



Feature Integration in Pattern Perception

Dennis M. Levi; Vineeta Sharma; Stanley A. Klein

Proceedings of the National Academy of Sciences of the United States of America, Vol. 94, No. 21 (Oct. 14, 1997), 11742-11746.

Stable URL:

<http://links.jstor.org/sici?sici=0027-8424%2819971014%2994%3A21%3C11742%3AFIIPP%3E2.0.CO%3B2-B>

Proceedings of the National Academy of Sciences of the United States of America is currently published by National Academy of Sciences.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/nas.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Feature integration in pattern perception

DENNIS M. LEVI*, VINEETA SHARMA*, AND STANLEY A. KLEIN†

*College of Optometry, University of Houston, Houston, TX 77204-6052; and †School of Optometry, University of California, Berkeley, CA 94720

Communicated by Bela Julesz, Rutgers University, Piscataway, NJ, August 7, 1997 (received for review June 10, 1997)

ABSTRACT The human visual system is able to effortlessly integrate local features to form our rich perception of patterns, despite the fact that visual information is discretely sampled by the retina and cortex. By using a novel perturbation technique, we show that the mechanisms by which features are integrated into coherent percepts are scale-invariant and nonlinear (phase and contrast polarity independent). They appear to operate by assigning position labels or “place tags” to each feature. Specifically, in the first series of experiments, we show that the positional tolerance of these place tags in foveal, and peripheral vision is about half the separation of the features, suggesting that the neural mechanisms that bind features into forms are quite robust to topographical jitter. In the second series of experiment, we asked how many stimulus samples are required for pattern identification by human and ideal observers. In human foveal vision, only about half the features are needed for reliable pattern interpolation. In this regard, human vision is quite efficient (ratio of ideal to real ≈ 0.75). Peripheral vision, on the other hand is rather inefficient, requiring more features, suggesting that the stimulus may be relatively underrepresented at the stage of feature integration.

The human visual system is able to effortlessly integrate local features to form our rich perception of patterns, despite the fact that visual information is discretely sampled by the retina and cortex. Feature integration has been previously studied in the context of figure-ground segregation or detection (1–6); however, we were interested in how the integration of supra-threshold features into a recognizable pattern depends upon the number and spatial arrangement of the features. This information would provide important clues about the mechanisms involved in feature integration and their topographical precision.

To investigate the mechanisms involved in the integration of features for pattern discrimination, observers judged the orientation (up, down, left, and right) of an E-like pattern (Fig. 1 *Inset*) that was constructed from circular Gabor or Gaussian features or “samples.” We chose Gabor patches, because they closely match the receptive field properties of neurons (or feature detectors) known to exist in primate visual cortex (7). In the first series of experiments, we investigated the tolerance to perturbation of the positions of the features defining the pattern; in the second experiment, we asked how many stimulus samples are required for pattern identification by human and ideal observers. Our results suggest that the neural mechanisms that bind features into coherent patterns are quite robust to topographical perturbations and that in the fovea, they are very efficient.

METHODS

The E-shaped pattern was composed of 17 Gabor features (i.e., the luminance distribution of each element is described by the

product of a circular Gaussian and an oriented sinusoid) presented on a display monitor with a mean luminance of 56 cd/m². The Gabor elements formed an E-shaped global pattern presented in one of four rotated orientations (up, down, left, or right), and the observer’s task was to identify the pattern orientation (i.e., a four-alternative forced choice). On each trial, an E-pattern was presented for ≈ 0.50 sec (accompanied by a tone) in the center of the screen, after which the observer gave her/his response by pressing one of four buttons, indicating the orientation of the global pattern. Visual feedback was provided after each response. To measure the “jitter” thresholds, we subjected the two-dimensional positions of each feature to position jitter. The magnitude of jitter was sampled from a single Gaussian distribution, with a random angle, by using the method of constant stimuli. For each patch, the angle was sampled from a uniform distribution of integer angles between 0 and 359 degrees. The radius of the patch center (from the ideal center) was determined from a normal distribution with mean = $[2 \times (3)^{1/2} \times \text{SD}] - 2$. Thus, the noise is annular, and the jitter threshold is specified as the annulus radius. The advantage of this method over Gaussian jitter is that trials are not wasted at near-zero jitter levels. Control experiments using two-dimensional Gaussian jitter gave qualitatively and quantitatively similar results (Fig. 1*A*, small thick symbols). Psychometric curves relating percent correct identification to the annulus radius obtained for each stimulus condition were fit with a Weibull function, and jitter thresholds were specified as the amount of jitter that reduced performance to 62.5% correct. Each threshold estimate was based on 125 trials. The reported thresholds represent the mean of three to five individual threshold estimates. Viewing was monocular. Sample thresholds were measured and analyzed in the same way as the jitter thresholds.

