

THE APERTURE PROBLEM—I. PERCEPTION OF NONRIGIDITY AND MOTION DIRECTION IN TRANSLATING SINUSOIDAL LINES

KEN NAKAYAMA and GERALD H. SILVERMAN

Smith-Kettlewell Institute of Visual Sciences, 2232 Webster St, San Francisco, CA 94115, U.S.A.

(Received 12 December 1986; in revised form 24 August 1987)

Abstract—To examine how local velocities from different regions of the visual field combine to form a coherent motion percept, we subjected a sinusoidal line stimulus to translational motion. Horizontal movement of a sinewave line along its axial direction is perceived as nonrigid if the angle at the zero crossing is smaller than a critical angle of about 15 deg. This angle is independent of spatial scale and the number of sinusoidal cycles. To extend the applicability of this concept of angle, we developed a mathematical model to predict an observer's sensitivity to small changes in motion direction based on two assumptions: (1) the computed velocity signal is obtained from the intersection of constraint lines defined by local velocity components, (2) local velocity components are contaminated by noise. Measurement of directional discrimination thresholds of moving targets confirmed our expectations. Thresholds varied as a function of the angle of the local contour independent of spatial scale and in quantitative accord with our assumptions.

Motion Velocity Aperture problem

INTRODUCTION

Motion is a specialized visual sense useful for a variety of visual tasks including a reconstruction of the third dimension, delineation of surface boundaries, visual kinesthesia, controlling of eye movements, mediation of size constancy, pattern vision, etc. (see Nakayama, 1985).

At first glance it would seem that computation of velocity at each region in the visual field would be easy. There are numerous cortical neurons in the visual field sensitive to particular directions of motion (Hubel and Wiesel, 1962; Schiller *et al.*, 1976). In addition many are preferentially tuned to a narrow range of velocities (Orban *et al.*, 1981). Recent work, however, suggests that the problem is more complicated and that a direct reading of the velocity vector for a given retinal position might require further stages of neural processing (Fennema and Thompson, 1979; Horn and Schunk, 1979; Adelson and Movshon, 1982; Ullman, 1983; see also Wallach, 1935; Burt and Sperling, 1981).

The problem can be appreciated by considering the properties of a cortical neuron with a spatially elongated receptive field. If the output of this cell varies with the velocity of a line moving at right angles to its preferred orien-

tation, one might assume that this output signal would be sufficient to measure the velocity of the moving line [see Fig. 1 (A)]. Such a measurement of velocity, however, is only a local reading of velocity. As such we designate it with the term V_L . The real velocity of a moving stimulus (designated as V) could be quite different. In fact an infinite number of velocities (V) could have given rise to any given local velocity (V_L) and these can be represented by an arrow falling on the constraint line shown in Fig. 1(B). Therefore a local reading of velocity by an orientation and directionally selective detector does not specify the actual target velocity.

Yet this sampling of a local velocity is highly informative, because it takes only two linearly independent readings of local velocity to reconstruct the true velocity (Fennema and Thompson, 1979). Each local reading defines a constraint line of possible velocity vectors which could have given rise to the local velocity vector [see Fig. 2(A)]. The intersection of two such lines in a hypothetical velocity space specifies the true velocity vector [see Fig. 2(B)].

In addition to outlining an exact and intuitive description as to how the aperture problem could be solved, Adelson and Movshon (1982) also provided supporting evidence using simple

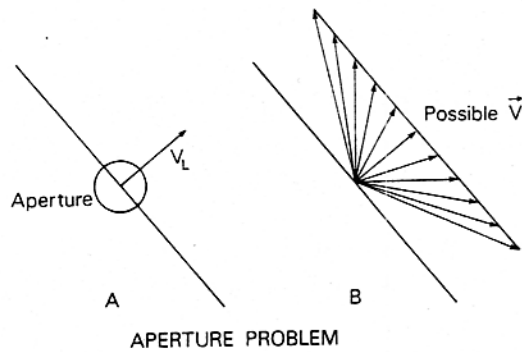


Fig. 1. The aperture problem. The direction and magnitude of motion cannot be determined by a single oriented unit. Assume that an oriented receptive field can read the component of velocity orthogonal to its preferred orientation, providing the nervous system with a local reading of velocity V_L at an aperture. This local velocity vector is generally not the same as the true velocity vector V . In (B) we see that V_L constrains but does not uniquely determine the velocity vector giving rise to V_L since a whole set of vectors could have generated V_L .

