Pergamon

0042-6989(94)00108-1

Vision Res. Vol. 34, No. 22, pp. 3057–3079, 1994
Copyright © 1994 Elsevier Science Ltd
Printed in Great Britain. All rights reserved
0042-6989/94 $7.00 + 0.00

Computational Analysis of Non-Fourier Motion

DAVID J. FLEET,* KEITH LANGLEY†

Received 22 November 1993; in revised form 25 April 1994

Non-Fourier motion is now commonplace in research on visual motion perception, yet lacks a computational framework. This paper examines this issue based on the observation that many non-Fourier motion stimuli have a simple characterization in the frequency domain, in terms of oriented power distributions that lie along lines (or planes) that do not pass through the origin. This provides a unifying theoretical framework for a very diverse class of non-Fourier phenomena. It also allows us to examine some central issues concerning the computational nature of non-Fourier models, and naturally occurring sources of non-Fourier motion. For example, it is shown that the orientation of power in frequency domain corresponds to the velocity of a multiplicative envelope, and may arise as a restricted form of lighting effects, translucency or occlusion. We also show that both the location and orientation of spectral power may be extracted from the phase and amplitude output of band-pass filters, consonant with existing non-Fourier models.

Motion perception Fourier analysis Multiplicative transparency Group velocity

1. INTRODUCTION

Fourier analysis plays an important role in psychophysics. It provides a powerful descriptive tool for a wide class of important stimuli, and it proves to be a convenient method of studying computational models. The domain of visual motion is no exception, the central idea being that a translating signal has all its non-zero power concentrated on a line (or plane) through the origin in the frequency domain (Adelson & Bergen, 1985; Fahle & Poggio, 1981; Fleet, 1992; Morgan, 1980; Watson & Ahumada, 1985; Watson, Ahumada & Farrell, 1986). There are however diverse classes of stimuli, the perception of which appears inconsistent with this Fourier-based perspective. Such signals, often referred to as non-Fourier stimuli, include drifting amplitude (contrast) envelopes, sinusoidal beats (e.g. Derrington & Badcock, 1985), sampled (aliased) motion (e.g. Nishida & Sato, 1992), drift-balanced stimuli (Chubb & Sperling, 1988), and theta motion (Zanker, 1993). Although they have become a major focus of research, a unifying theoretical framework is lacking.

To address this issue, this paper presents a computational perspective on non-Fourier motion. Contrary to current beliefs,‡ it is shown that a wide class of non-Fourier motion stimuli have a relatively straightforward characterization in the frequency domain. Current Fourier-based theories refer mainly to *location*s of power in frequency domain, where velocity is determined by the slope of the line (or plane) through the origin along which the power is concentrated. In examining various non-Fourier stimuli we have observed that it is also helpful to consider the *orientation* of the local power distribution in the frequency domain. When power is concentrated about a line that does *not* pass through the origin in the frequency domain, the orientation of the power distribution yields the velocity of a multiplicative amplitude envelope, analogous to the classical notion of group velocity in the theory of wave propagation (e.g. see Whitham, 1974).

Using this perspective we examine a variety of non-Fourier stimuli, and in doing so we suggest that the orientation of power in the frequency domain captures an essential property of these stimuli that is central to non-Fourier motion perception. The same perspective also allows us to examine the possible sources of non-Fourier motion that occur in natural images. It is shown for example that oriented power in the frequency domain occurs as a special case of nonlinear visual phenomena, including occlusion, lighting variations, and translucent effects, all of which involve multiplicative signal interactions. When the input involves the product of two signals, the orientation of spectral power gives the motion of one of the two signals.

*Department of Computing and Information Science, Queen's University, Kingston, Canada K7L 3N6.
†Department of Psychology, University College London, Gower Street, London, England.

‡For example, with reference to amplitude-modulated sinusoidal signals and beat patterns, Chubb and Sperling (1988, p. 1986) wrote "*certain sorts of apparent motion cannot be understood directly in terms of their power spectra*".

This paper also provides a computational basis for the measurement of both location and orientation of spectral power, based on the instantaneous phase and amplitude behaviour in the output of band-pass filters. Although it is shown that there are several ways to extract location and orientation of spectral power in frequency domain, in this paper we concentrate on the notion that the phase component of the filter output can be used to determine peak spectral locations, while the amplitude of the filter output provides information concerning the distribution of power about the peak spectral location. Amplitude is therefore shown to be an appropriate source of information from which to extract group velocity, which is consistent with the basic structure of current models of non-Fourier motion that involve some form of post-filtering nonlinearity (e.g. Wilson, Ferrera & Yo, 1992).

In this computational framework, constraints (measurements) on velocity may arise from both the phase and amplitude of single band-pass channels. It is shown that only in the case of coherent image translation will both the phase and amplitude velocities coincide. In other cases the interpretation of such constraints is more complicated. In some of the more interesting non-Fourier stimuli it appears that while different channels suggest different phase-based velocities, they all suggest a common group velocity; that is, although power is widespread throughout the frequency domain, it is strongly oriented in a single direction which corresponds to the amplitude envelope. In other cases the fine structure and amplitude envelopes are readily perceived transparently, suggesting that phase and amplitude structure of band-pass channels represent somewhat independent sources of information in the visual system.

The goal of this paper is to draw attention to the orientation of power in frequency domain as a tool to examine non-Fourier stimuli, the possible sources of non-Fourier motion, and non-Fourier models. We also suggest that the phase and amplitude components of the output of band-pass filters are primary sources of visual information that are relevant for motion perception across a wide range of stimuli. This does not mean that we believe that phase and amplitude are explicitly represented in the visual system. Rather, we suggest that they are relevant sources of information at the abstract computational level of analysis outlined by Marr (1982).

2. FREQUENCY ANALYSIS AND VISUAL MOTION

We begin with a review of the relevant Fourier theory, including an introduction to the classical notions of phase and group velocity from the theory of wave propagation (Brillouin, 1960; Whitham, 1974). With these basic tools we then examine a variety of non-Fourier motion stimuli, showing that Fourier analysis may indeed be an appropriate form of analysis in many cases.

Although we examine these issues with one-dimensional (1-D) signals, the main concepts generalize straightforwardly to 2 dimensions. The main difference in 2 dimensions, which is beyond the scope of this paper, concerns the aperture problem, the integration of different constraints to solve for the direction and speed of 2-D velocity.

2.1. *Image translation and phase velocity*

The main observation inherent in the use of Fourier analysis for visual motion has been the fact that a signal translates coherently with velocity v if and only if all its non-zero power lies along the line through the origin in frequency domain given by $kv + \omega = 0$, where k and ω denote spatial frequency and temporal frequency variables. (A proof of this given in Appendix A1 for completeness.) This means that all Fourier components, $\sin(k_0 x + \omega_0 t)$, of the translating signal have frequencies (k_0, ω_0) that satisfy $k_0 v + \omega_0 = 0$. The velocity $v = -\omega_0/k_0$ of a drifting sinusoidal grating is sometimes called phase velocity because it corresponds to the velocity of points of constant phase (e.g. the zero-crossings, crests, and troughs). To see this, note that the paths of constant phase for $\sin(k_0 x + \omega_0 t)$ are given by

$$k_0 x + \omega_0 t = c, \tag{1}$$

where c is a real scalar. Differentiating (1) yields the velocity: $v \equiv \mathrm{d}x/\mathrm{d}t = -\omega_0/k_0$. This illustrates the well-known relation between temporal frequency, spatial frequency, and phase velocity, as illustrated in Fig. 1.

Accordingly, one working hypothesis of perception research has been that motion perception is determined

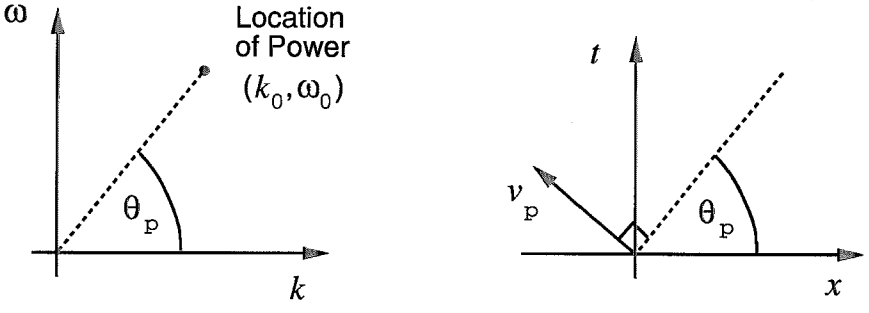

FIGURE 1. A drifting sinusoidal grating with spatiotemporal frequency (k_0, ω_0) corresponds to an impulse of power in the frequency domain. Its velocity is called phase velocity, and is determined by its location in frequency domain, as in $v = -\omega_0/k_0 = -\tan(\theta_p)$.

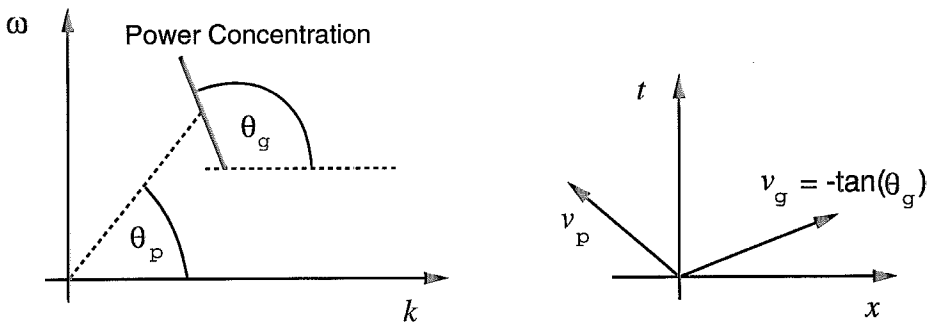

FIGURE 2. When power lies along a 1-D curve in the frequency domain (the solid line segment), phase velocity for each frequency component is determined by the orientation of the vector from the origin to that point in frequency domain: $v_p = -\tan(\theta_p)$. Group velocity is given by the orientation of the power distribution: $v_g = -\tan(\theta_g)$. The diagram on the right depicts these two velocities as direction vectors in space–time.

primarily from the location of power in frequency domain, that is, by the phase velocities of dominant spectral concentrations (Morgan, 1980; Adelson & Bergen, 1985; Watson & Ahumada, 1985). Chubb and Sperling (1988) have since referred to this as the *motion-from-Fourier-components* (MFFC) principle.

The facts that coherent image translation has a simple representation in the Fourier domain, and that several computational methods may be used to estimate it, are now well-known. The main methods include correlation-based methods (Reichardt, 1961; van Santen & Sperling, 1985), differential methods (Adelson & Bergen, 1986; Srinivasan, 1990), energy-based methods (Adelson & Bergen, 1985; Watson & Ahumada, 1985; Heeger, 1987) and phase-based methods (Fleet & Jepson, 1990; Fleet, 1992). Although these different approaches to the measurement or detection of image motion may differ, they are currently thought to be *broadly* equivalent insofar as they are based on a model of coherent image translation. In effect, they all measure some approximation to the line (or plane) through the origin of the frequency domain that contains most of the spectral power. In current terminology, coherent image translation has become known as *Fourier motion*, and these methods of detection or measurement are called Fourier-based methods.

2.2. Group velocity

But there remain several phenomena that appear inconsistent with this Fourier-based framework. These include the perceived motion of amplitude envelopes, sinusoidal beats, drift-balanced stimuli, second-order motion, sampled (aliased) motion and various multiplicative transparent phenomena. As an idealization of these phenomena, we consider a model that stems from the concept of group velocity from the theory of wave propagation in dispersive mediums (Brillouin, 1960; Whitham, 1974). While the notion of phase velocity is associated with the *location* of power, group velocity is related to the *orientation* of power distributions in the frequency domain (as in Fig. 2), and can be shown to correspond to the velocity of a multiplicative amplitude envelope.

In strict terms, as discussed in Appendix A2, group velocity is relevant when all power lies along a 1-D curve (called a dispersion relation) through the Fourier domain, the orientation of which determines group velocity. If the curve is expressed with temporal frequency written as a function of spatial frequency, i.e. $\omega(k)$, then group velocity is defined as

$$v_g = -\frac{d\omega(k)}{dk}. \tag{2}$$

Appendix A2 derives this definition and shows that v_g gives the velocity of a multiplicative amplitude envelope. The concentration of power along a 1-D curve in the frequency domain is of course an idealization, as is the notion of power concentrated along a line through the origin in the case of coherent translation. For our purposes the important observation is the oriented nature of the power distribution, the orientation of which corresponds to the motion of an amplitude envelope; Appendix A3 explains that similar conclusions hold if power is concentrated near a line in the frequency domain.

As a simple example of these ideas, consider a modified Gabor function (Gabor, 1946) in which the Gaussian envelope $G(x)$ moves with one velocity while the sinusoidal fine structure $\sin(k_0 x)$ moves with another:*

$$I(x, t) = G(x - v_1 t) \sin k_0(x - v_2 t). \tag{3}$$

As shown in Fig. 3, the amplitude spectrum of (3) consists of two 1-D Gaussian distributions centred at frequencies $\pm(k_0, k_0 v_2)$, oriented with slope $-v_1$.† The locations of the power distributions correspond to the scale and velocity of the sinusoidal fine structure,

*A wonderful example of this has been created by Freeman, Adelson and Heeger (1991) for computer graphics with a stationary envelope centred on the edges of a figure, with a time-varying phase component. Observers enjoy a sensation of motion due strictly to the local phase velocities, although the locations of power at the edges are stationary.

†To see this, remember from the convolution theorem that the Fourier transform of a product is the convolution of the individual Fourier transforms, and that the Fourier transform of the sinusoid is a pair of impulses.

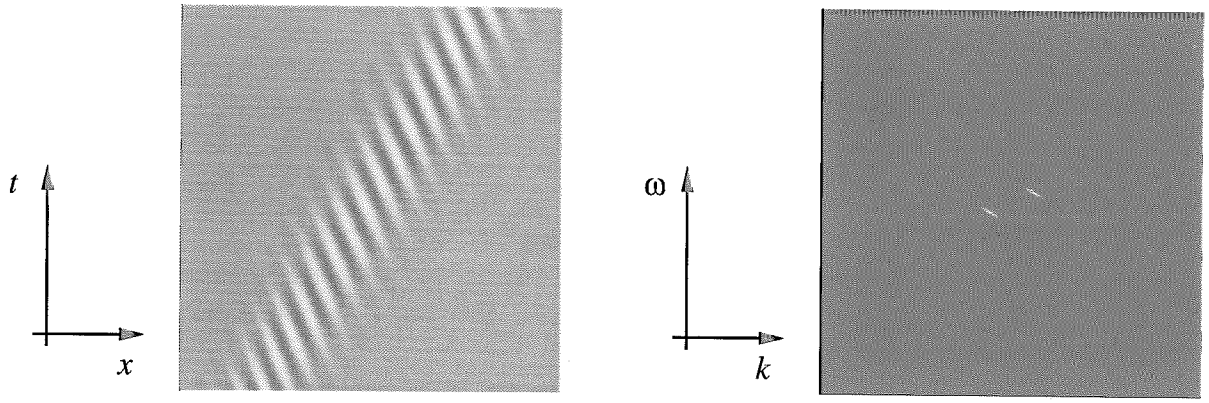

FIGURE 3. The modified Gabor function (A) has the Gaussian envelope drifting to the right while the sinusoidal fine structure drifts to the left. Its amplitude spectrum (B) shows power concentration along a line that does not pass through the origin. The origin is in the centre of the image.

while the orientation of the power distributions determines the envelope (group) velocity v_1.

In this case, because power lies along a line, the group velocity is constant for all frequencies.* However there may be a wide range of phase velocities. To see this, consider Fig. 4 which is an extreme example of (3) in which the Gaussian shrinks to an impulse:†

$$I(x, t) = \delta(x - v_1 t) \sin(x - v_2 t), \qquad (4)$$

where $\delta(x)$ is a Dirac delta function. In effect, this space–time sinusoidal profile could be a 1-D slice from any one of an infinite family of sinusoids, three of which are shown in Figs 4C–E. Every sinusoid in the family has a different space–time orientation (i.e. phase velocity) and a different wavelength.‡ Together, they represent all the frequencies that make up the Fourier transform of (4), shown in Fig. 4B. However, while the collection of non-zero frequencies include a wide range of leftward and rightward phase velocities, they all lie along lines with an orientation that is consistent with the space–time path of the amplitude envelope, that is, the group velocity v_1.

