Anisotropy in an ambiguous kinetic depth effect

Jeffrey B. Mulligan

Aerospace Human Factors Research Division, MS 262-2

NASA Ames Research Center

Moffett Field, CA, 94035

ABSTRACT

A set of animated stimuli (Lissajous figures) is described, each element of which is physically consistent with *two* different three-dimensional shapes undergoing rigid rotations about orthogonal axes. Human observers typically show a preference for one shape or the other; this preference may be biased by manipulating various parameters of the stimulus. Fairly good predictions of which shape will be seen are made using an adaptation of Hildreth's "smoothest velocity field" computation. When a given stimulus is rotated 90 degrees in the picture plane, the resolution of the ambiguity is often different, demonstrating *anisotropy* in the processing of the figures. The nature of this bias is such that for certain figures subjects see a three-dimensional object rotating about a vertical axis regardless of which two-dimensional orientation is used to present the stimulus. This bias is not predicted by the Hildreth model. One interpretation

of the results is that ambiguity in two-dimensional visual motion (i.e., the aperture problem) is not resolved prior to the interpretation of three-dimensional structure.

1. INTRODUCTION

Human perception of three-dimensional shape from a sequence of two-dimensional images is an extraordinary feat which is often taken for granted. The visual system must compute both a depth and a third component of velocity for each point in the scene, and do so in a way which produces overall consistency. An early observation of the recovery of three-dimensional structure from a sequence of two-dimensional silhouettes was reported by Miles [1]; Wallach and O'Connel [2] later dubbed the phenomenon the *kinetic depth effect* (KDE). For the case of rigid rotations of discrete points, Ullman [3] has determined the conditions that must be satisfied in order for the solution to be computationally realizable. Ullman's more recent models [4] allow for some departure from rigidity, as might arise either from actual non-rigid motion, or from noise in the input.

The complete problem becomes more difficult if one tries to extend this type of approach to images which are composed of line segments and smooth curves (not to mention gray-level images), instead of just isolated dots. This increased difficulty is due to the "aperture problem:" a motion detector viewing a moving line through a small aperture will be blind to motions of the line along its own length, and will therefore be unable to report the actual two-dimensional image velocity within its field of view.

The aperture problem complicates the analysis of structure-from-motion because algorithms like Ullman's require the two-dimensional image velocity at each point, while sensors with a small field of view can only report the local orientation and orthogonal velocity. Many models of the structure from motion problem assume that estimates of the two-dimensional velocities are available as input [3,4,5,6,7]. One elegant approach has been proposed by Hildreth [8,9] for figures comprised of closed curves. She suggested that a useful way to attack the problem would be to try to minimize the amount of variation in the hypothetical image velocities. Hildreth investigated a number of possible ways to define variation, but in most of her work she used the squared magnitude of the vector difference between velocities of adjacent points on the curve. By summing the local variation over the entire figure, a single number may be obtained for a given hypothetical solution of the aperture problem. Calculus may be used to find the solution which minimizes this quantity; Hildreth has called this solution the "smoothest velocity field." In this paper we shall use the term "roughness" to denote the variational quantity which is minimized.

One objection to Hildreth's approach is that although the solution of the aperture problem is subsequently used to analyze structure-from-motion, it is approached without regard for possible three-dimensional constraints. Intuitively, it seems more "information-efficient" to consider the ultimate three-dimensional interpretation when resolving the two-dimensional ambiguity, instead of being locked into a two-dimensional solution which could possibly be inconsistent with three-dimensional rigidity.

Hildreth has shown qualitatively that the algorithm makes mistakes in cases where humans also experience illusory percepts, such as in the "barber-pole" illusion. The experiments described in this paper were performed in an attempt to assess the validity of Hildreth's algorithm as a description of human perception. The approach was to use a stimulus which had two distinct physically plausible three-dimensional interpretations; each interpretation corresponded to a different solution of the aperture problem. Now, these ambiguous stimuli violate one of the conditions of Hildreth's algorithm: namely, that any intersections of the curve in the image should correspond to actual intersections of the three-dimensional generator. For this reason, no attempt was made to solve the minimization problem and find the smoothest velocity field. However, the "roughness" of a given hypothetical solution is still a well-defined mathematical quantity. Since we know that when humans are presented with the ambiguous stimuli they usually see one of the two rigid interpretations, we simply computed the roughnesses for the two rigid interpretations; if the visual system solves the aperture problem using a calculation like roughness, then subjects should perceive the "smoother" interpretation.

2. METHODS

2.1. Lissajous figures

The ambiguous stimuli were drawn from a class of curves known as Lissajous figures. Each figure has two distinct three dimensional interpretations which correspond to quite different shapes. Each of the two interpretations is a three-dimensional curve which lies on the surface of the cylinder, but the orientation of the cylinder is different for each of the two interpretations.

