To appear in **Nature Neuroscience**, January 2000.

Mechanisms of visual motion detection

Paul R. Schrater[1], David C. Knill[2], and Eero P. Simoncelli[3]

[1] Department of Psychology, University of Minnesota,
N218 Elliot Hall, 75 E. River Dr., Minneapolis, MN 55455, USA

[2] Department of Psychology, University of Pennsylvania
3815 Walnut St., Philadelphia, PA 19104, USA

[3] Center for Neural Science, New York University
4 Washington Place, 809, New York, NY 10003, USA

Visual motion is processed by neurons in primary visual cortex that are sensitive to the spatial orientation and speed of the visual input. Many models of local velocity computation are based on a second stage that pools the outputs of first-stage neurons selective for different orientations, but the nature of this pooling remains controversial. In a psychophysical detection experiment, we find that human subjects show near-perfect summation of image energy when it is distributed uniformly across all orientations, but poor summation when it is concentrated in specific orientation bands. The data are consistent with a model that integrates uniformly over all orientations, even when this strategy is sub-optimal.

When a person moves relative to the environment, the visual image projected onto the retina changes accordingly. Within small regions of the retina and for short durations this motion is commonly approximated as a two-dimensional translation. The field of velocity vectors associated with each such region is referred to as "optic flow". Physiological and psychophysical experiments have established that the mammalian cortex contains mechanisms sensitive to local image motion.[1] These mechanisms are generally considered to form the neural substrate for representing optic flow.

Neurons in primary visual cortex perform the first stage of cortical motion processing. They are selective for both the spatial orientation and speed of the visual input within a spatially localized region. Because of their orientation specificity, however, these neurons suffer from an ambiguity commonly known as the "aperture problem": Each neuron can only signal the speed of motion perpendicular to the orientation to which it is tuned.[2] The neural responses are insensitive to the velocity component parallel to this orientation. Many authors have suggested that a second stage of processing, commonly identified with visual area MT, computes an unambiguous representation of local pattern velocity by selectively combining the outputs of the first-stage detectors.[2–5] While such an integration stage appears to be consistent with much of the psychophysical and physiological data, the precise form of the combination rule remains unclear. This paper describes psychophysical experiments designed to compare the predictions of a number of hypothetical combination rules.

The simplest combination strategy is to integrate the responses of all those first-stage mechanisms consistent with a particular two-dimensional velocity.[2] It is convenient to describe this construction in the three-dimensional spatio-temporal frequency domain, where the spectral energy of a rigidly translat-

ing image is concentrated on an oblique plane (Fig. 1a–b). The orientation of this plane uniquely specifies the translational velocity (both speed and direction) of the luminance pattern.[6] It is well known that the first-stage motion detectors in the mammalian visual system can be described as computing the spectral energy within limited bands of spatio-temporal frequency.[7] Thus, a pattern-velocity detector may be constructed by summing the weighted outputs of first-stage mechanisms tuned to spatio-temporal frequency bands that lie on a common plane[8–10] (Fig. 1c). The population response of a family of detectors built in this way can determine the presence and velocity of local translating patterns, and the responses of individual detectors are well matched to the behaviors of a subset of neurons in visual area MT.[10]

Alternatively, the visual system could use an adaptive integration rule, selectively combining only those first-stage that are tuned to the spatial structure of the image. For example, one type of model makes initial robust estimates of the components of a pattern by selecting those detectors that are responding to the stimulus while ignoring those responding to noise.[11] These one-dimensional estimates are then combined to produce an estimate of the two-dimensional pattern velocity that is most consistent with the measured one-dimensional velocity components. In general, such adaptive pooling rules will produce more efficient motion detectors than fixed pooling rules because the detector is better matched to the signal. Human observers have been shown to adapt their spatial pooling to improve the detection of static images.[12] Thus, it is plausible that they may do the same for moving images.

