Nonlinear Preprocessing in Short-range Motion

ETHAN TAUB,* JONATHAN D. VICTOR,†‡ MARY M. CONTE†

Received 23 February 1996: in revised form 24 July 1996

The phenomenon of non-Fourier motion (visually perceived motion that cannot be explained simply on the basis of the autocorrelation structure of the visual stimulus) is well recognized, and is generally considered to be due to nonlinear preprocessing of the visual stimulus prior to a stage of standard motion analysis. We devised a sequence of novel visual stimuli in which the availability of a motion stimulus depends on the nature of the nonlinear preprocessing: an nth order stimulus \( P_n \) will generate a perception of motion if it is preprocessed by a nonlinearity of polynomial order \( n \) or greater, but not if preprocessed by a nonlinearity of polynomial order less than \( n \). We found that unambiguous motion direction was perceived for \( P_2 \), \( P_3 \), and \( P_4 \), but not for higher-order stimuli, and we measured the contrast thresholds for direction discrimination with superimposed noise. We found that an asymmetric compressive nonlinearity can, in a unifed fashion, account for these results, while a purely quadratic nonlinearity or a rectification of the form \( |p| \leq \alpha \) cannot. We compared velocity discrimination judgements for second-order non-Fourier stimuli (\( P_2 \)) with standard drifting gratings. Although velocity comparisons were veridical, uncertainties were greater for the non-Fourier stimuli. This could be reproduced by substituting a Fourier grating with superimposed noise for the non-Fourier grating. These findings are consistent with a single pathway which processes both Fourier and non-Fourier short-range motion, and are discussed in the context of other investigations which have been interpreted as demonstrating separate pathways. © 1997 Elsevier Science Ltd.

INTRODUCTION

Standard models of motion extraction, based on autocorrelation (e.g., Reichardt, 1961; van Santen & Sperling, 1984), Fourier analysis (e.g., Heeger, 1987), or detection of gradients (e.g., Sobe & Srinivasan, 1991), are sufficient to account for the perception of motion in many, but not all, stimuli that are perceived as moving. The existence of exceptional stimuli, so-called non-Fourier motion stimuli, implies that an elaboration of standard motion models is required. Many psychophysical experiments have been interpreted as implying that independent pathways must exist in the visual system for the perception of Fourier and non-Fourier motion stimuli. Nevertheless, a number of investigators have proposed that these two mathematically distinct types of stimuli can be perceived by a single pathway containing an early nonlinearity. In this paper, we discuss the nature of standard motion models and of Fourier and non-Fourier motion stimuli, and present new data obtained in our laboratory with a novel class of stimuli—"higher-order non-Fourier motion stimuli". Although some models of a separate non-Fourier pathway would appear to require yet a third pathway to account for the perception of motion elicited by these stimuli, we show that our data are in accord with the parsimonious hypothesis of a single pathway. Previously reported data that might appear to require independent processing of Fourier and non-Fourier motion are re-examined, and it is shown that a single pathway can account for these data as well.

STANDARD MOTION ANALYSIS

The velocity of an object is a measurable physical property, but the problem of determining velocity from a visual stimulus is not well-posed. When confronted with a visual stimulus that can be generated by two or more different configurations of moving objects, the visual system has no foolproof way to determine which configuration is the true (physical) one. The familiar barber-pole image, for example, might be generated by an oblique grating moving vertically or horizontally, or even diagonally (Fig. 1). The tendency of an observer to perceive the stripes as moving up the pole, i.e., along the long axis of the aperture through which they are seen, is a result of the way the visual system solves the "aperture problem". More generally, ambiguous stimuli serve as strong tests of models for visual motion analysis because of the multiplicity of a priori possible perceptions.

The problem of determining velocity, though not well posed, is readily formalized for our requirements. Formally, a dynamic visual stimulus may be described
by a function $I(x_1, x_2, t)$, the luminance at retinal location $(x_1, x_2)$ at time $t$. (The present treatment ignores color and depth.) We will find it useful to rewrite $I$ as

$$I(x_1, x_2, t) = I_0[1 + c P(x_1, x_2, t)],$$

where $I_0$ is the background luminance, $P$ describes the spatiotemporal stimulus pattern, and $c$ is the contrast. $P$, unlike $I$, can take both positive and negative values and is, by convention, normalized to have a maximum deviation of 1 (so that the maximum contrast is $c$). The visual system is required to perform a computation on $P$ that yields a planar velocity $v = (v_1, v_2)$ if the stimulus is indeed the image of a rigid object moving at velocity $v$. A number of different algorithms have been proposed that might perform this task. Our goal is to determine the form of the algorithm used by the visual system, and ultimately, to arrive at a neuroanatomic and neurophysiologic understanding of the visual structures that carry out this algorithm.

How can motion be detected from the stimulus function $P$? As a simple case, consider a full-field visual stimulus that moves rigidly at a constant velocity $v = (v_1, v_2)$. This means that for any $x_1, x_2, t$, and time interval $\Delta t$,

$$P(x_1, x_2, t) = P(x_1 + v_1 \Delta t, x_2 + v_2 \Delta t, t + \Delta t).$$  

We now ask what computations can be performed on $P$ that will yield $v$. We also want our computation to yield $v$ if condition (2) is “nearly” true, in some sense. One family of possible solutions starts by computing the autocorrelation of $P$, which is defined as

$$A(x_1, x_2, t) = \langle P(x_1', x_2', t') P(x_1' - x_1, x_2' + x_2, t' + t) \rangle,$$

where $\langle \rangle$ indicates an average over space $(x_1'$ and $x_2')$ and time $(t')$.

If condition (2) is met for velocity $v$, then $A$ takes on its maximal value, $\langle \langle P(x_1', x_2', t') \rangle^2 \rangle$, at all points of the form $(v_1 t, v_2 t, t)$. The task of detecting motion and finding $v$ thus becomes the task of detecting a line of constant values of $A(x_1, x_2, t)$. The motion extraction models of Reichardt (1961) and van Santen & Sperling (1984) are autocorrelational models of this type.

Another approach toward detecting motion begins with a Fourier transformation. The information contained in the stimulus function $P$ may be represented in another way by its Fourier transform, $\hat{P}$, which is defined as

$$\hat{P}(k_1, k_2, \omega) = \int \int e^{-i(k_1 x_1 - k_2 x_2 - \omega t)} P(x_1, x_2, t) dx_1 dx_2 dt.$$  

Here, $\hat{P}(k_1, k_2, \omega)$ represents the contribution to the visual stimulus made by its Fourier component at spatial frequency $k = (k_1, k_2)$ and temporal frequency $\omega$. The Fourier components of the visual stimulus are drifting sinusoidal gratings; the Fourier component corresponding to $k$ and $\omega$ is oriented perpendicular to $k$ and appears to move in the direction of $-k$ with velocity $\omega k/|k|^2$. The qualification “appears to” is necessary, because any additional movement of the grating parallel to its bars is equally consistent with this visual stimulus and therefore cannot be detected. That is, any drifting sinusoidal grating is compatible with a wide range of motion velocities. The gratings that are compatible with velocity $v$ are precisely those that satisfy $k \cdot v + \omega = 0$. These are the gratings for which $(k_1, k_2, \omega)$ lies on the plane through the origin perpendicular to the line of points $(v_1 t, v_2 t, t)$. The task of computing $v$ from $\hat{P}$ is therefore the task of identifying a plane which contains all points $(k_1, k_2, \omega)$ for which $\hat{P}(k_1, k_2, \omega)$ is nonzero. The motion perception model of Heeger (1987) is a Fourier model of this type.

Autocorrelational motion models are equivalent to Fourier motion models that ignore phase. This is so because of the relation between $\hat{P}$ and $\hat{A}$, the Fourier transform of $A$:

$$\hat{A}(k_1, k_2, \omega) = |\hat{P}(k_1, k_2, \omega)|^2.$$  

That is, $\hat{A}$ (and hence $A$) can be obtained from $\hat{P}$ by a calculation that ignores phase. The quantity $\hat{P}(k_1, k_2, \omega)|^2$ is a real number corresponding to the energy of the Fourier component of the stimulus at the spatiotemporal frequency $(k_1, k_2, \omega)$. Motion models based on a computation of the autocorrelation $A$, or, equivalently, the distribution of Fourier energies $|\hat{P}(k_1, k_2, \omega)|^2$, are called “standard” or “first-order” motion models, in the terminology of Chubb & Sperling (1988).

**NON-FOURIER MOTION**

We have seen that standard motion analysis provides a framework to understand the computation of velocity from visual stimuli. Standard motion analysis cannot, however, account for the perception of motion in all stimuli that are seen as moving. The exceptional stimuli have been called “second-order” or “non-Fourier” motion...
stimuli (Chubb & Sperling, 1988). We will use the term “non-Fourier” for visual stimuli which elicit the perception of motion, but for which the motion is inaccessible to standard motion analysis. We will use the terms “second-order”, “third-order”, and “fourth-order” for particular classes of non-Fourier stimuli, in a sense that will be defined. [Our use of the term “second-order” coincides with that of Chubb & Sperling (1988, 1989), but our use of the term “third-order” (Taub & Victor, 1993) differs from that of Lu & Sperling (1995a,b), who used the term to characterize motion stimuli based on their perceptual features, rather than on their statistical properties.]

