Vision Res. Vol. 33, No. 4, pp. 463–485, 1993
Printed in Great Britain.

0042-6989/93 $6.00 + 0.00

The Dimensionality of Texture-Defined Motion: a Single Channel Theory

PETER WERKHOVEN,*† GEORGE SPERLING,* CHARLES CHUBB*‡

Received 22 August 1991; in revised form 3 August 1992

We examine apparent motion carried by textural properties. The texture stimuli consist of a sequence of grating patches of various spatial frequencies and amplitudes. Phases are randomized between frames to insure that first-order motion mechanisms *directly* applied to stimulus *luminance* are not systematically engaged. We use ambiguous apparent motion displays in which a heterogeneous motion path defined by alternating patches of texture s (standard) and texture v (variable) competes with a homogeneous motion path defined solely by patches of texture s. Our results support a one-dimensional (single-channel) model of motion-from-texture in which motion strength is computed from a single spatial transformation of the stimulus—an *activity* transformation. The value assigned to a point in space–time by this activity transformation is directly proportional to the modulation amplitude of the local texture and inversely proportional to local spatial frequency (within the range of spatial frequencies examined). The activity transformation is modeled as the rectified output of a low-pass spatial filter applied to stimulus *contrast*. Our data further suggest that the *strength* of texture-defined motion between a patch of texture s and a patch of texture v is proportional to the product of the activities of s and v. A strongly counterintuitive prediction of this model borne out in our data is that motion between patches of different texture can be stronger than motion between patches of similar texture (e.g. motion between patches of a low contrast, low frequency texture l and patches of high contrast, high frequency texture h can be stronger than motion between patches of similar texture h).

Second-order motion Motion metamers Motion energy Motion correspondence

INTRODUCTION

First-order motion extraction

Drifting spatiotemporal modulations of various sorts of optical stuff (such as luminance, contrast, texture, binocular disparity, etc.) can induce vivid motion percepts; in each case "something" appears to move from one place to another. This introspective description, however, does not necessarily reflect the underlying processes in human visual motion processing.

The study of visual motion extraction mechanisms has traditionally focused on rigidly moving objects, projecting drifting modulations of *luminance*. Several physiologically plausible computational models have been proposed to extract motion information from drifting luminance modulations. Examples are the gradient detector (see Moulden & Begg, 1986) and the Reichardt or correlator detector (see Reichardt, 1961). These detectors are designed to detect drifting luminance modulations (or their linear transformations) and are therefore called *first-order* motion extraction mechanisms (Cavanagh & Mather, 1989)

Psychophysical experiments (e.g. van Santen & Sperling, 1984; Werkhoven, Snippe & Koenderink, 1990b) have shown that motion perception of drifting modulations of luminance is well explained by a first-order computation called *motion energy extraction*. Indeed, most current models of first-order motion detection (e.g. Reichardt detectors and gradient detectors) have now been shown to be equivalent or approximately equivalent to some variant of motion energy extraction (Adelson & Bergen, 1986; van Santen & Sperling, 1985). A standard approach to first-order motion energy extraction (e.g. Heeger, 1987; Adelson & Bergen, 1985) proposes that the visual system uses a battery of spatiotemporally oriented filters, each of which yields a real-valued function of the visual field over time. The output of each filter is squared at each location in space to obtain a measure of local energy at the spatiotemporal frequency to which that filter is tuned. The squared outputs of these filters (motion energies) comprise the input to a higher order process that computes a velocity flow field. For example, Heeger's (1987) model is built on the observation that the Fourier transform of a rigidly translating pattern has all its energy contained in a plane through the origin in frequency space. Each motion energy detector (narrow-band, spatiotemporal linear

*Department of Psychology and Center for Neural Science, New York University, New York, NY 10003, U.S.A.
†Present address: Utrecht Biophysics Research Institute (UBI), Buys Ballot Laboratory, Utrecht University, Princetonplein 5, 3584 CC, Utrecht, The Netherlands.
‡Present address: Department of Psychology, Rutgers University, New Brunswick, NJ 08903, U.S.A.

filter followed by squaring) has its energy confined to a Gaussian neighborhood of frequency space near the origin. The velocity vector assigned a given point in space at a given time is obtained by (i) weighting the energy spectrum of each detector by that detector's response, and (ii) finding the plane through the origin of frequency space that absorbs the greatest amount of this locally measured motion energy.

Second-order motion extraction

Chubb and Sperling (1988, 1989a, b, 1991) demonstrated broad classes of *drift-balanced* and *microbalanced* stimuli that clearly appeared to move but for which even complete knowledge of the energy of all their Fourier components would be useless in deciding whether their motion was to the left or to the right (see also Cavanagh, Arguin & von Grünau, 1989; Lelkens & Koenderink, 1984; Mather, 1991; Ramachandran, Rau & Vidyasagar, 1973; Turano & Pantle, 1989; Victor & Conte, 1990). Thus first-order motion energy extraction fails completely to account for the perception of motion in drift-balanced stimuli. Such stimuli are said to elicit *second-order motion perception* (Cavanagh & Mather, 1989; Chubb & Sperling, 1988). In second-order motion stimuli, what drifts is not a luminance modulation but modulation of contrast, or spatial frequency, texture type, flicker, or some other stimulus property.

Stages. Let L be the spatiotemporal luminance function defining a stimulus. The luminance at point (x, y) at time t is then denoted $L(x, y, t)$. In our analysis, we discriminate three stages for the extraction of motion information from L: preprocessing; flow field extraction; and decision.

First, a preprocessing stage in which one or more transformations T_i are applied to L yielding a set of real-valued, time-varying, "neural images" $T_i(L)$ (Robson, 1980). The value at point (x, y) at time t that results from applying T_i to L is thus denoted $T_i(L)(x, y, t)$. Usually, we think of i as referring to the dominant spatial frequency of a transformation—its scale.

Second, each time-varying, neural image $T_i(L)$ is the input to a motion-analysis stage $\vec{V}_i$ whose output is a (time-varying) velocity flow field $\vec{V}_i \bigcirc T_i(L) = \vec{V}_i[T_i(L)]$. For any point (x, y) in the visual field and every time t, the value $\vec{V}_i \bigcirc T_i(L)(x, y, t)$ is a two-dimensional vector that indicates estimated pattern velocity of the transformed image $T_i(L)$ in the neighborhood of (x, y) at time t. The scale of $\vec{V}_i$ corresponds to T_i. Associated with $\vec{V}_i \bigcirc T_i(L)$ is a real-valued function $S_i(L)$ that gauges the reliability or strength of the velocity estimate provided by $\vec{V}_i \bigcirc T_i(L)$. For instance, the velocity estimate obtained at point (x, y) and time t may have been computed from sparse or noisy data. In this case, irrespective of estimated direction or speed, the strength $S_i(L)(x, y, t)$ of the estimated velocity $\vec{V}_i \bigcirc T_i(L)(x, y, t)$ may be low.

Finally, all the velocity flow fields $\vec{V}_i \bigcirc T_i(L)$ and their associated strength maps $S_i(L)$ feed into a decision mechanism; its output determines the direction of apparent motion in ambiguous displays.

The preprocessing transformation T_i can be either linear or nonlinear. Generalizing previous terminology, we say that any system that employs linear preprocessing performs first-order motion extraction, whereas nonlinear preprocessing performs second-order motion extraction (e.g. Cavanagh *et al.*, 1989; Chubb & Sperling, 1988).

We refer to the transformations $\vec{V}_i \bigcirc T_i$ as motion channels. T_i is called the initial transformation and $\vec{V}_i$ the motion extractor. S_i is called the strength measure of the channel.

Motion-energy detection vs motion-correspondence detection. Both first- and second-order motion channels can be further classified by the type of motion extraction they use. A review of the literature on motion perception shows that two types of motion extractor have been considered and tested experimentally. We call these types of motion extraction motion energy extraction and motion correspondence extraction.

Motion *energy* extraction computes the directional energy of a Fourier representation of the drifting modulation signal, that is, the relative energy of "drifting" spectral components. Within the constraints set by frequency resolution, energy extraction is independent of the relative phase of the different spatial Fourier components of the modulation signal (van Santen & Sperling, 1984). In this respect, motion energy extraction computations are largely insensitive to similarities between items in a motion path. The first-order motion analysis models noted above (Reichardt, 1961; Adelson & Bergen, 1985; Marr & Ullman, 1981) all share this property.

Traditionally, however, psychophysicists have interpreted results of a wide range of motion experiments in terms of correspondence extraction. The metaphor of correspondence extraction describes motion as the convection of some invariant aspects of spatial structure over time. Thus, motion correspondence extraction depends on similarity of local features. The more nearly similar are two adjacent features that are separated by an interval in time, the greater will be the strength of motion between them.

The distinction between motion energy extraction and motion correspondence extraction can be summarized as follows: let α and β be two points separated by a brief interval in space and time, and let v_α and v_β be the stimulus intensities at α and β. Then motion energy extraction yields a motion strength that is a monotonically increasing function of the product $v_\alpha v_\beta$. Motion correspondence extraction yields a motion strength between α and β that is a decreasing, nonnegative function of $|v_\alpha - v_\beta|$.

Typically, motion channels using correspondence extraction yield higher motion strengths between similar textures than between dissimilar textures. In particular, a motion channel using a correspondence extractor can never yield motion strength between a patch of optical stuff A and a patch of different stuff B that is greater than the motion strength between two patches of stuff A. This can easily happen, however, for motion channels

using energy extractors. Suppose, for instance, that $v_A > v_B$, for v_A and v_B the respective values assigned stuff A and stuff B by a channel's initial transformation. Then, motion energy extraction yields greater strength of motion between a patch of A and a patch of B ($v_A v_B$) than between two patches of B ($v_B v_B$).

Motion-from-texture

The purpose of this paper is to characterize the mechanism of second-order motion perception in the subclass of drift-balanced stimuli for which motion is defined by a modulation of spatial texture properties. To reiterate, it is not produced by a moving texture patch—that would be rigid, luminance-defined motion. Texture-defined motion is most conveniently produced by a moving patch that is filled with a particular type of texture in which each successive frame represents a new, uncorrelated instance of that texture type (Chubb & Sperling, 1989a, 1991). As is true for all drift-balanced motion stimuli, an intriguing aspect of texture-defined motion perception is that (unlike perception of luminance defined or first-order motion) it cannot be explained by Fourier energy or autocorrelational motion analysis (standard motion analysis).

An early example of texture-defined motion was reported by Ramachandran *et al.* (1973). Detailed studies and analysis were recently presented by Chubb and Sperling (1988, 1989a, b, 1991), Cavanagh *et al.* (1989), Mather (1991), Turano and Pantle (1989), and Victor and Conte (1989).

We construct stimuli for which energy and correspondence mechanisms yield different predictions for the strength of texture-defined motion (Werkhoven *et al.*, 1990b). The resulting data demonstrate that texture-defined motion is computed by an energy mechanism, and not a correspondence mechanism. And we will show how psychophysical data can be used to discriminate between these two sorts of mechanisms in human perception of texture-defined motion. More importantly, these data indicate clearly that, for the class of textures we use (similarly oriented patches of random-phased sinusoidal grating with different spatial frequencies and contrasts), texture-defined motion perception can be modeled in terms of a single motion energy channel.

Energy channels

Texture grabbers. Chubb and Sperling (1989a, 1991) suggested a two-stage mechanism for extracting texture-defined motion. Under their model, texture-defined motion is computed by motion energy channels whose initial transformations are called texture grabbers. As discussed below (see Rectification), a texture grabber is a linear spatial filter followed by rectification. In Stage 2, the time-varying output (activity) from each texture grabber is subjected to motion energy extraction.

Rectification. By rectification we mean any function that is zero for an input of zero, and is monotonically increasing for both positive and for negative real inputs.

Previously, Chubb and Sperling (1989b) demonstrated stimuli displaying systematic second-order motion that could be easily explained in terms of a texture grabber that used fullwave rectification (e.g. absolute value, square, etc.). However, the motion of these stimuli was inaccessible to any mechanism whose texture grabber used halfwave rectification (nonzero output only for positive or only for negative inputs). These results suggest that at least some of the texture grabbers used in second-order motion perception use fullwave rectification. It remains to be seen whether there are second-order motion mechanisms that use halfwave rectification. In the present context, however, we do not distinguish between different kinds of rectification. The essential nonlinear characteristic of texture extraction processes has also been recognized by Bergen and Adelson (1988) and Caelli (1985).

The linear filter used by a texture grabber is presumed to be realized in the visual system by an array of linear neurons, all with the same receptive field profile, distributed across the visual field. The texture grabber output results from applying some fixed, rectifying nonlinearity (e.g. the absolute value or the square) to the output of each of these linear neurons. It is assumed that the spatial filter of Stage 1 operates on stimulus contrast (see Model), rather than on luminance, but this assumption is not critical to our arguments. The output of a linear filter may be positive or negative depending on the local phase of the sensed texture. Thus the expectation of the output of such a filter is zero over the phase-randomized texture patches from which our stimuli are constructed. The purpose of rectification is to produce a positive average output across the texture so that a texture grabber registers the presence or absence of texture, independent of local phase. Indeed, that is why the Stage-1 transformation (linear spatiotemporal filter followed by rectification) is called a texture grabber.

Activity. The output of a texture grabber in response to a particular texture is called activity.

Motion energy-channels. Together, a texture grabber followed by motion energy extraction form one (texture-defined motion) energy channel.

Motion correspondence-channels. Together, a texture grabber followed by motion correspondence extraction form one (texture-defined motion) correspondence channel.

Previous research in texture-defined motion

Historically, motion correspondence has been investigated with ambiguous motion displays in which motion is perceived as occurring along one or the other of several competing paths. Most studies have dealt with stimuli that stimulated the first-order motion system (e.g. Burt & Sperling, 1981; Kolers, 1972; Navon, 1976; Papathomas, Gorea & Julesz, 1991; Shechter, Hochstein & Hillman, 1989; Ullman, 1980; Werkhoven, Snippe & Koenderink, 1990a; Werkhoven *et al.*, 1990b) and these data are adequately explained by the first-order motion energy extraction models.

We consider here two recent studies that attempt to deal with motion correspondence in texture-defined motion stimuli. These studies illustrate the difficult

methodological issues that arise in attempting to determine motion correspondence, and thereby they indicate the necessity of the more complex paradigm which we use.

Watson's crossed-phi procedure. Watson (1986) attempted to measure the spatial frequency specificity of the perceptual mechanism responsible for texture-defined motion. He used a "crossed phi" method, in which two adjacent texture patches (A and B) in frame 1 exchanged positions in frame 2. The patches were Gaussian-windowed sine waves (Gabor patches). Observers reliably perceived apparent motion between the locations when A and B were different spatial frequencies. No apparent motion was reported when the patches were of similar spatial frequencies. Watson interpreted his results in terms of a model in which motion estimates are computed separately within different spatial frequency bands. He used the increasing probability of apparent motion with increasing differences in spatial frequency to estimate the spatial frequency selectivity of the motion channels. Furthermore, it was implicitly assumed that such a model was equivalent to a correspondence computation.