and compound patterns made up of sinusoidal gratings of differing orientations. They reported that two crossed gratings would be seen as moving in a direction which conformed to the intersection of such hypothetical lines in velocity space. They also demonstrated that the solution does not conform to a number of other schemes. For example, it cannot be predicted from the vector sum or the maximum velocity or the average of the component velocities (see also Daugman, 1981). From these considerations, Adelson and Movshon proposed that velocity encoding requires yet another stage of neural

processing beyond that seen in the measurement of local velocity vectors. "True" velocity units must combine information from at least two different component units. Figure 2(C) reproduces Adelson and Movshon's picture of a hypothetical higher order unit, consisting of components which have a $\pm 90^\circ$ range of local directions each having corresponding best velocity proportional to $\cos(\theta)$ of the preferred velocity. Examination of the receptive field properties of cells in area MT of the primate reveals a class of units which is consistent with this hypothesis. Such units respond to the overall pattern motion and not to local components (Movshon *et al.*, 1984; Albright *et al.*, 1984).

In this paper we extend the examination of the aperture problem in one unexplored direction. Adelson and Movshon emphasized the combination of orientated velocity signals at the same spatial locus. We deal explicitly with the possible combination of orientation and velocity signals over disparate regions of the visual field.

As an introduction to this approach, consider the planer curve as depicted in Fig. 2(A), moving to the right with a constant velocity V . Each portion of the curve has a local velocity V_L which is different from other local velocities and which is generally very different from the true velocity V , yet we usually do not see these different local motion vectors. Instead we see the coherent motion of a rigid object. This rather flawless perceptual synthesis suggests that the combination of orientation and velocity signals can occur across a spatially extended

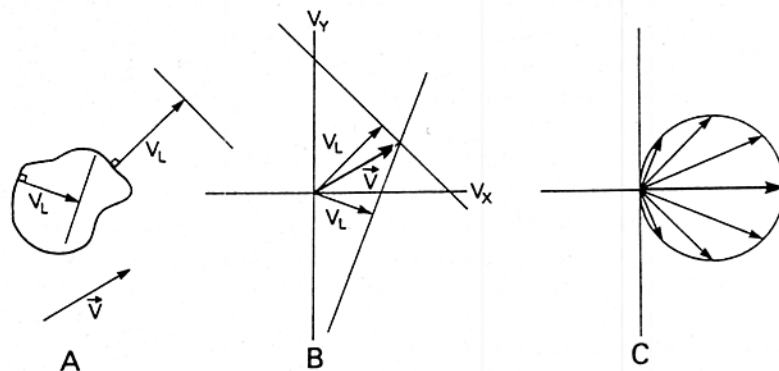


Fig. 2. (A) Figure which undergoes pure translation in the image plane with velocity designated by V . Arrows designated by V_L represents just two of the infinitely many local velocity vectors which accompany this image translation. Note that a give local velocity vector is orthogonal to its local contour and has a magnitude proportional to $V \cos \theta$ where theta is the difference in angle between the local velocity and the "true" velocity. (B) The local velocity vector have been replotted in a velocity space along with their corresponding constraint lines. Note that the intersection of these constraint lines represents the "true" velocity V .