In this paper, we describe a set of psychophysical detection experiments designed to rigorously test the predictions of three models for pooling of spectral energy: (1) a planar power detector, that sums the signal energy over all orientations in a fixed planar region of spatio-temporal frequency, (2) an adaptive planar power detector, that sums energy only over those regions of a plane containing the signal, and (3) an adaptive unrestricted power detector, that can sum energy over arbitrary regions of spatio-temporal frequency. Experimental stimuli were constructed by summing dynamic random noise patterns that were band-limited using spatio-temporally oriented filters analogous to those used to describe first stage motion mechanisms. Each of the three models makes distinct sub-threshold summation predictions for the detection of such stochastic signals embedded in white noise. The planar power detector predicts optimal summation performance when signal energy is distributed uniformly over all spatial orientations in a plane. But this model predicts suboptimal performance when the signal energy is concentrated in a subset of planar orientation bands, or distributed over non-coplanar regions of spatio-temporal frequency, since the model sums both signal and noise in these cases. Similarly, the adaptive planar power detector predicts optimal performance for any signal whose energy is concentrated in a plane, regardless of orientation content, but sub-optimal performance for stimuli composed of non-coplanar distributions of signal energy. Finally, the adaptive unrestricted power detector predicts optimal performance for any distribution of signal energy, with no improvement in performance for planar distributions.

Results

In a first experiment, we tested these predictions by measuring detection thresholds for three types of spatially localized stochastic signals embedded in white noise. Signals were generated by filtering

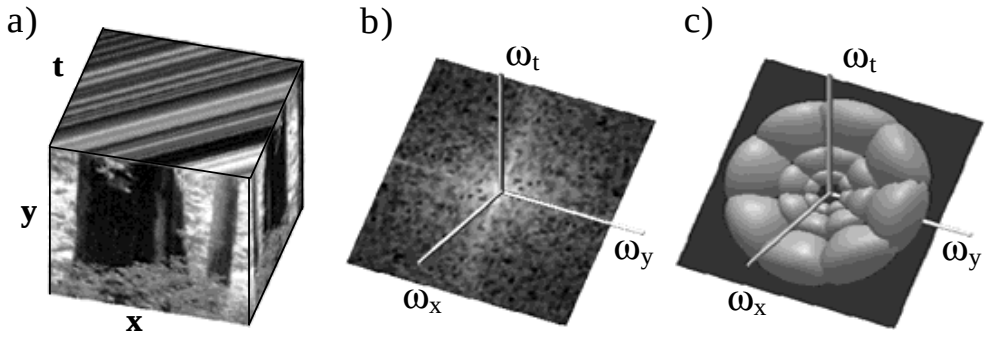

Fig. 1: A translational motion detector. **(a)** Space-time luminance pattern of an image translating to the right. This is a representation of the intensity information in the retinal image (the x- y plane) over time (t). The rightward motion can be inferred from the oriented pattern on the x-t face. **(b)** The Fourier amplitude spectrum of the luminance pattern, represented by the intensity of points in a three-dimensional spatio-temporal frequency domain. Non-zero Fourier amplitudes are constrained to lie on a plane through the origin. The orientation of this plane uniquely specify the direction and speed of translation. **(c)** Construction of a translation detector,[10] illustrated in the Fourier domain. Pairs of balls symmetric about the origin indicate the Fourier amplitude spectra of band-pass filters whose peak frequencies lie in the plane. A translation detector can be constructed by summing the squared outputs of such filters.

spatio-temporally white noise with different sets of band-pass filters. "Component" stimuli (Fig. 2a) had spectral energy confined to a single pass-band. "Plaid" stimuli (Fig. 2b) had two component bands on the same plane. "Planar" stimuli (Fig. 2c) were constructed using a set of component filters arranged in an annular ring on an oblique plane (designed so that the signal energy was uniformly distributed over an annulus in the plane). A two-alternative forced choice task was used to measure detection thresholds for each of the three stimuli. Stimuli were presented in two consecutive intervals, one containing the signal added to broad-band white noise and the other containing white noise alone. Subjects were asked to judge which interval contained the signal. Thresholds were estimated by fitting a psychometric function to the data and selecting the contrast energy that produced 81.1% detectability. Different stimulus types were presented in distinct experimental sessions, and subjects were told which stimulus type was being used. Prior to data collection, subjects were given several hours practice for each stimulus type over the course of several days until performance saturated. Further details are given in the methods section at the end of the paper.