Interestingly, if and only if the power is oriented along a line passing through the origin will the group velocity equal the phase velocity for all frequencies, and the signal therefore translates coherently. Otherwise, like the modified Gabor function above, or like a large spotlight moving over a scene, the orientation of the power in the frequency domain (group velocity) corresponds to the motion of the amplitude envelope.

*If the power were to lie along a curve rather than a line, as with many dispersion relations, then the group velocity is only relevant to a small band of frequencies where the dispersion relation is nearly linear.

†This is also an idealization of the randomly contrast-reversing bar described Chubb and Sperling (1988, Fig. 2).

‡The wavelength decreases as the cosine of the angle between the sinusoid orientation and the 1-D intensity profile.

2.3. *Relevance of frequency analysis*

Before continuing with an examination of non-Fourier stimuli it is essential that we qualify the use of Fourier analysis. Fourier analysis is global in nature and therefore most appropriate for stationary signals, where the expected behaviour in one region is similar to that in neighbouring regions. Although images are typically nonstationary, these simple Fourier descriptions provide a convenient idealization for the development and understanding of both stimuli and computational models. We do not expect amplitude spectra of natural images to be concentrated along 1-D curves. Rather, we view these cases as idealizations of local visual phenomena, as a convenient way of predicting the behaviour of the output of band-pass filters. Even the common idea of power concentrated along a line through the origin, for example, should be viewed as an idealization of local coherent translation.

As a consequence, because of various global effects, it is often not possible to easily predict local behaviour from a Fourier transform. For example, with random patterns there are very often dominant scales in local regions while globally the power appears flat, thus hiding dominant local spectral peaks. Another global artifact is the introduction of horizontal and vertical power caused by intensity discontinuities between top–bottom and side-to-side image borders (since the Fourier transform assumes a periodic signal). Thus, it is important to stress that Fourier analysis should only be used to analyze global idealizations of local phenomena, since we expect our computational models and measurement processes to be local. As discussed in Appendix A3, the occurrence of group velocity will only be evident in the Fourier domain when the envelope shape changes slowly through time, the underlying structure has band-pass spectral peaks, and the signal is stationary.

3. NON-FOURIER STIMULI

With the ideas of location and orientation in the frequency domain, corresponding to the notions of

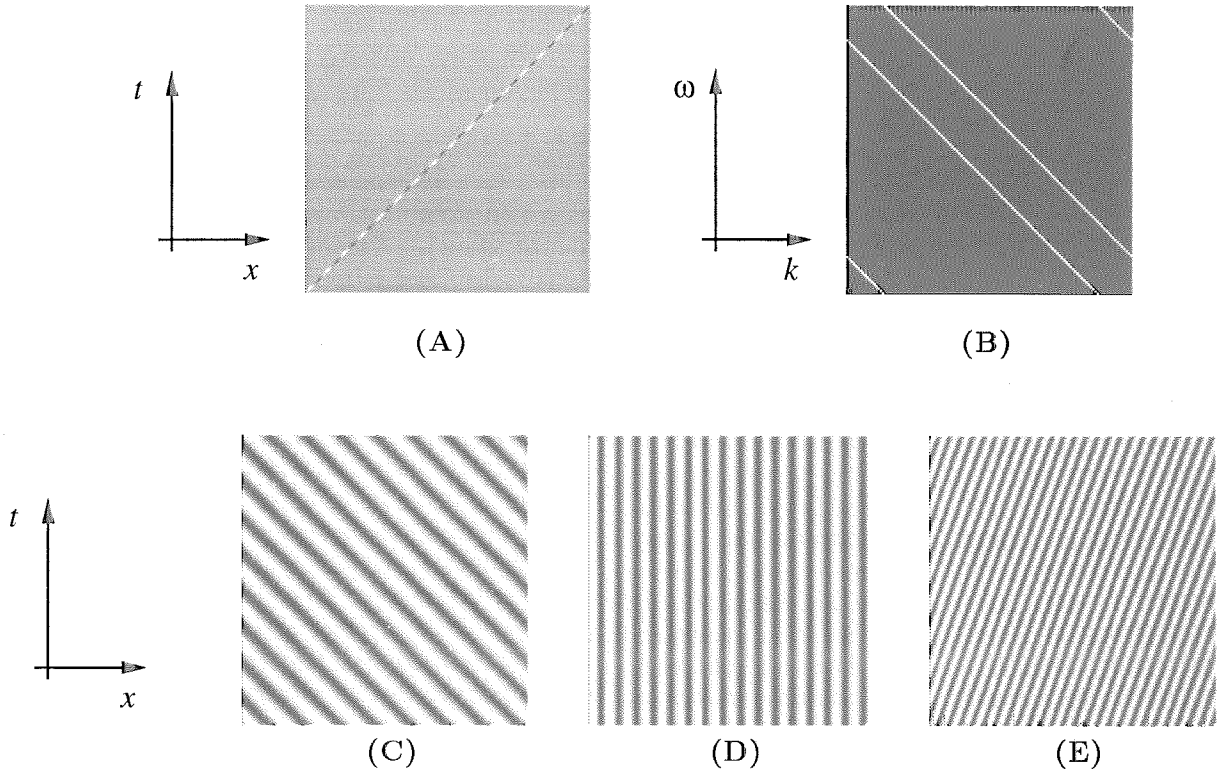

FIGURE 4. A sinusoidally-modulated translating impulse is shown (A) with its amplitude spectrum (B). The spectrum lies along a line that does not contain the origin. The origin is in the centre of the image. The orientation of the line determines the velocity of the path of the amplitude modulated line. The bottom row shows three of the Fourier components contained in the original signal.

phase and group velocity, we now examine several non-Fourier stimuli.

3.1. *Beats*

Sinusoidal beats (or interference patterns) are formed easily with a sum of two or three sinusoidal gratings, and have been used often in psychophysical experiments (e.g. Derrington & Badcock, 1985; Turano & Pantle, 1989). They are also used to introduce the concept of group velocity in elementary physics textbooks. In the simplest case, beats may be formed with the superposition of two sinusoidal signals:

$$I(x, t) = \cos(k_1 x + \omega_1 t) + \cos(k_2 x + \omega_2 t). \quad (5)$$

The Fourier transform of (5) consists of impulses at locations $\pm(k_1, \omega_1)$ and $\pm(k_2, \omega_2)$. Using trigonometric identities, it is also straightforward to show that (5) is equivalent to

$$I(x, t) = 2\cos(\Delta k x + \Delta\omega t)\cos(\bar{k}x + \bar{\omega}t), \quad (6)$$

where $\Delta k = (k_2 - k_1)/2$, $\Delta\omega = (\omega_2 - \omega_1)/2$, $\bar{k} = (k_2 + k_1)/2$, and $\bar{\omega} = (\omega_2 + \omega_1)/2$. As illustrated in Fig. 6, the difference frequencies, Δk and $\Delta\omega$, are referred to as beat frequencies, and correspond to the vector from (k_1, ω_1) to (k_2, ω_2), that is, the orientation of the power distribution. Accordingly the beat (or group) velocity is $-\Delta\omega/\Delta k$. The average frequencies, $\bar{k}$ and $\bar{\omega}$, are referred to as carrier frequencies, and represent the center of the power distribution. The carrier velocity is similarly given by $-\bar{\omega}/\bar{k}$.

Equations (5) and (6) show that there are two mathematically equivalent representations for such signals. Yet Fig. 5 shows that in certain cases our

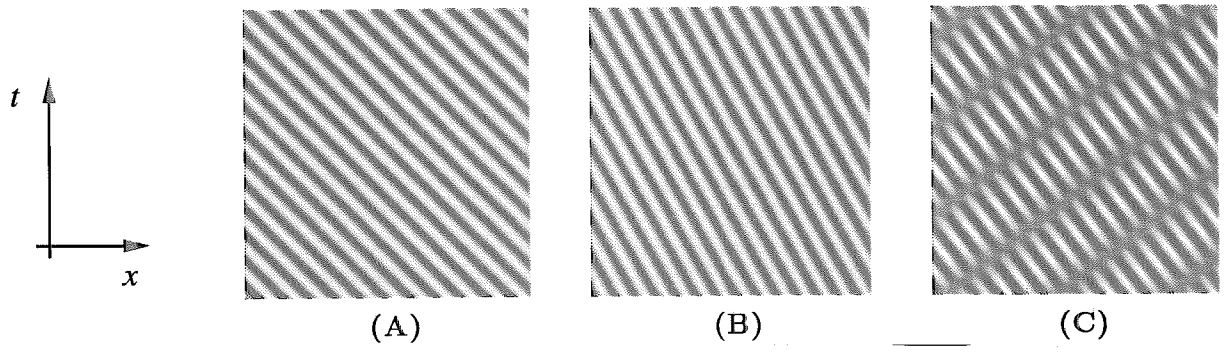

FIGURE 5. Two sinusoidal gratings are shown in (A) and (B), with their superposition in (C).

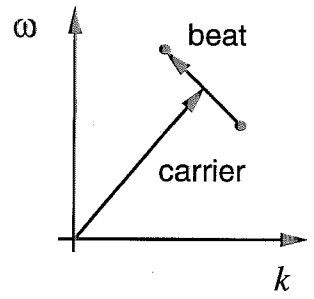

FIGURE 6. With two sinusoids of nearby frequencies, the centre of the power distribution is known as the carrier frequency, while the orientation of the power distribution is known as the beat frequency.

perception reflects a beat/carrier representation. In this case the low-frequency beats and the higher frequency carrier are clearly evident, even though the carrier has no corresponding frequency component in the Fourier transform of the signal (5). Furthermore, it is interesting to note that the perceptual dominance of the beat/carrier representation only occurs when the two frequencies are reasonably close. If the frequencies are sufficiently different, then the signal becomes somewhat ambiguous with the individual sinusoids appearing to dominate. As discussed in more detail in Section 4.6, this suggests a model in which the beats and carrier might be available only when the two sinusoids fall within the tuning range of a single channel.

Derrington and Badcock (1985) were among the first to notice that the human visual system can separate beats and carrier, with time-varying 1-D stimuli composed of two sinusoidal gratings as in (5). For certain combinations of frequencies humans perceive the modulating envelope (the beats) sliding coherently over a sinusoidal background, as suggested by the signal shown in Fig. 5. The perception of such stimuli is not easily accounted for by the MFFC principle because there is no sinusoidal component in the stimulus that corresponds to either the beat frequency or the carrier frequency. In fact, the beat frequency is often very far from the frequencies of the two component gratings, and appears to be extracted by a different mechanism than the sinusoidal fine structure (Derrington & Badcock, 1985).

Three component sinusoidal beats were used by Turano and Pantle (1989):

$$I(\mathbf{x}, t) = [1.0 + a\cos(\Delta k x + \Delta\omega t)]\cos(\bar{k}x + \bar{\omega}t), \quad (7)$$

where $|a| \leqslant 1.0$, $\Delta\omega \ll \bar{\omega}$, and $\Delta k \ll \bar{k}$.* They showed that detection thresholds for drifting envelopes (beats) can be substantially lower than those of single sinusoidal gratings (of similar frequency), while velocity difference thresholds for envelopes and sinusoidal gratings of the same frequency are very similar. Among the possible explanations outlined by Turano and Pantle (1989) is the hypothesis that there exist two motion channels, one that acts on image intensity and another that acts on the amplitude of the intensity modulation.

*This is often called a side-band signal since it contains three frequencies, the central one of which, (k_2, ω_2), coincides with the carrier. In some ways the two component case is more interesting because the perceived fine structure does not coincide with either of the two component sinusoids.

3.2. *Sampled motion stimuli*

In a very different context, Nishida and Sato (1992) suggested that non-Fourier motion may help explain human motion perception when a spatially band-pass signal is displaced by more than half a cycle of the lowest frequency in the pass-band. Recent debates have been focussed on whether or not the maximum displacement of band-pass noise that elicits a coherent motion percept (referred to as D_{max}) is larger than one half cycle of its lowest frequency (Cleary & Braddick, 1990; Braddick & Cleary, 1991; Bishof & DiLollo, 1990, 1991). Nishida and Sato (1992) showed that as spatial displacements increase beyond the half-cycle threshold, the behaviour of aftereffects is very different than that associated with coherent translation. The lack of aftereffects is more consistent with known results concerning envelope motion (Derrington & Badcock, 1985), and they therefore suggested this as evidence for the influence of a non-Fourier channel.

The stimuli used by Nishida and Sato (1992) involved a band-pass spatial signal that was held constant for a short period (128 msec in their case) and then shifted by about half a cycle of the lowest frequency, with essentially no delay between frames. As explained in detail in Appendix B, such signals can be viewed as temporally undersampled versions of continuous (coherent) image translation. The degree of undersampling (or aliasing) is a function of the magnitude of the displacement d, the duration of each frame Δt, and the spatial wavelengths present in the signal. The corresponding continuous velocity is $v_c = d/\Delta t$.

Figure 7A shows such a stimulus, with spatial wavelengths between 6 and 8 pixels, and a shift of 4 pixels (half of the longest wavelength) that occurs every 4 frames in time. The Fourier transform of this stimulus is shown in Fig. 7B, and reveals local concentrations of power, each of which is strongly oriented in a direction consistent with the continuous velocity v_c of the sampled signal. Appendix B shows that the positions and orientations of these power concentrations are determined by the continuous velocity v_c and the sampling interval Δt relative to the wavelengths in the signal. The multiple line segments of non-zero power are replicas of the power spectrum of the continuously moving signal (which satisfies $\omega + kv_c = 0$), caused by the discrete temporal sampling. Normally, with sufficiently high sampling rates relative to the velocity v_c, these replicas cannot be seen because of resolution limits of the visual system (Burr, Ross & Morrone, 1988; Watson *et al.*, 1986). However, by increasing the velocity v_c with the sampling interval held constant, or by increasing the sampling interval Δt, spectral power to which we *are* sensitive can be introduced far from the original line of non-zero power $\omega + kv_c = 0$. Finally, it is also evident, from Fig. 7B that power decreases with increasing temporal frequency; as explained in Appendix B, this is

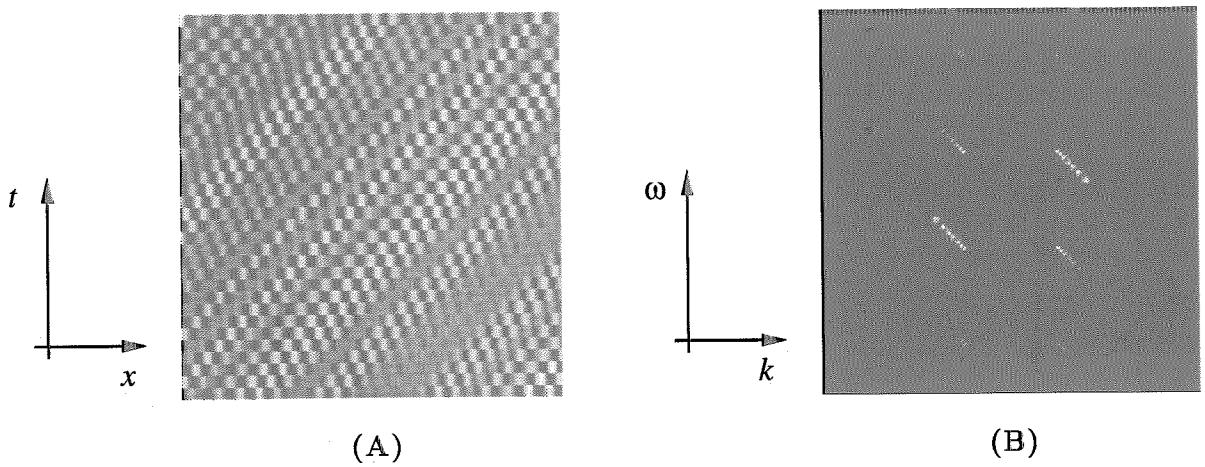

FIGURE 7. The amplitude spectrum (B) of this rightward stepping band-pass signal (A) contains several concentrations of power, each of which has an oriented distribution consistent with the rightward velocity. One of these power concentrations lies on the line through the origin that is also consistent with the rightward velocity. The concentration with the largest energy suggests leftward motion according the MFFC principle.

caused by each frame being held static over the sampling interval Δt.