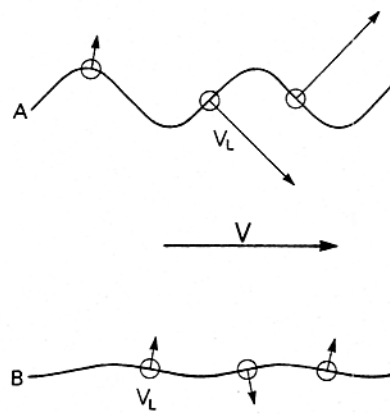


Fig. 3. Sinewave line stimuli which move at velocity V . Because the angle at the zero crossing for the upper sine wave (A) is larger than a critical angle, this waveform is seen as moving rigidly. The lower sine wave (B), however, is seen as moving nonrigidly because this critical angle is too small. V_L refers to the local velocity in each figure which has the largest magnitude. The small arrow in (A) represents a local component which has the same direction and amplitude as largest local component in (B).

region and suggests the integration of orientation and velocity signals from different topographic loci in the cortex. To explore this issue using psychophysical methods, we looked for a rigidly moving stimulus where one could consider the effects of two separate parameters, angle and spatial scale.

Rigidly moving sinusoidal lines with defined amplitude and spatial frequency provide a simple yet suitable stimulus (see Fig. 3). By varying the amplitude of such a sine wave for a given frequency, one can alter the difference in angle between angular components. By varying the spatial frequency of the sinusoidal line one can alter the distance (or scale) between angular components. It should be emphasized that this is a plane curve, not a luminance profile of a sinusoidal grating.

Initial qualitative observations on figure rigidity

We began our experiments with the rather surprising observation that rigid horizontal movement of a sinusoidal line does not necessarily lead to perceived rigid horizontal motion (see Fig. 3). In the extreme case, one only sees plastic undulation of the line with no hint of any translational motion. This was most dramatic for sine waves having low spatial frequency and low amplitude. To our knowledge this is the first reported description of figural nonrigidity elicited by pure translational motion in the image plane. Other nonrigid appearances of rigid motions have been reported (Wallach and

O'Connell, 1953; Braunstein and Andersen, 1984; Hildreth, 1984) but the motion always had a rotational component either in or outside the image plane.

EXPERIMENT 1: PERCEIVED RIGIDITY AS A FUNCTION OF AMPLITUDE AND SPATIAL FREQUENCY

To explore the conditions under which rigid and nonrigid motion were elicited, we varied various parameters of a sine wave figure. The sinusoidal line was displayed on a CRT monitor (Hewlett Packard 1332A with P31 phosphor) by synchronizing the output of a triggered function generator to the CRT sweep. The sinusoidal line was translated in the horizontal direction at constant velocity by adding a linear voltage ramp to the oscilloscope X-axis input. By over-scanning, that is by making the length of the electronically defined sinusoidal line longer than the width of the display screen, the right and left border of the stimulus was fixed and defined by the edges of the screen. The luminance of the sinusoidal line was approximately 100 cd/m^2 seen against the room illuminated screen which had a luminance of 2 cd/m^2 . Refresh rate was 100 Hz. The two authors served as observers.

For each trial, the sinusoidal line stimulus was flashed on a CRT screen. After 200 msec the line moved either to the left or to the right on a random schedule at a fixed velocity ($3^\circ/\text{sec}$). The duration of the motion was 200 msec, short enough to avoid tracking by the oculomotor pursuit system. The line disappeared 200 msec after the cessation of motion. The task of the observer was fixate the center of the display and to decrease the amplitude of the moving sine wave line until it was just seen as nonrigid. The sinusoidal amplitude corresponding to this transition between rigidity and nonrigidity was measured for different spatial frequencies of the sinusoidal line.

Before reporting on these results we deal with the possible contaminating factor of temporal frequency. Consider a sinusoidal line moving with pure horizontal motion. A given position along the retina will be exposed to the same spatial pattern repeating at a given temporal frequency and this temporal frequency will change with changes in spatial frequency if velocity is kept constant. Because it might be argued that this variation in temporal frequency could alter one's sense of rigidity, we first exam-