The ideal observer analysis is based on the joint probabilities of losing critical sample pairs (see Fig. 3 *Inset*). The probability of losing both of a pair of samples is P^2 . The probability of not losing both of a critical pair of samples is $(1 - P^2)$; thus, the probability of losing one pair, but not both of the other three critical sample pairs is $P^2 \times (1 - P^2)^3$. Because there are four critical pairs, the ideal observer computes the 16 combinatorial probabilities.

RESULTS AND DISCUSSION

Tolerance to Perturbation of the Positions of the Features.

The results of several experiments show that the jitter threshold is equal to about 0.5 times the feature separation. In the first experiment, the pattern was varied by changing the observers’ viewing distance (Fig. 1*A*, open symbols). This has the effect of changing the angular size (or standard deviation) of each individual patch, the spatial period of the carrier grating, and the separation between the features, in inverse proportion to the distance. Control experiments show that the jitter threshold is determined by the separation between the features. For example, fixing the separation while varying either the spatial period or the standard deviation of the Gabor patches has no effect on the jitter threshold (Fig. 1*B*). Varying the separation while fixing both the size and period of the features results in the jitter threshold equal to about 0.5 times

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0027-8424/97/9411742-5\$2.00/0 PNAS is available online at <http://www.pnas.org>.

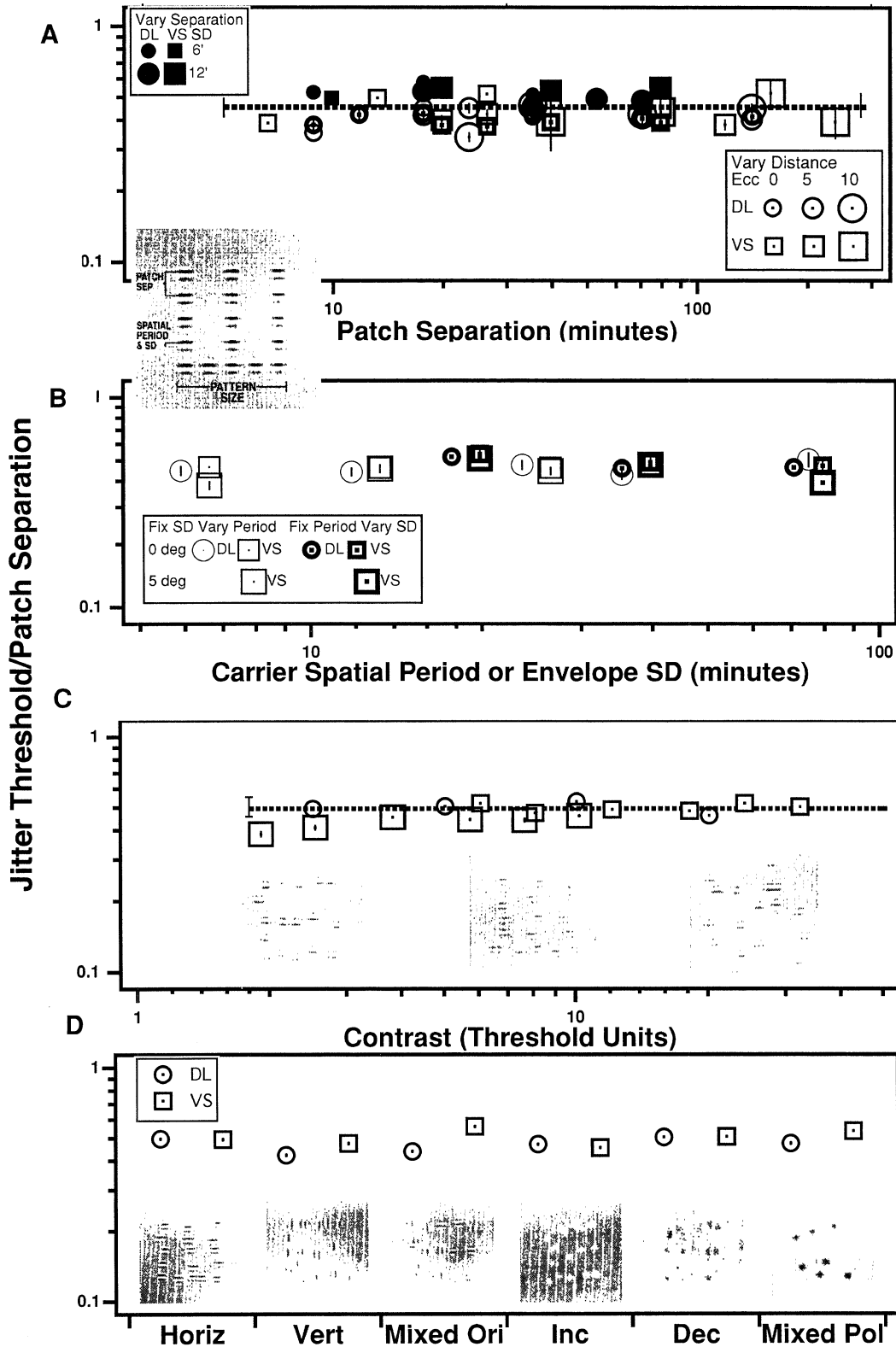


FIG. 1. (A) Jitter thresholds (specified as a fraction of the separation) of two normal observers at the fovea (small open symbols) and at 5 and 10 degrees (large open symbols) in the lower visual field plotted as a function of the center-to-center separation of the patches. The open symbols were obtained with high-contrast (80%) features by varying the viewing distance, so the stimuli were scaled replicas of each other. For these experiments, the patch SD and spatial period were equal to one-third of the separation. Small open symbols with heavy outline used two-dimensional Gaussian jitter. Solid symbols were obtained by varying the separation while fixing the spatial period and standard deviation (either 6 or 12 min coded by symbol size). The dotted line represents the mean threshold ± 1 SD. (Inset) Example of our E-like pattern, with 0 jitter. (B) Varying either the carrier spatial period or the envelope standard deviation (while fixing the separation) has no effect on jitter threshold. (C) Jitter threshold is independent of contrast in the fovea (small symbols) and the periphery (large symbols). The insets show three jittered stimuli with jitter equal to about one-sixth, one-third, and one-half of the patch separation. (D) For Gabor patches, the jitter threshold is independent of the carrier orientation (horizontal, vertical, or mixed orientations), and for Gaussian patches, it is also independent of polarity (dark patches, bright patches, or mixed polarities). (Insets) Samples of the stimuli.