Our second-order non-Fourier motion stimulus is a modified drifting sinusoidal grating constructed from unit-sized checks. The value of the stimulus function $P_2$ at each check is determined by multiplying the grating by either +1 or −1, with the value chosen randomly at each check. Thus, for integer values of $x_1$ and $x_2$,

$$P_2(x_1, x_2, t) = (-1)^{R(x_1, x_2)} \cos(k'_1 x_1 + k'_2 x_2 + \omega' t). \quad (6)$$

where $R(x_1, x_2)$ is chosen randomly and independently from {0, 1} at each pair of integers $(x_1, x_2)$. Equation (6) will be more readily generalizable in what follows if we rewrite it equivalently as

$$P_2(x_1, x_2, t) = \cos(k'_1 x_1 + k'_2 x_2 + \omega' t + \pi R(x_1, x_2)). \quad (7)$$

All checks are modulated sinusoidally at time at the frequency $\omega'$. An instantaneous snapshot of this stimulus, shown in Fig. 2, resembles a randomly colored checkerboard modulated by a sine wave, so that a grating of gray bars appears. With this temporal modulation, the gray bars appear to drift across the screen.

Standard motion analysis cannot account for the perception of motion in this stimulus. The key observation is that the random choice of $R(x_1, x_2)$ forces the autocorrelation of $P_2$ to be zero:

$$A_2(x_1, x_2, t) = \langle P_2(x'_1, x'_2, t') P_2(x'_1 + x_1, x'_2 + x_2, t' + t) \rangle$$

$$= \langle (-1)^{R(x', x'_2)} (-1)^{R(x'_1 - x_1, x'_2 - x_2)} \cos(k'_1 x'_1 + k'_2 x'_2 + \omega' t') \cos(k'_1 (x'_1 + x_1) - k'_2 (x'_2 + x_2) + \omega' (t' + t)) \rangle.$$

$$= 0 \quad (8)$$

unless $x_1 = 0$ and $x_2 = 0$, because $R(x'_1 + x_1, x'_2 + x_2)$ and $R(x'_1 + x_1, x'_2)$ are independent. For $x_1 = 0$ and $x_2 = 0$, the autocorrelation function reduces to that of a grating: $A_2(0, 0, t) = \cos(\omega' t)/2$.

It follows from Eq. (5) that the Fourier motion energy $|P_2(k_1, k_2, \omega)|^2$ is nonzero only at points where $\omega' = \pm \omega$, and its values at these points are independent of $k_1$, $k_2$, and $\omega$. Thus, $|P_2(k_1, k_2, \omega)|^2$ contains an equal amount of energy in opposite directions, and hence no net motion energy.

Until now, we have ignored any visual processing which might precede the extraction of motion energy. However, to the extent that we can characterize the “front end” as a linear filter, the above argument remains valid. Passing the stimulus $P_2$ through a linear transformation $L$ will have the effect of multiplying each Fourier component $P_2(k_1, k_2, \omega)$ of the stimulus by a complex number [the value of the transfer function $L(k_1, k_2, \omega)$], and will thus not introduce any new nonzero Fourier components into $|P_2(k_1, k_2, \omega)|^2$.

While standard motion analysis cannot account for the global motion of the $P_2$ stimulus, it can account for local motion in small areas of the stimulus. When the $P_2$ stimulus is viewed in a small region [less than one wavelength $(2\pi/k')$ in width], it no longer appears to be a drifting grating. Rather, it looks like temporally varying random speckle, and local motion in either direction is sometimes seen. The average in Eq. (8) is guaranteed to be zero only if computed globally. Over restricted ranges of time and space, an empirical measurement of the autocorrelation $A_2(x_1, x_2, t)$ may be nonzero. However, unless $x_1 = 0$ and $x_2 = 0$, these estimates will fluctuate evenly around zero. This estimate of the autocorrelation may thus provide a nonzero motion signal, but one of random direction. Thus, a subject viewing the entire stimulus and using standard motion analysis would be expected to perceive randomly directed local motion. It is only the unidirectional, global motion of the $P_2$ stimulus that cannot be perceived with standard motion analysis. In our psychophysical experiments on $P_2$ and other non-Fourier motion stimuli, we have avoided the confounding effect of local motion by measuring thresholds for global direction discrimination rather than detection of motion as such.

To account for the perception of steady, unambiguous global motion in $P_2$ (and other non-Fourier motion stimuli), an extension of the standard model is required. The standard motion model can be schematically summarized by

$$P \rightarrow \text{linear filter } L \rightarrow L[P] \rightarrow \text{standard motion analysis} \rightarrow \text{velocity } \mathbf{v}. \quad (9)$$

We suggest [as did Chubb & Sperling (1989)] that the visual stimulus is subjected to a nonlinear transformation before undergoing standard motion analysis:

$$P \rightarrow \text{linear filter } L \rightarrow L[P] \rightarrow \text{nonlinear transformation } T \rightarrow T[L[P]] \rightarrow \text{standard motion analysis} \rightarrow \text{velocity } \mathbf{v}. \quad (10)$$

This suggestion is biologically plausible, because strong nonlinearities early in the visual system are known to exist. For example, a defining property of a Y cell is that it increases its firing rate in response to the introduction of a pattern into its receptive field and to its withdrawal. This must be the result of nonlinear...
spatial summation (Enroth-Cugell & Robson, 1966; Hochstein & Shapley, 1976). As another example, many cells in the primary visual cortex have little or no maintained discharge. As a result, they act as rectifiers. An on-center cell will increase its firing rate (from zero to some positive value) when confronted with a bright center and a dark surround, but cannot decrease its firing rate by an equal amount when confronted with the inverse stimulus. In principle, photoreceptor nonlinearities might also contribute to $T$, but such nonlinearities, which are smooth functions in the neighborhood of a contrast of zero, are unlikely to account for our findings (see section entitled “Estimation of the form of the nonlinearity $T$”).

The nonlinear transformation $T$ is critical to the emergence of unambiguous motion energy in a non-Fourier stimulus. For the second-order stimulus $P_2$ of Eqs (6) and (7), any nonlinear transformation that eliminates the cancellation of the factors ± 1 in Eq. (8) will lead to a nonzero value of the autocorrelation $A_2$, and thus, provide a nonzero input to standard motion analysis. Thus, full-wave rectification, $(T(p) = |p|)$, half-wave rectification $[T(p) = |p|$ for $p > 0$ and $0$ for $p < 0$], and a squaring transformation $(T(p) = p^2)$ would all reveal unambiguous motion in $P_2$. Because the squaring transformation is analytic, its effect is readily calculated explicitly.

Initially, we ignore $L$:

$$
T[P_2](x_1, x_2, t) = \cos^2(k_1'x_1 + k_2'x_2 + \omega't + R(x_1, x_2))
$$

$$
= \frac{1}{2} + \frac{1}{2} \cos(2k_1'x_1 + 2k_2'x_2 + 2\omega't).
$$

$T[P_2]$ thus contains a sinusoidal grating at spatiotemporal frequency $(2k_1', 2k_2', 2\omega')$, and will therefore be accessible to standard motion analysis. Note that such a grating has the same speed as the original grating from which the stimulus $P_2$ was derived, but with twice the spatial and temporal frequency. This accords well with the perception of moving gray bars: the bars are at the zeroes of a sinusoid of spatiotemporal frequency $(k_1', k_2', \omega')$, and there are thus two bars perceived in $P_2$ for every

---

**Non-Fourier Motion Stimuli**

A. Second-order

B. Third-order

C. Fourth-order

D. First-order + noise

---

**FIGURE 2.** Individual frames of the motion stimuli used in these experiments.
period of the underlying sinusoid. The presence of a nontrivial initial filter $L$ complicates the analysis but has no major effect on the conclusions: $T[I[P_2]]$ will be dominated by a sinusoid at $(2k_1', 2k_2', 2\omega')$ whose amplitude is determined by the behavior of $L$ at the temporal frequency $\omega'$ and spatial frequencies comparable to the pixel frequency.

Although the idea that $T(p) = p^2$ is attractive from the standpoint of analytic simplicity, there are many things that it cannot account for. If the motion pathway were the only short-range analysis pathway present, then stimuli in which a standard grating shifted by one-quarter of a cycle on each frame would appear ambiguous—this is decidedly not the case (Nakayama & Silverman, 1985). On the other hand, if the motion pathway [equation (10)] were in parallel with a standard motion analysis pathway without an initial nonlinearity [equation (9)], then the many striking similarities in the processing of short-range Fourier and non-Fourier motion (Turano & Pand, 1989; Turano, 1991; Victor & Conte, 1992b; Witt et al., 1994; Lu & Sperling, 1995a; Ledgeway & Smith, 1995a,b) would require explanation. For these reasons, the idea that a single motion pathway of the scheme [equation (10)], with an asymmetric nonlinearity $T$, might account for observations concerning both Fourier and non-Fourier motion. Half-wave rectification, transformations of the form $T(p) = p + a_2 p^2$, and many other functional forms are potential candidates for this asymmetric nonlinearity. Experiments 1 and 2 constrain these possibilities, and attempt to define the form of $T$. Experiments 3 and 4 compare velocity judgments for Fourier and non-Fourier gratings, to determine whether the findings are consistent with a single-pathway model.

