Published in: **Nature**, volume 410, pages 816-819, 12 April 2001.

Perceiving visual expansion without optic flow

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

[1] Department of Neuroscience, University of Pennsylvania,
215 Stemmler Hall, Philadelphia, PA 19104, USA

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

[3] Howard Hughes Medical Institute, and
Center for Neural Science, New York University
4 Washington Place, 809, New York, NY 10003, USA

When an observer moves forward in the environment, the image on his or her retina expands. The rate of this expansion conveys information about the observer's speed[1] and time-to-collision.[2–4] Psychophysical[5–7] and physiological[8,9] studies have provided abundant evidence that these expansionary motions are processed by specialized mechanisms in mammalian visual systems. It is commonly assumed that the rate of expansion is estimated from the divergence of the optic flow field (the two-dimensional field of local translational velocities).[10–14] But it could also be estimated from changes in the size (or *scale*) of image features.[15] To determine whether human vision uses such scale-change information, we have synthesized stochastic texture stimuli in which the scale of image elements increases gradually over time, while the optic flow pattern is random. Using these stimuli, we show that observers can estimate expansion rates from scale change information alone, and that pure scale changes can produce motion after-effects. These two findings suggest that the visual system contains mechanisms that are explicitly sensitive to changes in scale.

To test the hypothesis that the rate of image expansion on the retina might be estimated from changes in scale (Fig. 1), we designed three types of stochastic expansion stimuli (Fig. 2). We created a sequence of uncorrelated white noise images, and convolve each one with a bandpass filter whose centre frequency decreases exponentially with time. The result is a sequence of random textures whose average spatial scale grows according to a fixed rate of expansion. Because the individual stimulus frames are uncorrelated, both the optic flow field and its divergence are (on average) zero, and thus cannot be used to compute expansion rate. Nevertheless, most people who have viewed these stochastic stimuli report a strong percept of expansion, emanating from the point of fixation. This suggests that humans have mechanisms sensitive to changes in scale per se.

To test this hypothesis more rigorously, we measured the ability of subjects to match the expansion rate of our stochastic stimuli to the expansion rate of a field of dots undergoing constant expansionary motion. The dots in the latter stimuli have a limited lifetime (Fig. 2(b), and Methods), and thus contain optic flow information about the rate of expansion but minimal scale-change information. On each trial, observers fixated on a point lying between a stochastic expansion stimulus and an expanding dots stimulus. The stimulus movies were repeated while the observer adjusted the rate of expansion of the dots until it appeared to match the expansion rate of the stochastic stimulus. We measured the matching expansion rates for four different stimulus conditions: two durations and two average texture scales. In order to provide a standard against which to compare these matches, observers were also asked to perform the same task using deterministic texture expansion movies with identical parameters (see Fig. 2(c)).