The three-dimensional appearance of animated Lissajous figures was observed some time before the term "kinetic depth effect" was first coined [10]. Philip and Fisichelli subsequently investigated the effects of various parameters on the spontaneous depth reversals in Lissajous figures [11,12]. It should be noted that these depth reversals are yet another ambiguity in the figures (similar to the depth reversals seen with the "Necker cube"), which is quite distinct from the ambiguity which is the topic of this paper; if one counts depth reversals as well as shape differences, then there are a total of *four* possible interpretations. From these early reports, it is impossible to determine whether the investigators were even aware that the two interpretations with different rotational axes were physically consistent with a single stimulus. For example, Fisichelli [12] describes changing *the* axis of rotation by simply interchanging the cables providing the deflection signals to the CRT. The parameter values used in these studies, however, were ones which produce an extremely strong bias in favor of one of the two interpretations, so perhaps it was simply a matter of chance that spontaneous changes of rotational axis were not observed.

Before giving the explicit formulae describing Lissajous figures, it may be illustrative to consider the problem of depicting an *unambiguous* curve which lies on the surface of a cylinder. Imagine that we have a vertically oriented cylinder of unit radius, and that we wish to paint on the surface of this cylinder a curve whose vertical position is defined as a function of angular position, $y = f(\theta)$. (We adopt a coordinate system in which x and y are the normal viewing screen coordinates, with z being a depth axis.) In this case, we can describe the curve parametrically by the following equations:

$$x(\theta) = \sin(\theta), \tag{1.a}$$
$$y(\theta) = f(\theta), \tag{1.b}$$
$$z(\theta) = \cos(\theta). \tag{1.c}$$

The curve defined by these parametric equations will lie on the surface of a unit radius cylinder regardless of the nature of the function $f(\theta)$. If we interchange the definitions of x and y, then we obtain a curve which lies on a *horizontal* cylinder. This observation is the key to understanding the ambiguity of Lissajous figures: if we let $f(\theta) = \sin(\theta)$, then the above equations are symmetric in x and y, and the resulting curve lies on *both cylinders.*

This is illustrated graphically in figure 1. The top row shows the situation of a generic function $f(\theta)$ painted onto a vertical cylinder. The leftmost panel shows the cylinder unrolled (i.e., a plot of $f(\theta)$). The successive panels show orthographic projections of the cylinders from a number of viewpoints, ending with the side view. The second row shows the same process applied for the special case of $f=\sin(2\theta)$. In the third row, the function $f=sin(\theta/2)$ is wrapped onto a horizontal cylinder. Note that the xy projections (the rightmost panels of the second and third rows) are identical. Relatively unambiguous views of the two shapes are obtained in the intermediate rotations shown in figure 1. Note that there are no three-dimensional self-intersections in the saddle-shaped curve shown in the second row, while for the pretzel-shaped curve depicted in the third row the intersection in the final projection corresponds to a self-intersection of the three-dimensional curve.

The three-dimensional curves which project to Lissajous figures are described by the following sets of parametric equations:

Case 1 (curve lies on vertical cylinder):

$$x_{\mathbf{V}}(\theta) = A_x \sin(\omega_x \theta), \tag{2.a}$$
$$y_{\mathbf{V}}(\theta) = A_y \sin(\omega_y \theta+\phi_y), \tag{2.b}$$
$$z_{\mathbf{V}}(\theta) = A_x \cos(\omega_x \theta). \tag{2.c}$$

Case 2 (curve lies on horizontal cylinder):

$$x_{\mathbf{H}}(\theta) = A_x \sin(\omega_x \theta+\phi_x), \tag{3.a}$$

$$y_{\mathbf{H}}(\theta) = A_y \sin(\omega_y \theta), \tag{3.b}$$

$$z_{\mathbf{H}}(\theta) = A_y \cos(\omega_y \theta). \tag{3.c}$$

The frequency parameters ω_x and ω_y must be integers for the curve to close on itself as θ runs from 0 to 2π. Changing the phase (ϕ_x or ϕ_y) corresponds to rotating the cylinder about its axis.

We can see the equivalence of the projected curves in these two cases by making the following substitutions:

$$\theta = \theta' + \phi_x / \omega_x, \tag{4.a}$$

$$\phi_y = -\phi_x \frac{\omega_y}{\omega_x}. \tag{4.b}$$

By substituting these values into equations 2.a and 2.b, it is easy to see that

$$x_{\mathbf{V}}(\theta') = x_{\mathbf{H}}(\theta) \quad \text{and} \quad y_{\mathbf{V}}(\theta') = y_{\mathbf{H}}(\theta). \tag{5.a,b}$$

The z function is irrelevant since we assume orthographic projection onto the xy plane.

In order to animate the figures we let the phase (ϕ_x in equation 3.a) be a function of time:

$$\phi_x(t) = 2\pi\omega_t t, \tag{6}$$

where the parameter ω_t represents the angular velocity in revolutions per unit time. Note that, from equation 4.b, the angular velocities in the two interpretations differ by the ratio of the frequencies of the generating functions. Figure 2 depicts a few frames of a sequence, together with oblique views of the two possible shapes. An example showing the velocities associated with each of the two interpretations is shown in figure 3.