METHODS

Stimulus delivery

Motion stimuli were generated on a custom-designed visual stimulator (modified from Milkman et al., 1980) interfaced to a PDP 11/73 computer. Stimuli were presented on a Tektronix 608 monitor (256 × 256 pixels covering 8.8 × 8.8 cm), which provides a monochromatic display of mean luminance 150 cd/m² at 270.3 Hz. We compensated for the nonlinear voltage–luminance characteristic of the CRT with a look-up table derived from photocell measurements. With this compensation, stimulus luminance was linear (as a function of desired stimulus value) to within approximately 1% over the entire range of values used.

Stimuli were composed of checks which measured 1 pixel (horizontal) by 4 pixels (vertical) on the CRT. At our viewing distance of 57 cm, checks measured 0.034 × 0.138 deg. (Checks are larger, for demonstrative purposes, in the still photograph, in Fig. 2.) Gratings were vertically oriented (i.e., $k_2 = 0$). Spatial frequency $k_i/2\pi$ for a standard grating, $k_i/\pi$ for the second-order non-Fourier grating of Eq. (6), or $nk_i/2\pi$ for the $n$-th order non-Fourier grating of Eq. (12]) was 1 c/deg.

Five subjects, ranging in age from 20 to 38 years, participated in these experiments. All were normal observers with visual acuity corrected to 20/20. Subjects were instructed to fixate on a small dot in the middle of the display, but eye movements were not measured.

Display nonlinearity

It is well known that CRTs are subject to nonlinearities that cannot readily be corrected via a look-up table (Pelli & Zhang, 1991; Naiman & Makous, 1992). We measured the effect of these spatial and dynamic nonlinearities by comparing the monitor’s light output during display of a uniform field set to background with its light output during display of square-wave gratings, for nominal contrasts of up to 0.5. For 128 cycle/screen gratings parallel to the raster (which required changes in the intensity signal as each raster line is scanned), there was less than a 1% change in the mean luminance. For gratings perpendicular to the raster at 128 cycles/screen (which required changes of the intensity signal as each pixel is scanned), there was an 8% decrease in the mean luminance at a contrast of 0.5, and a 3% decrease at a contrast of 0.25. The extent of this nonlinearity is expected to be proportional to spatial frequency (Naiman & Makous, 1992), so that for the check sizes (4 pixels along the “fast” axis) and contrasts (r.m.s. at most 0.5, typically 0.25) used, the dynamic nonlinearity was always less than 2%, and typically less than 1%. It is unlikely (see section entitled “Estimation of the form of the nonlinearity $T$”) that these errors or look-up table inaccuracies contributed either to the detection of motion in the non-Fourier stimuli used or to modelling errors.

Psychophysical methods

Contrast thresholds for direction discrimination were determined by a two-alternative forced-choice staircase method. Stimuli were presented for 1 sec, and subjects were asked to determine whether the dominant direction of motion was leftward or rightward. After two preliminary reversals with step sizes of 0.3 log units, the geometric means of eight reversals with step sizes of 0.05 log units (two correct answers in a row to decrease contrast, one incorrect answer to increase contrast) were averaged to estimate a threshold contrast for 71% correct performance. To avoid the possibility that subjects were cued by initial feature location, final feature location, or distance travelled, the initial phase of the grating was fully randomized, and spatial frequency, velocity, and duration were jittered by 10% about their nominal values.

Velocity judgments were assessed by a method of constant stimuli. Stimuli were presented sequentially for 1 sec each in pairs, with the first stimulus designated as the “standard” and the second stimulus designated as the “probe”. Subjects were asked to determine whether the probe stimulus velocity was faster or slower than that of the standard. In each presentation, spatial phase was randomized, duration was jittered by 25%, and spatial frequency was jittered by 20%. Velocity was jittered by 20%, but the same jittered values were applied to standard and probe gratings within the same trial. All
possible (standard velocity, probe velocity) combinations
were presented in randomized order within one block,
and 18–20 blocks of data were collected.

In both series of experiments, subjects were allowed
practice trials with feedback until performance had
stabilized. During data collection, no feedback was
given.

RESULTS

Experiment 1: Motion thresholds for second-order stimuli

Second-order non-Fourier motion stimuli were gen-
erated according to Eq. (6) and presented to three subjects
(JV, MC and ET) for pattern velocities ranging from 1 to
16 deg/sec (with 10% jitter). With no added noise,
subjects correctly gave the direction of motion at
relatively low contrast levels (Fig. 3)—though not as
low as those sufficient for detection of a first-order
(Fourier) stimulus (see Fig. 4 below).

We also measured direction discrimination thresholds
in the presence of noise. For these experiments, the
“noise” (defined precisely below) essentially consisted of
sinusoidal contrast modulation of each pixel, with the
phase of the modulation at each pixel chosen at random.
Noise was added to the second-order non-Fourier stimuli
by alternating noise frames with frames consisting of
the non-Fourier motion stimulus, and adjusting the contrasts
of the individual frames to compensate for the 1:1
interleave. As shown in Fig. 3, the addition of noise
Elevated the threshold contrast for direction discrimina-
tion. The amount of elevation was similar across subjects
and approximately proportional to the amount of added
noise.

This experiment shows that second-order non-Fourier
motion stimuli are seen as moving in an unambiguous
direction, even though they necessarily contain areas of
randomly appearing, randomly directed local first-order
motion. It follows that standard motion analysis cannot
be the only operative mechanism in motion processing,
and that an initial nonlinearity T [Eq. (10)] is required.

Experiment 2: Motion thresholds for higher-order non-
Fourier motion stimuli

As pointed out above, the perception of motion in the
second-order non-Fourier motion stimulus could be
accounted for by supposing that the nonlinear transfor-
mation T is of the form $T(p) = p + a_2 p^2$. We now test this
hypothesis by defining a new non-Fourier motion
stimulus $P$ such that $T[L[P]]$ (for $T(p) = p + a_2 p^2$) is
inaccessible to standard motion analysis. If $P$ is none-
thless seen as moving, then the hypothesis that $T(p)$ is
quadratic must be rejected.

We define an $n$th-order non-Fourier motion stimulus as
one that generates a motion signal on standard motion
analysis if it is initially passed through linear filtering and
some nonlinear transformation of order $n$, but does not
generate a motion signal if it is passed through a
nonlinear transformation of any lower order. The stimuli
of Eq. (6) satisfy this definition for $n = 2$, as do the non-
Fourier stimuli introduced by Chubb & Sperling (1988,
1989). We now construct higher-order non-Fourier
stimuli by generalizing Eq. (7), an equivalent form of
Eq. (6), to arbitrary $n$.

$$
P_n(x_1, x_2, t) = \cos\left(k'_1 x_1 + k'_2 x_2 + \omega'_t + \frac{2\pi}{n} R(x_1, x_2)\right),
$$

where $R(x_1, x_2)$ is chosen independently from \{0, 1, \ldots, $n-1$\} at each check $(x_1, x_2)$.

To show that the stimuli $P_n$ of Eq. (12) do, in fact, meet
the definition of $n$th-order non-Fourier motion stimuli, we
show that for all $m < n$, the autocorrelation of any $m$th-
degree polynomial in $P_n(x_1, x_2, t)$ is zero, except at zero
spatial disparity. However, since $[P_n(x_1, x_2, t)]^m$ will be
shown to have a nonzero Fourier component at

![FIGURE 3. Threshold contrast for discrimination of direction of motion for the second-order stimulus $P_2$ with varying amounts of added noise.](image)
FIGURE 4. Threshold contrast for discrimination of direction of motion for nth-order motion stimuli Pn, Pn, Pn, and Pn, with varying amounts of added noise. Velocity: 4 deg/sec.

spatiotemporal frequency \((nk_1', nk_2', no')\), it will yield a signal on standard motion analysis provided that \(T\) is a polynomial with a term of order \(n\).

Let \(Q[P_n(x_1, x_2, t)]\) be an nth-degree polynomial in \(P_n(x_1, x_2, t)\), and consider the jth-degree term in \(Q\):

\[
[P_n(x_1, x_2, t)]^j = \cos \left( k_1' x_1 + k_2' x_2 + \omega t + \frac{2\pi}{n} R(x_1, x_2) \right)
\]

\[
= \frac{1}{2} \left\{ \exp \left[ i \left( k_1' x_1 + k_2' x_2 + \omega t + \frac{2\pi}{n} R(x_1, x_2) \right) \right] 
+ \exp \left[ -i \left( k_1' x_1 + k_2' x_2 + \omega t + \frac{2\pi}{n} R(x_1, x_2) \right) \right] \right\}^j.
\]

(13)

Each term in the binomial expansion of the last expression has a phase offset of the form \(2\pi hR(x_1, x_2)/n\), with \(h\) in the range \(-j, -j + 2, -j + 4, \ldots, j - 4, j - 2, j\). The autocorrelation of \(Q[P_n(x_1, x_2, t)]\), which involves products of terms in \(Q[P_n(x_1', x_2', t')]\) and \(Q[P_n(x_1 + x_1, x_2 + x_2, t + t)]\), is a cross-product of such expressions, and therefore contains terms with nontrivial phase factors of the form

\[
\exp \left[ \frac{2\pi}{n} \left( hR(x_1', x_2') + hR(x_1 + x_1', x_2 + x_2') \right) \right].
\]

(14)

where both \(h\) and \(h'\) are no greater than the degree \(m\) of \(Q\). Provided that \(x_1\) and \(x_2\) are not both zero, the values of \(R(x_1', x_2', t')\) and \(R(x_1 + x_1, x_2 + x_2)\) are independent and equally distributed among the choices \(0, 1, \ldots, n-1\). Thus, an ensemble-average of any term involving the phase factor is necessarily zero, except where \(x_1 = x_2 = 0\). Consequently an m-th order polynomial \(Q\) cannot serve as a nonlinear transformation which extracts motion.