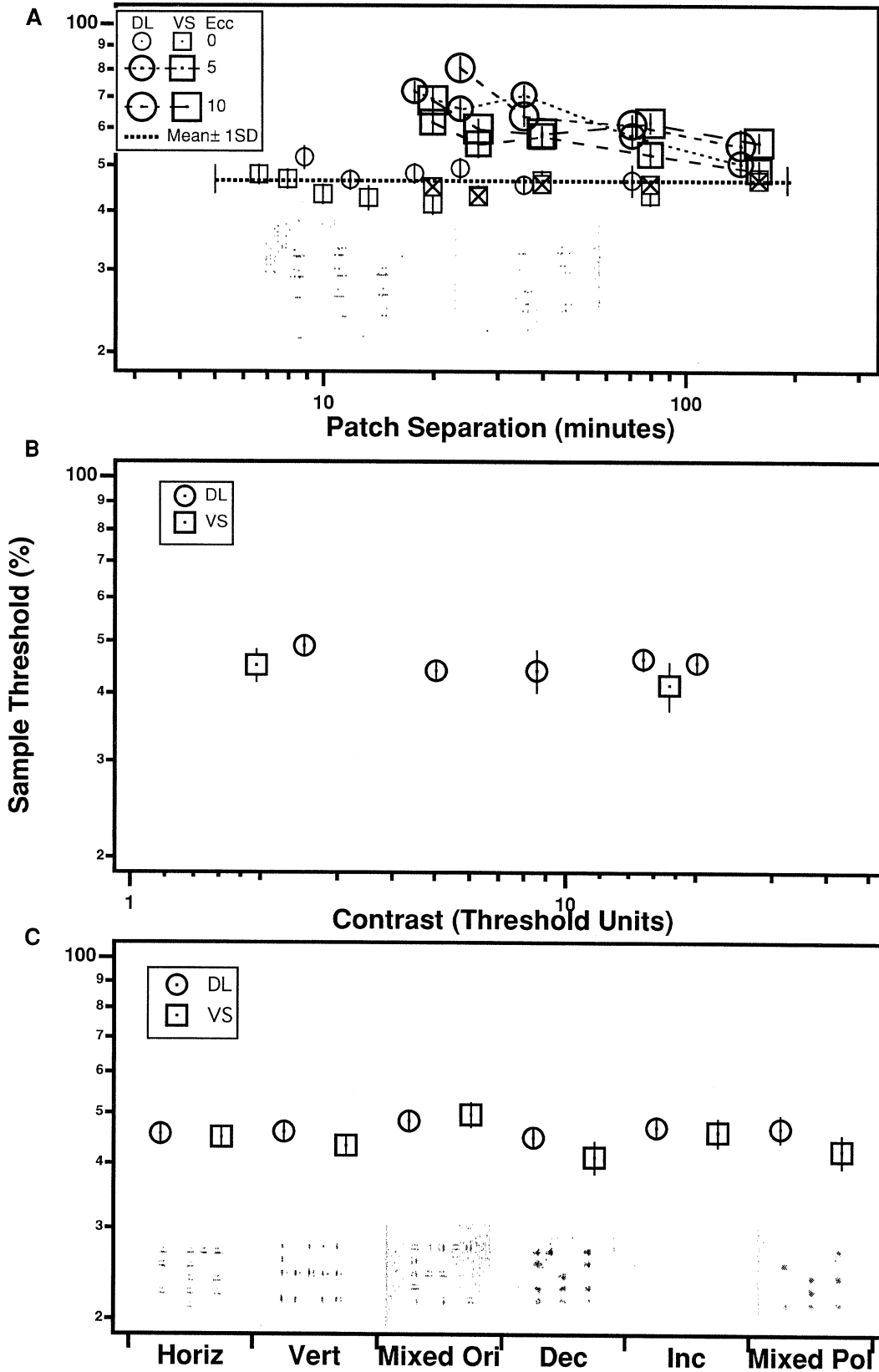


FIG. 2. (A) Sample threshold for two normal observers viewing foveally (small symbols) or at 5 and 10 degrees in the lower visual field (large symbols). Open circles and squares obtained with high contrast (80%) features and were obtained by varying the viewing distance and are plotted as a function of the feature separation. The dotted line shows the mean (± 1 SD) threshold of the normal fovea. Squares with crosses are foveal data obtained with the feature contrast adjusted to match the visibility of the periphery. (B) The sample threshold is also independent of contrast (specified relative to the contrast threshold for identifying the orientation of the pattern). (C) The sample threshold is independent of the carrier orientation (horizontal, vertical, or mixed orientations), and for Gaussian patches, it is also independent of polarity (dark patches, bright patches, or mixed polarities). This indicates the nonlinearity of the identification process. (*Insets*) Samples of the stimuli.