In our view, the ambiguous "crossed-phi" paradigm admits a simple alternative interpretation in terms of single energy channel model. Suppose there were just a single energy channel, and suppose that texture A happened to produce a bigger response from the texture grabber in this channel than texture B. Then, the change in position of patch A would produce a strong motion response in this channel; the change of position of patch B would produce a weak motion response in the opposite direction; net movement would be perceived in the A direction. The critical observation for a multichannel model is motion transparency—that motion of the A and B patches be seen simultaneously in opposite directions. Only then can we be sure that more than one channel is activated. In fact, such motion transparency was not reported by Watson, and, in our experience, it does not occur in such stimuli. Thus, Watson's experiment does not support a theory of multiple correspondence channels.

Green's Gabor patches. Green (1986) studied texture-defined motion with a rotating annular display similar to Navon (1976). The type of stimulus used by Green is schematized in Fig. 1. Call this stimulus *I*. One temporal period of *I* consists of four frames, as shown in Fig. 1.

Each of these frames is comprised of a circle of alternating patches of two types of texture, texture A and texture B. From frame to frame, these patches of texture take rotary steps clockwise around the circle. This rotary clockwise motion is equivalent to left-to-right motion in an analogous horizontal display, as indicated by the dotted lines connecting annular frames to horizontal frames.

Let T be an *arbitrary* texture grabber, and suppose that v_A is the average response of T to texture A and v_B is the average response of T to texture B. Then the output from texture grabber T in response to stimulus *I* is a spatiotemporal function whose average value over any patch containing texture A is v_A and whose average value over any patch containing texture B is v_B. Although there will certainly be variability to the T-output within a given texture patch, this intra-patch variability is not critical to the global motion percept elicited by *I*. What determines this global motion percept are the average T-output values, v_A and v_B, of patches of the two textures A and B.

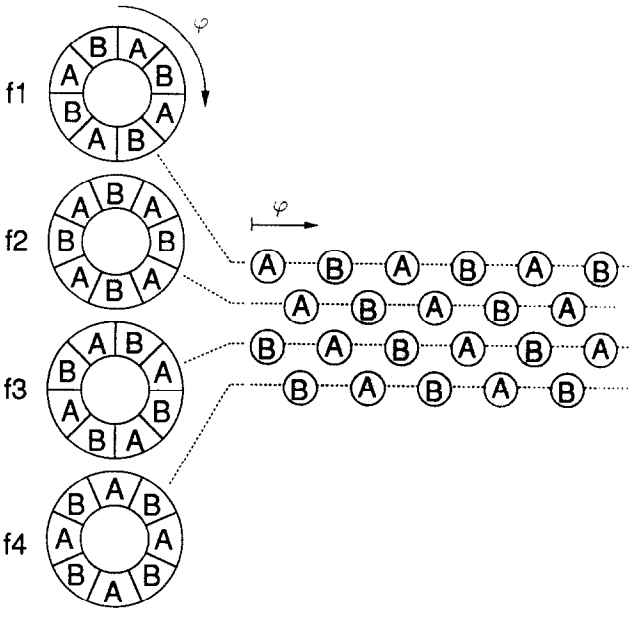

FIGURE 1. Green's stimulus, *I*. One temporal period of *I* consists of four frames. Each of these frames is comprised of a circle of alternating patches of two types of texture, texture A and texture B. From frame to frame, these patches of texture take rotary steps clockwise around the circle. This rotary clockwise motion is equivalent to left-to-right motion in an analogous horizontal display, as indicated by the dotted lines connecting annular frames to horizontal frames.

As many authors have observed (e.g. Adelson & Bergen, 1985; van Santen & Sperling, 1985), motion detection can be viewed as the detection of orientation in space-time. As is clear from inspection of Fig. 1(a), any motion detection mechanism that adheres to this general principle is bound to register clockwise motion in response to *I* whenever $v_A \neq v_B$.

In light of these observations, it is not surprising that observers in Green's experiment tended to perceive clockwise motion in displays such as *I*. In a critical sense, the clockwise motion of *I* is intrinsic to the format of the stimulus, and has little to do with the textures A and B comprising the patches of *I* (see Werkhoven *et al.*, 1990b). Nonetheless, Green took his results as support for the view that similar textures tend to match with each other in generating motion-from-texture.

Motion metamers

A psychophysical equivalence relation on a set Ω of physical stimuli is called a metamerism. Equivalent elements A and B of Ω are called metamers. Typically, metamerisms are defined using discrimination tasks. For example, if A and B are two illuminated patches that differ in spectral composition, we say they are metamers if an observer cannot distinguish between them.

In this paper, we focus on a different sort of metamerism that we call motion metamerism. Let Ω represent a set of texture patches that vary in spatial frequency, orientation, and contrast. The relation that we wish to capture is the following: for any two textures A and B in Ω, we call A and B motion metamers if and only if any occurrence of A in any dynamic visual display can be replaced by a patch of B without influencing the global motion percept elicited by that display. That is, A and B are motion metamers if and only if A and B are equivalent inputs to the mechanism that computes texture-defined motion. Obviously, A and B need not be equivalent inputs for other perceptual processes—as we shall show, motion metamers may appear quite different.

It is impractical to interchange A and B in all possible motion stimuli to verify that they are motion metamers. Instead, we use only two extreme test stimuli, in which any failure of metamerism would be most likely to appear. The essential core of the test we use is defined in terms of the stimuli I_1 and I_2 diagrammed in Fig. 2.

Each of these two stimuli pits two symmetrically opposite motion paths against each other. Stimulus I_1 pits a path comprised of a patch of texture A and a patch of texture B against a path comprised of two patches of texture A, whereas stimulus I_2 pits a path comprised of a patch of texture B and a patch of texture A against a path comprised of two patches of texture B. We presume that each of these paths has an associated motion strength, and that the global motion percept (left vs right) elicited by one of these stimuli depends only on which of its two paths has greater motion strength. In the case in which the global motion percept is ambiguous we assume that the strengths of the two component paths are equal.

For any textures A and B in Ω, we say A and B are transition invariant* if and only if the leftward vs rightward motion of each of I_1 and I_2 diagrammed in Fig. 2 is ambiguous (i.e. if each of I_1 and I_2 is equally likely to elicit a global rightward or leftward motion percept).

If textures A and B are transition invariant, then the motion strength of a match between A and A is equal to the motion strength of a match between A and B, and the motion strength of a match between B and A is equal to the motion strength of a match between B and B.

If A and B are motion metamers, then stimuli I_1 and I_2 are ambiguous in motion content; hence, A and B are transition invariant.

On the other hand, for practically all plausible texture-defined motion computations, if A and B are transition invariant, then they are also motion metamers. Indeed, the data we present make it clear that this is true of the computation that is actually used to compute texture-defined motion.

*The reason for this term will be clear in Transition Invariance and Motion Metamers.

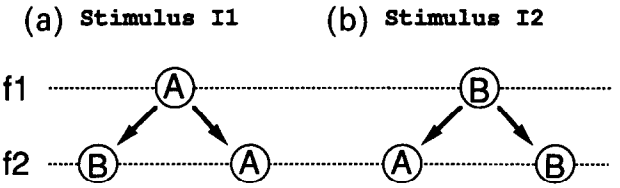

FIGURE 2. The binary relation ~ (transition invariance). (a) A schematic diagram of stimulus I_1. I_1 contains two frames. In the first frame there is a single patch of texture of type A. In the second frame, there are two patches of texture, one of type B and another of type A. These patches of texture are offset equal distances to the right and left of the location in frame 1 of the single patch of texture A. The stimulus I_1 sets up a competition between one motion path containing a patch of texture A and a patch of texture B and another, opposite motion path containing two patches of texture A. (b) A schematic diagram of stimulus I_2. For any textures A and B, we set A ~ B just if the stimuli I_1 and I_2 diagrammed in (a) and (b) respectively are both ambiguous in global motion content. That is, both stimuli I_1 and I_2 are equally likely to elicit global percepts of rightward or leftward motion. Any textures A and B for which A ~ B are said to be transition invariant. For a broad range of motion computations, it can be shown that, for any textures A and B, if A ~ B, then A and B are motion metamers in the strong sense (A and B can be freely traded for each other in any stimulus without changing the global motion percept elicited by that stimulus).

Motion competition schemes

The matching technique could be applied to a variety of ambiguous motion schemes for determining the dimensionality of the motion computation. However, not all of them have the power to discriminate between different types of motion channels (see e.g. the discussion on Green's display). We used an ambiguous motion scheme that was introduced by Werkhoven *et al.* (1990b). In this motion competition scheme, *one* heterogeneous motion path (between patches of texture s and texture v) competes *directly* with *one* homogeneous path (between patches of texture s).

By varying the properties of the textures v, we can determine the heterogeneous motion paths s, v that are equal in strength to a certain homogeneous path s, s.

Werkhoven *et al.*'s competition scheme not only allows to determine the dimensionality of the motion computation, but also allows to determine the number and type (energy vs correspondence) of channels involved in the motion computation. This requires a thorough analysis (given in the Model section).

However, an intuitively clear property of this scheme is that the two types of motion channels considered above (energy vs correspondence-channels) yield qualitatively different predictions for motion metamery and the relative strength of the heterogeneous and homogeneous motion paths. Hence, they are easily discriminated.

A preview

Dimensionality of the computation. In this paper, we discuss a general motion computation consisting of multiple motion channels, where each channel may be either an energy channel or a correspondence channel. By studying the above competition scheme with many

different pairs of texture patches (Expts 1 and 2), we can determine classes of transition invariant textures (motion metamers) and infer the dimensionality of the motion computation (Model section). The results strongly support the view that texture-defined motion is computed by a single energy channel.

METHOD

In this section we describe the ambiguous motion competition scheme used in the experiments. This scheme (proposed by Werkhoven *et al.*, 1990b) differs from other schemes (e.g. Burt & Sperling, 1981; Green, 1986; Navon, 1976; Shechter *et al.*, 1989; Ullman, 1980) in that it contains a single heterogeneous motion path (between patches of texture 1 and texture 2) that competes directly with a single homogeneous motion path (between identical patches of texture 2). Except for textural properties, the other parameters (such as step size and frame rate) of the motion paths are identical.

Instead of varying both textures 1 and 2, we sampled a subspace of possible textures resulting in two (similar) schemes: Scheme I and Scheme II. In Scheme I, we kept texture 2 constant (now texture s) and varied texture 2 (now texture v).

Stimulus

Motion competition Scheme I. In Expt 1, we used motion competition Scheme I. The motion stimulus consisted of a series of eight frames $(f_1, f_2, \ldots, f_8)$ shown successively in time. Figure 3 shows a sketch of the frames.

The first frame (f_1) contains an annulus of patches of alternating texture types s and v at regular positions (see Fig. 3, at the left side). Because the viewing distance was constant throughout the experiment, we will specify dimensions in degrees of visual angle. The annulus of texture patches has an inner radius of $r_1 = 1.04$ deg, and an outer radius of $r_2 = 2.08$ deg. The mean radius r is 1.56 deg. The patches (or *sectors*) are spatially contiguous. Since the annulus contains eight sectors, each sector has a width of 45 deg.

Frame f_2 was similar to frame f_1, except that patches of texture v are replaced by a uniform patch of background luminance. Furthermore, f_2 was rotated around the center of the annulus 22.5 deg with respect to frame 1 (see Fig. 3, left).

In a sequence of frames, the locations and types of patches in frame f_{n+2} were identical to frame f_n, except for a rotation around fixation of 45 deg.

The presentation time of a single frame ("frame-time") was 133.3 msec. Thus, the presentation time of the eight-frame sequence was 1.066 sec. The annulus revolved at an angular speed of 168.8 deg/sec, yielding a local velocity of the patch-centers of 4.6 deg of visual angle per second.

The ambiguous motion stimulus described above contains two motion paths. This can be understood most easily using a diagram in which we show the angular positions (φ) of the patches of texture for successive frames. Angular position is measured clockwise relative to the vertical. Such a diagram is shown in Fig. 3, at the right side. Note that the horizontal rows of patches correspond to frames 1, 2, 3 and 4 respectively. By definition, motion extraction is based on the dynamic properties of the stimulus, that is the spatiotemporal pattern of textures. In the diagram, possible motion paths are spatiotemporal (oblique) rows of elements. The arrows pointing to the left and right are examples of motion paths to the left and right respectively. In the following description of the stimulus, we will say that the neighboring elements in a motion path are spatiotemporally linked or "matched". Note that the term "matching" is used for the purpose of stimulus description only and that it does *not* refer to a "motion correspondence" computation.

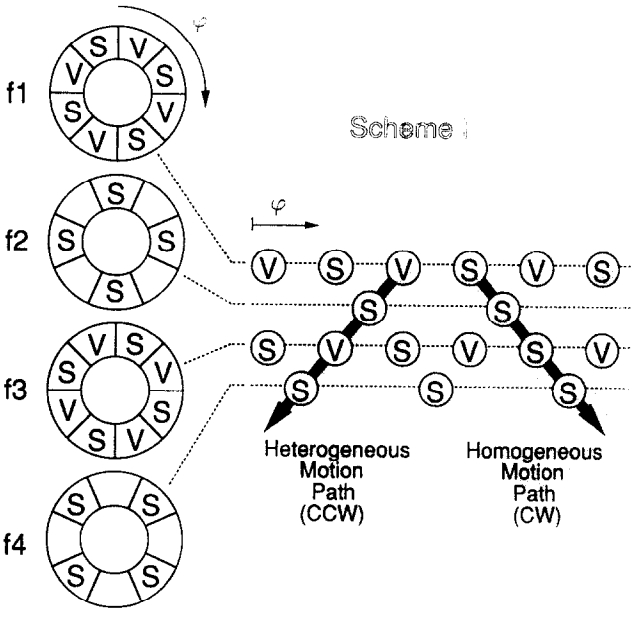

FIGURE 3. Motion competition Scheme I. Left: a series of frames $(f_1, f_2, \ldots)$ is shown successively in time (for details see Method section). The first frame (f_1) contains an annulus of patches of alternated texture type s and v at regular positions drawn against a uniform background. The annulus has an inner radius of $r_1 = 1.04$ deg of visual angle, and an outer radius of $r_2 = 2.08$ deg. The patches of texture s and texture v are spatially contiguous and alternate within the annulus. Since the annulus contains eight patches, each patch has a width of 45 deg. Angular position φ is measured clockwise with respect to the vertical. The second frame (f_2) is similar to frame f_1, except that the low frequent patches of texture v are now replaced by a uniform patch of background luminance. Furthermore, f_2 is rotated (clockwise) around the center of the annulus over an angle of 22.5 deg with respect to frame f_1. In a sequence of frames, frame f_{n+2} is identical to frame f_n, except for a rotation around the center over an angle of 45 deg (clockwise). Right: angular positions φ is along the horizontal axis. Patches of texture s and v are shown at their angular positions for frames $f_1 \cdots f_4$ yielding rows of patches. The top row of patches s and v corresponds to frame f_1. The second row of patches s corresponds to frame f_2. Hence, time (or frame number) is along the vertical axis. When frame f_n and frame f_{n+1} are presented in succession, two motion paths are *a priori* likely. A homogeneous motion path: clockwise matches (CW) between patches of identical texture s (indicated by the arrow pointing down and right). A heterogeneous motion path: counter-clockwise (CCW) matches between patches of texture s and patches of texture v (indicated by the arrow pointing down and left).