On the other hand, if the nonlinear transformation \(T\) is a polynomial which contains a term of degree \(n\), then \(T[P_n] \)

will yield an unambiguous motion signal when subjected to standard motion analysis. From Eq. (13), we find

\[
[P_n(x_1, x_2, t)]^n = \cos^n \left( k_1' x_1 + k_2' x_2 + \omega t + \frac{2\pi}{n} R(x_1, x_2) \right)
\]

\[
= \frac{1}{2^n} \left\{ \exp \left[ i \left( k_1' x_1 + k_2' x_2 + \omega t + \frac{2\pi}{n} R(x_1, x_2) \right) \right] 
- \exp \left[ -i \left( k_1' x_1 + k_2' x_2 + \omega t + \frac{2\pi}{n} R(x_1, x_2) \right) \right] \right\}^n
\]

\[
= \frac{1}{2^{n-1}} \cos \left[ n \left( k_1' x_1 + k_2' x_2 + \omega t \right) \right] + DC-term + \text{terms whose phase depends on } R(x_1, x_2).
\]

(15)

The terms whose phase depends on \(R(x_1, x_2)\) will vanish in the computation of the autocorrelation. The DC-term, which is present only if \(n\) is even, cannot contribute to a directional motion signal. However, the cosine term, which represents a sinusoid with spatiotemporal frequency \((nk_1', nk_2', no')\), will remain. The above reasoning is valid even if a linear filter \(L\) is applied to the stimulus \(P_n\) prior to application of the local nonlinearity \(T\), because \(L\) will neither create signal components at new temporal frequencies, nor introduce new kinds of spatial interactions into products of expressions similar to Eq. (13).

This shows that \(P_n\), as defined by Eq. (12), meets our definition of an nth-order non-Fourier motion stimulus: it will yield an unambiguous motion signal on standard motion analysis if initially acted upon by a nonlinearity of order \(n\), but will not if acted on by a nonlinearity of any lower order. In the limit that \(n\) is infinite, \(P_n\) becomes independent sinusoidal modulation at each check, with the phase at each check chosen randomly and indepen-
dently. Given our empirical finding (see below) that motion in \( P_5 \) is not perceptible for \( n \geq 5 \), we used \( P_5 \) as an approximation to random-phase noise in these experiments.

Note that in Eq. (15) the coefficient of the cosine term in \([P_n(x_1, x_2, t)]^m\) is \(1/2^{n-1}\). Thus, the amount of energy in its autocorrelation at the spatiotemporal frequency \((nk_1', nk_2', no')\) is \([1/2^{n-1}]^2\), which approaches zero rapidly with increasing \( n \). It is therefore expected that \( n \)-th order non-Fourier motion will not be visible for sufficiently large \( n \) because, even if the motion mechanism contains a nonlinearity of order \( n \), the Fourier motion energy of the transformed stimulus is likely to be subthreshold. Experiment 1 shows that \( P_2 \) produces an unambiguous motion signal even for relatively small contrasts. We next ask whether, for higher values of \( n \), \( P_n \) produces an unambiguous motion signal, and if so, how the threshold contrast for direction discrimination depends on noise contrast.

Higher-order non-Fourier motion stimuli were generated according to Eq. (12) and presented to subjects JV, MC and ET. In preliminary two-alternative forced-choice testing at a contrast of 1.0, all subjects saw unambiguous motion for \( P_3 \) and \( P_4 \), while no subject performed above chance for \( P_n \), \( n = 5, 6, 7 \) or \( 8 \). The appreciation of unambiguous motion for \( P_3 \) and \( P_4 \) was spontaneous, and did not require feedback or training. As in Experiment 1, we next determined thresholds for direction discrimination for the third- and fourth-order stimuli in the presence of noise.

Results are shown in Fig. 4 for a drift velocity of 4 deg/sec. Third-order stimuli, like second-order stimuli, are seen as moving by all observers, though at higher contrast thresholds. Fourth-order stimuli are seen as moving at only the highest contrast levels—the contrast threshold for direction discrimination is on the order of 0.5 for all observers. Thresholds in the presence of noise could not be reliably determined for \( P_4 \), because the maximum contrast available in an interleaved display was 0.5, which was comparable to the direction discrimination thresholds for these stimuli. Nevertheless, it is noteworthy that the fourth-order stimuli are perceived as moving at all, given the attenuation of motion energy by a factor of \( 64 = [1/2^{4-1}]^2 \).

It is interesting, but as yet unexplained, that one of the subjects (ET) consistently perceived unambiguous motion in \( P_4 \), but this motion was in the reverse of the “veridical” direction. One possibility is that this subject, who had the least experience as a psychophysical observer, was unable to suppress tracking movements, and reported the apparent direction of motion of the edges of the checks.

**Estimation of the form of the nonlinearity \( T \)**

We have shown that transformation \( T \) of Eq. (10), the hypothesized early nonlinearity in visual motion processing, cannot be a polynomial of order less than four—if so, non-Fourier motion stimuli of order 4 would not be perceptible as moving. We now derive an estimate of the form of \( T \) from the data obtained in Experiments 1 and 2.

Our strategy is as follows. For any stimulus \( P(x_1, x_2, t) \) and an assumed form for \( T \), one can calculate: (1) \( T[P](x_1, x_2, t) \); (2) the Fourier transform of \( T[P](x_1, x_2, t) \); and therefore (3) the Fourier motion energy in \( T[P](x_1, x_2, t) \), which is the motion signal that the nonlinearly transformed stimulus presents to the standard motion analyzer in the scheme. We assume that the detection of motion is based on a space–time average of the directional motion energy in \( T[P] \). The energy in the Fourier transform of \( T[P] \) which does not contribute to a directional motion signal is noise from the point of view of the motion analyzer.

If a single pathway such as the scheme [equation (10)] is correct, then (for stimuli of comparable spatial and temporal structure), contrast thresholds for direction discrimination should depend only on the directional motion energy and noise in \( T[P](x_1, x_2, t) \), and not on the order of the stimulus (i.e., not on \( P_1 \) vs. \( P_2 \) vs. \( P_3 \) etc.) nor on whether the noise was explicitly added to the stimulus, or a consequence of the \( R \)-dependent terms in Eq. (15). It is important to note that the signal and noise in the transformed stimulus \( T[P] \) depend in a complex (and not always intuitive) manner on the signal and noise in the original stimulus \( P \), as well as on the nature of the transformation \( T \). Since our estimate of \( T \) is essentially based on “silent substitution” of Fourier and non-Fourier stimuli, it is not likely to be influenced by processes such as large-area pooling of local motion signals or probability summation.

To implement this strategy, we assumed that the nonlinear transformation \( T \) was of the form

\[
T(p) = \begin{cases} \ |p|^n, & p \geq 0, \\ -g|p|^n, & p < 0 \end{cases}
\]

We kept the number of adjustable parameters in \( T \) low so that a good fit to all of the data would provide a strong test of our model. Nevertheless, the two parameters \( x \) and \( g \) cover many \textit{a priori} reasonable choices for \( T \), including a linear transformation \((x = 1, g = 1)\), half-wave rectification \((x = 1, g = 0)\), full-wave rectification \((x = 1, g = -1)\), squaring \((x = 2, g = -1)\), and other power laws.

In Fig. 5A, we present the relationship of the threshold contrast for direction discrimination to noise for subject MC, under the assumption that \( T \) is a linear transformation. The square root of the motion energy that reaches the motion analyzer is plotted along the ordinate, and the square root of the noise energy that reaches the motion analyzer is plotted along the abscissa. In this calculation, we have included the residual intensity–voltage nonlinearity in the display due to look-up table errors. Thus, the energy available for standard motion analysis is not exactly zero for \( P_2, P_3 \) and \( P_4 \). Not surprisingly, data for different kinds of stimuli do not lie along the same curve: direction is consistently detected for the non-Fourier stimuli \((P_2, P_3 \text{ and } P_4) \) at contrasts for which very little motion energy is available for standard motion analysis. The fact that the thresholds for motion
FIGURE 5. (A) Data for subject MC of Fig. 4, plotted in terms of the noise contrast present in the display. (B) Data from (A), plotted in terms of the noise contrast following a nonlinear transformation (see Fig. 6).

detection for these stimuli lie one to two orders of magnitude below the thresholds for $P_1$ indicates that intensity nonlinearities of the monitor do not contribute to the detection of these stimuli.

We then adjusted $T$ via the parameters $x$ and $g$ of Eq. (16) so that, in a least-squares sense, the threshold contrasts lay on a common line. The transformation $T$ was applied following the empirical residual intensity nonlinearity of the monitor, $T$ affects not only the amount of root-mean-squared motion energy that reaches the motion analyzer (plotted along the ordinate), but also the amount of root-mean-squared noise energy (plotted along the abscissa), and it affects both Fourier and non-Fourier stimuli. Following this transformation, which corresponds to Eq. (16) with $x = 0.72$ and $g = 0.08$, the thresholds for the four stimulus types lie along a common line [Fig. 5(B)]. That is, with this preprocessing stage, the thresholds for detection of non-Fourier and Fourier motion, with or without added noise, can be accounted for on the basis of a common nonlinear transformation and detection mechanism.