From the Fourier perspective, with the notions of location and orientation of power distributions in the Fourier domain, subsampled motion provides an interesting collection of stimulus conditions and effects, a detailed examination of which is somewhat beyond the scope of this paper. However, the notion of orientation in frequency domain helps show that there is information present in these signals about the original direction of motion in the form of group velocity. If the replications of power are sufficiently far apart relative to the tuning of different channels, the response in each channel will exhibit the same group velocity. Conversely, their phase velocities will indicate a wide distribution of power, inconsistent with a single Fourier motion. The case shown in Fig. 7B is also interesting because the line segment with the largest energy has phase velocities consistent with leftward motion, while all the group velocities suggest rightward motion. With this particular input, although the motion is somewhat ambiguous, there is a dominant percept of rightward motion.

3.3. Drift-balanced stimuli

Another well-known example of non-Fourier motion is the class of drift-balanced stimuli, popularized by Chubb and Sperling (1988, 1989). They constructed ensembles of drift-balanced stimuli so that, on average over each ensemble, there is as much Fourier power supporting leftward motion as rightward motion in the sense of phase velocity and the MFFC principle. That is, the expected (average) amplitude spectrum $E[A(k, \omega)]$, where $E[\cdot]$ denotes mathematical expectation, is symmetrical about the spatial frequency axis:

$$E[A(k, \omega)] = E[A(k, -\omega)]. \quad (8)$$

According to the MFFC principle there should be no preference for any particular direction of motion since phase velocities are symmetrically distributed over the ensembles for leftward and rightward motion. However, with several ensembles of drift-balanced stimuli, Chubb and Sperling showed that these stimuli can elicit a consistent motion percept (to varying degrees depending on the ensemble).

To understand the motion percept with drift-balanced stimuli it is instructive to examine the amplitude spectra of individual samples from an ensemble of inputs. In several of these ensembles one can show that the individual amplitude spectra exhibit significant elongation of the power concentrations. Moreover such power orientations are constant throughout the ensemble, and are consistent with the observed direction of motion. In other words, although power is symmetrically distributed and nearly flat throughout the frequency domain on average, it may still be highly oriented in individual samples or in local regions.

For example, consider the rightward-stepping randomly contrast-reversing bar depicted in Fig. 8A, following Chubb and Sperling (1988, Fig. 4a). This may also be described as a drifting *spotlight* that illuminates a sequence of locations at which intensity is a random variable. One can see from the Fourier transform in Fig. 8B that power occurs on both sides of the spatial frequency axis, suggesting that there are rightward and leftward phase velocities. One can also see clear oriented concentrations of power, which, as discussed in Appendix A3, suggests that there is a single envelope velocity that might be extracted from different band-pass channels.

As explained in Appendix C, the amplitude spectrum of Fig. 8A can be described in terms of the space–time orientation of the path of the squares, the sequence of intensity variations along the path, and the size of the squares. The space–time orientation of the path (at 45°) determines the dominant orientation of the amplitude spectrum (i.e. its level contours at −45°). The 1-D intensity variation along the path determines the profile of the amplitude spectrum perpendicular to its level contours; this causes the spectral peaks in Fig. 8B. Finally, the shape and size of the squares determines the

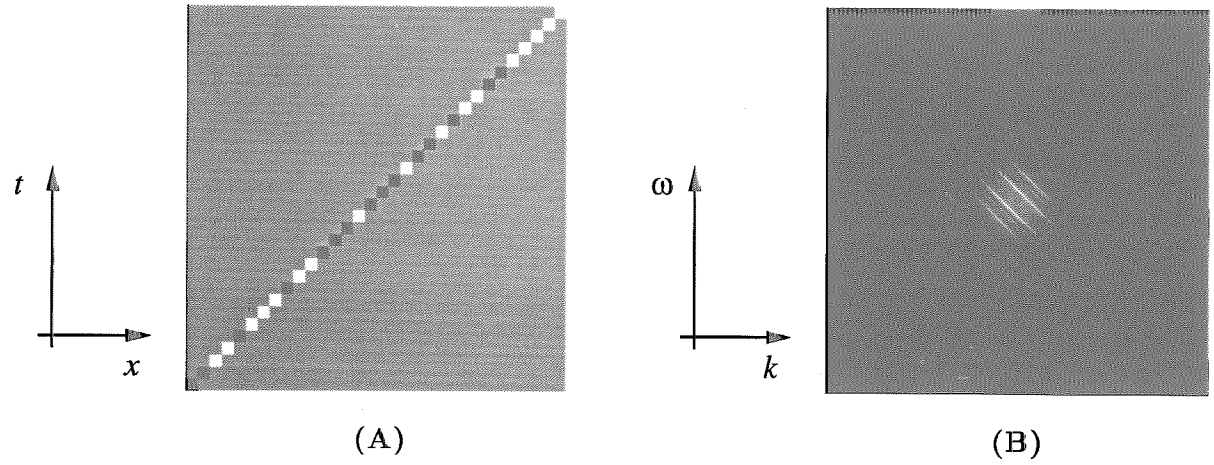

FIGURE 8. A sample from a drift-balanced ensemble is shown in (A) with its amplitude spectrum in (B). The spectrum shows no significant power concentration along a line through the origin, but it does exhibit peak concentrations with orientations consistent with the rightward motion of the bar.

overall patterns of zero power which remain the same for all stimuli in the ensemble.

Unfortunately, the drift-balanced stimuli of Chubb and Sperling do not lend themselves easily to Fourier analysis as a way of understanding the *local* behaviour of the signal. This is true for example of the sinusoidally-modulated noise as well as the contrast-reversing bar. As discussed in Section 2.3, for non-Fourier stimuli such as the drift-balanced stimuli used by Chubb and Sperling, we would only expect power to be concentrated about isolated lines in frequency domain when the fine structure has distinct peaks of spectral power, and when the shape of the envelope changes slowly through time. Although these conditions do not exist globally in most cases, they do exist in many local regions. For example, Fig. 9 shows two windowed portions of the stimulus in Fig. 8 and their amplitude spectra. The local structure in the first case suggests coherent translation since most of the power lies along a line through the origin in frequency domain. The structure in the second case does not reflect coherent translation since the power does not lie along a line through the origin. However, it is clear that the orientation of power remains the same in both cases, suggesting a common group velocity that is consistent with the motion of the bar.

It is important that one consider local analyses of such stimuli, or construct idealizations whose global structure is characteristic of local behaviour. One idealization of the rightward-stepping randomly contrast-reversing bar is the sinusoidally modulated path shown in Fig. 4A. Another idealization is given in Appendix C. Interestingly, not only do these idealizations follow from the model, we often find them to be more compelling demonstrations of non-Fourier motion.

3.4. *Sources of non-Fourier motion*

Non-Fourier stimuli such as these are often associated with the motion of texture boundaries, or the motion of motion boundaries. The sinusoidal beats and the modified Gabor function discussed above show that the non-Fourier components of the signal (the beats) can also be perceived transparently, sliding over some fine structure. From this perspective, it is important to examine the conditions under which oriented structure in the frequency domain will arise from similar naturally occurring visual signals, such as the occlusion of one surface by another, multiplicative transparency, and the non-linear transmission of light through textured translucent material, such as stained glass.

3.4.1. Transparency. Perhaps the simplest case to consider first is a multiplicative form of transparency (or translucency). For example, assume that a surface with radiance $I_1(x)$ is viewed through a (non-refractive)

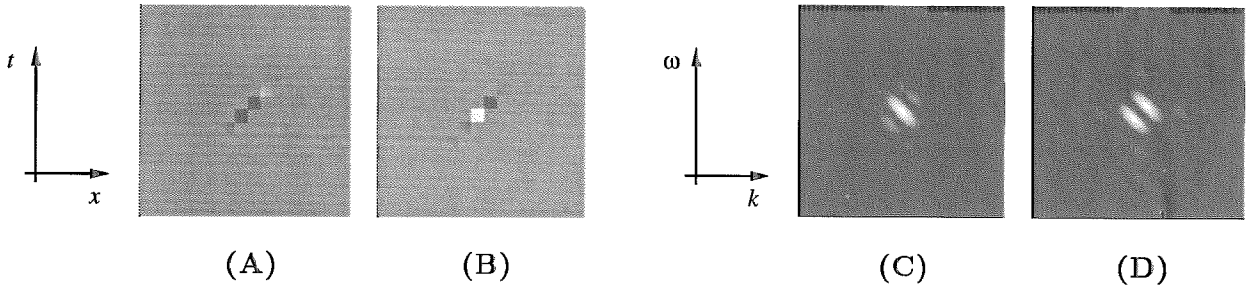

FIGURE 9. The left two images, (A) and (B), show Gaussian windowed regions of the sample of a drift-balanced ensemble shown in Fig. 7. The right two images, (C) and (D), show their amplitude spectra. It is clear that some regions such as that shown in (A) reflect coherent translation because power is concentrated along a line through the origin. In other regions, such as that in (B), power is still oriented in the same direction, but no longer passes through the origin.

translucent material with density $\rho_0(x)$. If both are moving, we might write the intensity input as

$$I(x, t) = \rho_0(x - v_0 t) I_1(x - v_1 t), \quad (9)$$

In practice we expect $0 < \rho_0(x) < 1$ so that sign changes in the radiance $I_0(x)$ are not caused by the translucent material. Thus, we can reexpress $\rho_0(x)$ as a constant γ plus a mean-zero term $\rho(x)$, yielding

$$I(x, t) = \gamma I_1(x - v_1 t) + \rho(x - v_0 t) I_1(x - v_1 t), \quad (10)$$

where γ is the mean value of $\rho_0(x)$, and $\rho(x) = \rho_0(x) - \gamma$. Interestingly, some of the non-Fourier stimuli above contain only the product term, and therefore violate this condition.

From the results of Appendix A1 and the convolution theorem one can show that the Fourier transform of (10) is

$$\hat{I}(k, \omega) = \gamma \hat{I}_1(k)\, \delta(\omega + v_1 k) + [\hat{\rho}(k)\, \delta(\omega + v_0 k)] * [\hat{I}_1(k) \delta(\omega + v_1 k)], \quad (11)$$

where $\delta(k)$ is a Dirac delta (impulse) function, and $*$ denotes convolution. The first term represents power concentrated along a line through the origin of the frequency domain that is consistent with the velocity v_1 of the background radiance pattern.

The velocity of the transparent material v_0 occurs only in the second term, which is the convolution of the Fourier transforms of $\rho(x - v_0 t)$ and $I_1(x - v_1 t)$. As shown in Fig. 10, this convolution in the frequency domain will introduce power relatively far from the lines $\omega = -v_0 k$ and $\omega = -v_1 k$. However, because it is a convolution of 1-D profiles it exhibits simple structure in certain situations. For example, notice that if the intensity $I_1(x)$ is slowly varying compared to $\rho(x)$, then the Fourier transform of the product in (10) will have power concentrated about the line $\omega = -v_0 k$, consistent with the velocity of the transparent signal. Second, if one of the two 1-D spectral profiles, for example, $\hat{I}_1(k)$, exhibits a spectral peak, then power will be spread along a line in frequency domain that is consistent with the motion of the other signal, in this case parallel to the line $\omega = -v_0 k$ as shown in Fig. 10. In these special cases the local power distributions will be highly oriented, and the orientation will correspond directly to the velocity of one of the two multiplicative signals.

When the local spectra of both signals are flat and nonzero, then power will be not be restricted to lines in frequency domain. In these cases the simple notion of group velocity becomes inappropriate, and the computational models that follow from it will not yield the velocities of the component signals. Therefore, it is the relative spectral distributions of the two signals that determines how readily their motions can be extracted based on the orientation of power in the frequency domain.

3.4.2. Occlusion. Occlusion can be modelled locally in a similar manner. Consider an object with velocity v_0 that occludes a background with velocity v_b. Let $I_0(x)$ and $I_b(x)$ denote the intensity profiles of the occluding object and the background, and let $\chi(x)$ denote a characteristic function that specifies the region of space occupied by the object:

$$\chi(x) = \begin{cases} 1 & \text{if location } x \text{ is occupied by the object} \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

The combined intensity pattern consists of the background, plus the object, minus the region of the background occluded by the object:

$$I(x, t) = I_0(x - v_0 t) + I_b(x - v_b t) - \chi(x - v_0 t) I_b(x - v_b t). \quad (13)$$

As above, the Fourier transform of (13) is

$$\hat{I}(k, \omega) = \hat{I}_0(k)\, \delta(\omega + v_0 k) + \hat{I}_b(k)\, \delta(\omega + v_b k) - [\hat{\chi}(k)\, \delta(\omega + v_0 k)] * [\hat{I}_b(k)\, \delta(\omega + v_b k)]. \quad (14)$$

The first two terms are the Fourier transforms of the object and background. The third term represents the multiplicative distortion caused by the occlusion. In this case the multiplicative term includes the intensity profile of the background and the characteristic function $\chi(x)$. The intensity profile of the foreground $I_0(x)$ does not appear in the multiplicative term.

The theta motion stimuli of Zanker (1993) provide interesting examples of non-Fourier motion that involve occlusion, although they are somewhat more complex

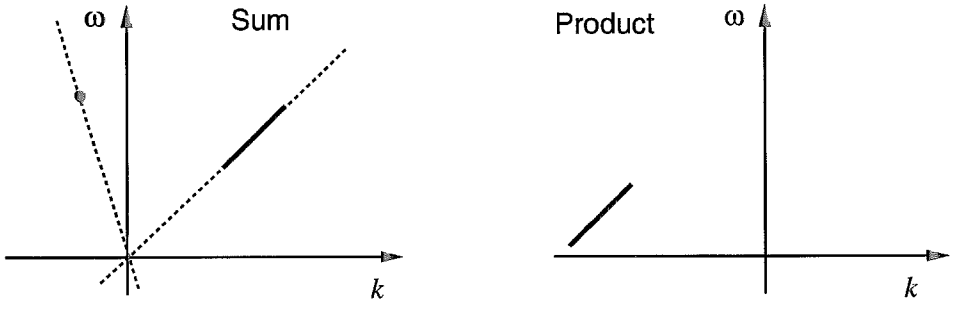

FIGURE 10. When two signals are combined additively, the resultant Fourier transform is the sum of the two individual Fourier transforms. When two signals are combined multiplicatively, the resultant Fourier transform is the convolution of the two individual Fourier transforms. This shows the Fourier transforms of a sum and a product of two signals, one of which has a distinct spectral peak.