When frame f_n and frame f_{n+1} were presented in succession, two matches between patches of frame f_n and patches of frame f_{n+1} were *a priori* possible. The first match is a homogeneous clockwise match between patches of identical texture s separated by +22.5 deg (indicated in the diagram by the arrow pointing down and to the right). The second match is a heterogeneous counter-clockwise match between patches of texture v and patches of texture s (−22.5 deg, indicated by the arrow pointing down and to the left). Matches between frames f_n and f_{n+2} are entirely ambiguous. Matches between patches of frames f_n and f_{n+3} involve large temporal separations (400 msec) relative to the equivalent matches between frames f_n and f_{n+1} (133.3 msec). It has been shown that motion strength decreases strongly and monotonically with temporal interval for intervals larger than approx. 30 msec (Burt & Sperling, 1981; Werkhoven & Koenderink, 1991). Therefore, the matches between frames f_n and f_{n+3} are unimportant for motion perception in these stimuli.

Scheme I displays contain homogeneous and heterogeneous motion paths in opposite directions. By randomizing the direction of rotation, the directions of the two motion paths (although still opposite) are randomized.

The annular pinwheel stimulus was used for various reasons. First, the motion stimulus was presented at a constant eccentricity in the parafovea, and the effects of anisotropy of the retina were averaged across equivalent areas of the visual field. Second, it was easier to maintain fixation so eye movements were better controlled.* Finally (with the use of circularly symmetric stimuli) a motion path does not end at the boundaries of the display, avoiding edge effects.

Motion competition Scheme II. Scheme II (used in Expt 2) is equivalent to Scheme I, except that textures s and v are interchanged. The motion stimulus and resulting motion paths for this experiment are sketched in Fig. 4.

Although the heterogeneous motion path (between patches of texture s and v) is identical to that of Scheme I, the homogeneous motion path is different from that of Scheme I. In Scheme II, the homogeneous motion path consists of patches of texture v. The critical importance of the two schemes for our paradigm concerns the question of whether, when a particular s and v are chosen so that motion paths are balanced in Scheme I, the paths will remain balanced when the same s and v are used in Scheme II. From the subjects' point of view, however, there is no difference between the two schemes because, for any stimulus generated by Scheme I, an identical stimulus can be generated by Scheme II.

*Torsional eye-movements induced by the rotating annuli (cyclo-induction) were not controlled in our experiment. Balliet and Nakayama (1978) reported the ability of extremely trained subjects to make stepwise eye torsions up to rotations of approx. 26 deg for large field stimuli (25–50 deg of visual angle). However, we do not expect torsional pursuit in our experimental conditions: small field stimuli, brief presentations, fast motion, unpredictable motion direction, and ambiguous or near-threshold motion stimuli.

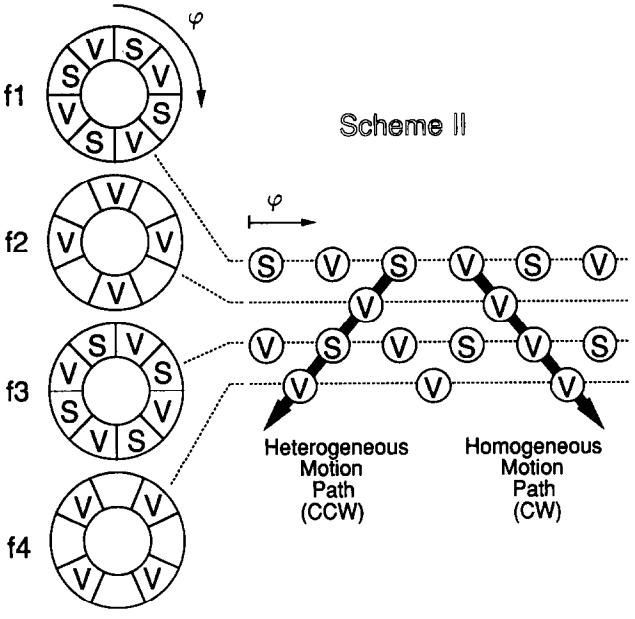

FIGURE 4. Motion competition Scheme II. This scheme is similar to Scheme I (see Fig. 3), except that textures s and v are interchanged. In Scheme II, the homogeneous motion path contains textures v.

However, during the course of a session, when v is varied between trials, different families of stimuli are generated by the two schemes.

Texture stimuli

The textures used to characterize texture-defined motion are patches of sinusoidally modulated gratings that differ in spatial frequency and amplitude. The grating patches were arranged in eight sectors of an annulus (pinwheel) around the fixation point with the grating extending radially in each sector. Two critical parameters that characterize a texture patch at a given location of the pinwheel are amplitude m and spatial frequency ω. Within a location, grating orientation was always radial. The phase γ of the grating was a random variable with a uniform distribution.

We use polar coordinates to further characterize the pinwheel. Let φ be the polar angle of a point in the image, and ρ be the distance to the origin (the center of the annulus). Then the luminance distribution at the point ρ, φ in sector j of frame i is:

$$L_{i,j}(\rho, \varphi) = L_0[1 + m_{i,j} \sin(2\pi r\varphi\omega_{i,j} + \gamma_{i,j})]. \quad (1)$$

We define the mean spatial frequency $\omega_{i,j}$ as the spatial frequency at mean radius r. The mean spatial frequency $\omega_{i,j}$ of a texture patch depends only on whether j is odd or even. That is, two spatial frequencies, ω_s, ω_v strictly alternate between adjacent patches on every frame of the display.

Within a trial, the amplitude $m_{i,j}$ of a sector i,j depended only on whether i and j were even or odd. On odd frames, $m_{0,j}$ was chosen as m_s or m_v according to whether the sector j was even or odd. On even frames, sector amplitude $m_{e,j}$ alternated between 0 and m_s in Scheme I and between m_v and 0 in Scheme II. Between trials, m_v and ω_v were changed. Sixteen values of

amplitude m_v from 0 to 1 were used increasing by steps of 0.0625: 0, 0.0625, 0.13, . . . , 1. Spatial frequency ω_v was varied over a range of three octaves: 1.2, 2.5, 3.7, 4.3, 4.9, 5.6, 7.4 or 9.9 c/deg. The amplitude m_s and spatial frequency ω_s of texture s were constant throughout the experiment: $m_s = 0.5$, $\omega_s = 4.9$ c/deg.

The phase $\gamma_{i,j}$, $0 \leqslant \gamma_{i,j} \leqslant 2\pi$, was chosen randomly and independently for every combination of i and j, that is, for every single patch. The phase randomization of every patch makes the motion of the stimulus inaccessible to any first-order (Fourier-based) mechanism. Phase randomization insures that motion mechanisms sensitive to correspondences in stimulus luminance were not systematically engaged (Chubb & Sperling, 1988).

Figure 5 shows an example of a series of frames for Scheme I. Texture s is a "medium" frequency grating and texture v is a "low" frequency grating. The regions inside and outside the annulus (background) were uniform gray and had a luminance value ($L_0 = 72$ cd/m^2). Within the annulus' texture patches the expected luminance value was equal to the background luminance.

Apparatus

The experiment was controlled by a IBM 386 PC compatible computer, driving a TrueVision AT-Vista video graphics adapter. A 60 Hz Imtec 1261L monitor with a P4-type phosphor was used to display the stimuli. The screen dimensions were 21.8 × 14 cm (640 × 480 pixels; 12.3 × 8.0 deg visual angle).* We used a look-up table to linearize the monitor's luminance values with the gray values of the computed stimulus patterns. The decay time to 10% and 1% intensity was about 1.3 and 6.2 msec respectively which is shorter than the temporal properties of retinal processing (Farrell, Pavel & Sperling, 1990; Sperling, 1976).

Subjects

Two subjects participated in the experiments: one of the authors (PW) and a colleague (JS). PW is emmetropic. JS is myopic (−0.5 D) but was in focus for the viewing distance used. Both subjects were experienced psychophysical observers. Natural pupils, binocular viewing, and spectacle corrections were used throughout. Several naive subjects confirmed the main findings for the experiments.

Procedure

Subjects indicated the dominant motion path (counter-clockwise/clockwise) by pressing one of two buttons. In both experiments, texture s (the standard texture) had amplitude $m_s = 0.5$ and spatial frequency $\omega_s = 4.9$ c/deg. From trial-to-trial, the spatial frequency ω_v and amplitude m_v of texture v was varied. The experiments determined the probability $P_i(m_v; \omega_v)$ of perceptual dominance of the heterogeneous motion path as a function of m_v for certain ω_v using the method of constant stimuli. The subscript i, $i = 1, 2$, indicates Expt 1 with competition Scheme I (Fig. 3) or Expt 2 with Scheme II (Fig. 4).

The probabilities $P_1(m_v; \omega_v)$ and $P_2(m_v; \omega_v)$ are estimated by the fraction of perceptually dominant heterogeneous motion paths out of 36 presentations. Spatial frequency ω_v was varied over a range of three octaves: $\omega_v = 1.2, 2.5, 3.7, 4.3, 4.9, 5.6, 7.4$ and 9.9 c/deg. Within a session, amplitude m_v was varied (pseudo-randomly from trial-to-trial; ω_v was varied only between sessions. For each spatial frequency ω_v, Expts 1 and 2 were both conducted within one session.

Subjects viewed the stimuli in a room with dimmed background illumination.

EXPERIMENT 1: SCHEME I

Results

By definition, the homogeneous path (consisting entirely of identical patches of texture s) does not change in this experiment when texture v is varied (see Scheme I, Fig. 3). The strength of the heterogeneous path, which is composed of alternate patches of textures s and v is varied by varying spatial frequency and amplitude, ω_v and m_v, of texture v. Figure 6 shows the probability $P_1(m_v; \omega_v)$ of reporting the heterogeneous motion path as dominant as a function of the amplitude m_v of texture v. Each panel shows $P_1(m_v; \omega_v)$ for a different value of spatial frequency ω_v.

The data show that the probability of reporting the heterogeneous path as dominant increases monotonically from 0 (for small m_v) to 1 (for $m_v = 1$) for all values of ω_v except the highest, where the probability of heterogeneous motion dominance has only reached about 65% when $m_v = 1$. A remarkable feature of these data is that in all eight panels, the probability $P_1(m_v; \omega_v)$ of heterogeneous motion dominance exceeds 50% for sufficiently high amplitude of patch v.

The upper left panel of Fig. 6 shows data for a two octave difference between the spatial frequency of texture s ($\omega_s = 4.9$ c/deg) and the spatial frequency of texture v ($\omega_v = 1.2$ c/deg). Heterogeneous motion is perceived in 50% of the presentations when the amplitude m_v of texture v is approx. 0.2. Note that at this balance point where both paths are equally likely, both the amplitudes and the spatial frequencies of textures s and v are markedly different. Once m_v exceeds 0.5, the heterogeneous motion path is dominant in 100% of the presentations. A 100% perceptual dominance of a heterogeneous over a homogeneous path demonstrates that the similarity between the textures in a motion path certainly is not essential for motion strength. Indeed, for sufficiently large m_v, the heterogeneous path is dominant over the homogeneous path for every combination of frequencies tested in Fig. 6.

*Due to the limited bandwidth of the video amplifier (30 MHz) of the monitor, an anisotropy was observed for the average luminance of differently oriented textures that contain high spatial frequencies. Therefore, we only displayed the pixels at column position m and row position n for which $(m + n)$ was even. The other pixels were dark. Hence, vertical and horizontal gratings share a common "carrier" component. This procedure forfeits maximum luminance and resolution in favor of eliminating anisotropy; the net resolution (320 × 240 pixels) was more than adequate for the displays.

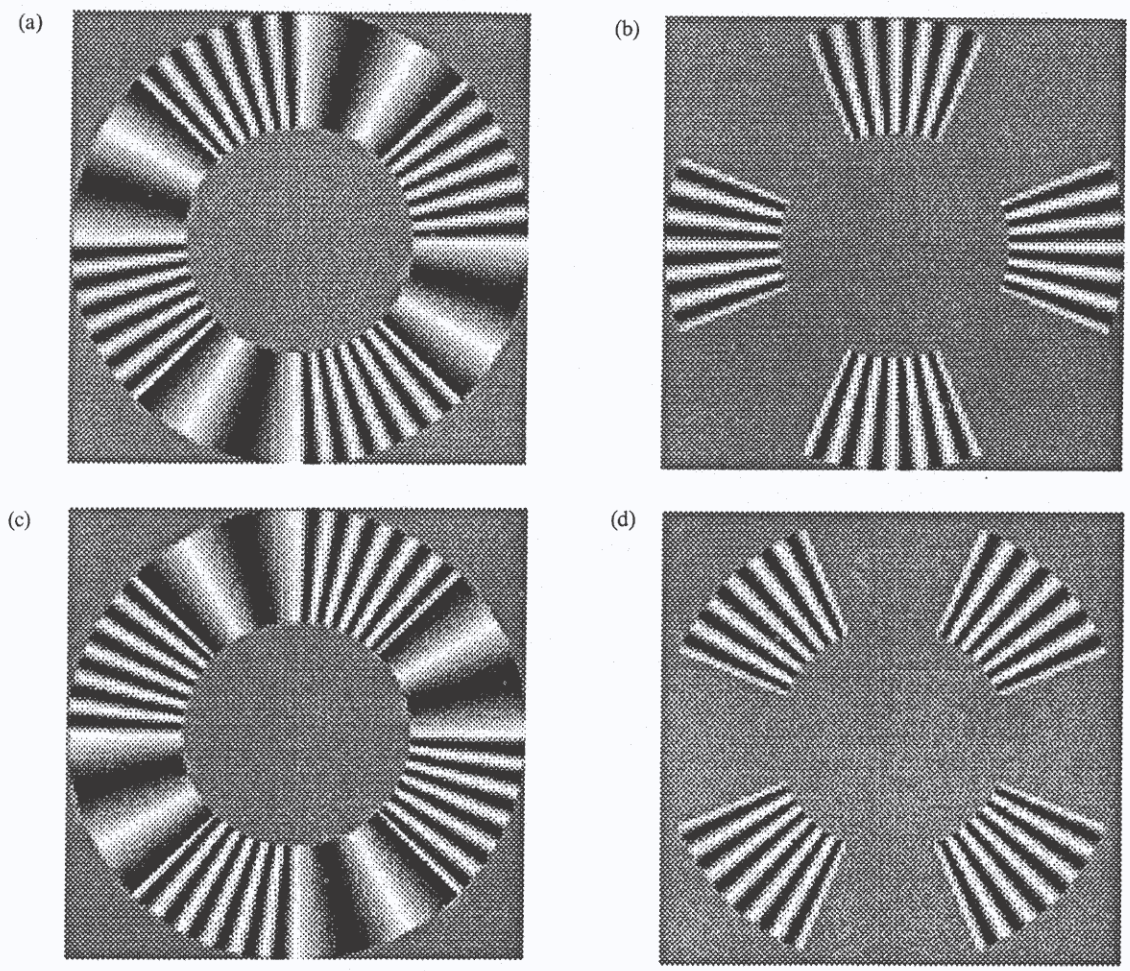

FIGURE 5. An example of the ambiguous motion display (as sketched in Fig. 3). Frames f_1, f_2, f_3, and f_4 (containing the patches of textures) are shown in (a), (b), (c) and (d) respectively. For this example, textures s and v differ only in their spatial frequency: the spatial frequency of texture s is two octaves higher than that of texture v.