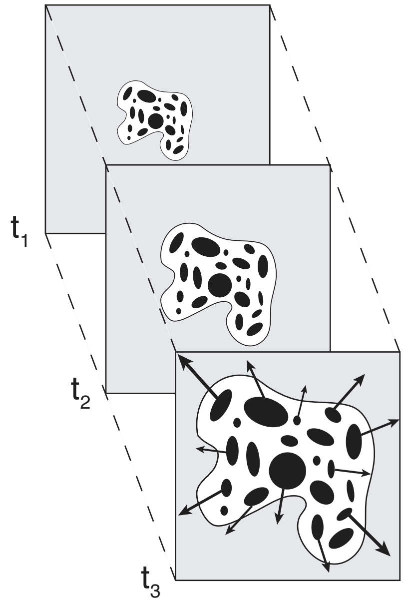
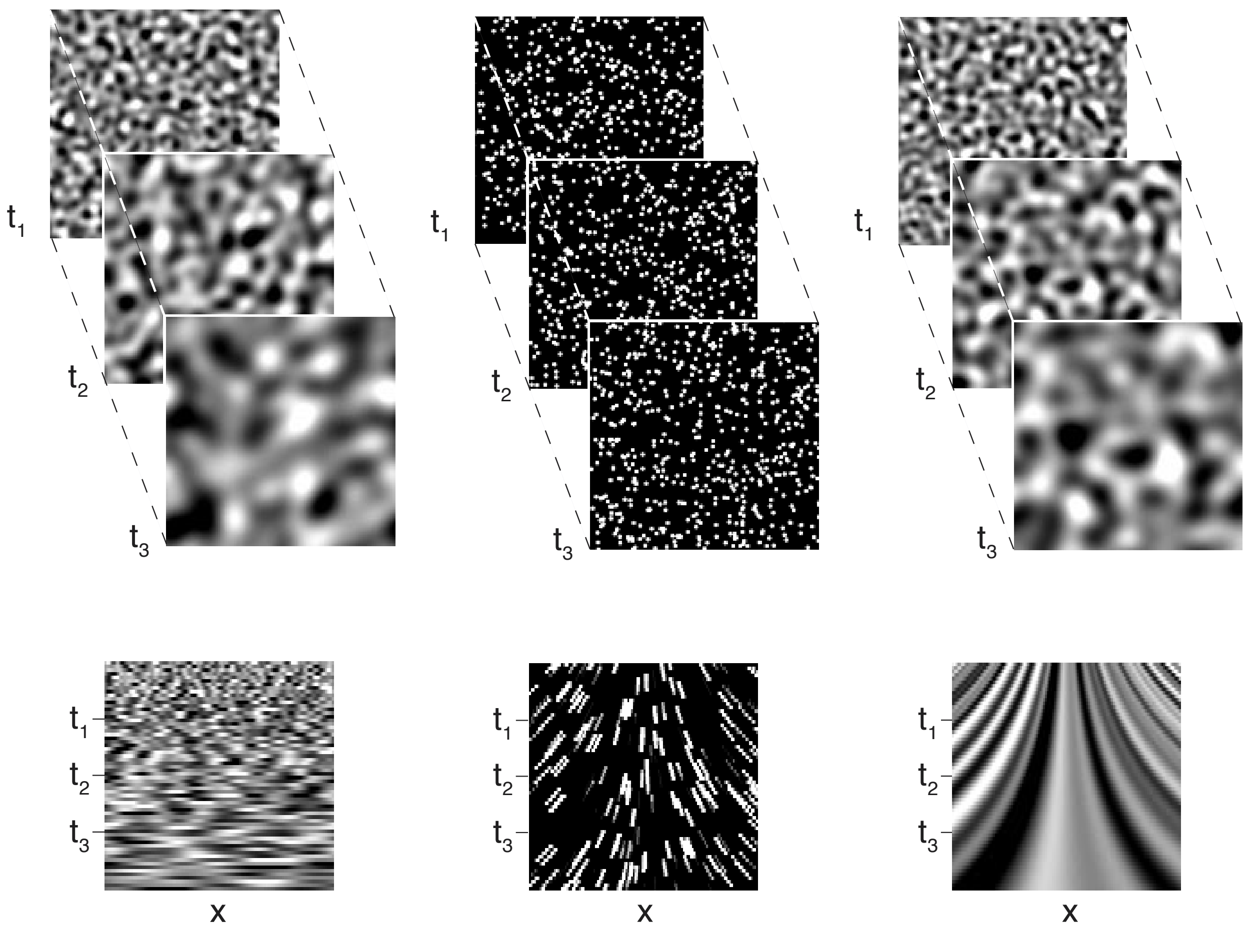


Figure 1: Illustration of visual input when an observer moves forward in the environment. Shown are three frames taken from a sequence of visual images (a movie). Superimposed on the last frame are example optic flow vectors, indicating the local translational motion of selected points in the scene. The radiating pattern of these vectors indicates that the observer is moving forward. The divergence of this vector field provides one source of information about the rate of expansion. The change in scale of the objects or texture elements provides another source of information.

Figure 3 shows the results of the matching experiment for both stochastic and deterministic texture expansion stimuli. All observers show consistent increases in matched dot expansion rates as a function of the expansion rate simulated by the texture stimuli. Although matches are sometimes quite far from veridical, they are no more so for the stochastic stimuli than for the deterministic stimuli. Linear regression on the matching data shows some variation in slope and intercept across the four different stimulus conditions, but no consistent pattern emerges across subjects.