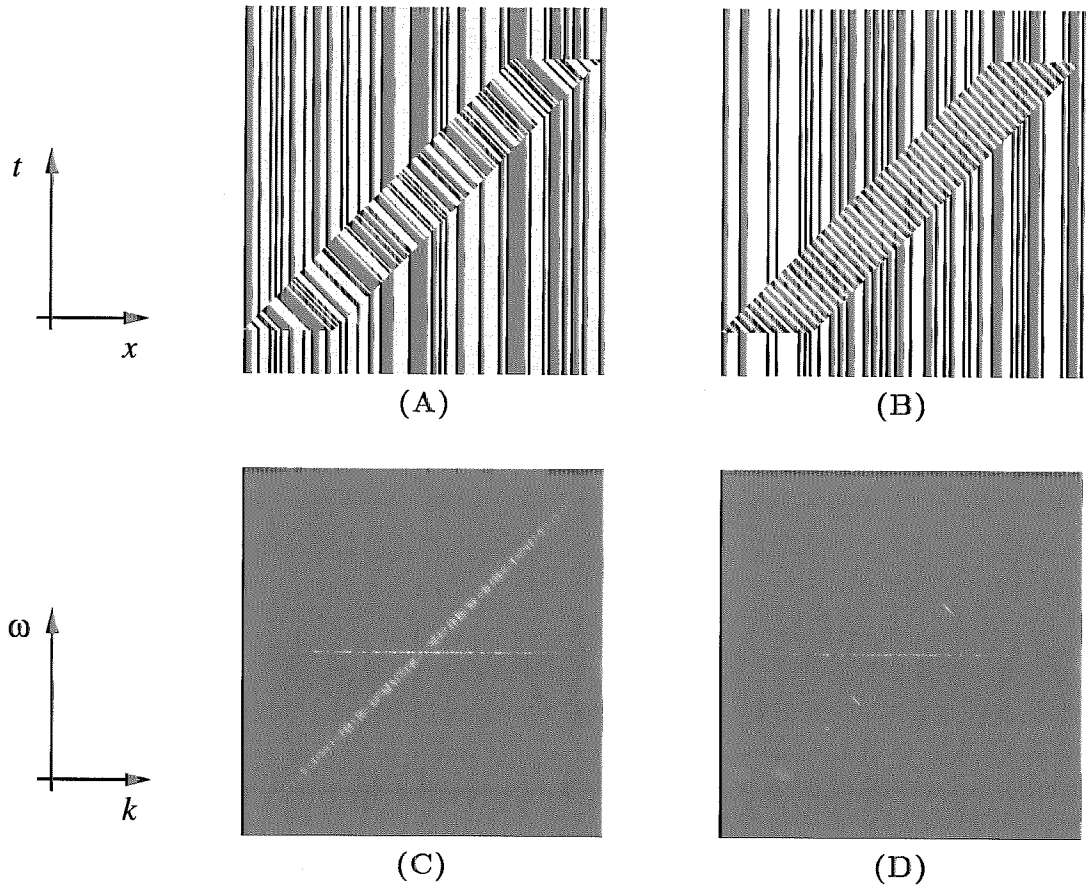

FIGURE 11. The first image (A) shows an instance of a drifting occlusion window, like the theta motion stimuli of Zanker (1993). The other stimulus (B) is an idealization of theta motion in which the random dots within the moving window are replaced by a sinusoidal grating. The amplitude spectra of the two images are shown below in (C) and (D).

than the model in (13). As shown in Fig. 11, the occlusion window in theta motion stimuli moves independently of the foreground and the background, and may be expressed as

$$I(x,t) = \chi(x - v_w t) I_0(x - v_0 t) + I_b(x - v_b t) - \chi(x - v_w t) I_b(x - v_b t). \quad (15)$$

There are now three velocities and two multiplicative terms. The Fourier transform of Fig. 11A is shown in Fig. 11C, and not surprisingly, does not clearly exhibit a group velocity that is consistent with the motion of the occlusion window. The apparent lack of group velocity is due to the variations in local scale throughout the stimulus that tends to flatten the amplitude spectrum. However, from the arguments above, group velocity should appear as we introduce a spectral peak into the foreground or the background. Toward this end, Fig. 11B shows a somewhat idealized version of theta motion in which the occluding foreground pattern is replaced by a sinusoid. The resulting Fourier transform in Fig. 11D now clearly shows a group velocity that is consistent with the moving occlusion window. Interestingly, we also find that such idealizations provide a more compelling percept of the window's motion.

In summary, it appears that group velocity may be viewed as an idealization of a restricted class of multiplicative transparency and occlusion due to multiplicative combinations of signals. It occurs when one of the multiplicative intensity patterns exhibits a local spectral peak, in which case the group velocity yields the velocity of the other multiplicative component. This raises the question: is the occurrence of group velocity (locally) important or necessary for multiplicatively combined signals to be perceived individually in a coherent manner? Does this characterize the class of multiplicative signal combinations that can be perceived transparently, or is the visual system able to interpret wider classes of multiplicative motion?

4. COMPUTATIONAL MODELLING

To model the perception of *non-Fourier* motion stimuli, various authors have suggested that there may be two channels (e.g. Chubb & Sperling, 1988; Turano & Pantle, 1989; Wilson *et al.*, 1992; Zanker, 1993), as illustrated in Fig. 12. One channel, referred to as the Fourier motion channel, involves one of the standard methods of measuring coherent image motion (called Fourier-based mechanisms), such as correlation or energy-based schemes, applied directly to intensity. The other channel, called the non-Fourier channel, involves spatiotemporal band-pass filtering followed by some

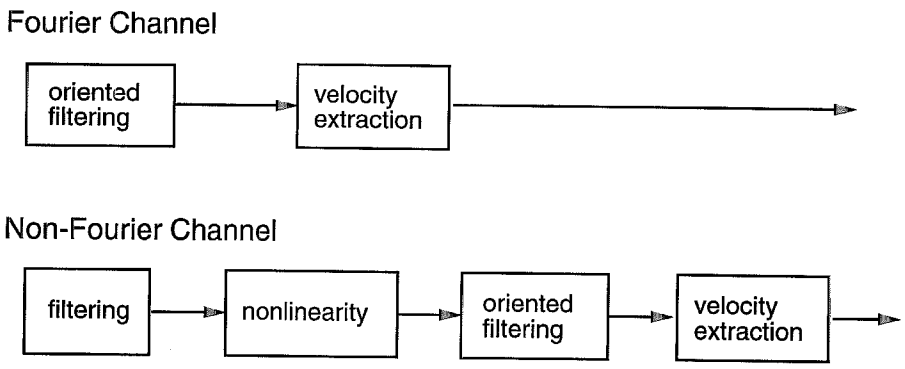

FIGURE 12. This illustrates the two-channel model.

form of non-linearity, the output of which is then fed to a standard Fourier-based mechanism.†

The principle differences among the models concern the nature of the initial filtering and the nonlinearity in the non-Fourier channel. Chubb and Sperling (1988, 1989) suggested that the initial filtering involves a limited form of band-pass filtering, followed by a rectifying nonlinearity. The filtering is relatively broad-band, and is used to remove mean illumination and to enhance band-pass temporal structure; they have suggested different types of filters that appear suited to different stimuli. Wilson *et al.* (1992) proposed a model in which the prefilters are the same oriented filters as those in the Fourier channel, and the nonlinearity is a simple squaring operation. Moreover, they extended the two channel idea to account for the extraction of 2-D velocity, combining 1-D constraints from both channels. The results appear to account for a wide variety of phenomena, along with several predictions that help to support the model (Wilson & Mast, 1993).

Closely related to the Fourier/non-Fourier distinction is the first-order/second-order distinction proposed by Cavanagh and Mather (1989). They describe first-order motion as motion available directly from intensity, and the second-order motion as the motion derived from other stimulus properties, such as the motion of textured boundaries, or occlusion boundaries. In practice this resembles the two-channel models mentioned above.

The remainder of the paper examines the relationship between such computational models and the notion of oriented power in the frequency domain. Towards this end, we consider the information available in band-pass filter responses. Our approach is based on quadrature-pair filters, assuming that velocity measurement is a local process at its primitive level. We focus mainly on the idea that the phase component of the filter output is related to phase velocity (the localization of spectral peaks), and the amplitude component is related to group velocity (the orientation of power). This view is consonant, but different in detail, with current models of non-Fourier motion (Chubb & Sperling, 1988, 1989; Wilson *et al.*, 1992).

It is important to remember that this is an abstract framework from a computational perspective (Marr, 1982). We do not suggest that phase and amplitude are explicitly represented at the neural level. Rather, the model makes these principal sources of information explicit, and therefore further aids understanding of the stimuli and their relation to the models. We begin by outlining the definition of the filters, and the instantaneous phase and amplitude components of their outputs. After discussing our reasons for examining a representation based on phase and amplitude we discuss the typical behaviour of phase and amplitude, especially in the context of coherent image translation and group velocity in a single channel.

4.1. *Spatiotemporal band-pass filters*

Although many properties of linear filters have been discussed in the context of early motion perception, here we assume only that the filters are band-pass, and occur in quadrature pairs (that are approximately $\pi/2$ radians out of phase with similar amplitude spectra). For convenience, we view the quadrature pair of real filters as a single complex-valued filter [where the imaginary part of the filter is the Hilbert transform of the real part (Papoulis, 1977)].

Let $K(x, t;\ k_0, \omega_0)$ be a complex-valued impulse response, such as a Gabor function (Gabor, 1946),‡ tuned to frequencies about (k_0, ω_0). Imagine that its real and imaginary parts are simply the cosine-phase Gabor and the sine-phase Gabor respectively. The response to $K(x, t;\ k_0,\ \omega_0)$ can be expressed as

$$R(x, t) = K(x, t; k_0, \omega_0) * I(x, t), \qquad (16)$$

where $*$ denotes the usual convolution operator, and $I(x, t)$ is the input. Because $K(x, t;\ k_0,\ \omega_0)$ is complex-

†These two channels are sometimes referred to as *linear* and *non-linear* respectively, referring mainly to the preprocessing that precedes the velocity computation, but this can be misleading since both channels, in their entirety, involve significant nonlinearities.

‡Although Gabor filters are not quadrature pairs, they provide a reasonably good approximation for sufficiently small bandwidths of about 1 octave or less.

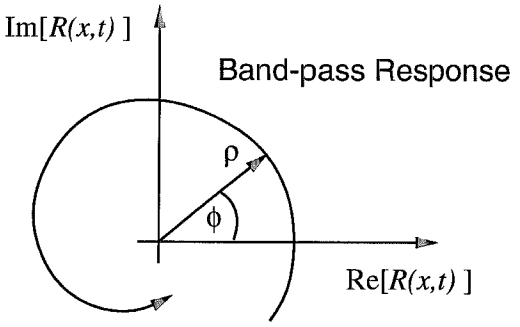

FIGURE 13. Response as a function of space–time is shown as a counter-clockwise function in the complex-plane. Phase and amplitude form a polar representation of the response, as a function of space–time.

valued, (16) involves the convolution of both its real and imaginary parts with $I(x, t)$, and therefore yields a complex-valued response $R(x, t)$:

$$R(x, t) = \mathrm{Re}[R(x, t)] + i\,\mathrm{Im}[R(x, t)], \qquad (17)$$

where $i^2 = -1$. It is also convenient to express the response in polar coordinates in the complex plane (see Fig. 13) in terms of its amplitude and phase components, $\rho(x, t)$ and $\phi(x, t)$, as in

$$\begin{aligned} R(x, t) &= \rho(x, t)e^{i\phi(x, t)}, \\ &= \rho(x, t)\cos\phi(x, t) + i\rho(x, t)\sin\phi(x, t), \end{aligned} \qquad (18)$$

where amplitude and phase are simple functions of the real and imaginary parts:

$$\begin{aligned} \rho(x, t) &= |R(x, t)| \\ &\equiv \sqrt{\mathrm{Re}[R(x, t)]^2 + \mathrm{Im}[R(x, t)]^2}, \end{aligned} \qquad (19a)$$

$$\begin{aligned} \phi(x, t) &= \arg[R(x, t)] \\ &\equiv \mathrm{Im}[\log R(x, t)] \in (-\pi, \pi]. \end{aligned} \qquad (19b)$$

To emphasize the local nature of these response properties, it is common to refer to $\phi(x, t)$ and $\rho(x, t)$ as *instantaneous phase* and *amplitude*.

The final concept we need here is *instantaneous frequency*, which is defined as the derivative of instantaneous phase. More precisely, the spatial and temporal instantaneous frequencies are

$$\tilde{k}(x, t) = \frac{\partial\phi(x, t)}{\partial x}, \quad \tilde{\omega}(x, t) = \frac{\partial\phi(x, t)}{\partial t}. \qquad (20)$$

To understand these definitions, remember that the frequency of a sinusoidal signal refers to the reciprocal of wavelength which is the distance between adjacent crests. Alternatively, frequency may be thought of as the rate of change of phase; where phase is a quantity that varies linearly between $-\pi$ and π over one wavelength. As frequency increases, so does the rate of phase variation as a function of spatial position. The definition given in (20) measures the instantaneous rate of phase change at each point, allowing frequency to change through space–time.

4.2. *Why phase and amplitude?*

There are many ways to represent a complex-valued band-pass signal, but two natural ones are apparent, namely, in terms of its real and imaginary parts, or in terms of its phase and amplitude. Our reasons for choosing phase and amplitude stem from their tendency to capture salient aspects of the signals. Phase gives the local structure of response while amplitude gives the strength of response (illustrated in Section 4.6.). We feel that, in an abstract sense, this particular representation of the filter output helps address questions concerning signal representation as well as motion perception.

The use of quadrature amplitude is not new to models of motion perception as it is central to energy-based frameworks (e.g. Adelson & Bergen, 1985; Emerson, Bergen & Adelson, 1992)]. The importance of phase has also been noted previously. For example, Morrone and Burr (1988) reported that phase coincidence across several scales plays a major role in perceived salience of image structure. This is especially true with phases that are integer multiples of $\pi/2$, that are perceived as edges and bars of different polarities. While the relative magnitude of amplitude depends in part on the degree of phase coincidence through scale, it is the local phase behaviour that determines the structure and polarity of the local signal. In the auditory domain, representations of band-pass channels in terms of phase and amplitude are commonplace. Interaural time differences for auditory lateralization are usually associated with phase information, but it has also been suggested that envelope may also play a significant role (Bernstein & Trahiotis, 1982, 1985). This view is also supported by physiological recordings (Yin, Kuwada & Sujaku, 1984). Phase has also been proposed recently to play a major role in the extraction of binocular disparity in visual cortex, with cells showing interocular phase-specific responses (Ohzawa, De Angelis & Freeman, 1990; DeAngelis, Ohzawa & Freeman, 1991). Finally, phase and envelope have also been found useful in machine vision for motion analysis, the measurement of binocular disparity, texture analysis, and the extraction of lines and edges (e.g. see Bårman, Haglund, Knutsson & Granlund, 1991; Bovik, Clark & Geisler, 1990; Fleet & Jepson, 1990; Fleet, Jepson & Jenkin, 1991; Heeger, 1987; Jenkin & Jepson, 1991; Langley, Atherton, Wilson & Larcombe, 1990; Sanger, 1988).

4.3. *Behaviour of instantaneous phase and amplitude*

Before considering specific cases of visual motion, it is useful to review some of the general properties of phase and amplitude. In particular, a common observation about band-pass signals, such as those shown in the simulations in Section 4.6, is that their local behaviour is approximately sinusoidal, with a frequency close to the filter tuning, and a slowly varying amplitude. In terms of instantaneous properties, this implies that $\phi(x, t)$ is predominantly linear and the amplitude $\rho(x, t)$ is

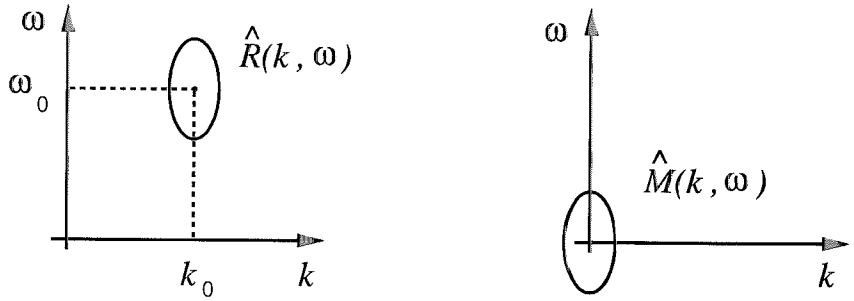

FIGURE 14. Multiplying the band-pass filter response $R(x, t)$ by a sinusoidal signal as in (21) will translate its amplitude spectrum to the origin.

low-pass. A pure sinusoid, for example, has linear phase and constant amplitude.

Recent results have shown that the spatiotemporal extent within which phase is predominantly linear (and hence the filter output nearly sinusoidal) increases as the bandwidth of the filter decreases (Fleet & Jepson, 1993). Moreover, the instantaneous frequency of the filter output [i.e. the phase derivative (20)] will usually be close to the filter tuning.* In fact it can be shown that the average instantaneous frequency (averaged over space–time) is a measure of the location of the band-pass signal's spectral centre of mass (or peak energy concentration) (Boashash, 1992). The specific relationship between the instantaneous frequency at a point in space–time, and local maxima in frequency domain of the amplitude spectra of the local neighbourhood is not precisely understood. Despite this, it is generally thought that instantaneous frequency provides a measure of local energy maxima in frequency domain.

Concerning the slowly varying nature of amplitude, Knutsson (1982) derived an autocorrelation function for $\rho^2(x, t)$ to show that it is low-pass, and that the mean frequency of $\rho^2(x, t)$ is zero. To see this from another perspective, note that we can transform the Fourier transform of $R(x, t)$ into a low-pass signal by multiplying it by a sinusoid, leaving its amplitude unchanged, that is

$$M(x, t) = R(x, t)\mathrm{e}^{-i(xk_0 + t\omega_0)} \tag{21}$$

where (k_0, ω_0) is the tuning frequency of the filter. As illustrated in Fig. 14, the Fourier transform of $M(x, t)$ is simply the transform of $R(x, t)$ shifted to the origin, $\hat{M}(k, \omega) = \hat{R}(k + k_0, \omega + \omega_0)$ (Papoulis, 1977). The demodulated signal $M(x, t)$, with amplitude $\rho(x, t)$, is a low-pass signal, whose highest frequencies depend on the bandwidth of the filter.