In order to compare detection performance across stimuli, we assume that performance for the component stimulus reflects the detection efficiency of a band-limited, first-stage motion energy detector mechanism (The component signal has a bandwidth similar to the psychophysically optimal bandwidths found by Watson and Turano[13]). Specifically, we assumed a signal detection model in which subjects were assumed to make detection judgments for the component stimulus based on the output of a filter matched to the component signal, with the output corrupted by additive internal noise. The

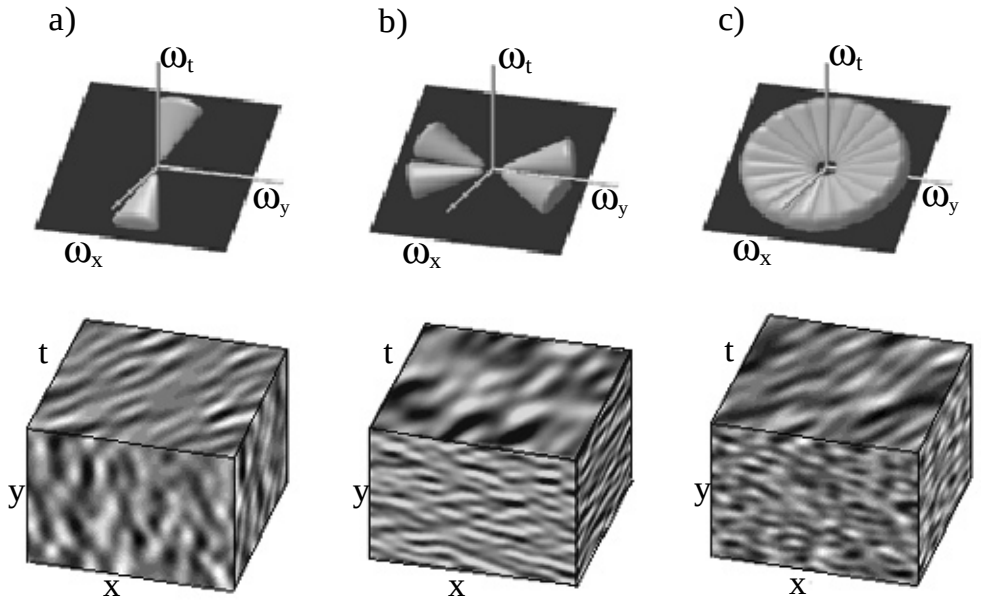

Fig. 2: Filter sets and examples of their corresponding signals. The top row depicts level surfaces (65% of peak response) of the three different filter sets used to generate stimuli. The bottom row depicts space-time luminance patterns of signals produced by passing spatio-temporal Gaussian white noise through the corresponding filter sets. **(a)** The "component" stimulus, constructed from a spatially and temporally band-pass filter. The x-y face of the stimulus shows structures that are spatially band-pass and oriented along the x axis. The orientation of structures on the x-t face indicates rightward motion. **(b)** The "plaid" stimulus, constructed from two "component" filters lying in a common plane. The x-y face of the stimulus shows a mixture spatial of spatial structures with dominant orientations close to the y axis. **(c)** The "planar" stimulus, constructed from a set of 10 "component" filters lying in a common plane. The stimulus is spatially band-pass and isotropic (x-y face), and exhibits rightward motion (x-t face).

internal noise embodies the added uncertainty created by neural noise, filter/signal mismatch and central decision noise. We computed an estimate of the internal noise by comparing subjects' thresholds with that of an ideal observer for the component stimulus. Assuming that the internal noise is approximately constant across all stimulus conditions, we derived optimal summation predictions for the plaid and planar stimuli. These predictions reflect the performance of ideal observers that sum energy only over the spatio-temporal frequency bands containing the signal in a stimulus and are corrupted by a constant level of internal noise (such a model gives the commonly cited "square-root law" for contrast summation).

Figure 3a shows subjects' detection thresholds for the component, plaid and planar stimuli. The heavy black lines on the plaid and planar plots show the predictions of ideal summation derived from the component data. The plaid data show little or no summation, while thresholds for the planar stimulus fall well within the range predicted by perfect summation. Figure 3b shows an alternative characterization of the results. We have plotted subjects' detection efficiencies for each of the three types of stimuli. Detection efficiencies are computed by comparing subjects' detection thresholds with that of ideal observers optimally tuned to the signals contained in the stimuli (note that the ideal observers are