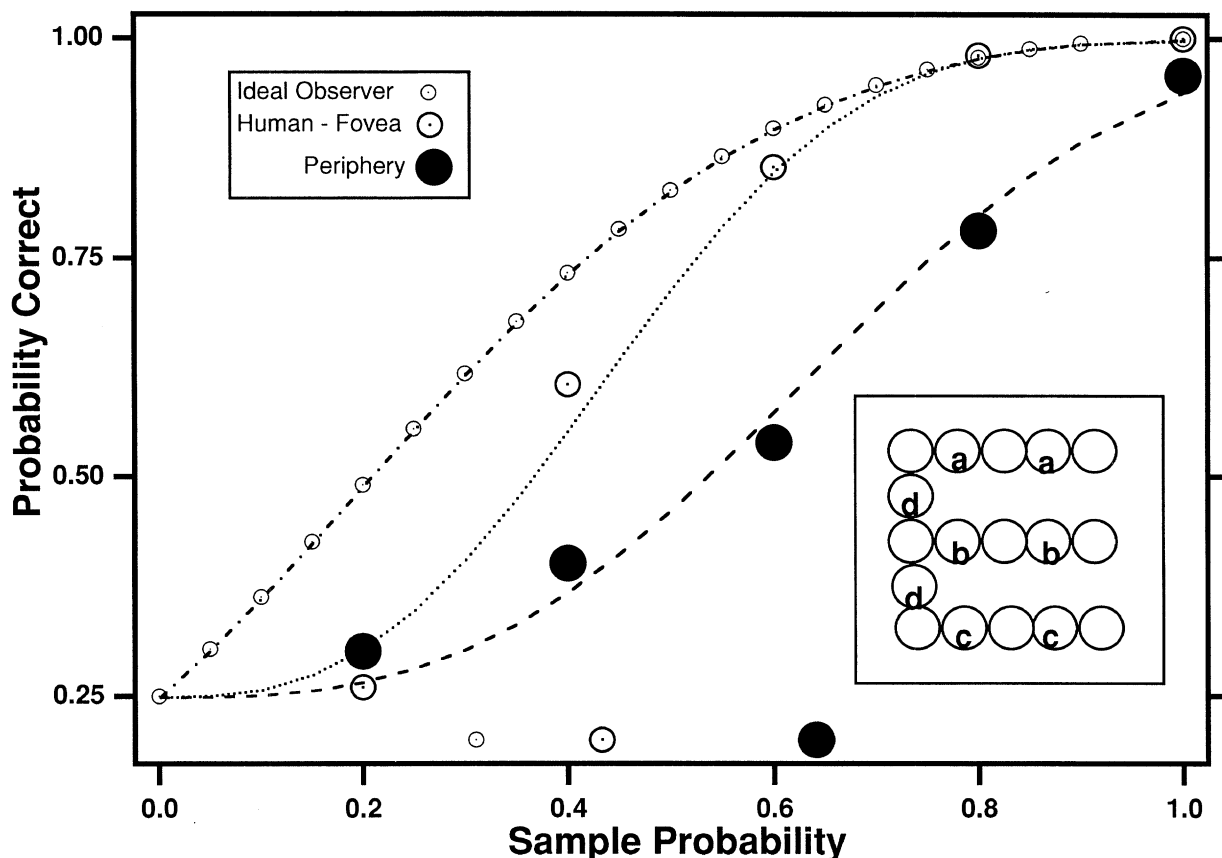


FIG. 3. Probability correct is plotted as a function of the probability of each sample being displayed. Small open circles and dot-dashed lines show performance of an ideal observer. Examples of the psychometric functions of the human fovea (medium open circles and dotted lines) and periphery (at 5 degrees, solid circles and dashed lines) are shown (with each point based on ≈ 100 trials). The data are sample psychometric functions of observer DL at a separation of 23.6 min. The symbols plotted along the abscissa represent the sample thresholds (corresponding to the 62.5% correct point). (Inset) The four critical sample pairs (labeled with letters) used by the ideal observer.