Although individual subjects show differences in the slopes of matching functions between stochastic and deterministic stimuli, group averages show no significant difference in the average slope for the two types of stimuli (ANOVA, F(1,38) = 0.49, p = 0.49, mean slopes 0.91 & 0.95, respectively). In addition, group averages of the r^2 values of the fits are not significantly different (T-Test, $t(38) = 1.6$, p = 0.11, mean r^2: 0.44 & 0.52, respectively). Note, however, that subjects' matches for the deterministic stimuli do have a lower variance than those for the stochastic stimuli (av. std. 0.53 & 0.65; F(1,158) = 12.7, p<0.001). This implies that observers can use the optic flow information available in the deterministic expansion stimuli to improve the reliability of their matches.

The matching experiment leaves open the possibility that subjects use a cognitive strategy to estimate the expansion rates of stochastic expansion stimuli (for example, by estimating the static texture size at two or more moments in time and dividing by the time interval between them). To rule this out, we performed an additional experiment to measure motion aftereffects induced by the stochastic stimuli, since these effects are generally believed to be insensitive to cognitive intrusion.[16,17] We used an experimental design analogous to one used previously to measure adaptation to fields of coherent translating dots.[18] In these experiments, adaptation stimuli are followed by test stimuli containing mixtures of leftward and rightward moving dots. Adaptation to rightward motion causes a balanced



(a) (b) (c)

Figure 2: Illustration of the three types of stimuli used in our experiments. Top, show example $x - y$ frames of the stimulus movie; bottom, $x - t$ cross sections. **(a)** Stochastic texture stimulus. Each frame contains a random texture synthesized by convolving a Gaussian white noise image with a bandpass filter. The peak spatial frequency of the bandpass filter decreases over time, producing a gradual change in scale of texture elements. The $x - t$ slice illustrates the lack of consistent displacement information. Because each frame is generated from independent white noise, the resulting optic flow field is random, and thus cannot be used to estimate expansion rate. **(b)** Limited-lifetime random dots stimulus. Dots appear asynchronously in random locations, move according to the global expansion pattern for five frames, and then disappear. For this stimulus, the optic flow field may be used to estimate expansion rate, but there is minimal scale-change information. **(c)** Deterministic texture stimulus. Each frame is a re-scaled replica of a single image of bandpass-filtered Gaussian white noise. Both optic flow and scale changes provide information about the expansion rate.

