elaborated Reichardt detectors. J Opt Soc Am A Opt Image Sci 2:300–321

Victor JD, Conte MM 1990 Motion mechanisms have only limited access to form information. Vision Res 30:289–301

Watson AB, Ahumada AJ Jr 1985 A model of human visual-motion sensing. J Opt Soc Am A Opt Image Sci 2:322–342

Watson AB, Ahumada AJ Jr, Farrell JE 1986 The window of visibility: a psychophysical theory of fidelity in time-sampled motion displays. J Opt Soc Am A Opt Image Sci 3:300–307

Werkhoven P, Sperling G, Chubb C 1993 The dimensionality of texture-defined motion: a single channel theory. Vision Res 33:463–485

## DISCUSSION

*Mollon:* You said nothing at all about the actual temporal relationships of the stimuli. It should be said of course that they're in **apparent** movement and therefore one might look back at all the apparent movement literature, where the perceptual organization of the moving stimuli depends very much upon their spatiotemporal properties (Pantle & Picciano 1976). Were you using equivalent spatiotemporal arrangements in your different experiments?

*Chubb:* Yes. We used the same spatiotemporal stimulus configuration in every trial. The only things that varied were the textures used to paint patches of the stimulus. Although I did not describe the stimulus configuration in detail in my presentation, a full description is given in the paper. For two textures $\alpha$ and $\beta$, we refer to the stimulus involving $\alpha$ and $\beta$ that we actually used in our experiments as $I_{annular}(\alpha,\beta)$. $I_{annular}(\alpha,\beta)$ is an annular analogue to the horizontally moving stimulus $I(\alpha,\beta)$ obtained by painting rectangles marked 'A' in Fig. 1 with texture $\alpha$ and rectangles marked 'B' with texture $\beta$.

*Graham:* How big were those boxes?

*Chubb:* In $I_{annular}(\alpha,\beta)$, the rotating texture annulus had a radial width of 2.5° and the radius of the mean luminant circle in the middle of the annulus was also 2.5°. The annulus itself was partitioned into six texture patches. Each patch comprised approximately 9.8° squared and contained roughly 900 texture elements. (Each texture element was a 6.25' square.) Motion was generated by rotary displacements equal to half the angular patch width.

*Morgan:* What was the stimulus onset asynchrony, and does it matter?

*Chubb:* The stimuli we used to generate the data I presented contained four frames. Each had a duration of 67 ms, yielding a total stimulus duration of 267 ms. However, the precise temporal characteristics do not seem to matter much.

*Mollon:* I would expect the contrary from the older literature on phenomena such as the Ternus effect (Pantle & Picciano 1976). The perceived motion is very dependent on the spatiotemporal conditions. Your frame rate seems rather high; I suspect you would find a very much larger number of apparent mechanisms if you slowed it down.

*Barlow:* Obviously you could get apparent motion, not from adjacent frames, but from alternate ones. Does this ever contribute?

*Chubb:* No, the displays we used are completely ambiguous across multiple frames. If you look at Fig. 1, you can see this quite easily by considering, for instance, just the first and third frames. The first frame contains patches marked alternately 'A' and 'B'. The third frame is simply the reversal of the first frame. Patches marked 'A' in frame 1 are marked 'B' in frame 3, and vice versa. A similar observation holds for frames 2 and 4.

*Movshon:* Does the outcome depend on the fact that you are only allowing the subjects to judge one motion? Peter Werkhoven's films look transparent to me.

*Chubb:* It is true that the motion stimuli used by Werkhoven et al (1993) do sometimes appear to elicit simultaneously both clockwise and counterclockwise motion. As a subject, you often feel as though you are an umpire trying to decide which of the opposite motion percepts is stronger. This does not seem to be so much the case with the textures used in the current experiments.

*Movshon:* So you didn't feel like you saw transparent motion?

*Chubb:* No. As a subject in these experiments, in any given trial you are remarkably certain of the global direction of motion. This point is suggested by the data in Fig. 3: the probability that a stimulus elicits heterogeneous motion is a very steep function of contrast.

*Snowden:* Concerning the stimulus, is the A in frame 1 of Fig. 1 precisely the same pattern as the A in frame 2?

*Chubb:* No, they are the same independently generated texture.

*Snowden:* So you are really just seeing between the whole block rather than the individual elements?

*Chubb:* Yes, we are forcing a second-order judgement. These are all microbalanced stimuli.

*Malik:* You have made a strong assumption that it is the content of the texture that is important, rather than the boundaries of the texture.

*Chubb:* Yes, that is an assumption.

*Malik:* On the other hand, we know from Christoph Nothdurft's data that the variation in structure over space is the crucial variable in explaining texture discrimination (see Nothdurft 1994, this volume).

*Chubb:* That's true, but it doesn't seem to be true in the case of motion. Think about how you go about extracting a texture boundary. You first need to apply a battery of up-front image transformations. Then you need to apply an operator that surveys the outputs of the up-front transformations in order to derive boundaries between different sorts of texture. Now, you want to take the output of these boundary-detecting computations and feed it into a motion-sensing mechanism. It seems more reasonable to think that the motion-from-texture computation does not sit on top of the boundary-detecting mechanism, but rather works analogously to it. Like the boundary-detection mechanism, the motion-from-texture mechanism must begin by applying some battery of up-front image transformations; however, whereas the boundary-detection system needs to survey the output of these up-front transformations to derive spatial boundaries, the motion-from-texture mechanism needs to extract the motion content of these up-front transformations.