4.4. *Coherent image translation (Fourier motion)*

We are now ready to examine the behaviour of instantaneous phase and amplitude when the input translates coherently, as in

$$I(x, t) = I(x - vt, 0). \tag{22}$$

In this case, one can show that the response of any convolution operator will also translate with the same speed:

$$R(x, t) = R(x - vt, 0) = \rho(x - vt, 0)\,\mathrm{e}^{-i\phi(x - vt, 0)} \tag{23}$$

where $R(x, 0)$ is the spatial response profile at time 0. Not surprisingly, this means that both the phase and amplitude of the filter output also translate with velocity v. In particular, it follows from (23) that

$$\phi(x, t) = \phi(x - vt, 0) \tag{24a}$$

$$\rho(x, t) = \rho(x - vt, 0). \tag{24b}$$

One can then differentiate these equations with respect to time to get the motion constraints

$$\phi_x(x, t)\, v + \phi_t(x, t) = 0, \tag{25a}$$

$$\rho_x(x, t)\, v + \rho_t(x, t) = 0, \tag{25b}$$

from which velocity of instantaneous phase is $v = -\phi_t(x, t)/\phi_x(x, t)$, and the amplitude velocity is $v = -\rho_t(x, t)/\rho_x(x, t)$.

This shows that both the amplitude and phase outputs yield consistent velocity information. Also note that because phase derivatives are just instantaneous frequencies, the phase-based constraint in (25a) resembles the relationship between Fourier components and phase velocity discussed in Section 2.1. In other words, given velocity v, all instantaneous frequencies lie on the same line through the origin in the frequency domain as do the Fourier frequencies of the original translating signal. We conclude from (25) that the coincidence of phase and amplitude orientations in space–time is a useful constraint on the local coherence of the signal.

4.5. *Multiplicative envelope velocity (non-Fourier motion)*

When all power in a single channel lies along a line segment in frequency space, analogous to the simple example of group velocity given in Section 2.2, one can show that instantaneous frequency will lie somewhere along the same line, depending on the distribution of power, and the amplitude velocity can be shown to equal

*One can derive a probability density function for the distribution of instantaneous frequencies for typical filters, thereby showing that probability density is typically concentrated about the filter's tuning frequency, and its spread increases with the bandwidth of the filter (Fleet, 1992).

the group velocity, which depends on the orientation of the line.

To see this, let all nonzero power in a filter output $R(x, t)$ lie on the line in frequency domain through the point (k_0, ω_0) in direction (d_k, d_ω). All frequencies on the line therefore satisfy

$$(k(\alpha), \omega(\alpha)) = (k_0, \omega_0) + \alpha\,(d_k, d_\omega) \tag{26}$$

for some real-valued scalar α. If we demodulate the filter output $R(x, t)$ as in Section 4.3,

$$M(x, t) = R(x, t)\mathrm{e}^{-i(k_0 x + \omega_0 t)}, \tag{27}$$

which shifts the line segment in frequency domain by $-(k_0 x, \omega_0)$, then the amplitude spectrum of $M(x, t)$ lies on a line through the origin with direction (d_k, d_ω).

Accordingly, Section 4.4. has already shown that the amplitude and phase velocities of $M(x, t)$ will be the same, equal to $v = -d_\omega/d_k$. Since the amplitude components of the $M(x, t)$ and $R(x, t)$ are identical, the amplitude velocity of $R(x, t)$ will also be $-d_\omega/d_k$. It therefore follows that

$$(\rho_x(x, t), \rho_t(x, t)) = \alpha\,(d_k, d_\omega), \tag{28}$$

for some real α. The space–time orientation of the amplitude response gives the group velocity.

The instantaneous phase velocity of $R(x, t)$ is given by instantaneous frequencies. From (27), with some algebraic manipulation, it follows that the instantaneous frequencies of $R(x, t)$ are related to the instantaneous frequencies of $M(x, t)$ by

$$(\phi_x(x, t), \phi_t(x, t)) = (k_0, \omega_0) + (\psi_x(x, t), \psi_t(x, t)), \tag{29}$$

where $\psi(x, t) \equiv \arg[M(x, t)]$ denotes the phase component of $M(x, t)$, and $\phi(x, t) \equiv \arg[R(x, t)]$. Because the power spectrum of $M(x, t)$ lies on a line through the origin in the direction (d_k, d_ω), its instantaneous frequencies lie along the same line. Therefore, we can replace $(\psi_x(x, t), \psi_t(x, t))$ in (29) by $\alpha\,(d_k, d_\omega)$ where α is a real scalar:

$$(\phi_x(x, y), \phi_t(x, t)) = (k_0, \omega_0) + \alpha\,(d_k, d_\omega). \tag{30}$$

This shows that all instantaneous frequencies lie along the original line of nonzero power (26). The distribution of instantaneous frequencies along this line depends on the concentrations of power (Fleet, 1992). Interestingly, this means that the distribution of instantaneous frequencies from a local collection of filter responses may provide another source of information concerning group velocity, and hence non-Fourier motion.

In practice we do not expect filter outputs to have all their power concentrated strictly along line segments in frequency domain. With this in mind, Appendix A3 shows that the above results generalize in reasonable ways. In particular, when power is concentrated near a line, it still follows that the dominant orientation of the amplitude output of the filter in space–time corresponds to the orientation of power in frequency domain.

4.6. Simulations

To complement the theoretical results, we now show simulations of the behaviour of instantaneous phase and amplitude from the output of band-pass filters when applied to the stimuli examined in Section 2. We used Gabor filters (Gabor, 1946) with isotropic Gaussian windows, and bandwidths of one octave (measured at 1 standard deviation of the amplitude spectrum in frequency domain). The simulations concentrate on the phase and amplitude behaviour of the filter outputs because they are directly related to phase velocity (location of power in the frequency domain) and group velocity (orientation of power in the frequency domain).

Figure 15 illustrates the response properties of two filters applied to the plaid pattern shown in Fig. 5. This signal consists of two sinusoidal components with identical wavelengths, and orientations of 120° and 135°. The first filter, tuned to 127.5°, is equally sensitive to both sinusoidal components, and therefore they have equal amplitude in the filter's response. The other filter, tuned to 140°, is more sensitive to the sinusoid with orientation 135°, and this simply changes the relative amplitudes of the two sinusoids in the filter's response. Figures 15A,D show the real parts of the filter outputs, while Figs 15B,C and E,F show the corresponding amplitude and phase components as functions of space–time. Upon brief inspection it is evident that the amplitude signals appear to capture the beat–structure, while the phase captures properties of the carrier (the fine structure).

Examining these and other filter outputs in more detail, one may observe that the space–time orientation of the amplitude signal is stable over a wide range of filter tunings, despite the variations in the relative amplitude of the two sinusoidal components in the filter responses. This is a consequence of the results in Section 4.5 that show that the amplitude orientation depends on the orientation of (non-zero) spectral power in the frequency domain, which does not depend on the amplitudes of the sinusoids. In practice we expect that the perceptual salience of the beats will however depend on their magnitude and frequency which do depend on the relative amplitudes and frequencies of the two sinusoids. This becomes evident as the two frequencies become increasingly far apart.

Interestingly, the phase structure in Fig. 15 significantly depends on the relative amplitudes of the two sinusoids in the filter output. When the amplitudes are equal (in Fig. 15C), the instantanteous frequency (the phase derivative) is constant, midway between the two frequencies (except for the lines of phase singularities along the amplitude troughs). When the amplitudes differ (in Fig. 15F), the instantaneous frequency varies as a function of space–time, with its mean closer to the frequency of the sinusoid with the larger amplitude. This is especially noticeable in the amplitude troughs. But remember that this variability is by no means random, for the instantaneous frequencies extracted from all filter outputs will lie along the line in frequency domain that

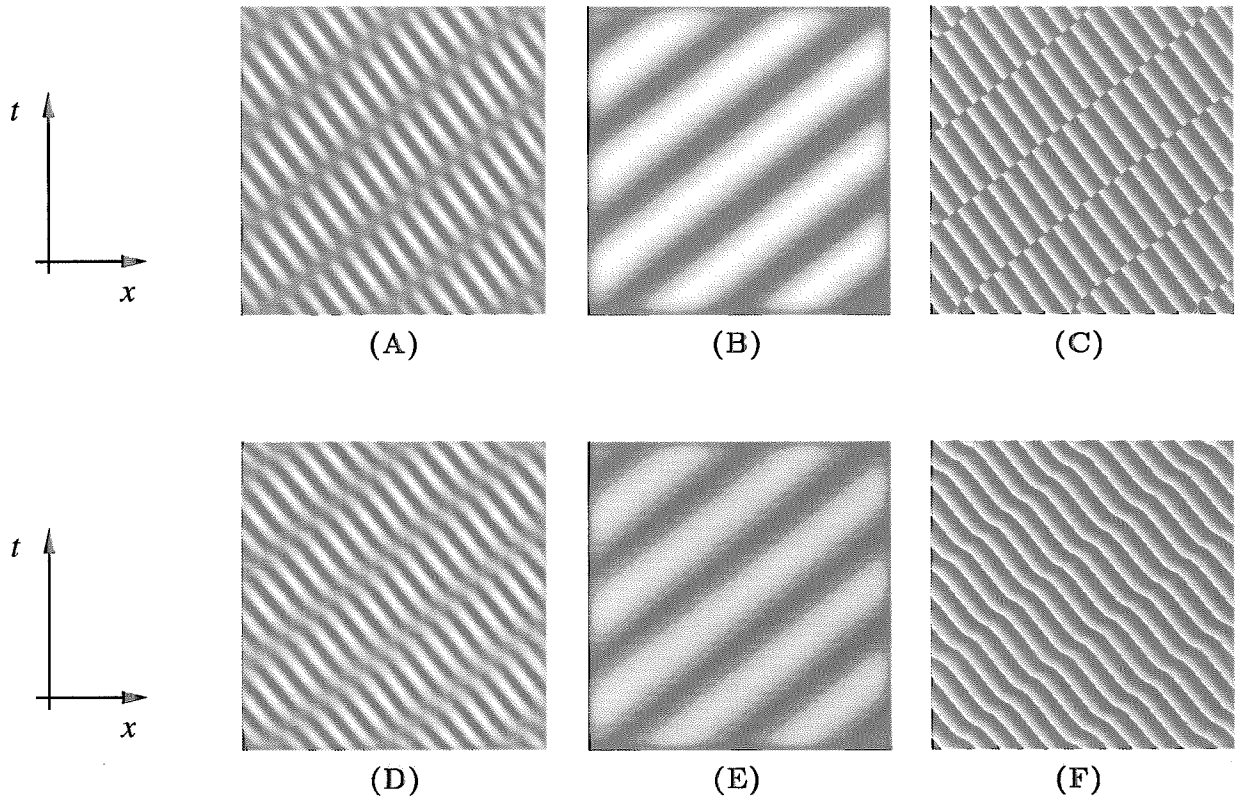

FIGURE 15. This figure shows responses of two one-octave Gabor filters applied to the beat pattern in Fig. 5, which consists of two sinusoids with orientations 135° and 120°. (A) shows the real part of the response of filter tuned to 127.5°. (B) and (C) show the corresponding amplitude and phase responses. (D), (E) and (F) show the real part, the amplitude, and the phase responses of a filter tuned to 140°.

passes through the two impulses of power, as explained in Section 4.5.

The next example, in Fig. 16, illustrates the amplitude and phase behaviour of a Gabor filter applied to the sampled motion stimulus from Fig. 7A. The filter is tuned near one of the dominant spectral concentrations shown in Fig. 7B, the frequencies of which have leftward phase velocities. Figure 16C shows that the instantaneous phase of the filter output also implies a leftward velocity. But the space–time orientation of the amplitude response is consistent with the rightward motion of the input pattern. For such stimuli, different filters are expected to yield different phase velocities, but the space–time structure of their amplitude responses can be very consistent across a wide range of filter tunings.

Of course, this stability of amplitude orientation depends on the shape and size of the filters' power spectra in relation to the power spectrum of the input (as discussed in Section 4.5 and in Appendix A3). Here, the filter's bandwidth is one octave, in which case the filter will respond significantly to only one concentration of non-zero spectral power of the input (shown in Fig. 7). Therefore, while each channel yields a different phase velocity, some leftward and some rightward, all amplitude velocities will be consistent with the rightward

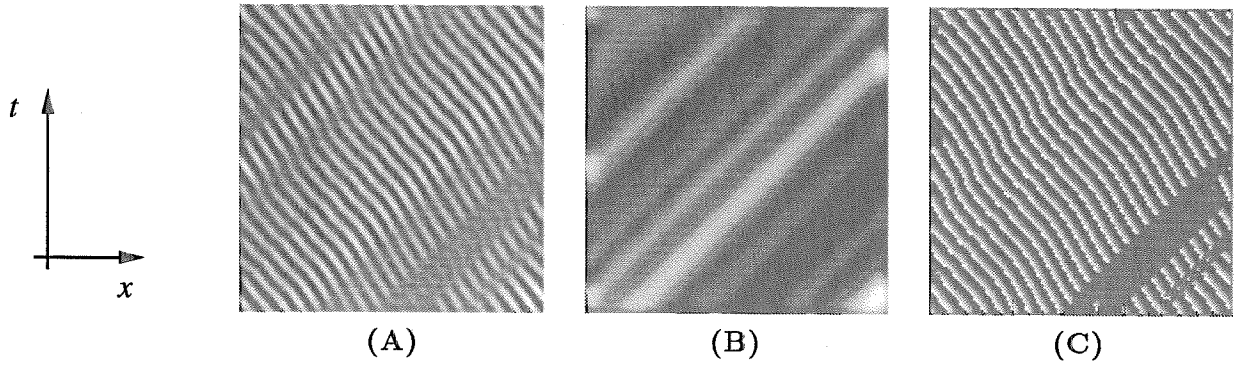

FIGURE 16. Simulations with Sampled Motion: (A) shows the real part of the response of a one octave Gabor filter tuned to 135° when applied to the subsampled motion stimulus in Fig. 6. (B) and (C) show the corresponding amplitude and phase responses. Regions of numerical phase instability are replaced with zeros in (C) as discussed in Section 5.1.

motion of the signal, which coincides with the dominant motion percept.

If the power spectra of the filters were much larger then they will be sensitive to more than spectral power concentration of the input. In this case, the power spectrum of the filter output would not be concentrated along a line in the frequency domain. In fact, following the results in Appendix A3, the amplitude component of response would then include a product of two interference patterns, namely, a low frequency term that reflects the orientation of the individual power concentrations, and a higher frequency component that reflects the vector between the two major power concentrations. As seen from Fig. 7B, this latter will be vertical in the frequency domain, and would appear as temporal contrast variation in space–time. In other words, we could no longer assume a simple amplitude envelope that translates coherently to the right.

It is worthwhile at this point to mention a major difference between the framework described here and current non-Fourier models. Here we have used quadrature filters, the amplitude and phase of which cleanly separate the two sources of information relating to location and orientation in the frequency domain, i.e. Fourier and non-Fourier motion. This separation is clear in Figs 15 and 16. By contrast, the non-Fourier channels of Chubb and Sperling (1988) and of Wilson *et al.* (1992) involve a half or full-wave rectification of the band-pass response directly. This will contain both phase and amplitude, since, the square of real part of $R(x, t)$ in (18) is equal to the product of squared amplitude and the squared cosine of the phase part. Since local phase orientation yields information concerning locations of spectral peaks in frequency domain (and hence Fourier motion), and amplitude is closely related to group velocity, such a non-Fourier channel contains information relevant to both phase and group velocity.

Our third example, in Fig. 17, shows the response of a Gabor filter applied to the rightward-stepping random contrast-reversing bar shown in Fig. 8. The simulations again show that the dominant space–time orientation of the amplitude component is consistent with the path of the contrast-reversing bar. The phase component of response however, reflects the filter tuning, which was perpendicular to the space–time orientation of the bar's path. One can also see that the level contours of response amplitude are not aligned perfectly with the path of the bar, since the amplitude decreases in various places along the path shown in Fig. 17B. In terms of Fourier analysis, as discussed in Appendix A3, this means that power does not lie strictly on a line in the frequency domain, and therefore the amplitude may be viewed as the product of two signals, but is still predominantly oriented in the appropriate direction.