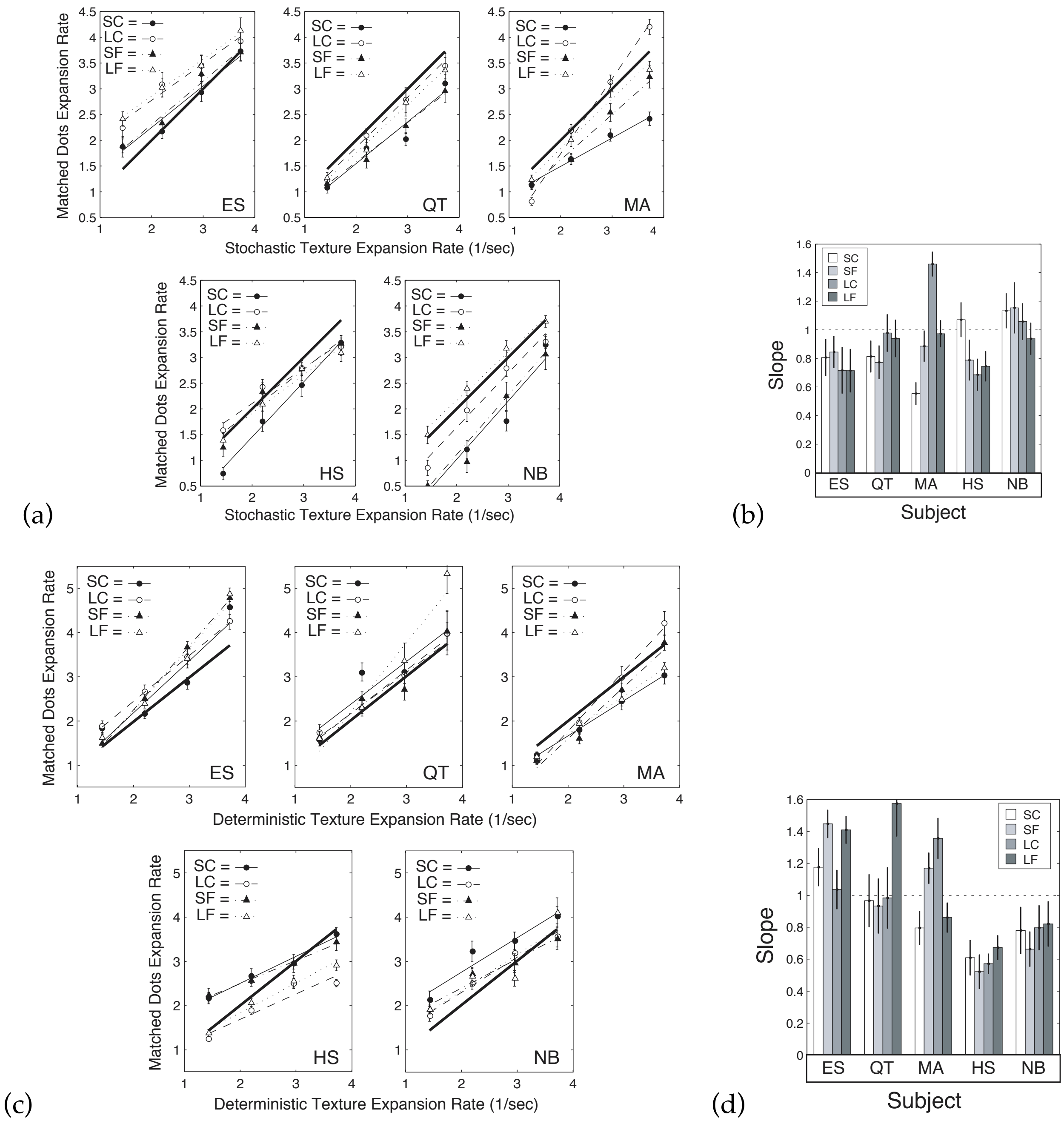


Figure 3: Perceptual matches of expansion rate between dot and texture stimuli. **(a)** Perceptually matched expansion rate of dot stimuli as a function of the stochastic texture expansion rate for five observers. Symbols indicate mean of 20-30 trials; bars indicate standard error. Stimulus conditions are denoted by different plot symbols: black = short duration (0.43 sec); white = long duration (0.73 sec); circles = course scale (geometric mean spatial frequency = 0.9 cyc/deg); triangles = fine scale (2.1 cyc/deg). Thin lines indicate least-squares linear fits for each condition; thick line indicates a veridical match (unit slope). **(b)** Slope estimates from (a). Error bars indicate 95% confidence intervals. Dashed horizontal line denotes a veridical match. **(c)** Perceptual matches between expansion rates of deterministic texture stimuli and dot stimuli (details as in (a)). **(d)** Slope estimates from (c).

mixture to appear to move leftward. The strength of this aftereffect is measured as the percentage of test stimulus rightward dots required to produce a percept of no net motion.

Our experiment used stochastic expansion textures as adaptation stimuli. Test stimuli were created by mixing frames from expanding and contracting stochastic texture movies. In particular, we first created a set of stochastic expansion movies with frequency and duration parameters identical to those of the adaptation stimuli. We converted half of these into contraction movies by reversing the frame order. Ambiguous stimuli were created by selecting each frame at random from either an expansion or contraction movie.

A mixture parameter, α, controlled the proportion of frames drawn from expansion movies (see Methods for details). Thus, $\alpha = 1$ corresponds to a stimulus whose frames are all drawn from expansion movies. On each trial, subjects viewed the adaptation stimulus for 0 (no adaptation), 1 or 5 seconds, followed by a test stimulus, after which they indicated whether the test stimulus appeared to be expanding or contracting. The α-value used to create the test stimuli was adjusted using a psychometric procedure to produce "expanding" responses 20%, 50%, and 80% of the time. The resulting data were fitted with a cumulative Normal function with a constant slope across adaptation conditions (nested model comparison tests did not support the use of multiple slopes for any of the subjects: $\chi^2(2) < 0.2$ for all subjects).