the separation (Fig. 1A, solid symbols). Note that patches with different feature sizes (SDs denoted by different solid symbol sizes) with the same separation have nearly identical thresholds. When expressed as a fraction of the patch separation, the jitter threshold is ≈ 0.5 times the patch separation. Thus, tolerance to positional jitter is determined mainly by feature separation.

The jitter threshold is extremely robust. For example, over a wide range of contrast levels, it is independent of contrast (Fig. 1C). It also shows little dependence on the patch details (Fig. 1D); for Gabor patches, the jitter threshold is independent of the carrier orientation (horizontal, vertical, or mixed orientations), and for Gaussian patches, the jitter threshold is also independent of polarity (dark patches, bright patches, or mixed polarities). This result is quite surprising, because detection thresholds for similar targets are about a factor of 2 worse when the patches have mixed (i.e., both horizontal and vertical) orientations (6), and spatial interactions at detection threshold are strongest when the elements are aligned (8–10).

Interestingly, the same degree of tolerance to jitter is evident in peripheral vision (Fig. 1A, large open symbols). The periphery has degraded spatial vision, and one explanation for the degradation is that peripheral vision suffers from uncalibrated topographical jitter of retinal cones and cortical receptive fields (11, 12). If the uncalibrated topographical jitter exceeded approximately half the patch separation, then performance should be degraded (even for unjittered targets). However, the close similarity in jitter thresholds in foveal and peripheral vision suggests that the computations involved in pattern discrimination are similar in fovea and periphery. In both foveal and peripheral vision, the neural mechanisms

involved in binding features into forms are evidently quite robust to topographical jitter.

How Many Stimulus Features (Samples) Are Required for Pattern Identification? We investigated the ability to interpolate the features into a pattern by varying the probability that each sample would be displayed (undisplayed samples were set to the mean luminance; see Fig. 2 Insets) and measuring the “sample threshold” (the proportion of samples required for 62.5% correct performance). In foveal vision, over the same broad range of stimulus conditions, the “sample” threshold is 40–50% of the maximally 17 samples forming the pattern. The sample threshold is independent of patch standard deviation, separation, and carrier spatial period, as determined either by varying the viewing distance (Fig. 2A, small open symbols) or by holding the patch size and spatial period fixed and varying the separation of the patches. Like jitter, it is also independent of the patch details (Fig. 2C, orientation or polarity) or contrast (Fig. 2B). Additional control experiments also show that in the fovea, it is largely independent of stimulus duration (at least between ≈ 55 msec and 2 sec). However, the periphery requires more samples than the normal fovea (Fig. 2A, large symbols), even when the pattern visibility is matched. The difference between the normal fovea and peripheral vision is largest with small patches. For the smallest sizes at which the observers could perform the task (i.e., the contrast of the pattern was at least twice the threshold for orientation discrimination), the sample threshold of the periphery was about 70%—an increase of about 55% relative to the fovea. It is worth noting that with sufficient magnification, the periphery is able to perform about as well as the fovea (Fig. 2A, at the largest separation, patch size and spatial period); however, the

critical comparison is that at small patch separations (e.g., ≈ 20 min), the periphery is severely compromised in the sampling task, but for the same stimulus, the jitter threshold is the same as that of the fovea.

The performance of an ideal observer (a machine, with perfect knowledge of the stimulus) would be limited by the presence or absence of one of four "critical" sample pairs (Fig. 3 *Inset*). The unlettered samples (Fig. 3 *Inset*) are uninformative, because they are common to all orientations. The performance of an ideal observer (Fig. 3, small open symbols) is slightly better than that of humans using their fovea. The ideal sample threshold (small open symbol near the abscissa) is $\approx 31\%$, compared with optimal human performance of $\approx 40\%$. Thus the normal fovea is quite efficient in using the samples (ratio of ideal to real ≈ 0.75). On the other hand, peripheral vision is rather inefficient (ratio of ideal to peripheral < 0.50).