Over a wide range of filter tunings we find that the amplitude structure continues to have this dominant orientation while the instantaneous phase component of response changes with the filter tuning. An interesting example of this occurs with the idealized stimulus shown in Fig. 4, the power spectrum of which is a line. In this case one can prove that the instantaneous frequency of the response of a Gabor filter (with an isotropic power spectrum) will equal the point on the input's power spectrum that is closest to the filter tuning. As the filter tuning varies smoothly, the instantaneous frequency of the response will vary along the line in a predictable fashion. This distribution of instantaneous frequencies could also be used to measure group velocity, as explained in Section 4.5.

Our final example, in Fig. 18A, shows an example of the theta motion stimuli of Zanker (1993), along with the amplitude and phase responses of two Gabor filters, tuned to 90° and 135° respectively in Figs 18B,C and D,E. These two filter tunings agree with velocities of the foreground and background. In both cases the amplitude component of response clearly shows evidence of the occlusion boundary where the filter responds to one of the two velocities. However, as expected from the results of Section 3.4, the velocity of the occlusion window is not everywhere evident in the output of the amplitude, since group velocity will only occur when the foreground or background exhibit local spectral peaks. Like the contrasting reversing bar discussed above, notice in the responses in Figs 18B,D that the local amplitude signal stems from a product of two signals,

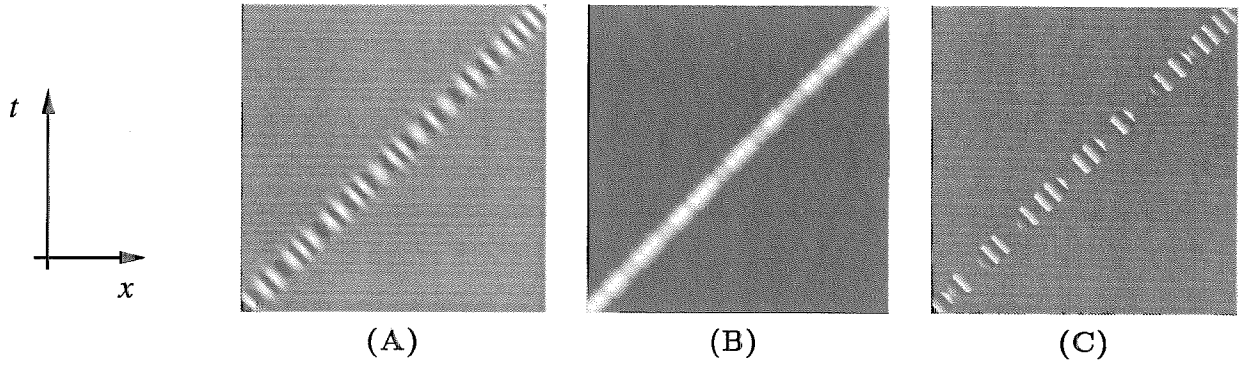

FIGURE 17. Simulation with Rightward Stepping Reversing Bar: (A) shows the real part of the response of a one octave Gabor filter tuned to 135° when applied to the drift-balanced stimulus in Fig. 7. The orientation tuning of filter is perpendicular to the space–time orientation of the envelope motion. (B) and (C) show the corresponding amplitude and phase responses. Regions of numerical phase instability are replaced with zeros in (C) as discussed in Section 5.1.

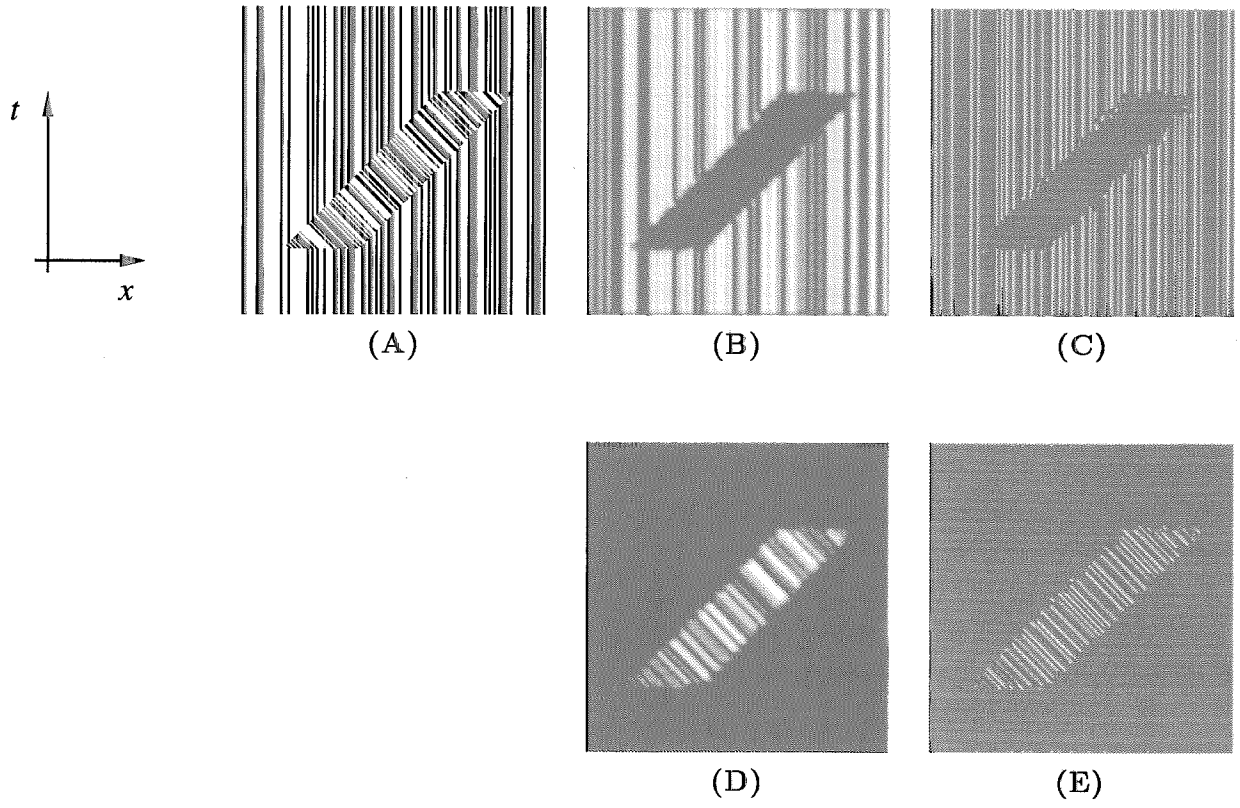

FIGURE 18. Simulation with Theta Motion: an example of Zanker's theta motion, similar to those shown in Fig. 10 is shown in (A). The remaining images show the amplitude responses in (B) and (D), and phase responses, (C) and (E), of Gabor filters tuned to 90° and 135°. Regions of phase instability are replaced with zeros in (C) and (E) as discussed in Section 5.1.

the orientations of which depend on the noise and the occlusion boundary, and therefore power does not lie strictly along a line in the frequency domain. When different scales are considered together, although not shown here, the segmentation of the occlusion boundary appears much more straightforward.

5. MEASUREMENT OF VISUAL MOTION

Before summarizing the main conclusions of this paper, it is of interest to provide some comments regarding some practical issues regarding methods for measuring visual motion. This is important since much of this paper concentrates on simple forms of idealized motion, based only on image translation. Furthermore, although a detailed computational treatment of methods for extracting Fourier and non-Fourier motion are outside the scope of this paper, some comments are warranted concerning alternative methods of measurement.

5.1. Phase velocity

There are several well-known methods of extracting coherent image translation (so-called Fourier motion). Perhaps the most common are the correlation-based, energy-based and gradient-based methods (e.g. van Santen & Sperling, 1985; Adelson & Bergen, 1986; Srinivasan, 1990; Bårman *et al.*, 1991). As discussed above, they are thought to be broadly equivalent. This paper suggests a closely related fourth approach, based on the phase behaviour in the output of band-pass channels.

Phase-based methods have been used successfully in machine vision (e.g. Fleet & Jepson, 1990; Fleet *et al.*, 1991; Fleet, 1992; Jenkin & Jepson, 1991; Langley *et al.*, 1990; Sanger, 1988), and produced accurate results in a recent comparative study (Barron, Fleet & Beauchemin, 1994). The preference for phase over the filter output, even in the case of smooth coherent motion, is the result of robustness considerations (Fleet & Jepson, 1993). Phase is relatively stable under deviations from image translation that commonly occur with projections of 3-D scenes (e.g. the dilation of the image as a camera approaches an object), while the amplitude of the filter response is not. This means that phase-based measurements can be accurate and local, based on the velocity of points of constant phase. For example, let space–time paths $x(t)$ along which phase is constant be

$$\phi(x(t), t) = c. \tag{31}$$

Assuming that $x(t)$ is smooth, we differentiate (31), which yields the constraint

$$\phi_x(x, t)\, v + \phi_t(x, t) = 0. \tag{32}$$

from which velocity can be measured.

Despite its robustness, instantaneous phase also exhibits regular instabilities that occur in the neighborhoods of phase singularities, where amplitude goes to zero (Jepson & Fleet, 1991). In these regions phase

is very sensitive to small changes in scale or spatial position, and does not produce reliable velocity estimates. But they can be detected: if the instantaneous frequency is not sufficiently close to the filter's frequency, or if the amplitude changes too quickly as a function of space–time, then the motion information from the channel in question should be discarded (Fleet & Jepson, 1993). It is also necessary to remove regions of phase behaviour where the amplitude of response is particularly small for numerical reasons. These constraints were important for the reliability of the optical flow technique described in (Fleet & Jepson, 1991). They were also used in Section 4.6, where noted, to remove regions of phase instability.

Finally, note that it is not necessary to explicitly represent phase to compute instantaneous frequencies. Rather, phase derivatives may be expressed in terms of the filter output as (Fleet & Jepson, 1990):

$$\phi_x(x, t) = \frac{\mathrm{Im}[R^*(x, t) R_x(x, t)]}{\rho^2(x, t)}, \tag{33}$$

where R^* denotes the complex conjugate of R. In terms of the real and imaginary parts of $R(x, t)$ this simplifies to

$$\phi_x(x, t) = \frac{\mathrm{Im}[R_x(x, t)]\mathrm{Re}[R(x, t)] - \mathrm{Re}[R_x(x, t)]\mathrm{Im}[R(x, t)]}{\mathrm{Re}[R(x, t)]^2 + \mathrm{Im}[R(x, t)]^2}. \tag{34}$$

5.2. Group velocity

Methods of computing group velocity, the orientation of the local power distribution in frequency domain, are not well established. The models proposed by Chubb and Sperling (1988) or Wilson *et al.* (1992), have not been explored in much depth computationally. As discussed below, the results in this paper show two methods of extracting orientation of power in the frequency domain, one based on response amplitude and the other based on instantaneous frequencies.

In ideal cases, where all power in the output of a filter lies along a line in frequency domain (not necessarily through the origin), the results of Sections 4.4 and 4.5 suggests that the most direct method of extracting the group velocity is to measure the velocity of points of constant amplitude. For example, we could use the constraint

$$\rho_x(x, t)\, v + \rho_t(x, t) = 0. \tag{35}$$

Moreover, like the phase derivatives above, amplitude derivatives can be simplified using

$$\rho_x(x, t) = \frac{\mathrm{Re}[R^*(x, t) R_x(x, t)]}{\rho(x, t)}. \tag{36}$$

Thus, both phase and amplitude derivatives may be computed in the same way, based on the real and imaginary parts of $R^*(x, t)R_x(x, t)$.

In practice, power may *not* lie strictly along a line, as is the case in some of the signals used in the simulations above. Moreover, amplitude derivatives will be noisy, in part because the amplitude signal is slowly modulating (in which case the derivatives are small near crests and troughs), and in part because amplitude is sensitive to geometric deformation through time that occur with dilation or rotation in the optical flow field (Fleet & Jepson, 1993). Therefore we cannot expect to get the same accuracy with local computations on amplitude, as we might with phase information, and it is therefore advisable to consider the dominant orientation/motion of amplitude over somewhat larger spatiotemporal extents than those of differential methods, perhaps with the use of band-pass prefilters, as suggested in Appendix A3. Interestingly, this approach is similar to that taken in the models of Wilson *et al.* (1992) and of Chubb and Sperling (1988), however, as noted above in Section 4.6, their models do not cleanly separate amplitude and phase as does the framework described here.

Sections 4.4 and 4.5 also suggest that another feasible method of extracting group velocity may be to consider the distribution of instantaneous frequencies in local neighbourhoods from similarly tuned filters. As explained in Section 4.5 this distribution of instantaneous frequencies should lie along a line in frequency domain, the orientation of which gives the group velocity. For example, one might consider a least-squares fit of instantaneous frequencies to a line based on the constraint

$$\phi_x(x, t)\, v_g + \phi_t(x, t) = c \tag{37}$$

rather than that in (32). This shows that one might extract information pertaining to non-Fourier motion via the orientation of power in frequency domain from either the amplitude or the phase of the response of band-pass channels.

Finally, it is also necessary to determine whether there is indeed a reasonably coherent amplitude velocity. As explained in Appendix 3 a single dominant envelope velocity will only occur when power is concentrated near a line in frequency domain. If the amplitude spectra of the filters are sufficiently large that more complex power distributions occur within the tuning of a single channel in a local region, then the model of group velocity and these methods of measurement would appear too simple. A careful examination of this issue is a very important direction for further research since it directly affects the relevance of oriented power in the frequency domain, as well as these computational models.

6. SUMMARY

This paper presents a characterization of non-Fourier motion using the orientation of power in the frequency domain, along lines that do not necessarily contain the origin. Like the concept of group velocity from the theory of wave propagation, the orientation of power in frequency domain is shown to correspond to the velocity of a multiplicative amplitude envelope.

Our goal in this paper is to examine how this observation relates to the stimuli and models of non-

Fourier motion. Towards this end, we considered a diverse collection of non-Fourier stimuli including drifting amplitude (contrast) envelopes, sinusoidal beats, sampled (aliased) motion, drift-balanced stimuli, and theta motion, most of which appear to exhibit oriented power in the frequency domain, especially in idealized cases. By characterizing stimuli in the Fourier domain we are introducing an additional level of description and prediction that may be used in the psychological exploration of visual motion perception.

We also explained some of the consequences of this perspective, by examining the sources of information concerning the orientation of power in the frequency domain that may be available in the output of a general class of band-pass filters. It was shown that there is a very simple complementary relationship between the phase and amplitude structure of the filter outputs. We viewed them as two distinct sources of information concerning Fourier and non-Fourier motion, where the phase component yields information about local spectral peak concentrations, and the amplitude component yields information concerning the orientation of power in frequency domain. These results help explain the apparent success of the current non-Fourier models of Chubb and Sperling (1988) and Wilson *et al.* (1992), that have concentrated to varying degrees on the amplitude part of band-pass versions of the stimuli. Our analysis also shows that alternative approaches may be equally well suited to extract Fourier and non-Fourier motion, including the use of energy techniques to extract Fourier motion, and the extraction of group velocity from the phase information in band-pass channels. The detailed behaviour of these somewhat different models, as they relate to the individual stimuli, is beyond the scope of this paper.

The other issue addressed here concerns the natural causes of non-Fourier motion. It is shown that oriented power in frequency domain can occur with multiplicative signal combinations, such as those caused by multiplicative transparency, lighting variations or occlusion. When two signals are combined multiplicatively, oriented power in frequency space occurs as a restricted case, when one of the two signals exhibits a distinct local spectral peak, in which case the orientation (or group velocity) yields the velocity of the other signal. This may therefore represent the class of multiplicative signal combinations that can be perceived transparently. If power in the local output of a band-pass channel were concentrated along a line in frequency domain, then more complex computational methods would be required to extract the relevant signal velocities. A detailed study of this issue remains a topic for future research.

In summary, using the notions of group velocity and oriented power in frequency domain we have examined a diverse class of non-Fourier stimuli, as well as computational models of non-Fourier motion and the potential sources of non-Fourier motion that occur in natural images. Together, these results provide an interesting new perspective on non-Fourier motion.