Figure 4 shows the null (50%) points estimated for each of five subjects as a function of adaptation time. With no adaptation, three of five observers perceived the ambiguous stimuli as expanding, consistent with previously reported biases for forward/backward motion perception.[19] As the adaptation time increases, subjects' expansion null points increase, indicating that adaptation to stochastic expansion stimuli produces a bias toward perceiving the test stimuli as contracting. After five seconds of adaptation, the bias is substantial: stimuli that would ordinarily have been judged to be expanding 99% of the time became completely ambiguous. In addition, the fact that the effect increases with adaptation duration indicates that the shifts are caused by a real aftereffect rather than an induction or priming effect.

The results of our matching and adaptation experiments provide compelling evidence for the hypothesis that the human visual system uses scale-change information to compute expansion rate.[15] The strength of the evidence is due to the stochastic texture stimuli, which neatly separate the optic flow information from the scale-change information. Related stereoscopic stimuli have been used to show that the percept of surface slant from stereopsis does not require point-wise disparity estimates,[20] and related sound signals have been used to show neural sensitivity to frequency modulations in the mammalian auditory system.[21]

The mechanism by which scale-change information is processed, and the method by which it is combined with optic flow information remains unknown. Two basic possibilities exist. The visual system might use a single mechanism that is sensitive to both scale-change and optic flow information. Alternatively, the visual system might have separate mechanisms for computing expansion from scale-change and optic flow information, combining the outputs of the two to arrive at a unitary estimate. For example, expansion can be computed separately from scale-change information by estimating the "motion" of the signal's Fourer energy pattern along the spatial frequency axis. Given a set of neurons whose responses compute local Fourier energy in a frequency band (e.g., V1 complex cells), this computation could be neurally implemented as a time-varying pooling of these neurons' responses adjusted at each moment to track the expected scale of the expanding image. These issues might be best addressed in the context of physiological studies using the stimuli introduced in this paper.

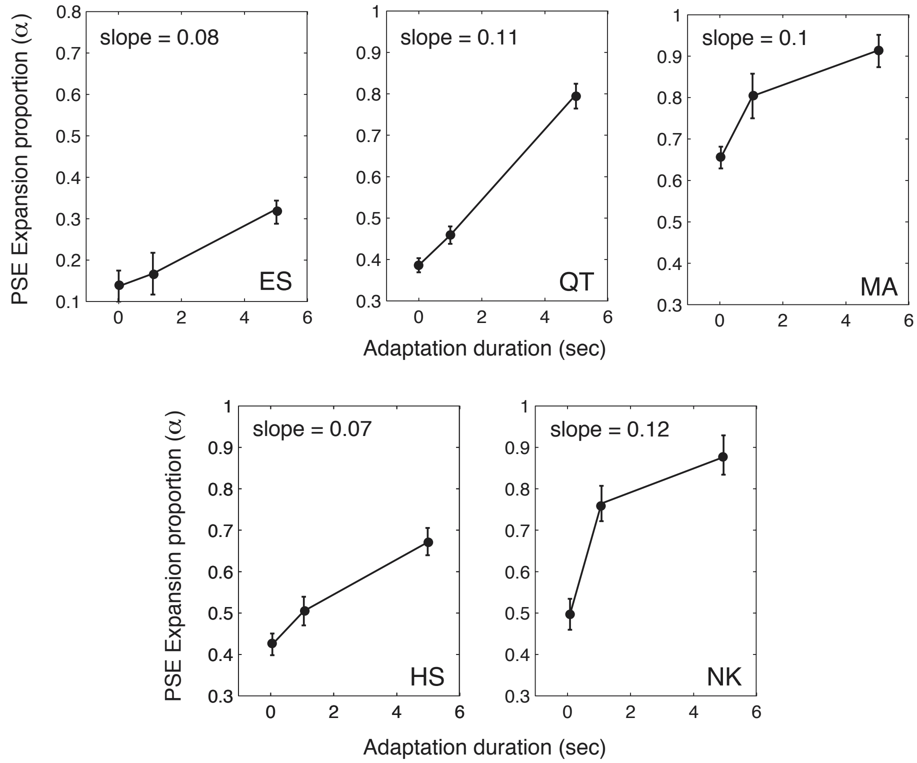


Figure 4: Effect of adapting to stochastic expansion stimuli on the perception of an ambiguous test stimulus for five subjects. The value of α yielding 50% expansion/contraction judgments is plotted against adaptation duration (zero corresponds to no adaptation). Bars indicate standard errors. Also indicated is the slope of the psychometric function fitted to the raw data, that was constant across different adaptation conditions.