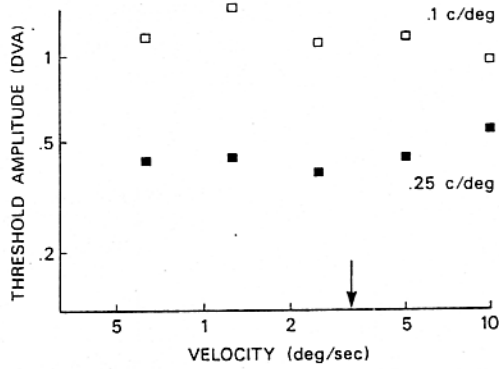


Fig. 4. Control experiment to deal with possible effects of temporal frequency variation. Amplitude thresholds to see rigidity for two spatial frequency sinewaves at different velocities. The solid dots refer to a spatial frequency of 0.1 c/deg and the solid circles refer to a spatial frequency of 0.25 c/deg. Note that there is no systematic difference in the amplitude to see rigidity for a given spatial frequency as velocity is varied. Arrow represents the velocity that was chosen for the data presented in Fig. 5.

ined the relationship between rigidity thresholds and temporal frequency.

To vary temporal frequency without changing spatial frequency, we varied velocity. Figure 4 shows the results of this control experiment indicating that at least over some reasonable range of velocities, the relation between threshold amplitude and rigidity is relatively flat, thus minimizing any subsequent interpretation which could be couched in terms of temporal frequency. Thus we felt confident to conduct the experiment at constant velocity and to vary spatial frequency.

Results of Experiment 1

We found a strong dependence of the amplitude needed to see rigidity in these moving sinusoidal lines with spatial frequency. The solid dots in Fig. 5 shows this relation. As spatial frequency is increased, it takes progressively less amplitude to see the line as moving rigidly in a horizontal direction. Furthermore, this line has a slope of -1 on double logarithmic coordinates. Sinusoidal lines along this reciprocity axis have a constant maximum angle at their zero crossings suggesting that a minimum angular difference in direction between local motion vectors is required for the observer to see rigidity. The reciprocal relation between amplitude and spatial frequency and its associated constant angle is illustrated in Fig. 6. For subject K.N. this angle is about 36° and for J.S. it is about 28° .

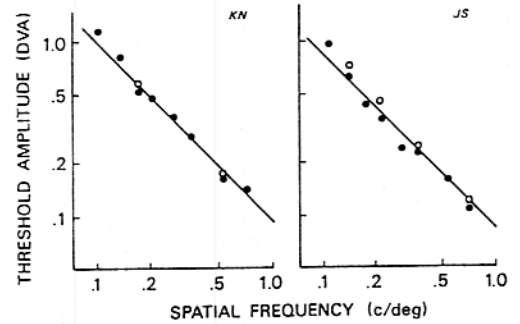


Fig. 5. Thresholds to see figural rigidity in sinusoidal lines. Solid dots represents thresholds for Experiment 1 where spatial frequency was varied and field size was kept constant. Open circles refer to thresholds where a constant number of cycles was employed (Experiment 2). Best fitting slope of -1 is shown corresponding to a critical angle of 18 deg. and 14 deg. for J.S. and K.N. respectively. Sinewave figures with amplitudes above fitted line are seen as rigid, those below are seen as nonrigid.

EXPERIMENT 2: CONSTANT NUMBER OF CYCLES

A potential contaminating factor when measuring thresholds as a function of spatial frequency is the variation in the number of cycles which are visible. For a given display size, increasing frequency is always accompanied by a proportionate increase in cycle number. If our notion regarding the maximum angle in the sinewave is correct, increasing the number of cycles should not change the results. We confirmed this supposition by conducting an experiment where the spatial frequency was varied but the number of cycles was kept constant. In this case it was restricted to two cycles. This limit in the number of cycles was accomplished by masking off the right and left portion of

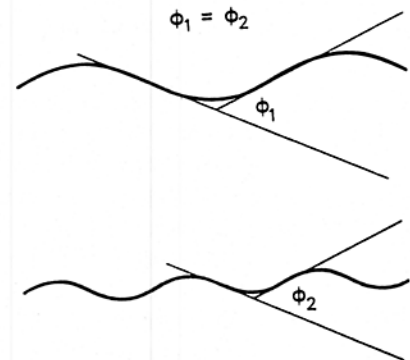


Fig. 6. Two sinusoids which have reciprocal relations between amplitude and frequency (having a slope of -1 on a double log plot as in Fig. 5) share the same maximum difference in contour orientation.