Figure 6 shows the nonlinear transformation $T$ derived from the data of Fig. 5. Note that $T$ is a compressive function, and that it treats positive and negative contrasts differently (and thus, can preserve luminance information for standard motion analysis).

A similar analysis (Fig. 7) was performed on data

FIGURE 6. The nonlinear transformation $T$ [Eq. (16)] with $x = 0.72$ and $g = 0.08$.

FIGURE 7. Parameters for the nonlinear transformation $T$ [Eq. (16)] estimated from experiments conducted at a range of velocities, across subjects.
obtained in all three subjects and at various pattern
velocities, including the $P_2$ data of Fig. 3. The power $\alpha$
was usually in the vicinity of 0.5 (mean, 0.45; range,
0.25-0.65). The gain for negative contrasts $g$ was usually
negative and always had absolute value $<0.5$ (mean,
$-0.1638$; range, $-0.4690$ to $+0.2185$). Although there is
some scatter in these values, there is no systematic
deendence on velocity or subject. Factors that contribute
to this variability probably include uncertainty in the
measurement of psychophysical thresholds, and the
constraints of the form [equation (16)] of $T$. Nevertheless,
the data clearly exclude a power $\alpha$ near 1, as well as a
nonlinearity which is nearly symmetric ($g$ near $1$) and
nearly antisymmetric ($g$ near $1$). Rather, these findings
suggest that $T$ may be thought of as partial rectification
with a fractional power law intensity–response function,
and may be approximated by

$$T(p) = \max(0, p^{0.5}). \quad (17)$$

The analysis so far has not considered a stage $L$ of early
linear filtering which precedes the posited nonlinearity $T$.
Although a detailed analysis requires knowledge of the
form of $L$, one can nevertheless understand the effects of
linear preprocessing without knowledge of its form.
Since all spatiotemporal Fourier components of $P_n$
have the same temporal frequency, the temporal filtering
associated with $L$ will simply produce an overall
amplitude change and phase shift, and thus not affect
the calculation of directional motion energy and noise
energy. The effects of spatial filtering due to $L$ are likely
to be more significant. In the extreme situation that $L$
combines values from a large number of nearby pixels,
the filtered stimulus will have nearly Gaussian-distrib-
uted values, independent of position in the stimulus
cycle, and no model can extract the motion signal. More
generally, to the extent that $L$ can be viewed as
combining luminance values from nearby pixels (whether
this combination is additive or subtractive), these nearby,
uncorrelated pixels will appear to be a source of noise. As
such, the “linearizing” effects (Spekreijse & Oosting,
1970) of this noise will diminish the effectiveness of any
nonlinearity in extracting the non-Fourier motion.
Consequently, spatial filtering due to $L$ will tend to make
$T$ appear more nearly linear. Models which omit explicit
consideration of $L$ thus will tend to underestimate the
nonlinearity $T$. That is, our estimates of $\alpha$ and $g$ will be
biased towards 1.

Experiment 3: Velocity discrimination for Fourier and
non-Fourier stimuli

To remain tenable, the single-pathway hypothesis must
account not only for direction-discrimination thresholds,
but also for other aspects of motion perception. The next
experiment asks whether Fourier and second-order non-
Fourier gratings of comparable contrast (Stone &
Thompson, 1992), spatial characteristics, and temporal
characteristics are perceived as having the same velocity,
and whether velocity discrimination thresholds for

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Psychometric functions for pairwise velocity comparisons
between probe Fourier gratings of five velocities with standard non-
Fourier gratings of five velocities, each in the range 2-4 deg/sec. Each
was presented at a contrast equal to 10 times its detection threshold
(0.041 for the Fourier grating, and 0.106 for the non-Fourier grating).
All 25 comparisons were randomly intermixed. Subject: RW.}
\end{figure}

Fourier and non-Fourier gratings can be accounted for
by a single pathway.

Three subjects (MC, NW and RW) participated in this
series of experiments. In the first set of experiments,
subjects compared Fourier and second-order non-Fourier
($P_2$) gratings at a nominal spatial frequency of 1 c/deg
$[\pi k_2/2\pi = 1$ in Eq. (12)]). For this experiment, stimuli
were presented at a contrast equal to 10 times the
subject’s detection threshold for the non-Fourier grating.
In each trial, subjects were asked to compare the velocity
of two stimuli; a “standard” and a “probe.” The 25 kinds
of trials (standard and probe stimuli presented at each of
five velocities ranging from 2.0 to 4.0 deg/sec) were
presented in random order, and responses were accumu-
lated across 18 of these blocks.

Figure 8 illustrates velocity comparison data obtained
from subject RW. The fraction of times that the probe
(Fourier) stimulus was judged faster than standard (non-
Fourier) stimulus, $f(r)$, is a sigmoidal function of the
velocity ratio, $r$. The velocity ratio at which the probe was
judged faster exactly half of the time indicates the point
of subjectively equal velocities, and the slope of the
psychometric function indicates the uncertainty of the
velocity comparisons—i.e., the range of probe velocities
that is accepted as approximately equal to the standard.
These data were fitted to a cumulative normal probability
distribution by a least-squares method:

$$f(r, b, \sigma) = \frac{1}{\sqrt{\pi \sigma}} \int_{-\infty}^{\log f} \exp \left[ -\left( \frac{u - \log b}{\sigma} \right)^2 \right] du \quad (18)$$

In this equation, $b$ is the judgment bias; it is the velocity
ratio at which subjectively equal velocities were reported.
The uncertainty is expressed by $\sigma$, and is essentially a just
noticeable difference for the log velocity ratio. At a log velocity ratio of 0.47σ, 75% of judgments will be correct.

As the standard velocity increases parametrically from 2.0 to 4.0 deg/sec, the subjectively equal probe velocity also increases, as is indicated by a rightward shift of the psychometric functions. Fitting all of the data to Eq. (18) reveals that there is no significant deviation of the bias ratio b from 1 (b = 1.041, log b = 0.040).

Figure 9 presents F vs F, NF vs NF, and NF vs F velocity comparisons for this subject as contour maps. These maps reveal an unexpected difference between the heterogeneous judgments (NF vs F) and the homogeneous judgments (F vs F and NF vs NF). If velocity judgments depended only on velocity ratio, then the contour lines (which represent constant frequency of judging the probe to be faster) would all have a slope of 45 deg (lines of constant velocity ratio). This expectation is met for F vs F judgments [Fig. 9(A)] and NF vs NF judgments [Fig. 9(B)]. However, for NF vs F judgments, the contour lines tend more towards the vertical, indicating that the velocity of the Fourier grating (the probe) is a more important determinant of the subject’s response than the velocity of the non-Fourier grating (the standard). Additionally, the uncertainty σ was greater for the NF vs F condition (σ = 0.18) than for the F vs F condition (σ = 0.07) or the NF vs NF condition (σ = 0.13). Subject NW, who had a velocity bias ratio b slightly less than 1 (b = 0.970, log b = −0.030), also showed a tilt of the contour lines towards vertical. The same pattern was seen for subject MC, for whom the bias ratio was essentially 1 (b = 0.983, log b = −0.017).

Thus, even though there is no bias in comparisons of velocity of Fourier and non-Fourier gratings, subjects do appear to rely more heavily on the velocity of the Fourier grating in judging which is faster. Furthermore, there appears to be greater uncertainty in the judgment of the velocity of a non-Fourier grating than in the judgment of the velocity of a comparable Fourier grating (Table 1).

**Experiment 4: Velocity judgments of Fourier gratings in noise**

The above findings of subjectively equal velocities for F and NF gratings are consistent with those of Ledegway & Smith (1995a), but appear inconsistent with a single-pathway scheme because of the difference in Weber fractions and the “tilt” of the contour lines of subjective equality [Fig. 9(C)]. However, a further test is required. This is because the nonlinear transformation T, when applied to a non-Fourier motion stimulus such as P2, yields not only a drifting luminance grating but also non-directional energy (“noise”). Such non-directional components in the autocorrelation are produced by the action of a local nonlinearity on any non-Fourier motion stimulus, and not just the stimuli used in these studies. However, this noise is not present when T acts on a Fourier grating. The next experiment determines whether

<table>
<thead>
<tr>
<th>Grating A</th>
<th>Grating B</th>
<th>Apparent velocity ratio (A:B)</th>
<th>Weber fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>F</td>
<td>1.005</td>
<td>0.17</td>
</tr>
<tr>
<td>NF</td>
<td>NF</td>
<td>1.008</td>
<td>0.26</td>
</tr>
<tr>
<td>F</td>
<td>NF</td>
<td>1.005</td>
<td>0.35</td>
</tr>
<tr>
<td>F</td>
<td>F, low contrast</td>
<td>1.096</td>
<td>0.16</td>
</tr>
<tr>
<td>F</td>
<td>equivalent</td>
<td>1.093</td>
<td>0.24</td>
</tr>
<tr>
<td>F</td>
<td>equivalent</td>
<td>0.949</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Results are averages across three subjects.
the apparent difference between the processing of a Fourier and a non-Fourier grating could be due to the “noise” inherent in a non-Fourier grating.

To test this possibility, we generated an “equivalent Fourier stimulus” $F_{eq}$, consisting of a standard Fourier grating and superimposed dynamic noise. The grating amplitude and noise amplitude in $F_{eq}$ were uniquely determined by the joint constraints that $T(F_{eq})$ and $T(P_2)$ had identical Fourier motion signals, and identical noise power. For this calculation, $P_2$ was taken to be the NF grating of Experiment 3, and the nonlinearity $T$ was given by Eq. (17), as suggested by the results of Experiment 2.