Methods

All stimuli were displayed at 60 Hz in 8 bits of precision on a calibrated gray-scale monitor. Stimuli were viewed monocularly from a distance of 55.7 cm, and subtended 16.5 degrees of visual angle. Stochastic textures were constructed by convolving independent frames of Gaussian white noise with bandpass filters that are described in the spatial frequency domain by: $F(\omega_r) = 0.5{+}0.5\,\cos_r \pi \log_2(\omega_r/2f(t))$, where $\cos_r$ is a cosine function restricted to one cycle, $f(t)$ is the peak frequency of the filter at time t, and $\omega_r = \sqrt{\omega_x^2 + \omega_y^2}$.

For a constant expansion rate a, $f(t) = \exp(\log(f_{\text{initial}}) - a \cdot t)$. Four different expansion rates were used: $(0.96, 1.44, 2.0, 2.75)$ sec^{-1}. For each rate, two different durations (0.43 and 0.73 sec) and two geometric-mean frequencies (0.9 and 2.1 cyc/deg) were used. For a given geometric mean frequency f_{mean}, the initial and final frequencies (f_{initial} and f_{final}) were defined by $\log(f_{\text{initial,final}}) = a \cdot T/2 \pm \log(f_{\text{mean}})$, where T is the duration. 125 different versions of each stochastic stimulus type were constructed by interleaving every third frame of 5 different movies. The deterministic stimuli were made from the stochastic texture stimuli by taking the first frame and performing band-limited sub-sampling and interpolation to digitally expand the initial image. For the dots stimuli, 200 constant size (4 pixels), randomly placed dots were moved according to $x(t{+}1) = x(t) + a \cdot x(t)$, which approximates $dx/dt = a \cdot x$. Dots had limited lifetimes of 5 frames (83 ms), and expired asynchronously. When the lifetime was exceeded, or when dots advanced past the edge of the image, they were randomly re-placed in the image. This procedure produces a constant expected density of dots across the image, and prevents the dots from carrying significant size change information.

Observers were asked to match the perceived expansion rate of the stochastic textures with that of the dots. Observers were given no feedback, and were queried at the end of the experiment regarding the phenomenal appearance of the stimuli and their matching strategy. Each trial consisted of the following events: observers fixated on a spot between a stochastic texture movie centered 9.1 degrees left, and a dots movie centered 9.1 degrees right of fixation. Both movies were repeated with a 0.33 sec delay between cycles. At any time observers could press keys to increase or decrease the expansion rate of the dots movie, or indicate that the expansion rates appeared to match. Between 24-34 matches were collected for each of the 16 different stochastic texture movies, the first 4 of which were discarded to remove early learning effects. The observer's matches for each condition were fit by least-squares linear regression.

Ambiguous test stimuli for the adaptation experiment were constructed by sampling frames from stochastic texture expansion movies, such that the nth frame of the test movie is copied from the kth frame of an expanding stochastic texture movie according to the distribution: $p(k|n, \alpha) = \alpha\phi(n) + (1 - \alpha)\phi(N - n + 1)$. $\phi(n)$ denotes a Normal density function with mean n and variance 1.5, rounded to the nearest integer, and N is the total number of frames in the movie. Observers viewed these stimuli centrally and after each presentation chose whether the stimuli appeared to be expanding or contracting. The QUEST adaptive psychophysical procedure[22] was used to estimate the α producing 20%, 50%, and 80% expansion responses. The test stimuli were preceded by adaptation stimuli of three different durations (0, 1, and 5 sec). The 5 sec adaptation condition also included an initial 20 sec adaptation period. For each adaptation condition, 400 trials were collected and all response data were fit by a cumulative Normal psychometric function via maximum likelihood. 50% null points were determined from these psychometric function fits, and standard errors of these estimates were estimated using 1000 fits to bootstrap re-samples of the response data. Differences in slope across adaptation duration were assessed using a nested-hypothesis test procedure.[23] The test determines whether the goodness-of-fit of a model with independent slopes for each adaptation duration is significantly better than a single slope model by computing $x = 2\log_e \frac{L_i}{L_r}$, where L_i & L_r are the likelihoods of the fit for the independent and reduced models, respectively. The statistic x is χ^2 distributed with 2 degrees of freedom, computed from the difference in the number of free parameters in the two models.