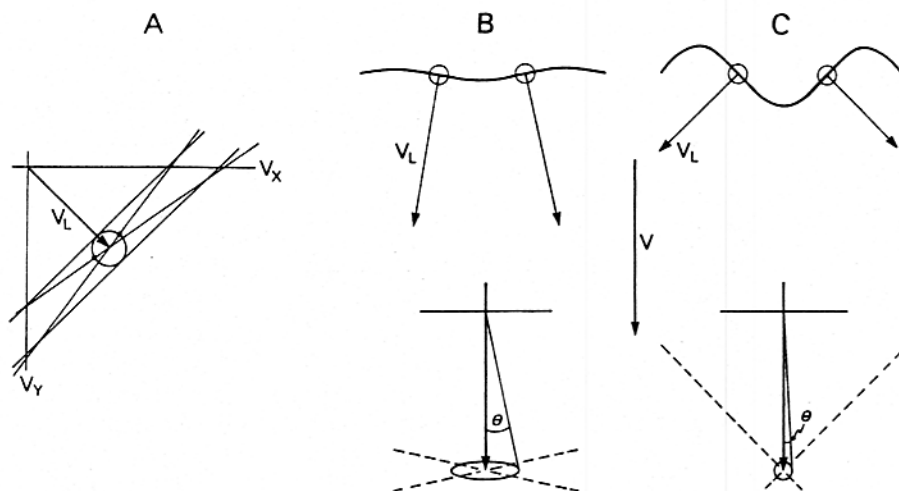


Fig. 7. Illustration of how error in the pickup of the local velocity signal can lead to different amounts of angular discrimination depending on the local angle. In (A), we show how a local velocity signal may be imprecise, having a random distribution of velocities centering around a mean value. The circle represents an iso-probability line for the local velocity signal. The four lines passing through the circle represent a corresponding subset of possible constraint lines which are defined by the "noisy" velocity signal. (B) and (C) provide a pictorial description of the importance of angle given a fixed amount of error. At the top of each section is a sinewave line stimulus which is moving in a downward direction. Each has a pair of local velocities which have the greatest deviation from downward and these are drawn on the figure and labeled V_L . The boundary drawn around the intersection of these constraint lines characterizes the zone of uncertainty. Note that the angle subtended by this zone is much greater for the case of (B) than for (C). Thus the predicted direction discrimination in (B) is worse than in (C).

the screen with an opaque occluder of the appropriate size for each spatial frequency. The open circles in Fig. 5 shows that the results were identical to Experiment 1, again falling close to a slope of -1 . This provides further evidence that it is indeed the maximum angle in the sinewave that dictates whether the figure will be seen as rigid vs nonrigid.

Discussion of Experiments 1 and 2 and a theoretical formalism

Because the preceding experiments suggested that the angular relationship between local motion vectors was decisive, we were motivated to make a quantitative model to predict the synthesis of local components into a perceived velocity vector.

*An elliptical error function, which takes account of the superior directional precision of local velocity signals in relation to magnitude signals was also considered (see Fig. 9 of Nakayama, 1985). Such properties were incorporated for the Monte Carlo simulation and its results had no significant change in the threshold predictions. Thus over the range of conditions considered here, velocity magnitude error largely determines the imprecision of the constraint line intersection.

The model extends the concept of Adelson *et al.* and Fennema *et al.*, by assuming that the estimation of the local velocity vectors V_L is imprecise, being contaminated by biological noise. Errors at this level will then be propagated to the solution of the aperture problem in a predictable fashion. To start, we assume that the noise is riding on the local velocity signal and is proportional to the signal itself with a 5% coefficient of variation. This is a figure which is consistently obtained from psychophysical measurements (Nakayama, 1981; McKee, 1981; McKee and Nakayama, 1984). Thus we represent each local velocity signal probabilistically, drawn from a radially symmetric gaussian distribution centred on the local velocity.* This leads to a range of possible constraint lines are illustrated in Fig. 7(A).