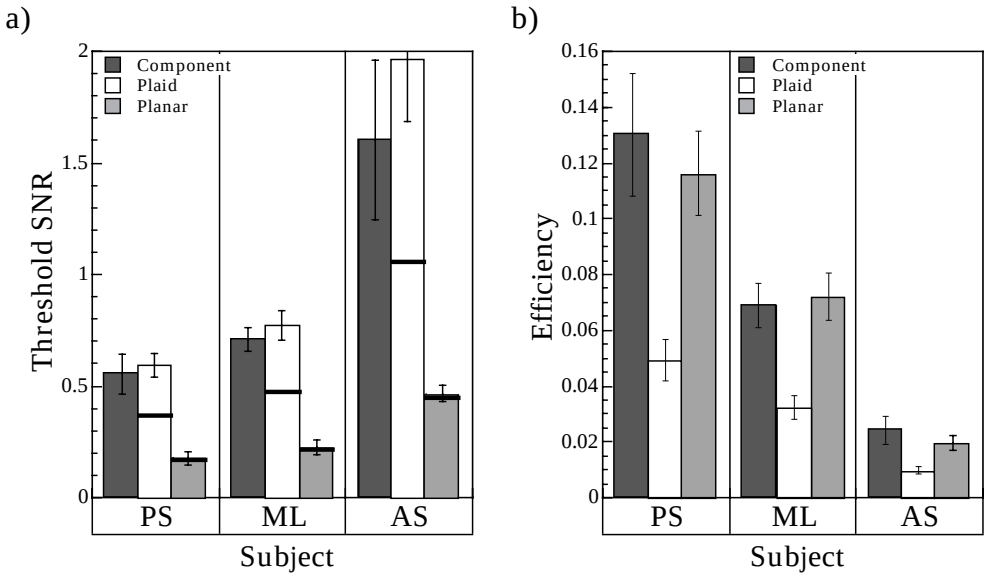

Fig. 3: (a) Detection performance of three subjects for the three stochastic signals of Fig. 2. Threshold signal to noise ratio (SNR) for 81.1% detectability. SNR is calculated as the ratio of the signal power to the background noise power. Heavy black lines indicate predictions for ideal summation, derived from the component condition thresholds. **(b)** Detection efficiencies for the three stimulus types. Efficiencies are plotted in proportions, with 1.0 reflecting perfect performance; that is, performance matching that of an ideal observer tuned to the structure of the signal in the stimulus (different for each stimulus type). The differences between the efficiencies of the pattern stimuli (plaid and planar stimuli) and the component stimulus provide a quantitative measure of summation of the pattern's components.

different for each of the three stimulus types). Efficiency provides a measure of the fraction of stimulus information effectively used by subjects in performing the detection task. Because perfect summation is ideal for these experiments, efficiency is also a measure of summation. The fact that detection efficiencies for the plaid stimulus were significantly lower than detection efficiencies for the component stimulus reflects the poor summation shown for the plaid. Conversely, the fact that detection efficiencies for the planar stimulus were equal to detection efficiencies for the component stimulus reflects the nearly optimal summation found for the planar stimulus.

Of the three detection models presented earlier, the data are clearly most consistent with the planar power detector model. The experiment, however, only compared summation performance for two types of "pattern" stimuli. In order to further test the model, we ran a second experiment, motivated by the observation that the planar power detector model predicts progressively improved summation when signal energy is distributed more broadly across spatial orientations in a plane, but not when energy is distributed out of a plane. We generated a new set of stochastic signals to test this prediction. Signals for the stimuli were created by passing spatio-temporal white noise through three configurations of filters, illustrated in Fig. 4a. The first is a plaid signal, similar to the plaid used in experiment 1. The second is a "planar triplet", created by adding a component band to the plaid, in the same plane as the plaid, and the third is a non-planar triplet, created by adding a component band out of the plane of the plaid. Detection thresholds were measured using the same method described for experiment 1.