*Bergen:* It just doesn't matter in this case. The kind of motion thing you are feeding it into doesn't really care.

*Chubb:* Maybe.

*Graham:* You don't think that by somehow looking at the borders instead of the contents of the patch, you could come out with the opposite conclusion, do you?

*Chubb:* No, I don't.

*Wilson:* You make a series of assumptions from which you're forced to conclude that there are at least two mechanisms here. One concern is the general issue of proliferation of mechanisms in the visual system. What do you think your weakest assumption is? Can you conceive of visual systems that could account for the data with just a single mechanism that was a little bit more sophisticated?

*Chubb:* I presented a model that's a little stronger than it has to be; there are ways of weakening it.

*Wilson:* Suppose that your low-contrast textures just generated a much longer latency. Would simple temporal aspects of a standard motion detector account for your data—short latencies in one case versus longer in the other, simply as a result of front-end filter characteristics?

*Chubb:* I don't know.

The point is, unless you think that there's also something else that's yielding another motion strength, you need two motion strength signals independently contributing in order to get a failure of transition analysis, irrespective of how you get them.

*Mollon:* Can I ask about the nature of multiple motion mechanisms? When you see ambiguous motion, are you supposing that there are two motion mechanisms, one that's secreting clockwise motion and one that's secreting anti-clockwise motion, and they're mixing in your consciousness?

*Chubb:* I wouldn't have put it that way.

*Mollon:* Are you then saying that there is a single higher-order movement mechanism, and if so, what are the inputs to that? Are the inputs really movement signals, or are they signals about the identity of textures?

*Chubb:* We are dealing with something that's quite distinct from the identity of textures. As I learned with Peter Werkhoven and George Sperling (Werkhoven et al 1993), you can have textures that are very distinct in terms of their capability to elicit texture boundaries. For instance, take a very high-contrast high-frequency texture and a low-contrast, low-spatial-frequency texture. They are quite distinct and elicit a very solid boundary, yet in the motion paradigm that we've been using, they yield completely ambiguous motion—motion that you can't see. This is because there is only one preprocessing transformation, that yields the same space average output to each of these two textures. It happens to be more sensitive to low spatial frequencies, but you can counteract this by cranking up the amplitude of the higher spatial frequency texture, so that by the time you finish preprocessing, you have something that's completely ambiguous to the motion mechanism that actually processes it.

*Sperling:* When the two directions of motion are balanced in this paradigm, the component texture patterns are motion metamers. In fact, there is a stronger

test: it must be possible to exchange the two kinds of patch—the high-contrast high-frequency patches and the low-contrast low-frequency patches—and still leave the competing motion directions balanced. That test is called transition invariance. When it holds it implies that there is only one motion mechanism— the equivalent to rod vision in the colour domain. Stimuli that produce equal responses of this mechanism are motion metamers. When transition invariance fails it means that there is more than one mechanism.

Given that there is only one mechanism, the strength of the response of this mechanism determines motion strength. Therefore, one can take a path composed of patches of high-frequency high-contrast sine waves alternating with patches of low-frequency low-contrast sine waves and the motion between these very disparate stimuli will overwhelm and win 100% of the time over the motion along a homogeneous path composed of the weaker one of the two.

That's a very strong counter-intuitive finding and it shows that it is not the similarity between textures that counts but how much they excite the motion-texture filter. The advantage of making some of the specific assumptions that Charlie [Chubb] mentioned is that it enables one actually to derive the properties of the filter and to predict the outcome of the experiments.

There's one more result about these metamers. The single-filter theory applies only to motion of similarly oriented patches. When grating orientation varies, there is more than one filter involved.

*Mollon:* Another way of putting this is that the motion-perceiving homunculus suffers certain colour-blindness-like states. Does that mean that there are multiple movement mechanisms, or are we talking about the selective blindnesses of a single mechanism?

*Chubb:* In the case that I was talking about here, one has to conclude that instead of monochromatic vision, you are dealing with dichromatic vision of a certain sort.

*Movshon:* Is there a single motion mechanism in these experiments because you are only allowing the subjects one judgement? It's perfectly clear that in any sensible neural sense there are many motion mechanisms in the stimulus, because there are receptive fields scattered all the way round the visual field. You are asking your subjects to integrate minimally over that collection of receptive fields and make a judgement. Suppose for some reason that the top half and the bottom half of the visual field gave different answers, it would none the less be the case in these experiments that you oblige the subjects to tell you only one thing, which would oblige them to adopt some strategy that would be equivalent and indistinguishable from saying there was only a single motion mechanism because they have to add or average or otherwise combine them into a single judgement.

*Chubb:* Our experiments can't answer that question.

## References

Nothdurft H-C 1994 Common properties of visual segmentation. In: Higher-order
    processing in the visual system. Wiley, Chichester (Ciba Found Symp 184) p 245–268
Pantle A, Picciano L 1976 A multistable movement display: evidence for two separate
    motion systems in human vision. Science 193:500–502
Werkhoven P, Sperling G, Chubb C 1993 The dimensionality of texture-defined motion:
    a single channel theory. Vision Res 33:463–485