The transition amplitudes between heterogeneous and homogeneous motion occur where the curves of Fig. 6 cross 50%. The transition amplitudes occur at a wide range of different amplitudes m_v for different spatial frequencies ω_v. Each P_1 curve is well characterized by two parameters: the transition amplitude $\mu_1(\omega_v)$ and the steepness $\sigma_1(\omega_v)$ at the transition amplitude (the subscript 1 indicates Scheme I). The transition amplitude $\mu_1(\omega_v)$ is defined as the amplitude m_v of texture v, necessary for balancing the motion paths [such that $P_1(m_v; \omega_v) = 50\%$], the steepness $\sigma_1(\omega_v)$ is defined as the derivative $\partial/(\partial m_v)P_1(m_v; \omega_v)$ with respect to m_v at the transition amplitude.

To estimate transition amplitude $\mu_1(\omega_v)$ and steepness $\sigma_1(\omega_v)$, we selected* data points of each probability curve around the transition amplitude. Within this selected range, the curve was assumed to be linear, and these data points were subject to a least square method of linear regression to estimate the regression coefficients $\mu_1(\omega_v)$ and $\sigma_1(\omega_v)$.

Estimates of $\mu_1(\omega_v)$ are shown in Fig. 7 as a function of the varied spatial frequency ω_v (open circles). The transition amplitude $\mu_1(\omega_v)$ increases systematically with increasing spatial frequency ω_v of texture v for both subjects. Together, the data of Figs 4 and 5 indicate that the strength of the heterogeneous motion path increases with increasing amplitude m_v but decreases with increasing spatial frequency ω_v.

Estimates of $\sigma_1(\omega_v)$ are shown in Fig. 8 as a function of the varied spatial frequency ω_v (open circles). The steepness $\sigma_1(\omega_v)$ of the probability curves at transition amplitude $\mu_1(\omega_v)$ decreases with the spatial frequency ω_v of texture v. In the Model section we elaborate on this finding.

Discussion

Sufficiency of a single energy-channel. In a single energy-channel, we assume that only one single type of texture grabber operates on the input yielding an activity representation of the input. Motion strength is the result

*In principle, we selected the three data-points around the transition amplitude (the crossing of the curves with the 50% guide line) that were closest to the 50% guide line. There were only two exceptions. First, at spatial frequency $\omega_v = 1.2$ c/deg, for subject PW, Expt 2, we selected the data points with amplitude $m_v = 0.19$, 0.25 and 0.31 (to avoid the low amplitude values, for which Scheme II becomes ambiguous). Second, at spatial frequency $\omega_v = 2.5$ c/deg, for subject JS, Expts 1 and 2, we selected the data points with amplitude $m_v = 0.38$ and 0.5 (since we had no data points close to the guide line).

of a motion energy analysis scheme applied to this activity representation. The motion strength of a path is computed from the product of activity measures between successive patches along the path in space-time. Motion strength of a heterogeneous path balances homogeneous motion strength when the responses (activities) to textures v and s are equal. Differences in textural properties between elements s and v are irrelevant as long as the activities are equal, just as, in scotopic vision, differences in wavelength are irrelevant as long as the rod response is the same.

The results for Scheme I suggest an activity transformation that is a monotonically increasing function of amplitude and a monotonically decreasing function of spatial frequency. For example, to balance the activity of texture s, with amplitude m_s and spatial frequency ω_s, with a lower spatial frequency texture v, $(m_v; \omega_v)$ requires a $m_v < m_s$. This pattern of results suggests a single class of texture grabbers consisting of a low-pass spatial filter followed by rectification.

We argued that a single energy-channel is sufficient to explain the results of Expt 1. It is important to note here, however, that our finding that heterogeneous motion can dominate homogeneous motion is also consistent with multiple energy-channels, as will be shown in the Model section. For example, the dominance of heterogeneous motion may well be the result of two independent energy-channels, both favoring heterogeneous motion. To uniquely determine the number of channels involved, we need the results for competition Scheme II together with a formal analysis (Model section).

Secondary contributions of a correspondence-channel. In the Discussion above, we argued that a single-channel model is sufficient to model the (amplitude/frequency dependent) dominance of heterogeneous motion found for Scheme I. However, we cannot exclude a possible secondary effect of texture similarity based on this scheme. To motivate Expt 2, we need to elaborate on this argument.

Although motion perception may be dominated by a single energy-channel, there may yet be a secondary contribution of a correspondence-channel.

The relative strength of the heterogeneous motion path would decrease as the differences between the spatial frequencies and amplitudes of successive patches of textures s and v increased. Suppose there were a

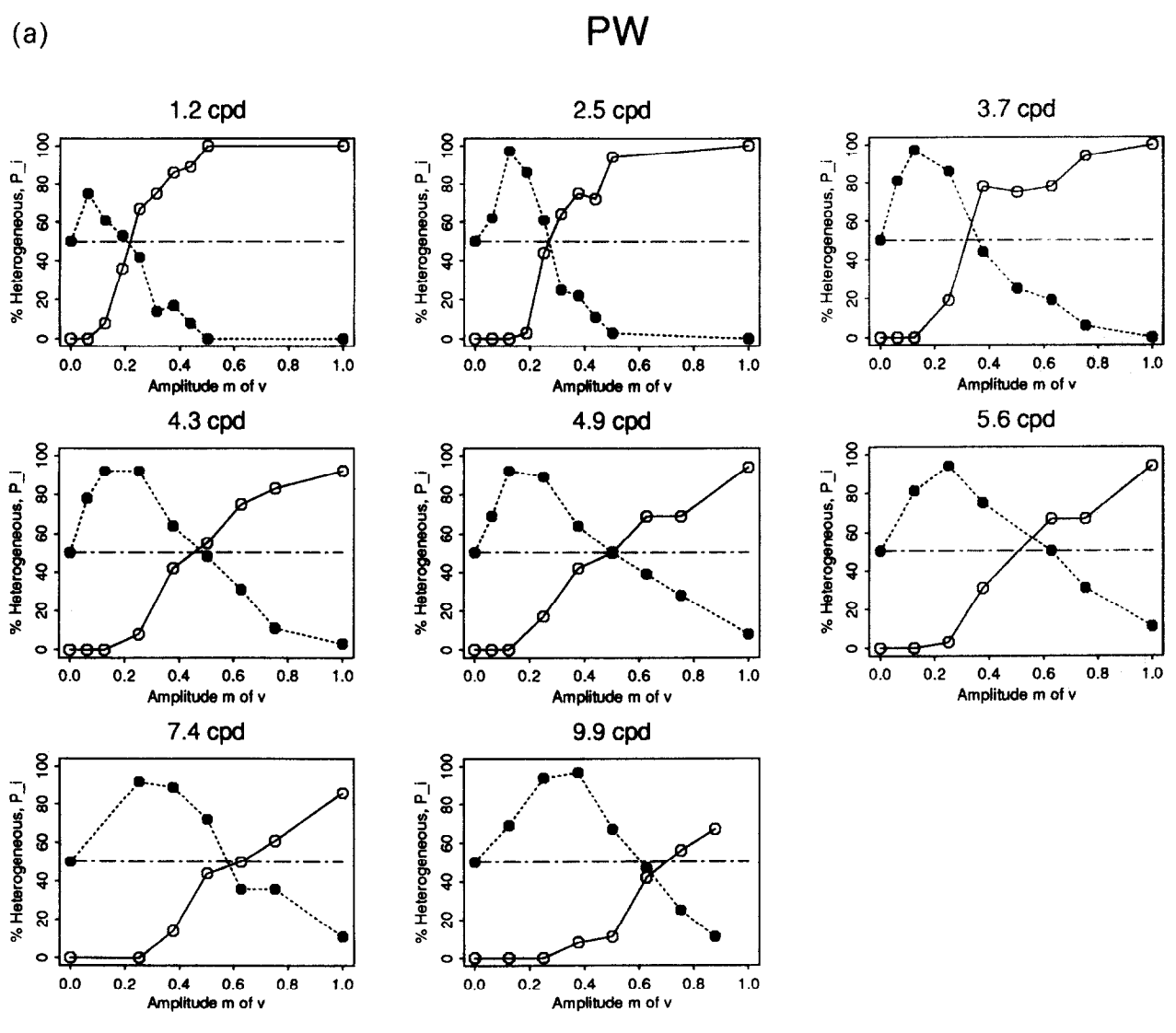

FIGURE 6(a). *Caption overleaf.*

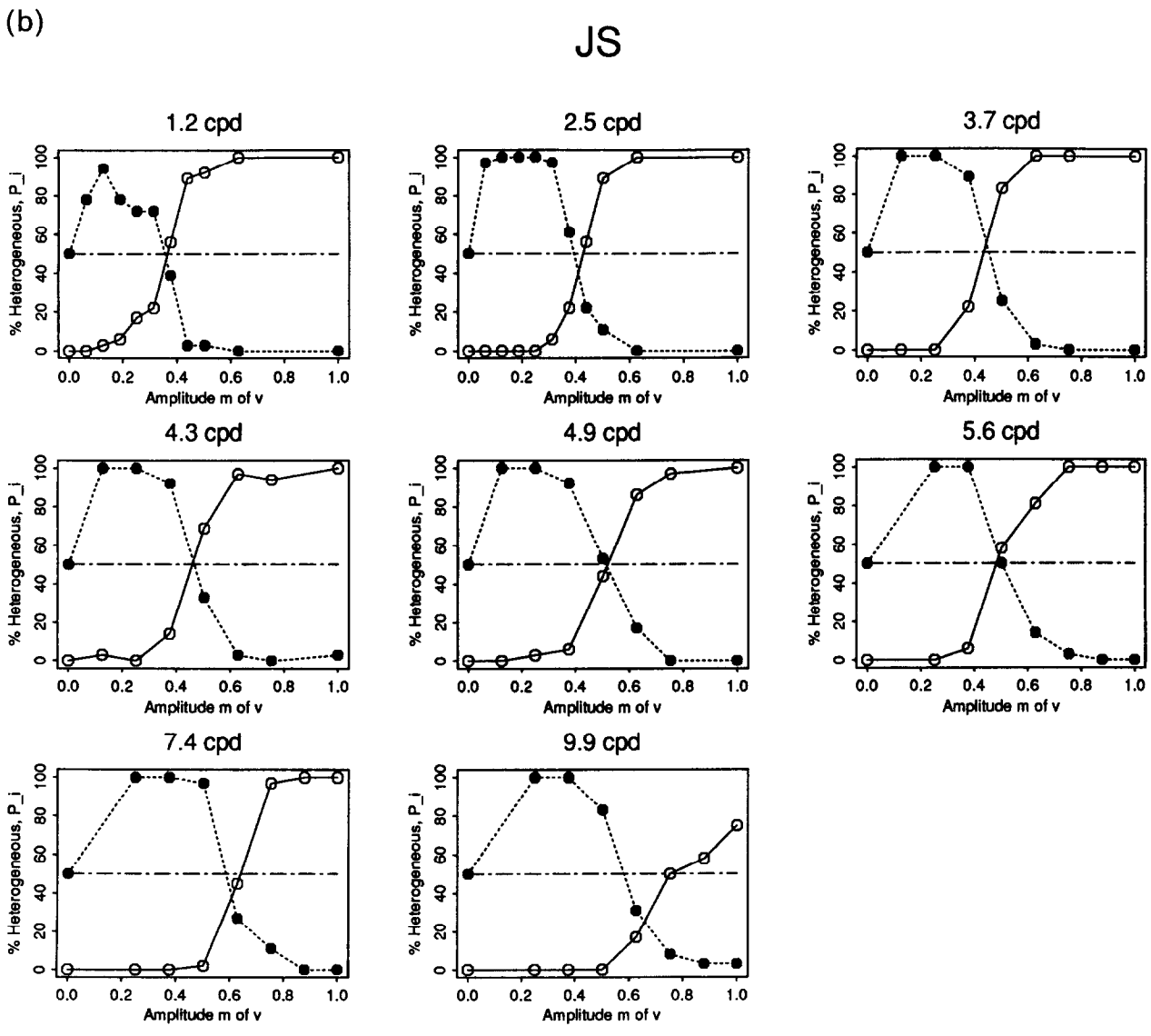

FIGURE 6. Probability $P_1(m_v; \omega_v)$ of dominance of a heterogeneous motion path over a homogeneous motion path is shown as a function of the amplitude m_v of texture v for different spatial frequencies ω_v of texture v for two subjects. Open circles represent the probability $P_1(m_v; \omega_v)$ for Scheme I (Fig. 3); solid circles $P_2(m_v; \omega_v)$ for Scheme II (Fig. 4). The horizontal dashed guide line indicates a 50% probability of heterogeneous motion dominance. The amplitude m_s and spatial frequency ω_s of texture s is the same for all panels: $m_s = 0.5$ and $\omega_s = 4.9$ c/deg. (a) Subject PW; (b) subject JS.

secondary contribution of a correspondence-channel in Expt 1, sensitive to differences between textures in either amplitude or frequency. Because the correspondence-channel favors the homogeneous path (by definition), motion balance requires v in the heterogeneous path to have a higher amplitude m_v to overcome the similarity in path s, s than if there were no correspondence-channel. Thus, in Scheme I, a secondary correspondence effect would displace transition amplitude $\mu_1(\omega_v)$ to higher values.

To test for a correspondence-channel, we introduce Scheme II in which s and v are interchanged (see Fig. 4). If there were a correspondence effect, in Scheme II it would favor the v, v path and the transition amplitude $\mu_1(\omega_v)$ would be shifted below $\mu_2(\omega_v)$ for any texture v.

When the homogeneous and heterogeneous motion paths remain balanced after interchanging textures s and v, this is called transition invariance. Transition invariance would imply that there is no contribution of a correspondence-channel.

EXPERIMENT 2: SCHEME II

Results

Figure 6 shows the probabilities $P_2(m_v; \omega_v)$ of the dominance of the heterogeneous motion path as a function of the amplitude m_v of texture v for different spatial frequencies ω_v of texture v. The data points for Scheme II are marked by a solid circle.

When $m_v = 0$, the display is physically as well as perceptually ambiguous. A value of 50% is shown for $m_v = 0$, though no data were collected at this point. By varying the amplitude of texture v in this experiment, the strength of both the heterogeneous motion path and the homogeneous motion path are varied. As the amplitude m_v increases, the probability of heterogeneous motion

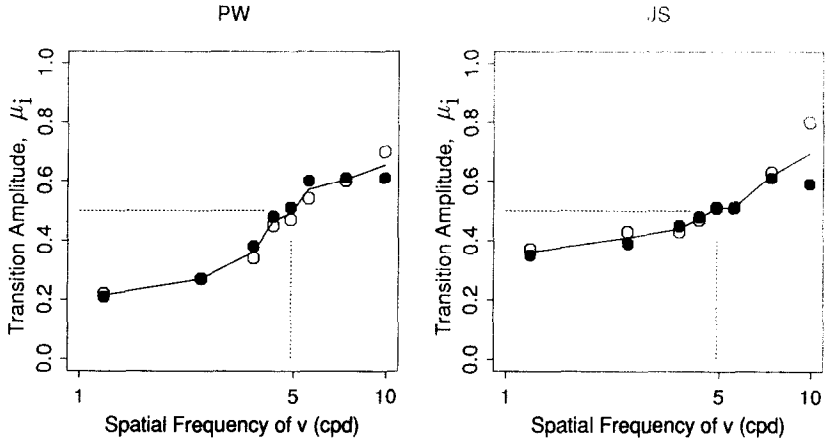

FIGURE 7. Transition amplitudes $\mu_i(\omega_v)$ as a function of spatial frequency ω_v. Open circles for Scheme I, solid circles for Scheme II. The vertical dashed line indicates the spatial frequency of texture s: $\omega_s = 4.9$ c/deg. The horizontal dashed guide line indicates the amplitude of texture s: $m_s = 0.5$.

dominance first increases to a maximum, then decreases to zero for high amplitude m_v. On the whole, for amplitudes above 0.1 or, in a few cases, 0.2, the Scheme I and Scheme II curves are mirror complementary, and seem to cross at exactly $P = 50\%$. That is, the two schemes produce remarkably similar transition amplitudes.