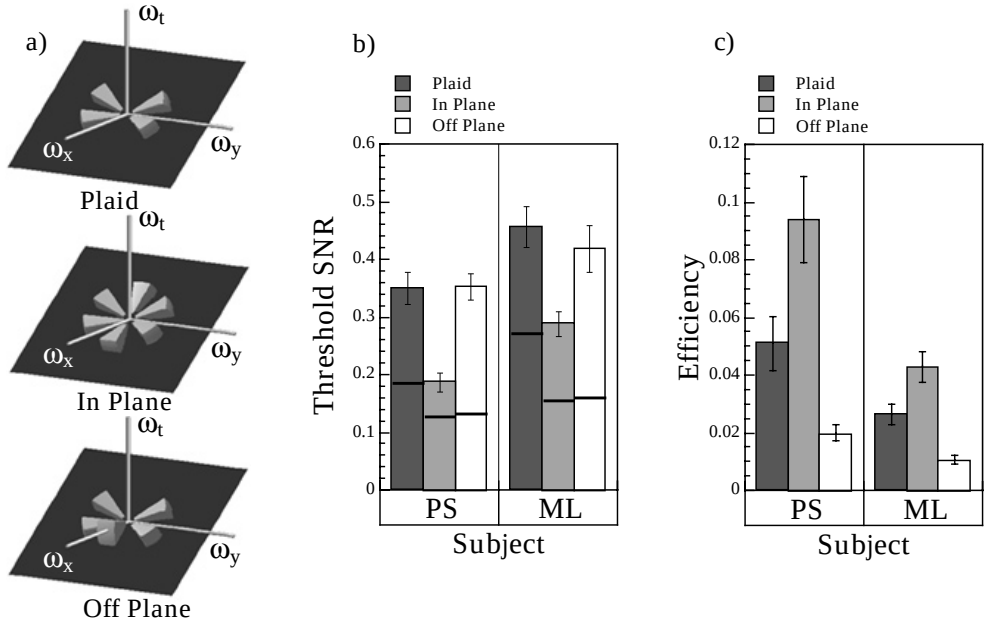

Fig. 4: (a) Threshold SNRs for detecting the three types of stimuli in experiment 2. Black bands indicate the predictions of ideal summation, based on subjects' detection thresholds for the component stimulus used in experiment 1. **(b)** Detection efficiencies for the three stimulus types. The results show a large increase in efficiency for the in plane triplet over the plaid and a decrease for the off-plane triplet, indicating increased summation with additional power on a common plane.

The planar power detector model predicts improved summation for the planar triplet, relative to the plaid, but no improved summation for the non-planar triplet.

Figure 4b shows detection thresholds for the three stimulus types used in experiment 2, along with the optimal summation predictions. Thresholds for the planar triplet were, as predicted, lower than for the plaid, while there was no significant difference between the the plaid and non-planar triplet thresholds. Figure 4c shows subjects' efficiencies for detecting each of the three signals used in the experiment. Detection efficiency was significantly better for the planar triplet than for the plaid - implying improved summation for the triplet stimulus. While detection efficiencies for the planar triplet are substantially better than for the plaid stimulus, subjects' efficiencies for the planar triplet remained lower than for the planar stimulus in experiment 1 (Fig. 3). This, again, is consistent with the planar power detector model, which predicts progressively better summation as signal energy is distributed more broadly over a plane in spatio-temporal frequency, with ideal summation reached only when the energy is distributed uniformly over the plane - a stimulus that best matches the putative detector.

Discussion

We can summarize the qualitative results of experiments 1 and 2 as follows: sub-threshold summation of signal contrast energy improves as the energy is distributed more and more broadly around different orientations in a plane in spatio-temporal frequency, but does not improve when the energy is distributed into non-planar regions of spatio-temporal frequency. The lack of summation for non-

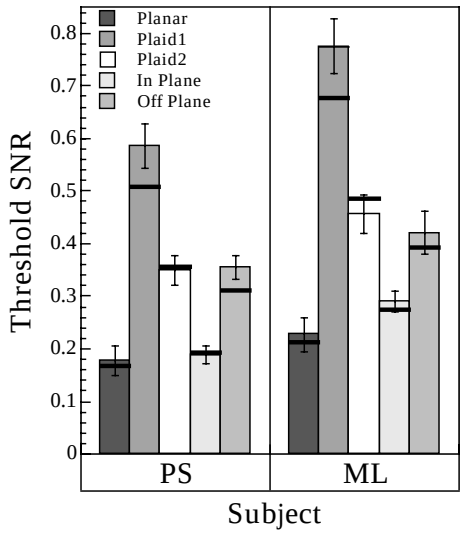

Fig. 5: Threshold SNRs for detecting the five types of pattern stimuli replotted from experiments 1& 2, where Plaid1 in the legend denotes the plaid from the first experiment and Plaid2 from the second. Plaid1 differs from Plaid2 in that its energy is more diffusely spread over frequency. Black bands indicate the predictions of a planar filter, based on subjects' detection thresholds for the component stimulus used in experiment 1.

planar regions agrees with previous studies suggesting a lack of generic summation for components moving in opposite directions.[14–16] Stated more plainly, detection efficiencies for moving stimuli improve as more orientations are added to a moving pattern , as long as the motion of those orientation components are consistent with the velocity of the pattern. Of the three detection models proposed in the introduction, these results are most consistent with the planar power detector model

We extended the analysis one step further by implementing a particular planar power detector and comparing its performance on the five different "pattern" stimuli used in the experiments with that of the subjects. The detector optimally summed energy over the band of frequencies contained in the planar stimulus from experiment 1 (using a matched "power" filter - see methods). We assumed that the output of this detector was corrupted by the internal noise levels estimated from subjects' detection thresholds for the component stimulus in experiment 1. Figure 5 shows the model predictions for the five pattern stimuli used. Given the assumptions built into the model concerning the exact spatio-temporal frequency band covered by the planar power detector, the match is surprisingly good. That is, not only do the qualitative results follow the predictions of the planar power detector model, but the quantitative results are well fit by a pre-defined instantiation of the model (without fitting the parameters of the model to the data).