Computer simulation of the model was straightforward. Using a Monte Carlo technique each of a pair of local velocity signals was drawn from its local Gaussian distribution. This defined two constraint lines. The intersection of these two lines was calculated from the appropriate pair of simultaneous equations. The process was repeated and the variance for the

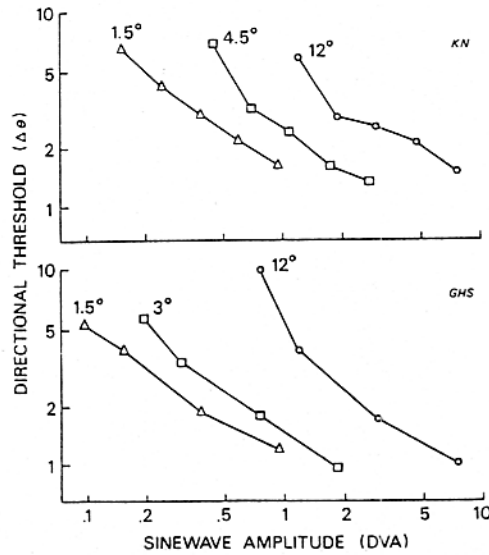


Fig. 8. Directional discrimination thresholds plotted as a function amplitude of a sinewave stimulus for three different spatial frequencies. Numbers next to each curve denote the spatial period of each sinewave (in degrees of visual angle).

direction of the synthesized velocity signal was computed for a given set of local velocities.

Given this very simple model with added "biological" noise we illustrate how the judgement of the direction of motion is influenced by the angle of local components in relation to the global velocity vector. Consider the case where two targets are moving in a near downward direction. In Fig. 7(B) the local velocity vectors are very close together in direction; whereas in

†The predictions of the model are essentially independent of the velocity of the target as long as the velocity of the local component is large enough to ensure that the velocity magnitude noise remains a constant ratio of the velocity magnitude signal.

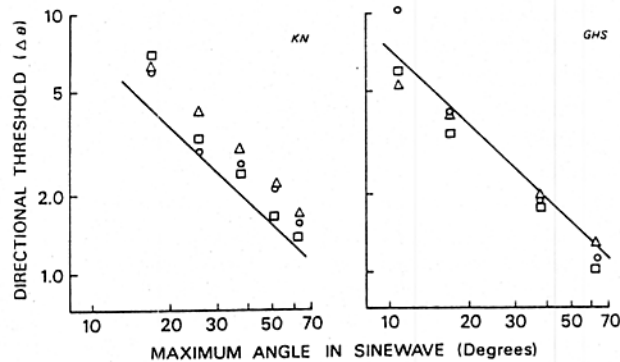


Fig. 9. Using the same data as in Fig. 8, we plot directional discrimination thresholds as a function of the angle at the zero crossing. Symbols are the same as in Fig. 8. Note that the data from a 4:1 range of spatial scale collapses to a simple dependence on angle. Solid line are predicted angular discrimination thresholds from model described in text.

Fig. 7(C) they are much farther apart. When the constraint lines defined by local components are very similar in orientation, our model predicts that the precision of the a motion direction estimate will be very poor. The quantitative prediction can be understood intuitively in terms of the intersection of straight lines at various angles. The intersection loci of velocity constraint lines which meet at extremely obtuse angles are likely to wander very significantly for small perturbations in the position of these lines. Thus small mis-estimates as to the magnitude of the local motion vector can have serious consequences in terms of the precision of estimating the direction of the true velocity vector. The predicted changes of velocity discrimination as a function of local angle can be seen by looking forward to Fig. 9 and noting the solid line.

The predictions of the model depend exclusively on the angular relationships between the components and its relation to the true direction. It does not depend on the size of the angular segments or their spatial separation nor does it depend critically on the velocity of the target*. Because this is a rather extreme simplification of the problem, we looked for an experimental situation to examine its plausibility. To accomplish this task we measured the ability of human observers to judge the direction of motion over a 4:1 range of spatial scale.