To examine the correspondence between the data from Schemes I and II, some definitions are needed. Let the transition amplitude $\mu_2(\omega_v)$ be the amplitude m_v of texture v for which the motion paths are balanced, and the probability of heterogeneous motion dominance $P_2(m_v; \omega_v)$ is 50%. The steepness at this transition amplitude is $\sigma_2(\omega_v)$. The transition amplitude $\mu_2(\omega_v)$ and steepness value $\sigma_2(\omega_v)$ are estimated as $\mu_1(\omega_v)$ and $\sigma_1(\omega_v)$ in the previous section.

To compare the transition amplitude $\mu_2(\omega_v)$ for Scheme II with transition amplitude $\mu_1(\omega_v)$ for Scheme I, they are presented together as a function of spatial frequency ω_v in Fig. 7. Transitions $\mu_2(\omega_v)$ are presented with solid circles. As in Scheme I, the amplitude $\mu_2(\omega_v)$ of texture v, necessary for balancing the motion paths, increases systematically with increasing spatial frequency ω_v of texture v. An exception for both subjects are the transition amplitudes for $\omega_v = 9.9$ c/deg.

To compare the steepness values $\sigma_2(\omega_v)$ for Scheme II with steepness values $\sigma_1(\omega_v)$ (for Scheme I), the absolute value of $\sigma_2(\omega_v)$ is shown as a function of the varied spatial frequency ω_v in Fig. 8 (using solid circles). It should be noted that the estimation is not very accurate: the standard deviation in the distribution of steepness coefficient $\sigma_i(\omega_v)$ is approx. 20%. However, like $\sigma_1(\omega_v)$, the steepness $\sigma_2(\omega_v)$ shows a tendency to decrease with increasing spatial frequency ω_v of texture v.

Discussion

Transition invariance and motion metamers. It is immediately clear that, for most spatial frequencies ω_v of texture v, the transition amplitude $\mu_2(\omega_v)$ is equal within measurement error to transition amplitude $\mu_2(\omega_v)$ (see Fig. 7). In fourteen of sixteen cases, the transition amplitudes are invariant when the textures s and v are interchanged. This we call transition invariance.

In two cases (the highest spatial frequency used—$\omega_v = 9.9$ c/deg—for both subjects), a small difference between transition amplitudes for Schemes I and II is observed. At the high spatial frequency of v, the amplitude of texture v necessary to balance the motion paths is slightly smaller for Scheme II than for Scheme I. This shift in transition amplitude suggests a small

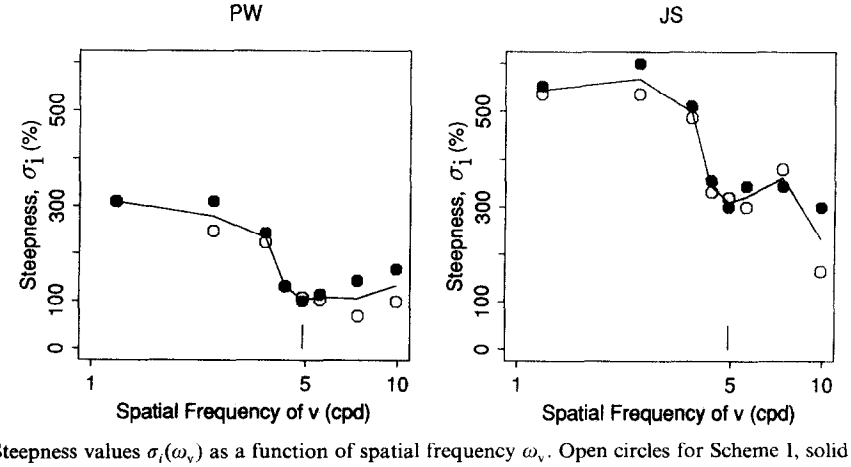

FIGURE 8. Steepness values $\sigma_i(\omega_v)$ as a function of spatial frequency ω_v. Open circles for Scheme I, solid circles for Scheme II. (Note that to facilitate comparison absolute values are given!) The vertical dashed guide line indicates the spatial frequency of texture s: $\omega_s = 4.9$ c/deg.

similarity effect (a small contribution of a correspondence-channel), and was discussed in the Discussion of Expt 1.

Transition invariance implies that textures s and v (at transitions) are equivalent with respect to motion processing and can be interchanged in any motion path (Scheme I and Scheme II) without affecting motion strength. This leads to the important conclusion that textures s and v are (texture-defined) motion metamers.

It is interesting to note that Green (1986, Fig. 7, p. 604) was unable to find an amplitude that could make a spatial frequency patch of 5.0 c/deg into a motion metamer of a 1.7 c/deg patch. We had no difficulty in finding metamers between even more disparate spatial frequencies. However, our data in Fig. 5 show that one of the two subjects would require the 5 c/deg stimulus to have more than two times the amplitude of the 1.7 c/deg stimulus, and this is outside the range of amplitudes that Green explored.

Necessity of a single energy-channel. The general finding of transition invariance strongly constrains the possible ways in which motion can be computed between textures in the class we are considering.

Transition invariance shows that there is *no* secondary contribution of correspondence-channels (see the discussion on this issue in Expt 1). The effect that a patch of texture v has on the strength of motion is independent of the other patches in the path. At a transition, the strength of motion path s, v is equal to that of v, v and that of s, s, although a correspondence-channel would yield stronger motion for the homogeneous paths.

The only alternative is a system of multiple energy-channels that must be combined and represented by a single scalar representation (e.g. summation of energy-channels). In the Model section, we prove (under the assumption of channel summation) that if multiple energy-channels were involved, the transition amplitude would generally shift when the textures s and v are interchanged in Schemes I and II. However, when motion perception is exclusively ruled by a single energy-channel (the product of the activity of a single type of texture grabber), the transition amplitude is invariant when the textures s and v are interchanged. Hence, transition invariance uniquely supports a single energy-channel model of texture-defined motion perception.

EXPERIMENT 3: AMPLITUDE LINEARITY

Motivation

In the above experiments, we have shown that the transition amplitude $\mu_1(\omega_v)$ increases systematically with increasing spatial frequency ω_v of texture v for both subjects. The strength of the heterogeneous motion path in Scheme I increases monotonically with increasing amplitude m_v but decreases with increasing spatial frequency ω_v. In order to further specify the dependency of motion strength on amplitude, we performed an experiment similar to that described above using competition Scheme I, and varied the amplitude of texture s.

Results

We kept the frequency of textures s and v constant ($\omega_s = 4.8$ c/deg and $\omega_v = 1.2$ c/deg) and measured the transition amplitude μ_1 as a function of amplitude m_s (Scheme I). Transition amplitude was estimated from the psychometric curves using the method described earlier.

Figure 9 shows the transition amplitude μ of texture v for three amplitude values of texture s ($m_s = 0.50, 0.75$ and 1.00) for three subjects. The data strongly suggest a linear dependence of the transition amplitude of texture v on the amplitude of texture s. The solid lines are the best fits (minimizing the sum of squares), accounting for at least 97% of the variance for each subject.

Discussion

We showed that the transition amplitude of texture v needed to balance the motion path s, v with the motion path s, s varied linearly with the amplitude of texture s. This dependency is easily accommodated in a model where the texture grabber is linear in the amplitude of the texture. In fact, one can easily show that amplitude linearity follows directly from the linear data under the assumption that the texture grabber is a separable function of spatial frequency and amplitude. A linear (low-pass) spatial frequency filter is a simple example of such a separable filter characteristic.

MODEL

Summary of model constraints

We used the analogy with colorimetry and some general assumptions about the possible motion computations involved to reach the conclusion that texture-defined motion strength is ruled by a single energy-channel. We summarize our reasoning.

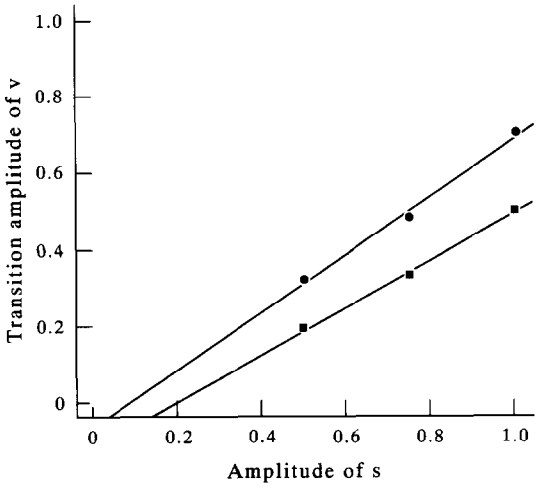

FIGURE 9. The dependence of transition amplitude μ_1 (ω_v) on amplitude m_s of texture s. The spatial frequency ω_s was 4.9 c/deg, and ω_v was 1.2 c/deg. Competition Scheme I was used. Circles, subject JS; squares, subject PW. The solid lines show the best linear fit (minimizing the sum of the squared deviations).

We discriminate two classes of motion computations: energy-channels and correspondence-channels, yielding different metrics for the strength of a motion path. Consider, a heterogeneous motion path composed of patches of texture s and v. The strength of an energy-channel for an s, v path is determined by the product of the activity of texture s and that of v. The activity of a texture is the output of some nonlinear transformation (texture grabber) that maps texture into a scalar. Energy-channels are insensitive to differences in textural properties and allow heterogeneous motion paths s, v to dominate over homogeneous paths. By definition, the strength of a correspondence-channel is determined by the similarity of the textural properties of textures s and v. That is, homogeneous paths s, s and v, v dominate heterogeneous paths s, v.

In theory, multiple channels of each type may be involved in a motion computation yielding a motion strength vector representation of arbitrary dimensionality. However, the experimental results impose the following constraints. First, the class of motion paths equal in strength for both Scheme I and Scheme II indicates that the computation is one-dimensional. Second, the invariance of transitions for Scheme I and Scheme II exclude correspondence-channels. This leaves us with a system of multiple energy-channels, that combine into a single scalar representation of motion strength.

Although we have shown that a single energy-channel is sufficient to model the data, we promised a proof for the necessity of a single energy-channel. This proof is based on the inconsistency of multiple energy-channels with transition invariance. We assume a system of multiple energy-channels that linearly combine to represent motion strength (summation of energy-channels). Such a system would result in different transitions for Scheme I and II. The proof is given and discussed in the Appendix.

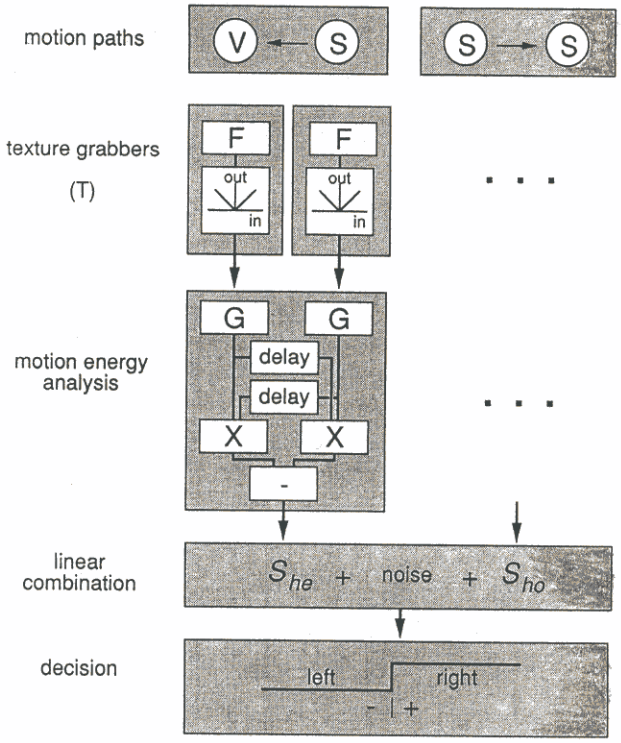

FIGURE 10. Diagram of a single channel motion computation. First stimulus amplitude is extracted followed by a linear spatial filter F and rectification. The spatial filter together with the rectification is called "texture grabber" (the first stage). The output of the texture grabber is called activity. The second stage (motion energy analysis) is basically a coincidence detector: it computes the product of the delayed activity at location 1 with the current activity at location 2. Response variability across trials is due to internal noise which is modeled by an additive noise having a standard model density function with mean 0 and standard deviation 1. The heterogeneous path is dominant whenever the net motion strength in the direction of the heterogeneous motion path (after adding noise) is positive (decision stage).

The energy-channel

In this section, we derive the characteristics of the single energy-channel. This energy-channel consists of two stages. The first stage is the nonlinear transformation (texture grabber). The simplest version of a texture grabber is a spatiotemporal linear filter followed by rectification (see Chubb & Sperling, 1989a, b). The output of this first stage (the texture activity) is fed into the second stage: motion energy analysis. Stages one and two are sketched in Fig. 10.

Stage 1: texture grabbers. It is now well-established (see review by Shapley & Enroth-Cugell, 1984), that early retinal gain-control mechanisms pass not stimulus luminance, but rather a signal approximating stimulus contrast, the normalized deviation of stimulus luminance from its local average. We assume that the spatiotemporal filters of Stage 1 operate on stimulus contrast.

The output magnitude of these filters varies over the visual field, depending on what textures happen to populate these regions. The output of a linear filter to a texture is variable and depends on the local phase of the texture. The purpose of rectification is to transform regions of highly variable response into regions of high average value, thus insuring that the rectified output registers the presence or absence of texture, independent of phase. Examples of rectification are half-wave rectification (setting negative values to zero) and full-wave rectification (anything that is symmetric with respect to input sign, such as absolute value or squaring).

The output of Stage 1 is called activity. The resulting transformation (accomplished by Stage 1) yields a spatiotemporal function whose value reflects the local texture preferences of the Stage 1 filter in the visual field as a function of time (see also Bergen & Adelson, 1988; Caelli, 1985). The activity transformation of the texture grabber depends on the amplitude m and spatial frequency ω of the textures involved.

In Expt 3, we have shown that texture activity is linear in texture amplitude. This is accommodated by a spatial filter that is linear in stimulus contrast. We can further characterize the spatial filter characteristics by the amplitude of its Fourier transform: $F(\omega)$. We assume that rectification is an absolute value operation. Thus, after

rectification, the activity transformation T is proportional to m and to $F(\omega)$:

$$T(m, \omega) = mF(\omega). \quad (2)$$

This texture activity T is fed into the second (motion energy analysis) stage.

Stage 2: motion energy analysis. The second stage (motion energy analysis) is a coincidence detector: it computes the product of the delayed activity at Location 1 with the current activity at Location 2 (van Santen & Sperling, 1984). For the displays we use in our experiments, the output of the second stage corresponds to motion strength.