Figure 10 shows the results of $F_{eq}$ vs F velocity judgments in a $5 \times 5$ design for subject RW. The main features of the NF vs F judgments are reproduced: there is no bias in velocity judgments ($b = 0.986$, log $b = -0.014$), but the velocity of the F grating is weighted more strongly, as indicated by a tilt of the contour lines towards vertical. Furthermore, velocity judgments for $F_{eq}$ vs $F_{eq}$ were characterized by an uncertainty $\sigma$ of 0.14 that was greater than that for F vs F judgments ($\sigma = 0.07$), but comparable to that for the NF vs NF judgments ($\sigma = 0.13$).

Results from Experiments 3 and 4 are summarized in Table 1. Apparent velocity ratios (derived from $b$) and Weber fractions (0.47, corresponding to the 75% correct point) are averaged across subjects. There is no (<1%) subjective difference in the apparent velocity of Fourier and non-Fourier gratings equated for visibility. However, velocity comparisons have a greater uncertainty for the NF vs NF comparison than for the F vs F comparison, and a still greater uncertainty for the NF vs F comparison. These findings, which might appear to indicate that a separate pathway processes the NF grating, are also seen when the NF grating is replaced by an “equivalent” F grating in noise, $F_{eq}$. The uncertainty in the F vs $F_{eq}$ comparison is not quite as large as in the NF vs F comparison. While this might indicate a need for a second pathway, it might also merely indicate errors in the estimation of $T$.

**DISCUSSION**

**Summary of results**

We defined an $n$th-order motion stimulus as one in which the underlying unidirectional motion is revealed only by an initial nonlinearity of formal order at least $n$, and showed how these stimuli may be constructed. For $n = 2$, 3 and 4, the constructed stimuli ($P_2$, $P_3$ and $P_4$) generated the perception of unidirectional motion. We determined contrast thresholds for motion detection for these stimuli alone and in the presence of added noise. We showed that these thresholds were consistent with each other and with measured thresholds for Fourier gratings, provided that the nonlinearity $T$ preceding standard motion analysis was asymmetric and compressive. We identified differences in the perception of Fourier and second-order non-Fourier gratings of comparable spatial and temporal structure, and showed that velocity perception for non-Fourier gratings could be mimicked by a Fourier grating with added noise, in which the amplitudes of the grating and noise were calculated from the nonlinearity $T$.

It is now well-established that certain visual stimuli which lack directional motion energy are perceived as moving. The accepted working model for this phenomenon (e.g., Chubb & Sperling, 1989), as diagrammed by the scheme, is that of a nonlinear preprocessor, followed by standard motion analysis. However, the nature of the nonlinearity, the extent to which which pathways which process non-Fourier motion are distinct from pathways which process Fourier motion, and the number of distinct non-Fourier pathways, are as yet unclear. We discuss our work with attention to these issues.

**The nature of the nonlinearity**

In general, non-Fourier motion stimuli will reveal unambiguous motion if they are subject to a stage of nonlinear processing prior to standard motion analysis. This nonlinear processing is not the very first stage of computation: prior linear spatial and/or temporal filtering is required (Chubb & Sperling, 1989). For most non-Fourier stimuli, the nature of the nonlinearity is not critical, as long as it is at least partially eliminates cancellation of negative and positive contrasts. Thus, quadratic (Emerson et al., 1992; Lu & Sperling, 1995a), linear (Chubb & Sperling, 1989; Victor & Conte, 1992b), and other power laws ($\alpha = 0.73$) (Anderson et al., 1991) have been proposed for the initial local nonlinearity.
Since full-wave and half-wave rectifiers differ only by a linear filter (i.e., a half-wave rectifier function can be synthesized by $T(p) = (p + |p|)/2$), it may be difficult to distinguish between these possibilities in the presence of linear spatiotemporal preprocessing. However, our results imply a non-integer exponent $x$ for the contrast-scaling behavior of $T$, with a power near 0.5 providing the best account of the data. Furthermore, if $T$ were symmetric, then there would be a smaller difference in thresholds for $P_3$ and $P_4$ than what we found.

Our method for identifying $T$ is only an approximate one and is based on an assumed functional form. We cannot be confident of the shape of $T$ in the low-contrast range (e.g., below 0.01). While there are undoubtedly other functional forms for $T$ that would provide a satisfactory fit, the major features of our results imply that any such $T$ would necessarily be both compressive and asymmetric.

Our model has a frank asymmetry: as proposed, increments and decrements of equal magnitude do not have the same effect on the output of the nonlinearity $T$. However, our data are consistent with the possibility that two pathways of the form exist [equation (10)], one with the proposed $T$, and one with its mirror-image $T(x) = T(-x)$, provided that each of these signals is processed separately (Watt & Morgan, 1985).

Recently it has been proposed (Fleet & Langley, 1994) that non-Fourier motion can be thought of in terms of extraction of oriented lines or planes in the spatiotemporal Fourier transform of the stimulus which need not pass through the origin. However, this is inconsistent with our findings that the stimuli $P_3$ and $P_4$ have an unambiguous direction. The extraction of oriented (but perhaps displaced) lines or planes in Fourier space is tantamount to a purely quadratic nonlinearity preceding standard motion analysis (extraction of oriented undispaced lines or planes). This is because a quadratic nonlinearity applied to the stimulus merely convolves the Fourier domain representation of the stimulus with itself. Thus, a slanted line which does not pass through the origin, once convolved with itself, generates a longer line, at the same slope, which does pass through the origin—and thus, could be detected by standard motion analysis. Conversely, motion mechanisms which extract sloping lines in Fourier space will identify motion in any stimulus whose convolution with itself contains a sloping line through the origin. However, the Fourier representations of $P_3$ and $P_4$ do not contain any such sloping lines—this is equivalent to the statement that they require preprocessing by a nonlinearity of formal order higher than 2 to bring out motion by standard motion analysis. Thus, while the latent motion in $P_3$ and $P_4$ is essentially that of an envelope, this envelope is not manifest by sloping lines in Fourier space. Similarly, while dynamic occlusion typically results in the generation of non-Fourier motion signals (Fleet & Langley, 1994; Stoner & Albright, 1995), not all non-Fourier motion stimuli (for example, $P_3$ and $P_4$) correspond to dynamic occlusion.

Plaids

Plaid stimuli, which consist of two superimposed gratings moving in nonparallel directions, are perceived over a wide range of experimental conditions as a single coherent image, whose velocity is the unique velocity consistent with coherent motion of the plaid—the so-called intersection-of-constraints (IOC) velocity (Adelson & Movshon, 1982). Human observers’ perceptions correspond only approximately to the IOC velocity (Stone, 1990; Yo & Wilson, 1992; Kim & Wilson, 1993). Heeger and coworkers (Heeger, 1987; Heeger & Simoncelli, 1995) have proposed detailed implementations of standard motion analysis based on a geometrical view of the stimulus representation in Fourier space, and several investigators (Movshon et al., 1985; Wilson et al., 1992) have suggested biological implementations of motion analysis. These models provide a reasonable account of a broad range of psychophysical phenomena.

These models generally do not include an early nonlinearity, but it is straightforward to understand the effects that such a nonlinearity will have. As described above, extraction of velocity from a two-dimensional stimulus amounts to determination of a plane in spatiotemporal Fourier space which contains the bulk of its motion energy. The IOC calculation yields the unique plane which contains the two vectors that correspond to the Fourier transforms of the gratings that constitute the plaid. The effect of a nonlinearity is to generate additional spatiotemporal Fourier components, which are sums and differences of the spatiotemporal frequencies that correspond to the original gratings. However, these new Fourier components are linear combinations of the two components of the plaid, and therefore lie in the same plane as the Fourier components of the original stimulus. Thus, they do not alter the constraints available to the IOC calculation. This is somewhat surprising, in that one might have expected that the presence or absence of a local nonlinearity (e.g., the extraction of “blobs” from crossed gratings) might make a difference in the information available for motion analysis.

Because non-Fourier pathway(s) do not generate any new constraints, this analysis does not reveal whether they have any role in the processing of plaid motion. However, the work of Derrington and coworkers (Derrington & Badcock, 1992; Derrington et al., 1992) provides clear evidence in this regard. They used two-dimensional plaid stimuli in which the component motion was defined only by non-Fourier signals. These non-Fourier signals combined to provide a coherent motion of the plaid, in a manner similar to the combination of Fourier component motion signals in standard plaids. This is consistent with the idea that the nonlinearity that extracts non-Fourier motion occurs early in visual processing. On the other hand, the main determinants of whether gratings cohere, or moving stimuli appear to be rigid, appear to be higher-level influences (Nakayama & Silverman, 1985; Krauskopf & Farell, 1990; Shiffrar &
Pavel, 1991; Kooi et al., 1992), rather than their Fourier or non-Fourier content as such.

How many motion pathways are necessary?