REFERENCES

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

Adelson, E. H. & Bergen, J. R. (1986). The extraction of spatio-temporal energy in human and machine vision. *Proceedings IEEE workshop on motion*, (pp. 151–156). Charleston.

Bårman H., Haglund L., Knutsson H., & Granlund G. (1991). Estimation of velocity, acceleration and disparity in time sequences. *Proceedings IEEE workshop on visual workshop*, (pp. 44–51). Princeton.

Barron, J. L., Fleet, D. J., & Beauchemin, S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision, 12*, 43–77.

Bernstein, L. R. & Trahiotis, C. (1982). Detection of interaural decay in high-frequency noise. *Journal of the Acoustical Society of America, 71*, 147–152.

Bernstein, L. R. & Trahiotis, C. (1985). Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities. *Journal of the Acoustical Society of America, 77*, 1868–1880.

Bishof, W. F. & DiLollo, V. (1990). Perception of directional sampled motion in relation to displacement and spatial frequency: Evidence for a unitary motion system. *Vision Research, 30*, 1341–1362.

Bishof, W. F. & DiLollo, V. (1991). On the half-cycle displacement limit of sampled directional motion. *Vision Research 31*, 649–660.

Boashash, B. (1992). Estimating and interpreting the instantaneous frequency of a signal. *Proceedings IEEE, 80*, 520–568.

Bovik, A. C., Clark, M. & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12*, 55–73.

Braddick, O. J. & Cleary, R. (1991). Is there is a half-cycle displacement limit for directional motion detection? *Vision Research, 31*, 761–763.

Brillouin, L. (1960). *Wave propagation and group velocity*. New York: Academic Press.

Burr, D. C., Ross, J. & Morrone, M. C. (1988). Smooth and sampled motion. *Vision Research, 26*, 643–652.

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

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

Chubb, C. & Sperling, G. (1989). Second-order motion perception: Space–time separable mechanisms. *Proceedings of the IEEE motion workshop*, (pp. 126–138). Irvine.

Cleary, R. & Braddick, O. J. (1990). Direction discrimination for band-pass filtered random dot kinematograms. *Vision Research, 30*, 303–316.

DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. (1991). Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature, 352*, 156–159.

Derrington, A. M. & Badcock, D. R. (1985). Separate detectors for simple and complex grating patterns? *Vision Research, 25*, 1869–1878.

Emerson, R. C., Bergen, J. R. & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research, 32*, 203–218.

Fahle, M. & Poggio, T. (1981). Visual hyperacuity: Spatiotemporal interpolation in human vision. *Proceedings of the Royal Society, London, B213*, 451–477.

Fleet, D. J. (1992). *Measurement of image velocity*. Norwell, Mass.: Kluwer Academic.

Fleet, D. J. & Jepson, A. D. (1990). Computation of component image velocity from local phase information. *International Journal of Computer Vision, 5*, 77–104.

Fleet, D. J. & Jepson, A. D. (1993). Stability of phase information. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*, 1253–1268.

Fleet, D. J., Jepson, A. D. & Jenkin, M. (1991). Phase-based disparity measurement. *CVGIP: Image Understanding, 53*, 198–210.

Freeman, W. T., Adelson, E. H. & Heeger, D. J. (1991). Motion without movement. *Computer Graphics, 4*, 27–30.
Gabor, D. (1946). Theory of communication. *Journal of IEE, 93*, 429–457.
Heeger, D. J. (1987). A model of the extraction of image flow. *Journal of the Optical Society of America A, 4*, 1455–1471.
Jenkin, M. R. M. & Jepson, A. D. (1991). Techniques for disparity measurement. *CVGIP: Image Understanding, 53*, 14–30.
Jepson, A. D. & Fleet, D. J. (1991). Phase singularities in scale-space. *Image and Vision Computing, 9*, 338–343.
Knutsson, H. (1982). Filtering and reconstruction in image processing. Ph.D. dissertation, Dept of Electrical Engineering, Linkoping University.
Langley, K., Atherton, T. J., Wilson, R. G. & Larcombe, M. H. E. (1990). Vertical and horizontal disparities from phase. *Image and Vision Computing, 9*, 296–303.
Marr, D. (1982). *Vision*. New York: Freeman.
Morgan, M. J. (1980). Analogue models of motion perception. *Philosophical Transactions of the Royal Society (London), B290*, 117–135.
Morrone, M. C. & Burr, D. C. (1988). Feature detection in human vision: A phase-dependent energy model. *Proceedings of the Royal Society London, B235*, 221–245.
Nishida, S. & Sato, T. (1992). Positive motion after-effect induced by band-pass filtered random-dot kinematograms. *Vision Research, 23*, 1635–1646.
Ohzawa, I., DeAngelis, G. C. & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science, 249*, 1037–1041.
Papoulis, A. (1977). *Signal analysis*. New York: McGraw-Hill.
Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In Rosenblith, W. A. (Ed.), *Sensory communication*. New York: Wiley.
Rosenfeld, A. & Kak, A. (1976). *Digital image processing*. New York: Academic Press.
Sanger, T. (1988). Stereo disparity computation using Gabor filters. *Biological Cybernetics, 59*, 405–418.
van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A, 2*, 330–321.
Srinivasan, M. V. (1990). Generalized gradient schemes for the measurement of two-dimensional image motion. *Biological Cybernetics, 63*, 421–431.
Turano, K. & Pantle, A. (1989). On the mechanism that encodes the movement of contrast variations. *Vision Research, 29*, 207–221.
Watson, A. B. & Ahumada, A. J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A, 2*, 322–342.
Watson, A. B., Ahumada, A. J. & Farrell, J. E. (1986). Window of visibility: A psychophysical theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America A, 3*, 300–307.
Whitham, G. B. (1974). *Linear and nonlinear waves*. New York: Wiley.
Wilson, H. R., Ferrera, V. P. & Yo, C. (1992). Psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience, 9*, 79–97.
Wilson, H. R. & Mast, R. (1993). Illusory motion of texture boundaries. *Vision Research, 33*, 1437–1446.
Yin, T., Kuwada, S. & Sujaku, Y. (1984). Interaural time sensitivity of high-frequency neurons in the inferior colliculus. *Journal of the Acoustic Society of America, 76*, 1401–1410.
Zanker, M. J. (1993). Theta motion: A paradoxical stimulus to explore higher-order motion extraction. *Vision Research, 33*, 553–569.

Acknowledgements—We are very grateful to Lief Haglund, David Heeger, and Allan Jepson for sharing their insights and providing comments on earlier drafts. DF is also grateful to David Heeger and Brian Wandell for their hospitality and facilities at Stanford University while this manuscript was being completed. This work has been funded in part by NSERC Canada, the IRIS Federal Centre of Excellence, and the British Council.

APPENDIX A

Fourier Analysis and Visual Motion

Here we derive some important results concerning the frequency analysis of visual motion. We first prove the well-known result that image translation is equivalent to power lying along a line through the origin in the frequency domain. We then introduce the notion of group velocity and show that the orientation of power in frequency domain gives the velocity of a coherently translating multiplicative amplitude envelope. The final section addresses the generalization of group velocity to the case where the oriented power distribution in frequency domain no longer lies strictly along a line.

A1. Coherent image translation

A signal translates with velocity v if and only if all its non-zero power lies along the line through the origin in the frequency domain given by $kv + \omega = 0$. To see this, note first that if all power lies along the line $kv + \omega = 0$ then (via Fourier analysis) the image is equivalent to a weighted sum of sinusoidal waves with spatial and temporal frequencies k_j and ω_j satisfying $k_j v + \omega_j = 0$. That is, all nonzero Fourier components travel with the same phase velocity $v = -\omega_j/k_j$ (see Section 2.1.), and hence the entire signal translates with no distortion with velocity v.

Conversely, assume that $I(x, t)$ is composed of a 1-D function $I_0(x)$ translating with velocity v

$$I(x, t) = I_0(x - vt). \tag{A1}$$

To show that its non-zero power lies on the line $kv + \omega = 0$, we take its Fourier transform, written $\hat{I}(k, \omega)$. Using the Fourier shift property (Papoulis, 1977), one can show

$$\begin{aligned}\hat{I}(k, \omega) &= \int\int I_0(x - vt)\mathrm{e}^{-i(\omega k + t\omega)}\mathrm{d}x\ \mathrm{d}t \\ &= \hat{I}_0(k) \int \mathrm{e}^{-itvk}\mathrm{e}^{-it\omega}\mathrm{d}t \\ &= \hat{I}_0(k)\delta(\omega + kv),\end{aligned} \tag{A2}$$

where $\hat{I}_0(k)$ is the Fourier transform of the 2-D intensity pattern $I_0(x)$. Here, $\delta(x)$ is a Dirac delta function satisfying $\delta(x) = 0$ when $x \neq 0$. Therefore, $\hat{I}(k, \omega)$ is only nonzero when $\omega = -kv$; i.e. on the line through the origin.

A2. Group velocity

For dispersive mediums, such as water, the velocity of a disturbance (or signal) is not simply given by the phase velocities of elementary (sinusoidal) signal components. Sinusoidal waves of different wavelengths travel with different phase velocities, and the velocity of the signal is given by the group velocity (Brillouin, 1960; Whitham, 1974). In effect, the group velocity is defined to represent the velocity of a modulating amplitude envelope, the motion of regions of significant amplitude.* Our interest in group velocity stems from its relation to amplitude velocity.

The relationship between wavelength and phase velocity in dispersive mediums is usually characterized by a dispersion relation $\omega(k)$, that maps spatial frequency to temporal frequency. The phase velocity associated with a spatial frequency k_0 remains $v_\mathrm{p}(k_0) = -\omega(k_0)/k_0$. The group velocity associated with a band of frequencies close to k_0 is defined by

$$v_\mathrm{g} = -\frac{\mathrm{d}\omega(k)}{\mathrm{d}k}\Big|_{k=k_0} = v_\mathrm{p} + k_0\frac{\mathrm{d}v_\mathrm{p}}{\mathrm{d}k}\Big|_{k=k_0}, \tag{A3}$$

*The motion of amplitude, given by group velocity has been used to approximate particle velocity in quantum mechanics, the velocity of energy transport, and the velocity of disturbance fronts (or signal velocity) (Brillouin, 1960). Group velocity as defined below provides an accurate model of the motion of a disturbance so long as the variation of phase velocity with wavelength is sufficiently smooth, and the medium is not dissipative.

that is, the negative derivative (slope) of the dispersion relation at k_0.

There are several ways to derive this relation and its meaning, one of which relies the notion of stationary phase (Papoulis, 1977). Below we take a somewhat different approach. Consider a dispersion relation $\omega(k)$ that is predominantly linear over a band of frequencies near k_0. We can then expand $\omega(k)$ as a Taylor series about k_0, with the first two terms providing a good approximation:

$$\omega(k) \approx \omega_0 + (k - k_0)\,\omega'(k_0), \tag{A4}$$

where

$$\omega_0 \equiv \omega(k_0), \quad \text{and} \quad \omega'(k_0) \equiv \frac{\mathrm{d}\omega(k)}{\mathrm{d}k}\Big|_{k=k_0}.$$

Furthermore, let $A(k)$ denote the distribution of power as a function of spatial frequency. With these preliminaries, we can write the signal as a collection of sinusoidal components with spatial frequencies near k_0, their associated amplitudes $A(k)$, and their corresponding temporal frequencies, $\omega(k)$:

$$I(x, t) = \int_{-\infty}^{\infty} A(k) \mathrm{e}^{i[kx + \omega(k)t]}\,\mathrm{d}k. \tag{A5}$$

To see that this signal has a group velocity v_g as defined above, substitute the approximate dispersion relation (41) into (42), yielding

$$I(x, t) \approx \int_{-\infty}^{\infty} A(k) \mathrm{e}^{i[kx + (\omega_0 + (k - k_0)\omega'(k_0))t]}\,\mathrm{d}k. \tag{A6}$$

With further algebraic manipulation this becomes

$$\begin{aligned} I(x, t) &\approx \mathrm{e}^{it[\omega_0 - k_0\omega'(k_0)]} \int_{-\infty}^{\infty} A(k) \mathrm{e}^{ik[x + \omega'(k_0)t]}\,\mathrm{d}k \\ &= I(x + \omega'(k_0)t, 0) \mathrm{e}^{it[\omega_0 - k_0\omega'(k_0)]}. \end{aligned} \tag{A7}$$

The last step follows from the Fourier shift theorem (Papoulis, 1977).

Finally, taking the magnitude of $I(x, t)$,

$$|I(x, t)| = |I(x + \omega'(k_0)t, 0)|, \tag{A8}$$

one can see that the amplitude component of $I(x, t)$ translates in an undistorted manner with velocity $v_g = -\omega'(k_0)$. Thus, as illustrated in Fig. 2, while phase velocity depends on the location of power in the frequency domain, group velocity reflects the orientation of power.

A3. Oriented power in frequency space

We now consider the case where power is oriented in the frequency domain, but not strictly on a line. For example, consider the following amplitude spectrum

$$A(\mathbf{k}) = A_1(\mathbf{u} \cdot \mathbf{k}), \tag{A9}$$

where, for convenience $\mathbf{k} \equiv (k, \omega)$, and $\mathbf{u} \equiv (u_k, u_\omega)$ is a unit direction vector (i.e. $\|\mathbf{u}\| = 1$) that is perpendicular to the level contours (i.e. the orientation) of the $A(\mathbf{k})$. Our primary concern is the case where the 1-D profile $A_1(k)$ has most of its power concentrated near some k_0.

In practice, when we consider the output of band-pass filters, the amplitude spectra will not extend infinity far in any direction. Locally we might therefore approximate the amplitude spectrum as the product of $A_1(k)$ and a window. Since $A_1(k)$ is generally assumed to concentrated in the direction $\mathbf{u}$ let us assume that the window limits the spectral extent in the perpendicular direction $\mathbf{u}^\perp = (-u_\omega, u_k)$, as drawn in Fig. A1:

$$A(\mathbf{k}) = A_1(\mathbf{u} \cdot \mathbf{k}) W(\mathbf{u}^\perp \cdot \mathbf{k}). \tag{A10}$$

Although this has a symmetric mathematical form, we will assume that $W(k)$ is broad relative to the elongation of the amplitude spectrum in the direction of $\mathbf{u}^\perp$. This means that the predominant orientation in frequency space remains perpendicular to $\mathbf{u}$.

With some mathematical manipulation [exploiting the separability of (A10)] it can be shown that the inverse Fourier transform of $A(\mathbf{k})$ is the product of two complex-valued signals:

$$I(x, t) = a_1(u_k x + u_\omega t)\, w(-x u_\omega + t u_k), \tag{A11}$$

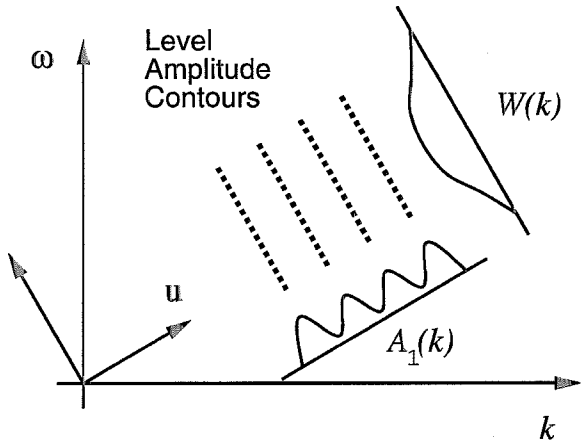

FIGURE A1. This depicts an amplitude spectrum which represents the product of two 1-D spectra. One has its level contours perpendicular to $\mathbf{u}$, where the profile of the spectrum in the direction $\mathbf{u}$ is called $A_1(k)$. In the case of pure group velocity, all non-zero power should be along a line in which case $A_1(k)$ would be an impulse. In practice, there will also be some variation of power along the level contours of $A(k)$, and therefore we introduce a slowly varying modulation in the direction $\mathbf{u}^\perp$, denoted by $W(k)$.

where $a_1(x)$ and $w(x)$ are the inverse Fourier transforms of $A_1(k)$ and $W(k)$. Here, (A11) implies that $a_1(x)$ drifts with velocity $-u_\omega/u_k$, while $w(x)$ drifts with velocity u_k/u_ω. Note that group velocity occurs as a special case of this, in which the 1-D spectral profile in the direction of $\mathbf{u}$, that is $A_1(k)$, is an impulse function. Group velocity is then given by u_k/u_ω.