To simplify the computation in the model, we assume that the first-stage spatiotemporal filter is space–time separable. Indeed, space–time separability seems to be the rule in apparent motion (Burt & Sperling, 1981; van de Grind, Koenderink & van Doorn, 1986).* Given space–time separability, we can ignore the temporal component of filtering because temporal patterns were not varied in our stimuli.

We proceed as follows. The perceived direction of motion is considered to be the outcome of a competition in motion strength between motion paths. Within a path the strength of motion between a patch of texture v and a patch of texture s is determined by the product of the activities of the first stage. We assume that the strengths of detectors for all paths are additive in the final motion percept, and adopt a linear combination model (Dosher, Sperling & Wurst, 1986). Additive internal noise determines the shape of the psychometric functions for motion direction as a function of amplitude.

Consider the strength model with respect to competition Scheme I (Fig. 3). In one direction there is a homogeneous motion path containing patches of identical texture s. In the opposite direction, there is a heterogeneous motion path containing patches of different textures s and v. For sine wave stimuli, a half-Reichardt model (simple product) is equivalent to the whole Reichardt model (difference of products) (van Santen & Sperling, 1985), so we need to consider just a simple product rule.

The strength of the heterogeneous motion path is:

$$S_{1,\mathrm{he}}(m_\mathrm{v}, \omega_\mathrm{v}, m_\mathrm{s}, \omega_\mathrm{s}) = m_\mathrm{v}F(\omega_\mathrm{v})m_\mathrm{s}F(\omega_\mathrm{s}). \quad (3)$$

The motion strength $S_{1,\mathrm{ho}}$ for the homogeneous motion path is equal to:

$$S_{1,\mathrm{ho}}(m_\mathrm{s}, \omega_\mathrm{s}) = -m_\mathrm{s}^2F^2(\omega_\mathrm{s}) \quad (4)$$

(strength in the opposite direction has opposite sign).

Linear combination of both components with equal weights yields a net motion strength D_1 in the direction of the heterogeneous path:

$$D_1(m_\mathrm{v}, \omega_\mathrm{v}, m_\mathrm{s}, \omega_\mathrm{s}) = S_{1,\mathrm{he}}(m_\mathrm{v}, \omega_\mathrm{v}, m_\mathrm{s}, \omega_\mathrm{s}) + S_{1,\mathrm{ho}}(m_\mathrm{s}, \omega_\mathrm{s}). \quad (5)$$

Response variability across trials is due to additive internal noise which is assumed to be distributed as a standard normal density function with mean 0 and standard deviation λ (Fig. 10). A linear addition of noise yields the internal decision variable i which has a normal distribution N with mean D and standard deviation λ.

According to signal detection theory (Green & Swets, 1966) the probability P of heterogeneous motion dominance is:

$$P_1(m_\mathrm{v}; \omega_\mathrm{v}) = P(i > 0) = \frac{1}{\sqrt{2\pi\lambda^2}} \int_0^\infty N\{D_1(m_\mathrm{v}, \omega_\mathrm{v}, m_\mathrm{s}, \omega_\mathrm{s}), \lambda\}\, \mathrm{d}i. \quad (6)$$

Substituting motion strengths [expressions (3) and (4)] into the additive linear combination [expression (5)] and then substituting [expression (5)] into the noise-driven decision process [expression (6)] yields:

$$P_1(m_\mathrm{v}; \omega_\mathrm{v}) = \frac{1}{\sqrt{2\pi\lambda^2}} \int_0^\infty N\{[m_\mathrm{v}F(\omega_\mathrm{v})m_\mathrm{s}F(\omega_\mathrm{s}) - m_\mathrm{s}^2F^2(\omega_\mathrm{s})], \lambda\}\, \mathrm{d}i, \quad (7)$$

for the probability of heterogeneous motion dominance for Scheme I (Fig. 3).

Similar reasoning yields the net motion strength D_2 and the probability $P_2(m_\mathrm{v}; \omega_\mathrm{v})$ of heterogeneous motion dominance in Scheme II (see Fig. 4):

$$D_2(m_\mathrm{v}, \omega_\mathrm{v}, m_\mathrm{s}, \omega_\mathrm{s}) = S_{2,\mathrm{he}}(m_\mathrm{v}, \omega_\mathrm{v}, m_\mathrm{s}, \omega_\mathrm{s}) + S_{2,\mathrm{ho}}(m_\mathrm{v}, \omega_\mathrm{v}) \quad (8)$$

$$= m_\mathrm{v}F(\omega_\mathrm{v})m_\mathrm{s}F(\omega_\mathrm{s}) - m_\mathrm{v}^2F^2(\omega_\mathrm{v}) \quad (9)$$

and

$$P_2(m_\mathrm{v}; \omega_\mathrm{v}) = \frac{1}{\sqrt{2\pi\lambda^2}} \int_0^\infty N\{[m_\mathrm{v}F(\omega_\mathrm{v})m_\mathrm{s}F(\omega_\mathrm{s}) - m_\mathrm{v}^2F^2(\omega_\mathrm{v})], \lambda\}\, \mathrm{d}i. \quad (10)$$

This model predicts the transition and steepness at transitions of the probability curves for both the experiments.

Predictions for Scheme I

For different spatial frequencies ω_v of texture v, we measured the probability $P_1(m_\mathrm{v}; \omega_\mathrm{v})$ of heterogeneous motion dominance as a function of the amplitude m_v of texture v. Our model predicts that the probability P_1 of heterogeneous motion dominance is an error function of the net motion strength D_1 [see equation (6)]. In this experiment, the net motion strength D_1 is linear in m_v. Hence, we expect an error function for the probability function $P_1(m_\mathrm{v}; \omega_\mathrm{v})$ as a function of m_v [see equation (7)].

*It is reasonable to consider that the linear filter in the texture grabber may itself be composed as a weighted sum of many filters, i.e. filters that also are in the processing path for first-order motion. A linear filter composed as the sum of component filters would be space–time separable if each of its component filters were space–time separable and had the same temporal function, independent of spatial scale. This seems to be the case in motion processing (Burt & Sperling, 1981; van de Grind *et al.*, 1986).

Transition amplitude. The transition amplitude $\mu_1(\omega_v)$ is defined as the amplitude m_v of texture v at which the probability of heterogeneous motion dominance $P_1(m_v; \omega_v)$ is 50% for a given spatial frequency ω_v of texture v. Hence, for $m_v = \mu_1(\omega_v)$, the strength of the heterogeneous and homogeneous motion paths are balanced and we have $S_{1,he} = -S_{1,ho}$ or [see expressions (3) and (4)]:

$$\mu_1(\omega_v)F(\omega_v) = m_s F(\omega_s) = \kappa, \tag{11}$$

where κ is a constant equal to the activity of standard texture s. If $F(\omega_v)$ is a low-pass filter, $\mu_1(\omega_v)$ will be a monotonically increasing function of ω_v (as supported by our experiments):

$$\mu_1(\omega_v) = \kappa F^{-1}(\omega_s). \tag{12}$$

Steepness. The steepness $\sigma_1(\omega_v)$ is defined as the derivative of $P_1(m_v; \omega_v)$ with respect to m_v at transition amplitude $\mu_1(\omega_v)$:

$$\sigma_1(\omega_v) = \frac{\partial}{\partial m_v} P_1(m_v; \omega_v)|_{m_v = \mu_1(\omega_v)} = \frac{1}{\sqrt{2\pi\lambda^2}} \kappa F(\omega_v). \tag{13}$$

Thus, the steepness $\sigma_1(\omega_v)$ is expected to decrease as a function of the spatial frequency ω_v for low-pass filters (as supported by our experiments).

In conclusion we expect error functions for the probability $P_1(m_v; \omega_v)$ of heterogeneous motion dominance as a function of amplitude m_v with (a) a transition amplitude $\mu_1(\omega_v)$ that is inversely proportional with $F(\omega_v)$ and (b) a steepness $\sigma_1(\omega_v)$ that is proportional with $F(\omega_v)$. If we have low-pass filters, $F(\omega_v)$ decreases monotonically with spatial frequency ω_v.

Predictions for Scheme II

For different spatial frequencies ω_v of texture v, we measured the probability $P_2(m_v; \omega_v)$ of heterogeneous motion dominance as a function of the amplitude m_v of texture v. $P_2(m_v; \omega_v)$ is an error function of D_{II} [see equation (10)]. However, for Scheme II (unlike for Scheme I) D_{II} is not linear with the varied amplitude m_v of texture v. As we increase the amplitude m_v of texture v, D_{II} shows a quadratic dependence on m_v. Therefore, we do not expect an error function for $P_2(m_v; \omega_v)$.

If amplitude m_v of texture v is zero, the probability of heterogeneous motion dominance P_2 will be 50% (the motion stimulus is purely ambiguous!). Starting at $m_v = 0$, it first increases linearly with m_v, is maximal for $m_v = m_s F(\omega_s)/[2F(\omega_v)]$, and decreases again with further increases of m_v. Obviously, there may exist an amplitude $m_v = \mu_2$ (between the "optimal" amplitude, that yields a maximal D_2, and a very high amplitude, that yields a negative D_2) for which $P_2 = 50\%$.

Analogous to the derivation in the previous section, one can find the analytic expressions for the transition $\mu_2(\omega_v)$ and steepness $\sigma_2(\omega_v)$ of the probability curves for Scheme II. The expressions for the transition amplitudes are equal: $\mu_2(\omega_v) = \mu_1(\omega_v)$. The expressions for the steepness of the transitions for Scheme I and II differ only in sign: $\sigma_2(\omega_v) = -\sigma_1(\omega_v)$.

The texture grabber

We can simply find the Fourier transform $F(\omega)$ of the low-pass filter from the reciprocal transition $\mu_i^{-1}(\omega_v)$ [see expression (12)] and from the steepness $\sigma_i(\omega_v)$ as a function of spatial frequency ω_v [see expression (13)].

The reciprocal transition amplitudes are expected to be proportional to the function $F(\omega_v)$. Estimates of the reciprocal transition amplitudes $\mu_i^{-1}(\omega_v)$ are shown in Fig. 11.

From the reciprocal transitions in Fig. 11, it follows that $F(\omega)$ is a low-pass filter in the range of frequencies examined.

The model predicts that the steepness of the probability function is proportional with the function $F(\omega_v)$ and inversely proportional with λ (the strength of the internal noise). Thus, unlike the transition amplitude, the steepness is biased by the internal noise contribution. If the relative strength is constant and independent of the spatial frequency and amplitude of the patches of texture

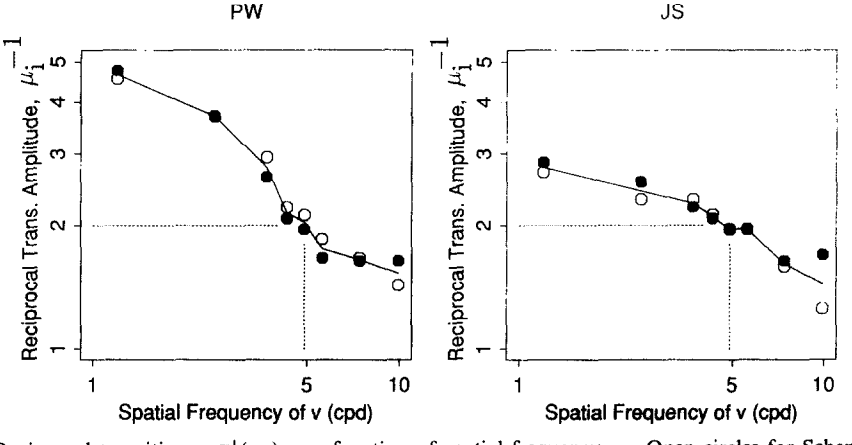

FIGURE 11. Reciprocal transitions $\mu_i^{-1}(\omega_v)$ as a function of spatial frequency ω_v. Open circles for Scheme I; solid circles for Scheme II. The vertical dashed guide line indicates the spatial frequency of texture s: $\omega_s = 4.9$ c/deg. The horizontal dashed guide line indicates the reciprocal amplitude of texture s. The solid line curve is the mean of the reciprocal transitions. In terms of the model, this curve shows the amplitude of the Fourier transform of the spatial filter $F(\omega)$ of the texture grabber involved [see equation (2)].

involved, the steepness $\sigma_i(\omega_v)$ is expected to be proportional with $F(\omega_v)$. Estimates of $\sigma_i(\omega_v)$ are shown in Fig. 8. The steepness shows a tendency to decrease with increasing spatial frequency. However, we find some nonmonotonicity, in particular for higher spatial frequencies. This may reflect a certain variability of the internal noise for different spatial frequencies.

EXPERIMENT 4: PERCEIVED CONTRAST

We have discussed texture grabbers and motion energy analysis in terms of objective amplitude of patches of texture. The experiments implied that the activity of the texture grabber increases monotonically with objective amplitude and decreases monotonically with spatial frequency. An interesting question is whether this relation is consistent with the subjective amplitude of static grating contrast as a function of spatial frequency. In other words, is the activity of a texture grabber simply proportional to the subjective amplitude?

To answer this question, we performed an amplitude discrimination experiment.

Method

In a two interval presentation subjects looked at an annulus containing either gratings s or v. In one interval we showed an annulus of gratings s (see frame f_2 of Fig. 3), with fixed amplitude $m_s = 0.5$ and fixed spatial frequency $\omega_s = 4.9$ c/deg. In the other interval we showed an annulus of gratings v (see frame f_2 of Fig. 4), with amplitude m_v and spatial frequency ω_v. The order of presentation of the intervals was randomized. Each annulus was shown for 133 msec (which is equal to the frame display time in the motion stimulus). The intervals were separated by a time interval of 133 msec in which the screen was uniform with background luminance. Apparatus, viewing conditions, and other aspects were identical to the motion experiment.

Procedure

The task of the subject was to indicate the interval that contained the patches of grating with the highest amplitude. We measured the probability $P_c(m_v; \omega_v)$ that observers judge the grating v as the grating with the highest amplitude as a function of the objective amplitude m_v of grating v. In the amplitude matching experiment, we examined two spatial frequencies: $\omega_v = 1.2$ c/deg, and $\omega_v = 7.4$ c/deg of grating v. These were the lowest and highest spatial frequencies for which we found transition invariance in our modern experiment. From these probability curves, we estimated the matching amplitude of grating v for which the perceived amplitude of grating s and v was equal. The precise estimation of the matching amplitude was analogous to the estimation of transition amplitude in the motion competition experiments.