We have provided a quantitative account for the thresholds for the detection of motion direction in \( P_t \) in terms of a single nonlinear pathway that processes both Fourier and non-Fourier motion. Furthermore, this model accounts for the observation that velocity judgments for non-Fourier stimuli are veridical, but associated with a greater uncertainty. Adaptation and sensitivity studies provide independent psychophysical evidence that Fourier and non-Fourier pathways are processed by the same mechanism (Turano & Pantle, 1989; Turano, 1991; Ledgeway & Smith, 1995b), and this view is also in accord with direct physiological evidence (Albright, 1992). As in models for standard motion perception (e.g., Heeger & Simoncelli, 1995), we postulate that the basic computational unit is present at many spatial scales (and retinal eccentricities). Furthermore, although we did not examine this issue here, the effective balance of linear and nonlinear contributions [i.e., the parameter in Eq. (17)] may well vary at the extremes of spatial scale—for example, as the envelope spatial frequency becomes higher, the relative contribution of nonlinearities might be expected to decrease, simply because of optical and receptoral factors limiting the sensitivity of nonlinear subunits.

Most workers have concluded that Fourier and non-Fourier motion signals must be processed by separate pathways (e.g., Chubb & Sperling, 1989; Boulton & Baker, 1993a,b; Mather & Tunley, 1995; Lu & Sperling, 1995a). The specific functional form for the initial nonlinear stage that we postulate has implications, not all of which are obvious, for the way that Fourier and non-Fourier signals are processed and interact. We therefore re-examine many of these studies, to determine to what extent they are consistent with a one-pathway model.

There is ample evidence that the perception of motion in displays in which components move in large discrete steps is extracted by a pathway which is distinct from that which processes short-range motion (Braddick, 1974, 1980). Importantly, this long-range vs short-range distinction was identified in displays in which motion correspondence is driven by untransformed luminance (Fourier) signals. Motion driven by texture elements, depth, and “features” also appears to be processed by a pathway distinct from the pathway which processes short-range luminance-driven motion, on the basis of its slower dynamics, spatial coarseness, binocularity, and attentional modulation (Anstis, 1980; Braddick, 1980; Cavanagh, 1992; Lu & Sperling, 1995a,b). We too have found that stimuli constructed to contain textural tokens but balanced for features that could be extracted by a single nonlinear preprocessing stage generate a qualitatively different motion signal (Victor & Conte, 1990), which is substantially slower than processing of luminance-defined motion stimuli. Additionally, a single-intensity-based nonlinearity cannot account for interactions between color and motion, such as the apparent slowing of motion in near-isoluminant displays (Cavanagh et al., 1984), the role of color in transparency (Krauskopf & Farell, 1990; Kooi et al., 1992), or the role of color correspondence in disambiguation of motion (Papathomas et al., 1991). Motion based on more local motion tokens (Zanker, 1993) also likely represents a kind of feature-based motion, but it is worth pointing out that some of “theta-motion” can indeed be extracted by a computation of the form [equation (10)], provided that the initial linear filter contains spatial and temporal components and the nonlinearity is not simply second-order.

Recognizing that motion based on feature extraction and long-range steps is processed separately, we therefore restrict our attention to perception of motion in achromatic stimuli presented continuously or in frame-by-frame displays with brief interframe intervals and short interframe displacements. Even within this restricted domain, there are many claims that two (or more) kinds of motion pathways are required to account for experimental data. These pathways are considered to be intrinsically different (i.e., distinguished by the presence of, or nature of, the nonlinear preprocessor), and not simply different in spatial scale. If the initial nonlinearity in the non-Fourier pathway (or pathways) were well-modeled by a quadratic transformation, then it would be relatively straightforward to determine whether this pathway existed in parallel to standard motion analysis. A more complex initial nonlinearity of the sort that we propose necessarily mixes linear and nonlinear components, because of its asymmetry. Furthermore, because the nonlinearity is not simply a quadratic one, it produces higher harmonics, and these may have non-intuitive effects. For both of these reasons, it is necessary to re-examine previous lines of evidence that appear to require multiple short-range motion pathways. Fundamentally, our claim of a single kind of pathway can never be proven by psychophysical or physiologic means; it can only be supported by parsimony and a lack of evidence to the contrary.

Stimuli with scale-dependent reversal in apparent motion

Chubb and Sperling’s (1989) study of visual displays with conflicting Fourier and non-Fourier motion components provided some of the early evidence for separation of these two pathways. They found that such displays appear to have motion in the non-Fourier direction on close viewing, but in the Fourier direction when viewed from a greater distance. The shift from predominantly nonlinear preprocessing at short viewing distances to predominantly linear preprocessing at large distances is, however, consistent with a single initial filter with both linear and nonlinear components. Consider a receptive field element which subsumes a fixed visual angle and whose nonlinear nature consists of partial rectification of the form [equation (16)]. At short viewing distances, this receptive field element might encompass only a single stimulus pixel, and thus its nonlinear nature would be directly manifest in its output. However, at longer
viewing distances, it would subtend multiple stimulus pixels, many of which would be uncorrelated. These uncorrelated pixels would act as a source of noise, and thus reduce the apparent effects of the nonlinearity, as discussed above in connection with the effects of spatial summation on the estimate of \( T \). That is, a single pathway with linear filtering followed by a nonlinearity can appear to be roughly linear for high stimulus densities, but highly nonlinear for low stimulus densities.

The nature of the “T” stimulus used by Chubb & Sperling (1989) suggests that there is at least partial full-wave rectification in the initial nonlinearity; i.e., that the nonlinearity is of the form [equation (16)] with a somewhat negative value of \( g \). Although we have used \( g = 0 \), this estimate was derived from an assumption that there was no spatial filtering prior to the initial nonlinear stage. As we have pointed out, unmodelled spatial pooling prior to the nonlinearity biases the estimate of the form of \( P \) towards linearity. Thus, our data are consistent with spatial pooling along with a somewhat negative value for \( g \).

Evidence from micropattern displays

Boulton & Baker (1993a,b) examined the maximum displacement for apparent motion (\( D_{\text{max}} \)) for two-frame displays composed of Gabor micropatterns. They found that under conditions of high micropattern density or brief interstimulus intervals, psychophysical performance was consistent with the Fourier components of the stimulus, but under conditions of low density in space or time, performance was best explained on the basis of a second, highly nonlinear, mechanism. This argument that the change in the apparent nature of processing with spatial scale implies separation of motion pathways for Fourier and non-Fourier motion is akin to that of Chubb & Sperling (1989), and again overlooks the expected behavior (apparent linearization) of a nonlinear filter confronted with multiple uncorrelated inputs.

Evidence from random-dot kinematograms

Mather & Tunley (1995) used a clever random-dot kinematogram display to examine characteristics of motion processing. Even though their stimuli contained directional motion energy without a nonlinear preprocessor, they showed that rectification was required to account for the robust perception of motion despite contrast inversion. They posited that first-order motion was processed by a half-wave rectifying pathway, and argued that a separate full-wave rectifying pathway was required to prevent intrusion of reversed-phi motion. However, this second conclusion was based on an assumption that direction judgments were based on motion energy at low spatiotemporal frequencies (their Fig. 6); without this assumption, half-wave rectification would suffice. Furthermore, their argument does not exclude a single pathway based on an asymmetric full-wave rectifier. This is provided by the form [equation (16)] with a somewhat negative value of \( g \), and is

consistent with our findings, provided that spatial summation is included in the preprocessing stage.

Evidence from interleaved apparent-motion displays

“Transition-invariance” is a powerful technique developed by Chubb and coworkers (Werkhoven et al., 1993) to determine whether a single class of detectors can account for detection of motion in apparent-motion stimuli. This approach is based on the apparent motion stimuli whose space–time diagrams are shown in Fig. 11. In each stimulus, there is the possibility to see apparent motion to the right or to the left. The observer adjusts some parameter of the token A until motion appears ambiguous. In stimulus I, this equates the apparent strength of the homogeneous motion path \( A \rightarrow A \rightarrow A \rightarrow \ldots \) with that of the heterogeneous motion path \( B \rightarrow A \rightarrow B \rightarrow A \rightarrow \ldots \); in stimulus II, this equates the apparent strength of the homogeneous motion path \( B \rightarrow B \rightarrow B \rightarrow \ldots \) with that of the heterogeneous motion path \( A \rightarrow B \rightarrow A \rightarrow B \rightarrow \ldots \). If motion balance is achieved at the same parameter setting for token A, then A and B are said to obey “transition invariance”.

Assuming that standard motion analysis is characterized by crosscorrelation, motion balance in stimulus I implies that \( T(A)T(A) = T(B)T(A) \), and thus \( T(A) = T(B) \). This implies that \( T(B)T(B) = T(A)T(B) \), which in turn implies that motion balance is present in stimulus II. Thus it would seem that in any single-pathway model, transition-invariance must hold. However, subtle but definite violations of transition-invariance for stimuli of the sort shown in Fig. 11 have been found (McGowan & Chubb, 1995; Papathomas et al., 1995). This would appear to indicate that perceived motion in first- and second-order stimuli could not be accounted for by a single pathway.

There are, however, several problems with this argument. One problem is that it may not be valid to assume that processing at each successive time interval is independent. Since spatiotemporal filtering (at the very least, in the outer retina) necessarily precedes any

\[ \text{FIGURE 11. A diagram of stimuli used in the “transition-invariance” technique (Werkhoven et al., 1993).} \]
nonlinearity, the local response to the alternate ABABAB rows will not be identical in the two stimuli—in one case, it follows a row of As; in the other, it follows a row of Bs. A related issue is that the two displays differ in overall spatiotemporal frequency content: the spatiotemporal frequency content of stimulus I is dominated by that of A, and the spatiotemporal frequency content of stimulus II is dominated by that of B. As indicated below, the processes by which motion information is integrated across different spatiotemporal frequencies are not as yet clear, and thus, differences in the overall spatiotemporal frequency content could contribute to differences in the equivalence points.