EXPERIMENT 3: PRECISION OF MOTION DIRECTION AS A FUNCTION OF MAXIMUM ANGULAR DIFFERENCES IN A SINUSOIDAL LINE FOR VARIOUS SPATIAL FREQUENCIES

Methods

In this experiment we presented the same

sinusoidal line stimulus, except we moved it mostly in a vertical direction rather than horizontally. As a consequence it moved orthogonal to rather than along its baseline.

Before presenting the exact methods and results in quantitative terms, qualitative aspects should be mentioned. Consider the limiting case where the sinusoidal amplitude of the wavy line is so small that it appears as straight. In this case, all of the different directions of motion will look equivalent, an experimental example of the aperture problem. All will appear as if the line were moving at right angles to its orientation. With a much larger amplitude in the sinusoidal line, it should be clear that the sideways components of motion can be encoded and the observers angular thresholds should fall.

To obtain a quantitative measure of the observer's angular discrimination thresholds, we used the method of constant stimuli. At random interval the target line would appear, immediately move in a downward or nearly downward direction for a period 200 msec, and then disappear. On any given trial, 1 out of 5 possible directions of motion were presented. The observers task was to respond as to whether the target appeared to move to the right or left in its downward traverse. The percent "rightward" response was tabulated and used to compute a directional discrimination threshold by probit analysis. Threshold angles were defined as those which corresponded to a d' of 0.675. Figure 8 shows these thresholds plotted as a function of the amplitude of the sinewave for 3 different spatial frequencies.

Because our model predicts that directional discrimination depends on the difference in angle between local motion components, without regard to differences in distance between these angles, the data in Fig. 8 can be transformed so that abscissa represents the maximum angular difference in the sinusoidal line regardless of spatial frequency. Figure 9 shows the angular discrimination thresholds plotted as function of angle rather than of amplitude in the sinewaves. Despite the very large difference in spatial frequency for the three conditions, it should be clear that the data was essentially coincident in terms of precision versus maximum angle. The predictions of our simulation are summarized by the solid line. It should be evident that the data are in substantial agreement with the predictions, particularly for slope of the relation between maximum local angle and the discrimination threshold.

Discussion

Before providing an interpretation of our results in terms of oriented local velocity signals, we consider one alternative explanation. Instead of synthesizing a velocity signal from oriented components, one could argue that our visual system encodes the movements of partially defined nodes or blobs. For the case of the sinewave, this would be the peaks and troughs of the sinusoid itself. Suppose for example that it is the curvature of the contour which determines such blob visibility and that this rather than lines supplied the input to the motion system. Formally curvature is defined as

$$\kappa = d\theta/ds$$

where κ is the curvature, θ is the local direction of the contour and s is the distance along the arc. The maximum curvature is at the peaks and troughs of the sinewave and this is proportional to the square of spatial frequency (Spiegel, 1963). Thus a mechanism based on curvature alone would predict a slope of -2 for the rigidity data seen in Fig. 5 which was inconsistent with the present data.

In contrast, our results support a dependence on local angle difference. First we show that it is the maximum angular difference between local orientations of a sinusoidal line which determines whether one sees rigidity. Second, we show that this local angle alone predicts direction discriminability.

One fact remains unexplained—our observations of perceived nonrigidity itself. From an ideal mathematical point of view, all figures undergoing pure translational motion should be seen as rigid since there is a unique rigid interpretation. The constraint lines for all local motions meet a single point in velocity space (see Hildreth, 1984). This idealized formulation, however, ignores noise. Yet our results on discrimination thresholds indicate that "noise" cannot be ignored accounts for the observers ability to see differences in motion direction. Thus, when the angle between two constraint lines is very acute, the synthesized magnitude becomes extremely uncertain and essentially useless.