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

Acknowledgements: We thank the anonymous reviewers for many helpful suggestions. This work was begun when all three author were at the University of Pennsylvania. Paul Schrater was supported by training grants from NEI at the University of Pennsylvania and later at the University of Minnesota, Eero Simoncelli was supported by the National Science Foundation and the Alred P. Sloan Foundation, and David Knill was supported by a research grant from NIH.

References

1. C. W. Clifford, S. A. Beardsley, and L. M. Vaina. The perception and discrimination of speed in complex motion. *Vision Res*, 39(13):2213–27, 1999.

2. D N Lee. A theory of visual control of braking based on information about time-to-collision. *Perception*, 5:437–459, 1976.

3. W. Schiff and M. L. Detwiler. Information used in judging impending collision. *Perception*, 8(6):647–58, 1979.

4. M. K. Kaiser and H. Hecht. Time-to-passage judgments in nonconstant optical flow fields. *Percept Psychophys*, 57(6):817–25, 1995.

5. D Regan and K I Beverley. Illusory motion in depth: aftereffect of adaptation to changing size. *Vis. Res.*, 18(2):209–212, 1978.

6. D Regan and K I Beverley. Binocular and monocular stimuli for motion in depth: changing-disparity and changing-size feed the same motion-in-depth stage. *Vis. Res.*, 19(12):1331–1340, 1979.

7. M C Morrone, D C Burr, S Di Pietro, and M Stefanelli. Cardinal directions for visual optic flow. *Current Biology*, 9:763–766, 1999.

8. GA Orban, L Lagae, A Verri, S Raiguel, D Xiao, H Maes, and V Torre. First-order analysis of optical flow in monkey brain. *Proc. Natl. Acad. Sci.*, 89:2595–9, 1992.

9. C J Duffy and R H Wurtz. Sensitivity of mst neurons to optic flow stimuli: I. a continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, 65(6):1329–1345, 1991.

10. J J Gibson. *The Perception of the Visual World*. Houghton Mifflin, Boston, 1950.

11. W H Warren, M W Morris, and M L Kalish. Perception of translational heading from optical flow. *J Exp Psychol Hum Percept Perform*, 14(4):646–660, 1988.

12. W H Warren, A W Blackwell, K J Kurtz, N G Hatsopoulos, and M L Kalish. On the sufficiency of the velocity field for perception of heading. *Biol Cybern*, 65(5):311–320, 1991.

13. J A Perrone. Model for the computation of self-motion in biological systems. *J Opt Soc Am A*, 9(2):177–194, 1992.

14. J A Crowell and M S Banks. Perceiving heading with different retinal regions and types of optic flow. *Percept Psychophys*, 53(3):325–337, 1993.

15. K I Beverley and D Regan. Texture changes versus size changes as stimuli for motion in depth. *Vis. Res.*, 23(12):1387–1400, 1983.

16. S M Anstis. The perception of apparent movement. *Phil. Trans. R. Soc. Lond. B Biol. Sci.*, 290:153–68, 1980.

17. P Cavanagh and G Mather. Motion: the long and short of it. *Spatial Vision*, 4:103–29, 1989.

18. R Blake and E Hiris. Another means for measuring the motion aftereffect. *Vis. Res.*, 33:1589–1592, 1993.

19. J. A. Perrone. Anisotropic responses to motion toward and away from the eye. *Percept Psychophys*, 39(1):1–8, 1986.

20. C Tyler and E Sutter. Depth from spatial frequency difference: An old kind of stereopsis? *Vis Res*, 19:859–865, 1979.

21. R Tian and J P Rauschecker. Processing of frequency-modulated sounds in the cat's posterior auditory field. *J. Neurophysiol.*, 79:2629–2642, 1998.

22. A B Watson and D G Pelli. Quest: A Bayesian adaptive psychophysical method. *Perception & Psychophysics*, 33:113–120, 1983.

23. M K Kendall and A Stuart. *The advanced theory of statistics: Design, analysis and time series*, volume 3. Hafner Publishing Co., New York, 1966.