Yet a third, and perhaps most interesting, problem with this argument is raised by the work of Chubb & Darcy (1995). At the “standard motion analysis” stage, two signals (say a and b) derived by sampling the inputs at points separated in space and time are compared by a cross-correlator C(a, b). In motion analysis based on a Reichardt (1961) model or any of its computational equivalents (e.g., van Santen & Sperling, 1984), this cross-correlator is assumed to be strict multiplication: C(a, b) = ab. Chubb & Darcy (1995) point out that motion signals will be generated even if this cross-correlator need not strictly be multiplication, and have devised a psychophysical approach to investigate C directly based on two-frame displays. This investigation revealed that C is not symmetric, and furthermore, the symmetric component of C is not even multiplicative. Additionally, there is some recent direct physiological evidence (Kontsevich & Ferster, 1995) that the neural implementation of the Reichardt detector may indeed be characterized by an asymmetric cross-correlator. These deviations permit violations of transition-invariance, even in a one-pathway model. Suppose that a represents the result of the initial nonlinearity acting on token A, and b represents the result of the initial nonlinearity acting on token B. The condition for motion balance in stimulus I is 2C(a, a) = C(a, b) + C(b, a). The condition for motion balance in stimulus II is 2C(b, b) = C(a, b) + C(b, a). If C were simply a product, then either equation would imply that a = b, and hence, that the other equation held as well. But as Chubb & Darcy (1995) have shown, C has a more general form, and probably includes terms such as a^2b. With these terms, neither motion-balance condition necessarily implies that a = b, and neither motion-balance condition necessarily implies the other one. This allows violation of transition-invariance within the context of a single-pathway model.

Evidence from superimposed Fourier and non-Fourier motion displays

Lu & Sperling (1995a, b) identified three motion systems: a first-order motion system (standard motion analysis), a second-order motion system (nonlinear preprocessing followed by standard motion analysis), and a “third-order” system (feature-tracking). The second-order system shared many of the attributes that served to distinguish the first-order system from the feature-tracking system, including rapid dynamics, high sensitivity, monocularity, and resistance to pedestals. The main evidence that the first-order and the second-order systems were not identical comes from an experiment in which they examined direction judgments for superimposed Fourier and non-Fourier gratings as a function of relative spatial phase (their Experiment 4). They argued that if such judgments were mediated by a single combined pathway, then there should be a dependence of psychophysical performance on relative phase. (For example, two standard drifting gratings, when superimposed in antiphase, generate a spatially uniform display and thus no motion signal.) Since only a small phase dependence was observed, they reasoned that these two pathways are independent.

We have simulated this experiment with an initial nonlinearity corresponding to Eq. (17). For a stimulus consisting of a superimposed Fourier and non-Fourier grating in which the envelope spatiotemporal frequency of the non-Fourier grating matches the spatiotemporal frequency of the Fourier grating, motion energy varies by 16% (average fractional deviation) with relative spatial phase [corresponding to Lu & Sperling, 1995a; Fig. 10(a, b)]. When the envelope spatiotemporal frequency of the non-Fourier grating is twice that of the Fourier grating, motion energy varies by 7% (average fractional deviation) with relative spatial phase [corresponding to Lu & Sperling, 1995a; Fig. 10(b, c)]. While the simulations indeed reveal that there is a dependence on relative phase, this dependence is surprisingly small. The unexpectedly small size of the phase dependence is a consequence of the nature of the nonlinearity we propose. In contrast, a full-wave rectifier leads to phase-dependent interactions of 24% (NF grating whose envelope spatial frequency is equal to that of the F grating) and 43% (NF grating whose envelope spatial frequency is twice that of the F grating). A purely quadratic nonlinearity would lead to phase-dependent interactions of 4% and 58% in these two cases.

These simulations thus indicate that an asymmetric power-law rectifier is consistent with the small phase dependence observed by Lu & Sperling (1995a). Not only is the expected phase dependence of motion energy small, but also there are two other factors which further reduce the likelihood that it would have been observed psychophysically. The authors measured fraction of correct judgments, not motion energy, and the measurements were made in the range of 85–95%-correct performance. In this range, the fraction correct is likely to be a compressive function of motion energy. Furthermore, the Fourier and non-Fourier gratings were presented in alternate rows of the raster display; this spatial separation would reduce interactions among the two kinds of gratings simply because they might tend to stimulate separate detectors.

Fourier and non-Fourier gratings: what is required for equivalence?

Although many investigators have embraced the
concept that non-Fourier gratings are processed by a pathway in which a preliminary nonlinearity precedes standard motion analysis, they have often assumed that one can investigate the relationship of these two pathways by asking whether the non-Fourier grating is equivalent, in some sense, to the envelope that is extracted by the nonlinearity (e.g., Albright, 1992; Solomon & Sperling, 1994; Ledewey & Smith, 1995a). The problem with this approach is that an initial nonlinearity will not only generate the envelope, but it will also generate other spatiotemporal frequencies—particularly if the nonlinearity has higher-order components, such as the form [equation (17)]. Many of these additional Fourier components will not contain motion information, and thus behave like a mask. That is, a single-pathway model does not predict that a non-Fourier grating will behave in the same way as its envelope, but rather, that a non-Fourier grating will behave in the same way as a Fourier grating plus a noise mask, provided that these stimuli, after nonlinear preprocessing, have the same spatiotemporal spectra.

In Experiment 4, we have shown that explicit inclusion of this mask can account for much of the apparent difference between the perception of a non-Fourier stimulus and the perception of its Fourier envelope. We were able to construct this “equivalent” stimulus because the stimulus $P_1$ has a simple structure and we assumed a definite form for the nonlinearity. For other non-Fourier stimuli (Chubb & Sperling, 1988; Petersik, 1995), the effective mask may be complex in spatiotemporal structure, and it may or may not be possible to create “equivalent” stimuli which contain Fourier motion, yet provide the same spatiotemporal components following nonlinear transformation. Furthermore, in experiments in which the non-Fourier grating is presented along with a Fourier grating, the additional spatiotemporal components generated by the nonlinearity will depend on the Fourier grating as well. Finally, interpretation of responses to Fourier motion stimuli necessarily entail consideration of the additional spatiotemporal components generated by the nonlinearities involved in the processing of non-Fourier motion.

Because of these complications, dissection of Fourier and non-Fourier stimuli requires a thorough understanding of how the visual system integrates motion energy across spatial frequencies (Ramachandran & Anstis, 1983; Zhang et al., 1993; Nishida et al., 1995; Smith & Derrington, 1995), how contrast is used (Thompson, 1982; Stone et al., 1990; Stone & Thompson, 1992; Cast et al., 1993), and how ambiguous stimuli with complex spectra are interpreted as coherent, transparent, (Adelson & Movshon, 1982; Movshon et al., 1985; Ferrera & Wilson, 1990; Stoner et al., 1990; Stoner & Albright, 1992; Victor & Conte, 1992a; Wilson et al., 1992; Kim & Wilson, 1993; Wilson & Kim, 1994) or non-rigid (Shiffrar & Pavel, 1991; Yo & Wilson, 1992). These considerations, which are largely independent of whether separate Fourier and non-Fourier pathways exist, make it problematic to interpret superposition experiments as straightforward evidence for the presence of multiple short-range motion pathways (Solomon & Sperling, 1994).

Given these complexities, it is reasonable to wonder whether motion analysis, which entails complex interactions across spatial and temporal frequencies, would take a simpler form if analyzed in the spatiotemporal domain rather than in the Fourier domain. Many workers (e.g., Fennema & Thompson, 1979; Sobey & Srinivasan, 1991; Heeger & Simoncelli, 1995) have taken the alternative view, and proposed models for motion processing based on the extraction of a local spatiotemporal gradient. Such models will necessarily entail interactions across spatial frequencies. Johnston and Clifford have elaborated this view, and have shown that a unified model of this sort correctly predicts a variety of apparent motion illusions (Johnston & Clifford, 1995a) as well as some aspects of interactions between contrast and perceived velocity (Johnston & Clifford, 1995b). Furthermore, their model (Johnston & Clifford, 1995b) includes rectification at the front end, which is introduced to improve performance for luminance grating stimuli and for contrast-modulated gratings alike. Our work is consistent with this kind of model, provided that the rectification is not a simple square-law device. Most likely, the form of the rectification is not critical to its purpose. Additionally, we echo Johnston and Clifford’s view (1995b) that the requirement for multiple motion mechanisms (Kim & Wilson, 1993; Chubb & Sperling, 1988, 1989; McGowan & Chubb, 1995; Solomon & Sperling, 1994) at multiple spatial scales engenders the problem of how these multiple signals are then integrated.

REFERENCES

Turano, K. & Pantle, A. (1989). On the mechanism that encodes the
Victor, J. D. & Conte, M. M. (1990). Motion mechanisms have only limited access to form information. Vision Research, 30, 289–301.

Acknowledgements—This material was presented in part at the Association for Research in Vision and Ophthalmology annual meetings in 1993 and 1994. We thank Manisha Waran and Rochelle Witt for their assistance, and gratefully acknowledge the Frontiers support received by Rochelle Witt. This work was supported by NIH grant EY7977 to J. Victor.