Our experiments critically test the predictions of a simple model for velocity-tuned motion detectors in the mammalian visual system. The model detectors sum the outputs of early motion energy detectors with different preferred orientations, and whose spatio-temporal frequency bands intersect a common plane. The result is a family of velocity-tuned detectors that measure the stimulus energy in fixed, orientationally broad-band, planar regions of spatio-temporal frequency. Such mechanisms have been plausibly linked to the subclass of "pattern" neurons in the middle temporal (MT) visual area.[10] The

model may be contrasted with adaptive velocity estimation models that are dynamically tuned to the spatial orientation content of the image. Our sub-threshold summation results strongly support the fixed pooling model over adaptive models.

As with all psychophysical studies, we have had to make certain assumptions in order to draw conclusions about the underlying neural mechanisms. In particular, we have assumed that the detection of single component signals is mediated by the responses of first stage energy mechanisms, but that the detection of the compound signals is mediated by second stage responses. In addition, we assume that observers can learn (through training) to attend to the most sensitive mechanism for each stimulus.

Although our experiments strongly support the existence of second-stage mechanisms that pool over orientation, further processing is required to account for the full range of motion perception phenomena. For example, while planar pooling models are capable of representing more than one motion in a given local region,[10] they are not able to provide a consistent account of whether stimuli appear to move coherently. Additional computations are necessary to account for such phenomena. Finally, although our experimental results strongly suggest that pooling over *orientation* is fixed, they do not preclude the possibility of adaptive integration over spatial position, time, or spatial frequency (scale). Adaptive integration over spatial position, for example, would be a useful strategy for preventing the combination of information across occlusion boundaries. The experiments described herein may be naturally extended to examine the integration of first stage mechanisms over these other domains.

Methods

Stimuli in the experiment subtended 2.2 degrees of visual angle, had a duration of 0.43 sec, and were displayed in 12 bit precision on a calibrated gamma-corrected grayscale monitor. Observers fixated a small black dot 0.2 deg above the stimulus and were given several hours practice for each stimulus type over the course of several days until performance saturated. Observers were given auditory feedback after every trial during all experimental sessions. Only one signal type was used in a given 1 hour experimental session. Average signal energy was varied using the method of constant stimuli. The average background noise energy was fixed at 0.26 deg^2·sec.

Thresholds were estimated from the maximum likelihood fits of two-parameter Weibull functions, for which the threshold parameter α is a measure of signal to noise power at 81.1% correct for a 2AFC task. Standard errors were computed from both the Hessian of the likelihood function of the fit, and using a Monte Carlo sampling method.

The component filter has an amplitude spectrum given by $C(\omega_r, \omega_\theta, \omega_\phi) = R(\omega_r)|\cos(\omega_\theta)|^9 |\cos(\omega_\phi - \omega_{\phi_0})|^9$ in spherical frequency coordinates. The frequency radius ω_r is given by $\omega_r = \sqrt{\omega_x^2 + \omega_y^2 + (\omega_t/2.1)^2}$ (cyc / deg, cyc / deg, cyc / sec). The elevation shift $\omega_{\phi_0} = 36.9$ deg acts to rotate the filter out of the spatial frequency plane into a plane corresponding to a 1.93 deg / sec translation downward. The radial frequency function R is a smooth box function, whose transitions from 0 to 1 are given by 1/2 cycle of a cosine function. This function has low / high cutoffs of (0.49,7.6) cyc / deg and the cosine transition regions have widths of 1.45 cyc / deg. The plaid signal is constructed by adding two component filters, rotated by +70 and -70 degrees within the common plane. The planar signal is constructed using a sum

of 10 component filters, rotated by multiples of 18 degrees and constrained to lie in a common plane. For the second experiment, the filters were modified to decrease the spectral spread. They have the form $BP(\omega_r, wq, \omega_\phi) = W_r(\omega_r)W_\theta(\omega_\theta)W_\phi(\omega_\phi)$ where W_x is a smooth box function on the variable x. W_r had a transition region width of 1.45, and low-high frequency cutoffs of (0.49,7.6). W_θ and W_ϕ had a transition widths of 8 degrees, and the high low cutoffs that spanned 36 degrees. These filters were rotated to lie in the same positions as the component and plaid filters for the plaid and in plane triad. The Off plane component, however, had an elevation shift corresponding to a 0.26 deg / sec translation downward.