Functionally, therefore, the visual system is placed in a dilemma. Should it utilize these local velocity signals to generate highly unreliable conclusions or should it just accept the local velocity vectors *as is* and skip the conclusions? We suggest that it chooses the latter option even

though this means seeing the local components and comcomittant nonrigidity. Thus in Fig. 3(B) the two local components move in opposite directions and the observer sees nonrigidity.

One final question comes to mind. Consider the local component labeled V_L in Fig. 2(B). It has the same amplitude and direction as a corresponding component in Fig. 2(A) (see short unlabeled arrow). In the latter case the contour is seen as rigid whereas it is seen as nonrigid in the former case. Why is this so? We think it is the context of other motion components that is of importance. In our companion paper (Nakayama and Silverman, 1988), we show how these ordinarily nonrigid sections of contour can be pulled along by velocity information from neighbouring loci along the contour so that they appear as rigid.

Acknowledgements—Supported in part by grants 5P30 EY-01186 and 1R01 EY-05408 from the National Institutes of Health and The Smith-Kettlewell Eye Research Foundation. A preliminary abstract of this paper was published as: Nakayama K. and Silverman G. H. (1983) Perception of moving sinusoidal lines. *J. opt. Soc. Am.* **72**, 1862. Author's address: 2232 Webster Street, San Francisco, CA 94115, U.S.A. We thank Steven Chung for the design and construction of specialized electronic hardware.

REFERENCES

- Adelson E. H. and Movshon J. A. (1982) Phenomenal coherence of moving visual patterns. *Nature, Lond.* **300**, 523-525.
- Albright T. D., Desimone R. and Gross C. G. (1984) Columnar organization of directionally selective cells in visual area MT of the macaque. *J. Neurophysiol.* **51**, 16-31.
- Burt P. and Sperling G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychol. Rev.* **88**, 171-195.
- Braunstein M. L. and Andersen G. J. (1984) A counterexample to the rigidity assumption in the visual perception of structure from motion. *Perception* **13**, 213-217.
- Daugman J. (1981) The visual spatio-temporal fundamental. *J. opt. Soc. Am.* **771**, 1632.
- Fennema C. L. and Thompson W. B. (1979) Velocity determination in scenes containing several moving images. *Comput. Graph. Image processing* **9**, 301-315.
- Hildreth E. C. (1983) The measurement of visual motion. *ACM Distinguished Dissertation Series*. MIT Press, Cambridge, Mass.
- Hubel D. H. and Wiesel T. N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol., Lond.* **160**, 106-154.
- McKee S. P. (1981) A local mechanism for differential velocity detection. *Vision Res.* **21**, 491-500.
- McKee S. P. and Nakayama K. (1984) The detection of motion in the peripheral visual field. *Vision Res.* **24**, 25-32.
- Movshon J. A., Adelson E. H., Gizzi M. and Newsome W. T. (1985) The analysis of moving patterns. *Expl Brain Res. (Suppl.)* **11**, 117-151.
- Nakayama K. (1981) Differential motion hyperacuity under conditions of common image motion. *Vision Res.* **21**, 1475-1482.
- Nakayama K. (1985) Biological image motion processing: a review. *Vision Res.* **25**, 625-660, 1985.
- Nakayama K. and Silverman G. H. (1988) The aperture problem—II. Spatial integration of velocity information along contours. *Vision Res.* **28**, 747-753.
- Orban G. A., Kennedy H. and Maes H. (1981) Response to movement of neurons in areas 17 and 18 of the cat: velocity sensitivity. *J. Neurophysiol.* **45**, 1043-1058.
- Schiller P. H., Finlay B. L. and Volman S. F. (1976) Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *J. Neurophysiol.* **39**, 1288-1319.
- Spiegel M. (1963) *Differential Geometry*, Schaum, New York.
- Ullman S. (1983) The measurement of visual motion. *Trends Neurosci.* **6**, 177-179.
- Wallach H. (1935) Uber visuell wahrgenommene Bewegungsrichtung. *Psychol. Forschung.* **20**, 325-380.
- Wallach H. and O'Connell D. N. (1953) The kinetic depth effect. *J. exp. Psychol.* **45**, 205-217.