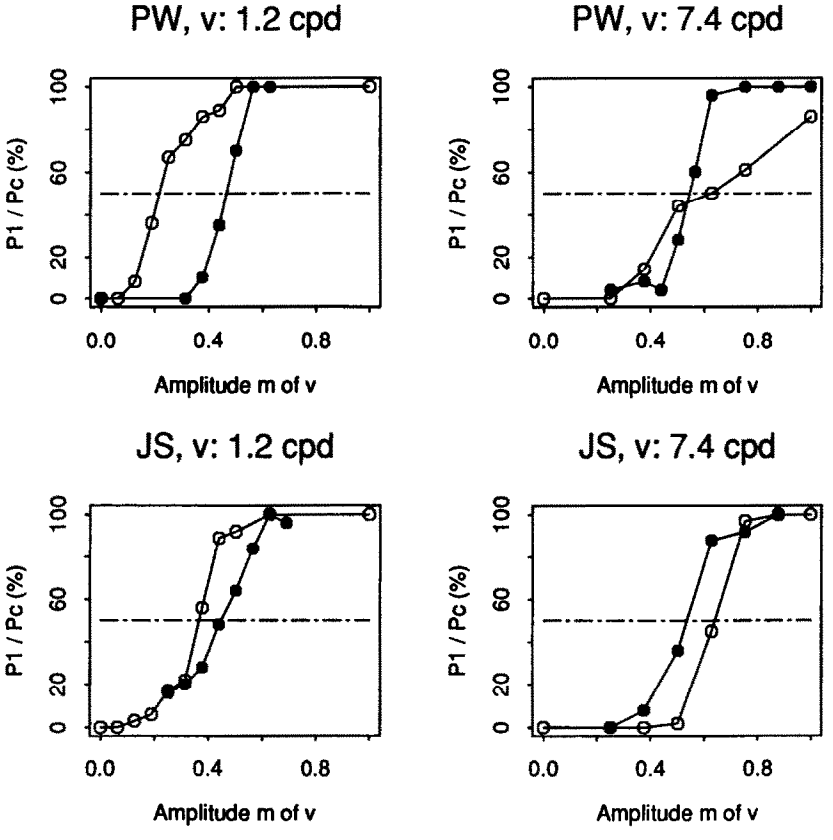

FIGURE 12. Results of the perceived amplitude experiment. Observers compared the amplitude of a grating v (spatial frequency ω_v and amplitude m_v) with the amplitude of texture s ($m_s = 0.5$, $\omega_s = 4.9$ c/deg). Shown are the probabilities P_c for judging the amplitude of v higher than that of s (solid circles). The matching amplitude for texture v is the crossing of the curve with the dashed 50% line. To compare the matching amplitude with the transition amplitude in the motion experiment, we have shown the probabilities $P_1(m_v)$ for Scheme I (open circles).

Results

In Fig. 12, we show the probabilities of judging the amplitude of grating v higher than that of grating s (with $m_s = 0.5$) as a function of objective amplitude m_v (solid circles). For all conditions and subjects, the perceived amplitude of texture v increases monotonically with its objective amplitude m_v. The amplitude m_v where the curve crosses the 50% guide line is the matching amplitude. For a "low" spatial frequency grating v ($\omega_v = 1.2$ c/deg), we find that the perceived amplitudes of s and v are matched when $m_v = 0.47$ for subject PW and $m_v = 0.44$ for JS. This matching amplitude is close to the objective amplitude $m_s = 0.5$ of grating s. For a "high" spatial frequency grating v ($\omega_v = 7.4$ c/deg), the matching amplitudes are $m_v = 0.54$ for PW and M_v 0.53 for JS.

The comparison of the matching amplitude with the transition amplitude in the motion experiments, we have also shown the probabilities to perceive heterogeneous motion using Scheme I as a function of m_v in the corresponding panels.

Discussion

Interestingly, the matching amplitudes for low and high spatial frequency gratings are approximately equal to the objective amplitude of grating s, for the range of amplitudes and spatial frequencies of grating v examined. That is, perceived amplitude does not depend on spatial frequency. However, the amplitude of grating v for balancing the motion paths when $\omega_v = 1.2$ c/deg for Scheme I was: $m_v = 0.22$ for subject PW and $m_v = 0.36$ for JS. Obviously, at the transition amplitude for the motion experiment, the perceived amplitude of grating s and v are markedly different. That is, the activities of the grating v are matched even when *both* spatial frequency and perceived amplitude are different from grating s. In conclusion, activity cannot be a function that depends solely on perceived amplitude.

EXPERIMENT 5: DICHOPTIC PRESENTATIONS

Motivation

We have successfully modeled the strength of motion-from-texture in terms of a texture grabber followed by motion energy analysis. Motion energy analysis is a type of motion computation that is *not* sensitive to correspondences in textural features. An interesting property of first-order motion energy analysis is that the neural substrate for such a process is organized so as to require successive stimulation to the same eye. When monocular motion information is not available to the observer first-order motion energy analysis fails.

The motion system that extracts first-order motion information of both eyes (when motion is presented dichoptically) has been classified as a correspondence-channel. For example, Pantle and Picciano (1976) studied apparent motion with a three-dot stimulus and reported element movement for monocular and binocular presentation, but group movement for dichoptic presentation. The group movement suggests a representation of features or shapes precedes the extraction of motion. Also, Georgeson and Shackleton (1989) show that drifting squarewave gratings with missing fundamental (MF) moved backwards while presented monocularly (following the third harmonic) but moved forwards when presented dichoptically. They suggested that the perceived direction of dichoptic apparent motion was consistent with a system that combines information across spatial frequency channels to identify local features and then tracks the location of corresponding features over time.

Generalizing the above reasoning to second-order motion, the motion mechanism for dichoptic presentations of our (second-order) stimuli would be sensitive to the similarity of the textures involved. Thus, the contribution of what we call correspondence-channels might be more pronounced when our competition schemes are presented dichoptically (sofar viewing has been binocular in our experiments). We tested our energy-channel model for motion-from-texture for both dichoptical and monocular presentations of our motion stimuli. This test may also locate the motion extraction process involved in our stimuli in terms of different levels in the visual nervous system (before or after the sites of binocular combination).

Results

The ambiguous motion competition Schemes I and II can be presented dichoptically in two different modes. In the first mode, the odd frames are presented in one eye and the even frames in the other. In this way, the spatiotemporal stimulus is purely ambiguous in each eye. Both the heterogeneous and the homogeneous paths are processed by dichoptic mechanisms. In this mode, dichoptic mechanisms are not competing with monocular mechanisms.

In the second mode, the patches of one texture type are presented in one eye and the patches of the second type of texture type are presented in one eye and the patches of the second type of texture in the other eye. In this way the homogeneous motion path (textures s for Scheme I) is presented in one eye, while the textures v in the other eye form a purely ambiguous stimulus. In this mode, dichoptic mechanisms processing the heterogeneous path have to compete with monocular mechanisms processing the homogeneous path.

We determined the psychometric functions for both competition schemes for a condition where the texture s and v differ two octaves in spatial frequency ($\omega_s = 4.9$ c/deg and $\omega_v = 1.2$ c/deg) for subject PW. The binocular results were presented in top-left panel of Fig. 6. As discussed for Expts 1 and 2, a difference between the transition amplitudes μ_1 and μ_2 indicates the involvement of additional (correspondence) channels. The results for monocular presentation were identical (within measurement error) to the results for binocular presentation. For both conditions, we find transition invariance: $\mu_1 = \mu_2 \approx 0.2$.

The results for both modes of dichoptic presentation were very similar to those for binocular presentation. That is, dichoptic presentation yields psychometric

functions for Schemes I and II similar to those for binocular presentation. For adequate amplitude m_v heterogeneous motion dominated homogeneous motion for both modes of dichoptic presentation suggesting the dominance of an energy-channel even when monocular motion information was absent. However, the contribution of a correspondence-channel is noticeable for dichoptic presentations: transition invariance no longer holds. We found $\mu_1 \approx 0.2$ and $\mu_2 \approx 0.1$ for both modes of dichoptic presentation.

Discussion

Motion perception between patches of nonsimilar texture is easily perceived for both modes of dichoptic presentation (as predicted by our energy-channel). Even in the second mode, where a dichoptic heterogeneous motion path competes with a monocular homogeneous path, heterogeneous motion can easily dominate for small amplitude of texture v (e.g. $m_v > 0.2$ for Scheme I). These results suggest that dichoptic processing of our motion stimuli is dominated by the same mechanisms as monocular processing and that motion strength is not predicted by the similarity between textural features such as spatial frequency.

However, although dichoptic presentation leaves transition amplitude μ_1 for Scheme I unaffected, transition μ_2 for Scheme II decreases. This difference from the binocular results indicates a significant contribution of other channels when monocular information for the heterogeneous path is ambiguous. A more detailed investigation might be useful.

GENERAL DISCUSSION

Fallacy of correspondence matching

The experiments presented in this paper provide cogent evidence that texture similarity is not relevant to the texture-defined motion computation (within the range of spatiotemporal parameters varied in this experiment). As an example it was shown that motion between patches of texture that differ by two octaves in spatial frequency and a factor of 2 in amplitude can be stronger than motion between patches of identical texture.

The correspondence matching metaphor to explain visual processes in several visual domains seems to have lost predictive power. Correspondence matching fails to explain the dominance of (1) heterogeneous motion paths composed of textures that differ in spatial frequency and amplitude (this paper), (2) heterogeneous motion paths composed of elements that differ in size, orientation and luminance (Werkhoven *et al.*, 1990a, b), and (3) stereoscopic matches between elements that differ in size and luminance (Gulick & Lawson, 1976).

The visual motion system does not seem to be designed to establish correspondence between similar features in a motion sequence. This should not come as a surprise given the inherent difficulties in designing correspondence matching mechanisms. Such mechanisms would look for "similar features" in "successive" time samples of the spatiotemporal stimulus. However, what constitutes a feature, and how strict should similarity be taken?

Recently developed stimulus (motion) energy models for motion extraction bypass the correspondence problem and are more likely candidates for the kind of visual processing early in the visual system (Adelson & Bergen, 1985; Heeger, 1992). The energy-channel described in this paper is equivalent to such a motion energy computation, applied to a nonlinear transformation of the stimulus (van Santen & Sperling, 1984).

Contrast and motion

In Expt 3, we showed that the transition amplitude of texture v needed to balance the motion path s, v with the motion path s, s varies linearly with the amplitude of texture s. In the context of our model, this means that the activity of a texture grabber is approximately linear in texture amplitude. In fact, we find linearity even for high amplitudes in the range of 50–100%. As a consequence of this amplitude linearity, motion strength varies linearly with the amplitude of each of the texture inputs. That is, the strength of motion between two textures with identical texture amplitude is quadratic with this amplitude. Approximate amplitude linearity of the input lines for first-order motion energy analysis was also found for experiments with spatiotemporal modulations of luminance Werkhoven *et al.* (1990b).

It should be noted, that the linear amplitude dependency is at odds with the amplitude thresholds for motion direction discrimination reported by Nakayama and Silverman (1985). They measured the smallest phase shift (yielding threshold direction discrimination performance) of sinusoidal gratings as a function of grating amplitude. The smallest phase shift yielding threshold performance leveled off for grating amplitudes exceeding 5%. They interpreted their finding in terms of a amplitude saturation function. However, their results are open to a different interpretation in which the minimum phase shift is limited by other (spatial) properties of the motion extraction mechanism leaving the amplitude dependency unknown.

A shared motion analysis stage?

An intriguing question is how mechanisms for the extraction of motion carried by the spatiotemporal modulation of luminance relate to those for extracting motion carried by the spatiotemporal modulation of texture type. To discriminate both mechanisms we have to compare the characteristics of the perception of both motion types. For example, Turano and Pantle (1989) studied velocity discrimination performance for both types of motion stimuli and showed similar discrimination characteristics. Their results support the hypothesis of a higher order (motion analysis) mechanism that accepts input from both the luminance domain as well as texture domain.

A shared motion energy analysis stage for the two types of motion is also supported by our finding that strength of motion-from-texture is ruled by the same metric as motion in the luminance domain. Motion

strength is the covariance (or product) of local activities. This activity is simply the luminance itself when the motion is carried by luminance (van Santen & Sperling, 1984) or a nonlinear transformation of the luminance pattern for motion-from-texture (this paper).

In conclusion, the extraction of motion from the spatiotemporal modulations of luminance and that of texture types seems to be mediated by a shared motion energy analysis stage. However, additional experiments with different paradigms may weaken this idea. For example, Mather (1991) showed that both motion types produce motion after effects, but that the duration of the aftereffects were significantly different.

Transitivity and additivity

Under the assumption of energy channels and channel summation, the transition invariance of a pair of textures s and v implies that s and v are (texture-defined) motion metamers. That is, all such textures v in this metameric class yield identical motion strength when embedded in a motion path s, v.

Metamery yields two strong predictions. First, metamery predicts transitivity: if textures a and b are metameric with s, then a is metameric with b. Second, metamery predicts additivity: if textures a and b are metameric with s, then any linear combination $\alpha a + \beta b$ (with $\alpha + \beta = 1$) is metameric with s.

These predictions have not yet been tested.

Motion transparency

The energy-channel proposed in this paper computes the difference between left- and rightward motion. This implies that motion transparency (the simultaneous detection of left- and rightward motion) is not readily accommodated in this model. Because the motion analysis component of the energy-channel is a Reichardt-correlator, the motion energy of the left- and rightward motion path are no explicit intermediate results). However, occasionally, observers reported transparency for stimuli that were nearly balanced.

Adelson and Bergen (1985) addressed this issue by pointing out that although their energy detector was functionally equivalent to correlation detector, the intermediate results are not. Specifically, the energy of left and rightward motion are explicit intermediate results in energy detectors, but not in correlation detectors (the output of a half Reichardt-correlation is the half-phase opponent energy!). Although our conclusions do not depend on the specific choice of motion model, a further study of transparency in this context might reveal the specific type of detector involved.

Extension of the parameter space

It is important to remember that we have shown the one-dimensionality of the motion-from-texture computation only with respect to parallel sinewave patches that differ in spatial frequency and amplitude. Chubb and Sperling (1991) found that motion-from-texture could be carried by differences in spatial orientation, although differences in orientation did not produce as vigorous motion as did differences in spatial frequency. This observation indicates that orientation (and possibly other properties) are relevant to motion-from-texture. It would be interesting to determine the dimensionality of the computation for a larger class of stimuli.

Although motion strength at a "frame time" τ of 8/60 sec is exclusively determined by the product of activities, we can not exclude that effects of texture similarity are stronger at longer frame time. In fact, the temporal frequency of texture modulation in our experiments is 1.9 Hz (one cycle consists of four frames of 133 msec each). At slower temporal frequencies, the processing time for the textures increases, perhaps enabling more elaborate "texture grabber" filters or correspondence-channels to contribute to motion strength.

Effects of other properties (e.g. orientation) and temporal parameters are currently under investigation.

REFERENCES

Adelson, E. H. & Bergen, J. R. (1985). Spatio-temporal 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. In *Proceedings: Workshop on motion: Representation and analysis* (pp. 151–155). IEEE Computer Society Press.

Balliet, R. & Nakayama, K. (1978). Training of voluntary torsion. *Investigative Ophthalmology and Visual Science, 17,* 303–314.

Bergen, J. R. & Adelson, E. H. (1988). Early vision and texture perception. *Nature, 333,* 363–364.

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

Burt, P. & Sperling, G. (1981). Time, distance and feature trade-offs in visual apparent motion. *Psychological Review, 88,* 171–195.

Caelli, T. (1985). Three processing characteristics of visual texture segregation. *Spatial Vision, 1,* 19–30.

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

Cavanagh, P., Arguin, M. & von Grünau, M. (1989). Interattribute apparent motion. *Vision Research, 29,* 1197–1204.

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. (1989a). Second-order motion perception: Space/time separable mechanisms. *Proceedings: Workshop on visual motion, Irvine, California, 1989* (pp. 126–138). IEEE Computer Society Press.