In the case of ideal group velocity, where power is restricted to a line, $A_1(k)$ is an impulse function, and $a_1(x)$ would be a complex exponential. Because the magnitude of the complex exponential is unity, it follows that

$$|I(x, t)| = |w(-x u_\omega + t u_k)|, \tag{A12}$$

and therefore $|I(x, t)|$ can be used to measure the group velocity [which is the velocity of $w(x)$ as discussed above]. This shows again that amplitude provides information about group velocity in the ideal case. It also shows nicely that the structure of the amplitude spectrum of $W(k)$ along the orientation of the power concentration determines the nature of the envelope in question. Furthermore, when $W(k)$ is broad with its frequency components in phase, the envelope $w(x)$ will be narrow and localized (as in the Gabor example in Fig. 3, and in the drift-balanced stimuli described in Section 3.3). Alternatively, if phase varies wildly along $W(k)$, or if there are a small number of distinct frequencies spread along the lines of power then the window will not be localized (as in the case of the beats in Section 3.1.).

In a more general case, power may be concentrated in the neighbourhood of a line, rather than strictly on it. In this case, we might view $A_1(k)$ as a blurred impulse rather than an impulse, in which case the magnitude of $I(x, t)$ is

$$|I(x, t)| = |a_1(u_k x + u_\omega t)|\,|w(-x u_\omega + t u_k)|, \tag{A13}$$

where $|a_1(x)|$ is now slowly varying because its corresponding amplitude spectrum if highly concentrated near the line. In these cases, so long as $|w(x)|$ contains sufficiently high frequencies, its orientation will still dominate the orientation of $I(x, t)$. Again this justifies the application of band-pass oriented filters to the amplitude component of the response of an initial layer of filters, which will be somewhat insensitive to the slow variations in $a_1(x)$, as in the model of Wilson *et al.* (1992).

Finally, the more general case exists when neither $A_1(k)$ nor $W(k)$ is concentrated near a point, in which case power will not be concentrated about a single line. This situation might exist for example with the sampled stimuli illustrated in Fig. 7 when the replications occur too close together relative to the filter tuning. In this case there would be a fundamental ambiguity since amplitude is the product of two signals with different velocities. Moreover, methods like those in this paper

and those of Chubb and Sperling (1988), and Wilson *et al.* (1992) would be inappropriate; they are not even guaranteed to yield one of the two component velocities. We are currently considering extensions to the model of group velocity that allows for the extraction of multiple velocities from the amplitude component of the filter output that would allow this case to be handled adequately. Whether biological systems have this capability is unclear.

APPENDIX B

Fourier Analysis of Sampled Motion

Our goal here is to derive the form of the Fourier transform of sampled motion stimuli, in which a signal is shifted by a certain distance every Δt frames, and is otherwise held constant between shifts. We express the stimulus as a discretely sampled version of a smoothly translating signal, the sampled version of which is blurred by a temporal pill-box function to make it static between samples. The effective displacement d can be expressed as a product of the continuous velocity v and the sampling interval Δt (between displacements).

Let $f(x, t) = f_0(x - vt)$ denote the continuously translating 1-D signal with velocity v, and imagine that it is sampled every Δt in time and Δx in space. That is, we define a discrete signal $h(x, t) = f(x, t)s(x, t)$, where $s(x, t)$ denotes a sampling function which may be written as a sum of Dirac delta functions:

$$s(x, t) = \sum_m \sum_n \delta(x - m\Delta x, t - n\Delta t). \tag{B1}$$

As discussed in Appendix A1, the Fourier transform of $f(x, t)$ is

$$\hat{f}(k, \omega) = \hat{f}_0(k)\delta(vk + \omega). \tag{B2}$$

The Fourier transform of $s(x, t)$ is another sampling function (Rosenfeld & Kak, 1976),

$$\hat{s}(k, \omega) = \Delta k \Delta\omega \sum_m \sum_n \delta(k - m\Delta k, \omega - n\Delta\omega), \tag{B3}$$

where $\Delta k = 2\pi/\Delta x$ and $\Delta\omega = 2\pi/\Delta t$. Thus the Fourier transform of the discrete signal $h(x, t)$ is

$$\begin{aligned}\hat{h}(k, \omega) &= \hat{f}(k, \omega) * \hat{s}(k, \omega) \\ &= [\hat{f}_0(k)\delta(vk + \omega)] * \sum_m \sum_n \delta(k - m\Delta k, \omega - n\Delta\omega).\end{aligned} \tag{B4}$$

The convolution of $\hat{f}_0(k)\delta(vk + \omega)$ with a single impulse, $\delta(k - m\Delta k, \omega - n\Delta\omega)$, amounts to a translation of the function to the location of the impulse $(m\Delta k, n\Delta\omega)$ in frequency domain. Thus (B4) represents the amplitude spectrum replicated periodically at intervals of Δk and $\Delta\omega$. Moreover, note that the orientation of each replication (i.e. each group velocity) is consistent with v.

For the stimuli of interest here, we assume that there is a moderate amount of temporal aliasing and very little spatial aliasing relative to the range of spatial and temporal frequencies to which the visual system is sensitive [cf. the window of visibility (Watson *et al.*, 1986)]. In other words, let $\Delta t \gg \Delta x$, in which case we can simplify (B4) to

$$\hat{h}(k, \omega) = [\hat{f}_0(k)\delta(vk + \omega)] * \sum_n \delta(\omega - n\Delta\omega). \tag{B5}$$

This situation is depicted in Fig. B1.

Finally, to ensure that the signal remains constant between samples we convolve the discrete signal with a constant averaging window $c(t)$ defined by

$$c(t) = \begin{cases} \dfrac{1}{\Delta t} & \text{for } |t| \leq \Delta t/2 \\ 0 & \text{otherwise.} \end{cases} \tag{B6}$$

The resulting stimulus may be written as

$$I(x, t) = c(t) * h(x, t) = c(t) * [f(x, t)s(x, t)], \tag{B7}$$

where $*$ is the usual convolution operator. The Fourier transform of $c(t)$ is given by

$$\hat{c}(\omega) = \frac{\sin(\omega\Delta t)}{\omega\Delta t}, \tag{B8}$$

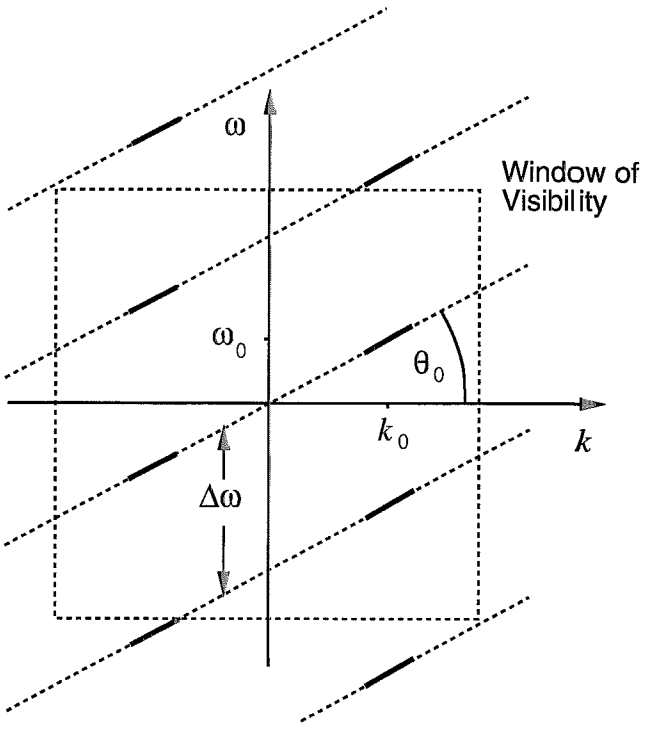

FIGURE B1. The amplitude spectrum includes the original line through the origin, plus its replications that occur because of temporal sampling. The small bold line segments indicate the regions of non-zero power that lie along the dotted lines. The important parameters include the velocity $v = -\tan(\theta_0)$, the spatial frequencies k_0, and the replication distance $\Delta\omega = 2\pi/\Delta t$. The window of visibility, corresponding to the range of human sensitivity, shows that multiple line segments of non-zero power should be visible in this case.

in which case the Fourier transform of $I(x, t)$ can be written

$$\begin{aligned}\hat{I}(k, \omega) &= \hat{c}(\omega)\hat{h}(k, \omega) \\ &= \frac{\sin(\omega\Delta t)}{\omega\Delta t}\left([\hat{f}_0(k)\delta(vk + \omega)] * \sum_n \delta(\omega - n\Delta\omega)\right).\end{aligned} \tag{B9}$$

In other words, the blurring acts as a low-pass filter, attenuating high temporal frequencies. The larger the sampling interval, Δt, the greater the attenuation of high frequencies. Therefore, despite the underlying continuous motion, the line segment in the Fourier spectrum with the largest energy is the one closest to the spatial frequency axis, with the smallest temporal frequencies. In the example given in Fig. 7B, the concentration with the maximum energy has leftward moving phase velocities, but a rightward moving group velocity, consistent with the original velocity v.

APPENDIX C

Analysis of Rightward-Stepping Random-Contrast Bar

Like the sampled motion stimuli above, we view the rightward stepping bar as a sampled version of a continuously drifting amplitude envelope, the samples of which are convolved with a spatiotemporal pill-box to create the square spatiotemporal regions seen in Fig. 8. Its Fourier transform can be described in terms of the direction of the amplitude envelope, the 1-D intensity pattern in this direction, and the squares which represent the sampling and pill-box blurring.

In the ideal case, the amplitude envelope is an impulse function moving rightward $\delta(x - vt)$ which modulates an arbitrary intensity pattern that varies in the space–time direction of the impulse envelope $f(vx + t)$, the continuous motion is given by

$$I(x, t) = f(vx + t)\delta(x - vt). \tag{C1}$$

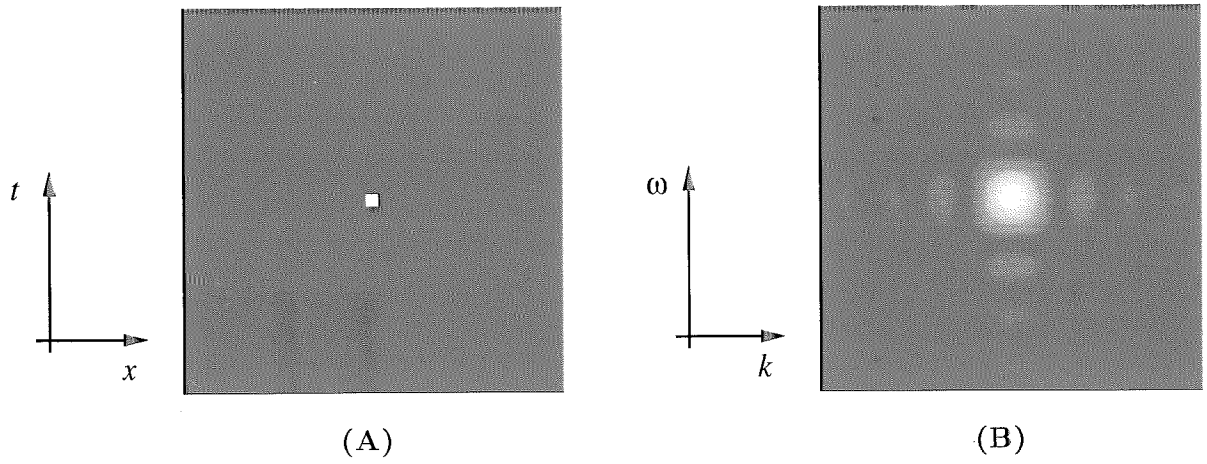

FIGURE C1. This shows the amplitude spectrum of a simple 8 pixel wide square. It illustrates the general pattern of power that occurs on average over the ensemble of rightward-drifting contrast-reversing bars.

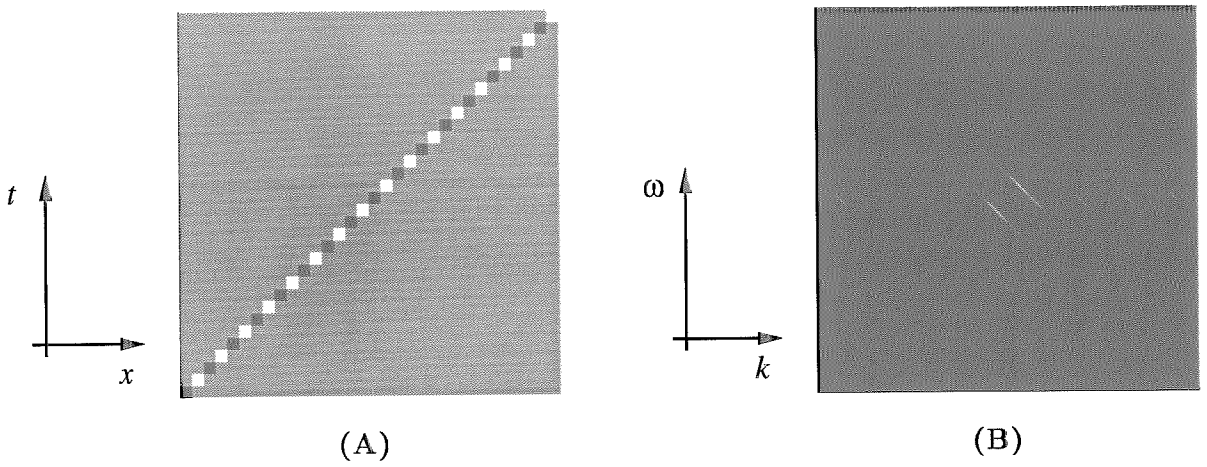

FIGURE C2. If we simplify the rightward contrast-reversing bar so that all local regions exhibit similar scales, the Fourier spectrum provides a clearer indication of the dominant distribution of power that is relevant to local operators.

With some algebraic manipulation, the Fourier transform of $I(x, t)$ can be shown to be

$$\hat{I}(k, \omega) = \frac{1}{1+v^2}\,\hat{f}(kv + \omega). \tag{C2}$$

Thus the level contours of the amplitude spectrum will be aligned in the same direction, given by

$$kv + \omega = c. \tag{C3}$$

where c is a real-valued scalar. Whenever $f(x)$ has some form of peak spectral concentration, the amplitude spectrum $|\hat{I}(k, \omega)|$ will be strongly oriented in frequency space, with an orientation consistent with velocity v.

For the rightward stepping bar considered here, $f(x)$ is a noise sample. On average, or for a sufficiently large sample, the spectrum will be flat. However, for smaller samples such as those in local windows, and even for that shown in Fig. 8, there will generally be spectral peaks. This helps to emphasize that although there are peak spectral concentrations locally, they may not be visible globally through the Fourier transform. In their sample stimuli, Chubb and Sperling used relatively large squares with a small number of time steps, which will exhibit peak concentrations at low frequencies.

When we take the pill-box blurring into account, the amplitude spectrum of the sampled signal is effectively multiplied by the amplitude spectrum of the pill-box. Figure C1 shows a space–time pill-box alone and its amplitude spectrum. The pill-box spectrum determines the locations of significant power in frequency space as shown in Fig. 8B. The orientation of the streaks of power in Fig. 8B is determined by the envelope orientation, and the profile of the spectrum perpendicular to this orientation is given by the spectrum intensity profile along the path of the envelope.

There are several ways to consider idealizations of these stimuli that retain their essential character, yet clearly show the appropriate behaviour in the frequency domain. For example, a global idealization of these stimuli should have the same scale everywhere, e.g. a periodic narrow-band pattern, and the envelope should be smoothly drifting in space–time instead of the jumping squares. Figure C2 shows one step toward an idealized version of the rightward drifting bar in which the noise sample is replaced with a periodic pattern. In this way the pattern in stationary in terms of distribution of local scales. Notice that this case also contains no repeated luminance values which would locally stimulate a Fourier-based mechanism as shown in Fig. 9. A further idealization might involve a smooth envelope (rather than a discrete sequences of squares), and a sinusoidal fine structure, which reduces to something like the signal in Fig. 4.