Ideal observers for the task are matched filter power detectors.[17] Let $F(\omega_x, \omega_y, \omega_t)$ denote the spectrum of the signal normalized to one, and $N(\omega_x, \omega_y, \omega_t)$ the (complex) spectrum of the noise. Let $|X(\omega_x, \omega_y, \omega_t)|^2 = X(\omega_x, \omega_y, \omega_t)X(\omega_x, \omega_y, \omega_t)^*$ denote the inner product of the complex function with its complex conjugate. Then the spectrum of the received signal $|S(\omega_x, \omega_y, \omega_t)|^2$ is $s^2|F(\omega_x, \omega_y, \omega_t)|^2 + N(\omega_x, \omega_y, \omega_t)^2$ on the signal plus noise interval and $|N(\omega_x, \omega_y, \omega_t)|^2 = \mathbf{N}^2$ on the noise alone interval. The best possible performance can be achieved by computing the energy

$$E = \int |F(\omega_x, \omega_y, \omega_t)|^2 |S(\omega_x, \omega_y, \omega_t)|^2 d\omega_x d\omega_y d\omega_t$$

over both signal plus noise and noise intervals and uses the difference of these values as a decision variable.[12] In practice the set of frequencies are quantized, and the integral is computed as a sum over 1.3×10^5 frequency samples.

Efficiencies are approximately given by the ratio of ideal to human squared energy thresholds. This approximation was corrected through simulations of the ideal, and the simulation values are shown in the figures.

Performance of the matched filter depends on the probability $p(E_s - E_n > 0)$, that the difference in energies measured by the filter on two intervals is greater than zero. To compute this probability, we derive an approximate expression for the distribution of $E_s - E_n$. For the stimuli used, the amplitude spectrum of the signal $S(\omega_x, \omega_y, \omega_t)$ consists of independent (real and imaginary) Gaussian distributed random variables at each frequency with zero mean and variance given by either $s^2|F(\omega_x, \omega_y, \omega_t)|^2 + |N(\omega_x, \omega_y, \omega_t)|^2$ on the signal interval or $|N(\omega_x, \omega_y, \omega_t)|^2$ on the noise alone interval. Now the power spectrum, as the product of the signal amplitude with its complex conjugate, consists of Chi-squared random variables with two degrees of freedom ($\chi^2(2)$) at each frequency, which are scaled by the squared amplitude spectrum of the signal. Because the energy is the sum of an extremely large number of $\chi^2(2)$ random variables (one for each frequency), weighted by the product of the filter and the signal power spectrum, we can use the law of large numbers to generate an extremely good bound on performance ($<<$ 1% error from true performance). By the law of large numbers, $E_s - E_n$ is Gaussian distributed, and hence can be characterized by its mean and variance. The mean and variance of $E_s - E_n$ are simply the sums of the means and variances of the $\chi^2(2)$ variables at each frequency. Computing the means and variances of the $\chi^2(2)$ variables that have been scaled by $|F(\omega_x, \omega_y, \omega_t)|^2 \cdot (s^2|F(\omega_x, \omega_y, \omega_t)|^2 + |N(\omega_x, \omega_y, \omega_t)|^2)$ on the signal interval and $|F(\omega_x, \omega_y, \omega_t)|^2 \cdot |N(\omega_x, \omega_y, \omega_t)|^2$ on the noise interval, and letting $F^m S^n = \sum_i |F(\omega_{xi}, \omega_{yi}, \omega_{ti})|^m |S(\omega_{xi}, \omega_{yi}, \omega_{ti})|^n$, then with a little algebra, the mean and variance of $E_s - E_n$ are given by: $\mu_{id} = 2s^2F^2F^2$ and $\sigma_{id}^2 = 8(s^4F^4F^4 + 2s^2F^4F^2\mathbf{N} +$

$2F^4\mathbf{N}^2$) respectively. Then $p(E_s - E_n > 0) = 1 - \Phi(0, \mu_{id}, \sigma_{id}^2)$, where Φ is the cumulative Gaussian distribution.