Chubb, C. & Sperling, G. (1989b). Two motion perception mechanisms revealed by distance driven reversal of apparent motion. Proceedings: Workshop on visual motion, Irvine, California, 1989. *Proceedings of the National Academy of Sciences U.S.A., 86,* 2985–2989.

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

Dosher, B. A., Sperling, G. & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research, 26,* 973–990.

Farrell, J. E., Pavel, M. & Sperling, G. (1990). The visible persistence of stimuli in stroboscopic motion. *Vision Research, 30,* 921–936.

Georgeson, M. A. & Shackleton, T. M. (1989). Monocular motion sensing, binocular motion perception. *Vision Research, 29,* 1511–1523.

Graham, N. (1992). Complex channels, early local nonlinearities, and normalization in texture segregation. In Landy, M. S. & Movshon,

J. A. (Eds), *Computational models of visual processing* (Chap. 18). Cambridge, Mass.: MIT Press.
Green, M. (1986). What determines correspondence strength in apparent motion? *Vision Research, 26*, 599–607.
Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
van de Grind, W. A. Koenderink, J. J. & van Doorn, A. J. (1986). The distribution of human motion detector properties in the monocular visual field. *Vision Research, 26*, 797–810.
Gulick, W. L. & Lawson, R. B. (1976). *Human stereopsis: A psychophysical analysis*. New York: Oxford University Press.
Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A, 4*, 1455–1471.
Heeger, D. J. (1992). Nonlinear model of neural responses in cat visual cortex. In Landy, M. S. & Movshon, J. A. (Eds), *Computational models of visual processing* (Chap. 9). Cambridge, Mass.: MIT Press.
Kolers, P. A. (1972). *Aspects of motion perception*. Oxford: Pergamon Press.
Lelkens, A. M. M. & Koenderink, J. J. (1984). Illusory motion in visual displays. *Vision Research, 24*, 1083–1090.
Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London, 200*, 269–294.
Mather, G. (1991). First-order and second-order visual processes in the perception of motion and tilt. *Vision Research, 31*, 161–167.
Moulden, B. & Begg, H. (1986). Some tests of the Marr–Ullman model of movement detection. *Perception, 15*, 139–155.
Nakayama, K. & Silverman, G. H. (1985). Detection and discrimination of sinusoidal grating displacements. *Journal of the Optical Society of America A, 2*, 267–274.
Navon, D. (1976). Irrelevance of figural identify for resolving ambiguities in apparent motion. *Journal of Experimental Psychology, Human Perception and Performance, 2*, 130–138.
Pantle, A. & Picciano, L. (1976). A multistable movement display: Evidence for two separate motion systems in human vision. *Science, 193*, 500–502.
Papathomas, T. V., Gorea, A. & Julesz, B. (1991). Two carriers for motion perception: Color and luminance. *Vision Research, 31*, 1883–1891.
Ramachandran, V. S., Rao, M. V. & Vidyasagar, T. R. (1973). Apparent movement with subjective contours. *Vision Research, 13*, 1399–1401.
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.
Robson, J. G. (1980). *Neural images: The physiological basis of spatial vision*. In Harris, C. (Ed.), *Visual coding and adaptability* (pp. 177–214). Hillsdale, N. J.: Erlbaum.
van Santen, J. P. H. & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America A, 1*, 451–473.
van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A, 2*, 300–321.
Shapley, R. & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. *Progress in Retinal Research, B3*, 263–346.
Shechter, S., Hochstein, S. & Hillman, P. (1989a). Size, flux and luminance effects in the apparent motion correspondence process. *Vision Research, 29*, 579–591.
Sperling, G. (1976). Movement perception in computer-driven visual displays. *Behavior Research Methods and Instrumentation, 8*, 144–151.
Turano, K. & Pantle, A. (1989). On the mechanism that encodes the movement of contrast variations: velocity discrimination. *Vision Research, 29*, 207–221.
Ullman, S. (1980). The effect of similarity between bar segments on the correspondence strength in apparent motion. *Perception, 9*, 617–626.
Victor, J. D. & Conte, M. M. (1990). Motion mechanisms have only limited access to form information. *Vision Research, 30*, 289–301.
Watson, A. B. (1986). Apparent motion occurs only between similar spatial frequencies. *Vision Research, 26*, 1727–1730.
Werkhoven, P. & Koenderink, J. J. (1991). Reversed rotary motion perception. *Journal of the Optical Society of America A, 8*, 1510–1516.
Werkhoven, P., Snippe, H. P. & Koenderink, J. J. (1990a). Effects of element orientation on apparent motion perception. *Perception and Psychophysics, 47*, 509–525.
Werkhoven, P., Snippe, H. P. & Koenderink, J. J. (1990b). Metrics for the strength of low level motion perception. *Journal of Visual Communication and Image Representation, 1*, 176–188.

Acknowledgement—This work was supported by the AFOSR Grant 91-0178.

APPENDIX

Multiple Energy-Channels and Transition Invariance

A system of multiple energy-channels

We propose a multi-channel model (multiple energy-channels) for computing the strength of motion-from-texture. The model consists of two stages, as shown in Fig. 13.

Stimulus transformation: texture grabbers. Stage 1 consists of n types of texture grabbers—where each type of texture grabber i is described by nonlinear spatiotemporal transformations T_i, $i = 1 \ldots n$, of the optical input. Each transformation yields a spatiotemporal function $T_i(\varphi, t)$ whose value reflects the local texture preferences of the Stage 1 filters in the visual field as a function of position φ and time t. (We use φ for position because, in our essentially one-dimensional stimulus, the texture position is determined by the angle φ.) The output of these texture grabbers is called activity. The n different transformations T_i of Stage 1 transform the optical input into n activity representations.

Motion detection. Stage 2 is a set of motion detectors. For specificity, but without loss of generality (see van Santen & Sperling, 1984; Chubb & Sperling, 1988, 1991) we adopt Reichardt's scheme for standard motion analysis (Reichardt, 1961) which consists of two oppositely tuned coincidence detectors. Motion detectors operate on the outputs of the texture grabbers. Each type of texture grabber (transformation T_i) has its own, unique set of motion detectors. A transformation T_i together with its motion detectors is called a motion channel i.

A coincidence detector performs a multiplication operation on the current activity $T_i(\varphi, t)$ at position φ at time t and the (delayed) activity $T_i(\varphi - \Delta\varphi, t - \Delta t)$ at position $\varphi - \Delta\varphi$ and time $t - \Delta t$. Hence, the output of the coincidence detector is: $T_i(\varphi - \Delta\varphi, t - \Delta t)T_i(\varphi, t)$. The outputs of two coincidence detectors

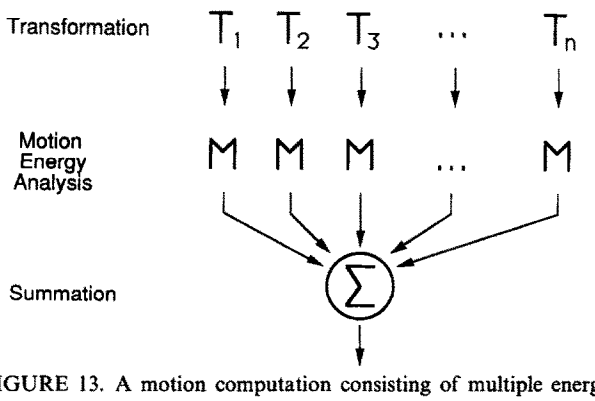

FIGURE 13. A motion computation consisting of multiple energy-channels. The first stage consists of n independent transformations T_i (the texture grabbers). Transformation T_i is a nonlinear tranformation (e.g. spatial filtering followed by rectification). The output of each transformation is called an activity representation of the optical input. Motion energy analysis (M) is applied to each of the activity representations of the input. Finally the motion strength is summed across the different channels.

tuned to identical velocities but opposite directions are subtracted to yield a net motion strength $D_i(\varphi, t)$:

$$D_i(\varphi, t) = T_i(\varphi - \Delta\varphi, t - \Delta t)T_i(\varphi, t) - T_i(\varphi - \Delta\varphi, t)T_i(\varphi, t - \Delta t). \tag{14}$$

Channel i has a positive output for motion in the direction of positive φ and a negative output for motion in the opposite direction.

Summation. In a one-dimensional motion computation, the outputs of a system of energy-channels described above (represented in an n dimensional channel space) are essentially mapped to a single (decision) dimension: the final net motion strength. This mapping maps an $(n-1)$-dimensional manifold in the channel space to a single point in the one-dimensional decision space (final motion strength). For example, channel summation maps a planar surface in the channel space to zero final motion strength (for Scheme I). For other combination rules than summation, other (nonplanar) surfaces will map to zero final motion strength. However, when we assume that this mapping is continuous and differentiable, these true manifolds are in first order approximated by a planar surface for small channel signals at transition points. Channel summation is a sufficient first-order combination rule.

Summation of channels D_i yields net motion strength D:

$$D(\varphi, t) = \sum_{i=1}^{n} D_i(\varphi, t). \tag{15}$$

Predictions for competition schemes

We apply the multi-channel computation to competition Schemes I and II (see Figs 3 and 4). Consider first Scheme I. The heterogeneous path is the motion between texture s (at time $t - \Delta t$ and position $\varphi - \Delta\varphi$) and texture v (at time t and position φ). Let $T_{i,s}$ be the activity of texture grabber T_i for texture s, and $T_{i,v}$ the activity of texture grabber T_i for texture v. The output of channel i for this path is the product of the delayed activity $T_{i,s}$ of texture s and the current activity $T_{i,v}$ of texture v. For simplicity, we will use the vector notation:

$$\vec{T}_s = \begin{bmatrix} T_{1,s} \\ T_{2,s} \\ \cdot \\ \cdot \\ T_{n,s} \end{bmatrix} \quad \text{and} \quad \vec{T}_v = \begin{bmatrix} T_{1,v} \\ T_{2,v} \\ \cdot \\ \cdot \\ T_{n,v} \end{bmatrix}. \tag{16}$$

The vectors $\vec{T}_s$ and $\vec{T}_v$ are the activity vectors of textures s and v respectively. An activity vector represents the activity of a texture in the n-dimensional transformation space (T-space) defined by transformations $T_1 \cdots T_n$.

For Scheme I, the motion strengths $S_{1,he}$ summed over all channels for the heterogeneous path can be written as the vector product:

$$S_{1,he} = \vec{T}_s \cdot \vec{T}_v = \sum_{i=1}^{n} T_{i,s} T_{i,v}. \tag{17}$$

We have arbitrarily assigned a positive sign to motion strength in this direction. Motion in the opposite direction has a negative sign [see equation (14)]. The output of channel i for the homogeneous path (between textures s) is the squared output of transformation $T_{i,s}$. The motion strength $S_{1,ho}$ of the homogeneous path is (after summing all channels) is:

$$S_{1,ho} = -\vec{T}_s \cdot \vec{T}_s. \tag{18}$$

Adding equations (6) and (7) gives the net motion strength D_1 in the direction of the heterogeneous path for Scheme I:

$$D_1 = \vec{T}_s \cdot (\vec{T}_v - \vec{T}_s). \tag{19}$$

Analogously, the net motion strength D_2 in the direction of the heterogeneous path for Scheme II is:

$$D_2 = \vec{T}_v \cdot (\vec{T}_s - \vec{T}_v). \tag{20}$$

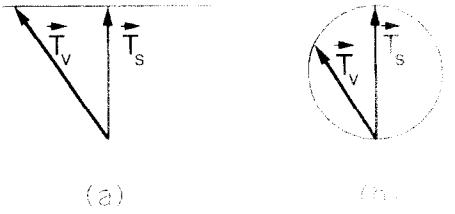

FIGURE 14. Solutions for transitions (path equality) in a two-dimensional T-space. Each texture in a motion path is processed by different texture grabbers. Vector $\vec{T}_v$ represents the activity of texture v in T-space, vector $\vec{T}_s$ that of s. The collection of activity vectors $\vec{T}_v$ that satisfy the constraints for path equality are given by the thin line in (a) for Scheme I and by a thin circle in (b) for Scheme II.

Transitions: Scheme I

At a transition for Scheme I, the net motion strength D_1 is zero:

$$D_1 = \vec{T}_s \cdot (\vec{T}_v - \vec{T}_s) = 0. \tag{21}$$

There exists an $(n-1)$-dimensional plane of $\vec{T}_v$ vectors in T-space for which the motion strength of the heterogeneous and homogeneous motion paths are balanced (the vectors $\vec{T}_v$ for which the difference vector $\vec{T}_v - \vec{T}_s$ are orthogonal to vector $\vec{T}_s$).

Consider, for example, a two-dimensional T-space (a two-channel motion computation). The vectors $\vec{T}_v$ in T-space that satisfy equation (21) for a certain vector $\vec{T}_s$ must end on the thin guide line in Fig. 14(a).

It should be noted in passing, that the net heterogeneous motion strength $D_1 = \vec{T}_s \cdot (\vec{T}_v - \vec{T}_s)$ can be positive. Hence, even in a multi-channel computation, the strength of the heterogeneous motion path can dominate.

Transitions: Scheme II

Similarly, at a transition for Scheme II (Fig. 4), the net motion strength D_2 is zero:

$$D_2 = \vec{T}_v \cdot (\vec{T}_s - \vec{T}_v) = 0. \tag{22}$$

The $(n-1)$-dimensional solution of $\vec{T}_v$ vectors in T-space for which the motion strength of the heterogeneous and homogeneous motion paths are balanced is not a plane. For example, we consider again the two-dimensional T-space. The vectors $\vec{T}_v$ in T-space that satisfy equation (22) for a certain vector $\vec{T}_s$ end on a circle containing $\vec{T}_s$ [see Fig. 14(b)].

Transition invariance

Using only the result for Scheme I, we cannot discriminate between a single-channel ($n = 1$) and multi-channel computations ($n > 1$): either single- or multi-channel computations might yield solutions to equation (21). To resolve the issue, we need the constraint of transition invariance.

Transition invariance means that once the motion strength of the heterogeneous path and that of the homogeneous motion path are balanced for a particular pair of textures s and v for Scheme I, this balance is not disturbed by interchanging the textures s and v (yielding Scheme II). We now show that transition invariance is inconsistent with a multi-channel computation.

The transitions are invariant if the activity vector $\vec{T}_s$ simultaneously satisfies equations (21) and (22). Because the difference vector $\vec{T}_s - \vec{T}_v$ is always in the plane defined by vector $\vec{T}_s$ and vector $\vec{T}_v$, the only vector $\vec{T}_v$ that satisfies both equations is $\vec{T}_v = \vec{T}_s$.

Vector $\vec{T}_v$ is equal to vector $\vec{T}_s$ if each transformation T_i involved in the motion computation has an equal output for both textures v and s:

$$T_{i,s} = T_{i,v} \quad (i = 1 \cdots n). \tag{23}$$

Equation (23) represents a very strong constraint for the ensemble of transformations that might be involved in a multi-channel computation. Every transformation T_i must have an isoactivity contour as a function of all textural properties (e.g. frequency–amplitude space) that contains both the activity of texture s and that of texture v. Furthermore, transition invariance holds for different texture pairs (s, v); the iso-activity contours of each transformation T_i must be identical for all these pairs. Transformations that are identical at arbitrarily many observable points, are identical in the range of observable points. To say that all T_i are identical is equivalent to saying that there is only one T_i, that is, the T-space is one-dimensional.