Thus, our experiments imply that the mechanisms involved in feature integration of suprathreshold isolated forms are scale invariant and nonlinear (independent of phase and contrast polarity). By using rather different methods, others have arrived at similar conclusions (2–5). Indeed, the rules governing form perception have long been of interest to psychologists and sensory physiologists, and many of these rules were understood by the Gestalt psychologists (13). Parameters such as separation, continuity, collinearity, etc., were all shown to play an important role in grouping and in figure–ground segregation (1, 2, 5, 13). Indeed, one plausible explanation for the scale invariance that we found is that jitter may introduce a curvature of the (otherwise straight) line, which is inversely proportional to the element separation. However, there are clear differences in the rules governing our task compared with figure–ground segregation, where there are strong joint constraints of position and orientation (2). The present results show that 1) in foveal vision, the perception of a suprathreshold isolated form is remarkably robust to variations in the nature, positioning, and number of features which define the form; and 2) that peripheral vision requires more samples for reliable form discrimination. The mechanisms of feature integration appear to operate by assigning "place tags" to each sample. The positional tolerance of these place tags is constrained by the separation of the features (about one-half

of the feature separation) and, in foveal vision, each feature requires only about a 50% probability of being present for reliable discrimination. Our results provide an upper limit for stimulus jitter for accurate form perception in normal vision and suggest that any uncalibrated intrinsic positional jitter in peripheral vision does not exceed this limit. The increased "sample" threshold in peripheral vision (even after scaling the target visibility) suggests that the stimulus may be relatively underrepresented at the stage of feature integration, perhaps due to undersampling (14). Thus, in peripheral vision, every sample counts, and removing samples degrades performance (i.e., the periphery lacks the redundancy present in foveal vision).

We are grateful to Hope Marcotte for programming; Harold Bedell, Laura Frishman, Scott Stevenson, and two anonymous referees for helpful comments on an earlier version of the paper; and Bridgitte Shen for help with data management. This research was supported by research grants (RO1 EY01728 and RO1 EY04776) and a Core grant (P30EY07551) from the National Eye Institute, National Institutes of Health.

1. Beck, J., Rosenfeld, A. & Ivry, R. (1989) *Spat. Vision* **4**, 75–101.
2. Field, D. J., Hayes, A. & Hess, R. F. (1993) *Vision Res.* **33**, 173–193.
3. Kovacs, I. & Julesz, B. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7495–7497.
4. Kovacs, I. & Julesz, B. (1994) *Nature (London)* **370**, 644–646.
5. Moulden, B. (1994) in *Higher-Order Processing in the Visual System*, Ciba Foundation Symposium 184 (Wiley, New York), pp. 170–192.
6. Saarinen, J., Levi, D. M. & Shen, B. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 8267–8271.
7. De Valois, R. L. & De Valois, K. K. (1988) *Spatial Vision* (Oxford Univ. Press, New York).
8. Polat, U. & Sagi, D. (1993) *Vision Res.* **33**, 993–999.
9. Polat, U. & Sagi, D. (1994) *Vision Res.* **34**, 73–78.
10. Kapadia, M. K., Ito, M., Gilbert, C. D. & Westheimer, G. (1995) *Neuron* **15**, 843–856.
11. Hess, R. F. & Field, D. (1993) *Vision Res.* **33**, 2663–2670.
12. Watt, R. J. & Hess, R. F. (1987) *Vision Res.* **27**, 661–74.
13. Wertheimer, M. (1923) *Psychol. Fortsch.* **4**, 301–350.
14. Levi, D. M. & Klein, S. A. (1986) *Nature (London)* **320**, 360–362.