To compute summation predictions, we modeled component performance as ideal but with the variance given by $\sigma_{id}^2 + \sigma_{internal}^2$, so that $p(E_s - E_n > 0) = 1 - \Phi(0, \mu_{id}, \sigma_{id}^2 + \sigma_{internal}^2)$. Setting $p(E_s - E_n > 0) = 81.1$ and plugging in the subject's threshold into μ_{id} and σ_{id}^2, we solved for $\sigma_{internal}^2$. To generate predictions, the $\sigma_{internal}^2$ was then added to the ideal variance for each of the other signal types, and a predicted threshold for 81.1% was computed. Predictions for the planar model were generated in a similar manner. If we represent the planar filter amplitude spectrum as $P(\omega_x, \omega_y, \omega_t)$, then the planar model computes the energies $E_{pl} = \sum |P(\omega_x, \omega_y, \omega_t)|^2 |S(\omega_x, \omega_y, \omega_t)|^2$ on both intervals. The decision variable mean and variance are then given by: $\mu_{id} = 2s^2P^2F^2$ and $\sigma_{id}^2 = 8(s^4P^4F^4 + 2s^2P^4F^2\mathbf{N} + 2P^4\mathbf{N}^2)$. Threshold predictions were generated from these expressions using the internal noise estimated from the observer's component stimulus performance.

References

1. Nakayama, K. Biological image motion processing: a review. *Vis. Res.*, **25**, 625–660, (1985).

2. Adelson, E. H. & Movshon, J. A. Phenomenal coherence of moving visual patterns. *Nature*, **300**, 523–525, (1982).

3. Albright, T. D. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiol.*, **52**, 1106–1130, (1984).

4. Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. in *Exp. Brain Res. Supplementum II: Pattern Recognition Mechanisms* (eds. Chagas, C., Gattass, R., & Gross, C.) 117–151, (Springer-Verlag, New York, 1986).

5. Morgan, M. J. Spatial filtering precedes motion detection. *Nature*, **355**, 344–346, (1992).

6. Watson, A. B. & Ahumada, A. J. in *Motion: Perception & representation* (ed. Tsotsos, J. K.) 1–10, (Association for Computing Machinery, New York, 1983).

7. Adelson, E. H. & Bergen, J. R. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A*, **2**, 284–299, (1985).

8. Heeger, D. J. Model for the extraction of image flow. *J. Opt. Soc. Am. A*, **4**, 1455–1471, (1987).

9. Grzwacz, N. M. & Yuille, A. L. A model for the estimate of local image velocity by cells in the visual cortex. *Proc. Royal Society of London A*, **239**, 129–161, (1990).

10. Simoncelli, E. P. & Heeger, D. A model of neuronal responses in visual area MT. *Vis. Res.*, **38**, 743–761, (1998).

11. Derrington, A. M. & Suero, M. Motion of complex patterns is computed from the perceived motions of their components. *Vis. Res.*, **31**, 139–149, (1991).

12. Kersten, D. Statistical efficiency for the detection of visual noise. *Vis. Res.*, **27**, 1029–1043, (1987).

13. Watson, A. B. & Turano, K. The optimal motion stimulus. *Vis. Res.*, **35**, 325–336, (1994).

14. Levinson, E. & Sekuler, R. The independence of channels in human vision selective for direction of movement. *J. Physiol.*, **250**, 347–366, 1975.

15. Watson, A. B., Thompson, P. G., Murphy, B. J. & Nachmias, J. Summation & discrimination of gratings moving in opposite directions. *Vis. Res.*, **20**, 341–347, (1980).

16. Dobkins, K. R. & Teller, D. Y. Infant contrast detectors are selective for direction of motion. *Vis. Res.*, **36**, 281–294, (1996).

17. van Trees, H. L. *Detection, Estimation, and Modulation Theory. Part III*, (Wiley, New York, 1971).

Correspondence: All correspondence should be addressed to Paul Schrater (Email: schrater@eye.psych.umn.edu)

ACKNOWLEDGEMENTS. P. S. was supported by an NIH training grant, D. C. K. was supported by a grant from the NIH, and E.P.S. was supported by a Sloan Fellowship, an NSF CAREER grant, and the Sloan Program in Theoretical Neurobiology at NYU.