2.2. Psychophysical procedures

The stimuli were presented on a cathode ray tube (Tektronix model 611). Signals for the X and Y deflections were produced by digital-to-analog converters, or DAC's (ADAC models 1023AD & 1023EX), under the control of a PDP11-23 computer. The x and y gains of the display scope were carefully adjusted to provide the same spatial displacement for a given DAC increment, thereby correcting for any possible gain differences between the two DAC's.

The digital-to-analog converters (DAC's) incorporated a direct memory access (DMA) controller (ADAC model 1620DMA), which allowed lists of coordinate pairs to be rapidly transferred from memory. After each pair of coordinates was transferred, the interface generated a brief pulse which was used for the Z (brightness) input to the CRT. This was a TTL logic pulse which had a duration approximately 1 microsecond. The front panel controls were adjusted to produce the maximum possible luminance. The time needed to write a single point with this apparatus was approximately 10 microseconds. Curves described by equations 3.a and 3.b were produced by plotting 512 points at uniformly sampled values of the parameter θ. Unfortunately, this resulted in a nonuniform spacing of the points along the curve; although the point spacing was always less than the spot size (so the curves appeared continuous), this did result in small intensity variations along the curve, inversely proportional to $\frac{\partial s}{\partial \theta}$. The coordinate lists for the sequence of frames comprising a single stimulus were computed (and resident in memory) before the onset of the stimulus; the DMA transfers for the individual frames were initiated following interrupts from a real-time clock. The frame rate was 100 Hz. The computation of the coordinate lists for each stimulus was speeded by using table look-up to access pre-computed values of sine and cosine; before each trial, these tables were scaled by the aspect ratio factors to reduce the number of multiplies needed.

Each trial consisted of a 2 second presentation of a figure, defined parametrically by equations 3.a and 3.b. For a given experimental condition, the frequency parameters ω_x and ω_y were fixed, but the aspect ratio A_x/A_y was varied from trial to trial. The product of A_x and A_y (and therefore the swept area of the stimulus) was held constant. The temporal frequency ω_t which determined the rotation frequencies was set so that the faster of the two rotations had a rate of 1 Hz.

After each trial the subjects were instructed to report whether the figure was perceived in "rolling pin" motion or "merry-go-round" motion. Although the subjects almost universally reported that the stimuli appeared three-dimensional, they were informed that in the event that they did not see a three-dimensional figure, they could make the judgement on the basis of whether the two-dimensional motion was primarily up and down (rolling pin) or side to side (merry-go-round). The subjects were also instructed that in the event that the percept changed during the course of the stimulus presentation, they should base their response on the appearance at the end of the presentation. Subjects entered their responses using the detached keyboard of the computer console.

It was noted in pilot experiments that elongation of the figure in one dimension tended to cause the rotation axis to be perceived in the same dimension as the elongation, i.e. large values of A_x/A_y produced a "rolling-pin" percept, while small values produced a "merry-go-round" percept. An up-down staircase was therefore used to control the selection of successive values of A_x/A_y, such that a "rolling pin" response decreased the value of A_x/A_y by a constant factor, while a "merry-go-round" response would increase it by the same factor. The factor used was 0.1 log units, or approximately 1.26.

Subjects were tested under six conditions, consisting of two orientations of three pairs of values for ω_x and ω_y. These were: ω_x=2, ω_y=1; ω_x=3, ω_y=1; and ω_x=3, ω_y=2. The remaining three conditions were obtained by simply exchanging ω_x and ω_y. Corresponding pairs of conditions were always run together. Two pairs of conditions were combined to make a block. Each of the three possible blocks was run twice, resulting in 4 replications of each condition. Within a block, each condition was assigned a single staircase; the trials were clustered into groups of four consisting of one trial from each staircase. Within each cluster of trials the order of the conditions was controlled by a pseudo-random number generator. Within each block, 50 judgments were made for each condition.

The subjects consisted of the one experienced psychophysical observer (the author), and an undergraduate student who had some practice in making pycho-physical judgements, but was naive with respect to the purpose of the experiment. Additional subjects were tested in individual conditions, but did not complete the full experimental protocol; the (incomplete) results from these subjects were similar to that shown for the two subjects who completed the full regimen of observations. More recently, and additional experienced subject (JAP) was run using a different apparatus. This apparatus consisted of a raster graphics system with a frame rate of 60 Hz. Subject JAP completed 3 blocks in which all 6 conditions were interleaved. All other parameters were identical to those described above.

Typical data from a single run of a single condition are shown in figure 4. The percentage of "rolling pin" responses is plotted against the log of A_x/A_y. The raw data from each block were fit with a cumulative normal using a weighted least-squares fitting procedure, described in detail by Mulligan and MacLeod [13]. The inflection point of the curve is located at aspect ratio for which we would expect to receive an equal number of "rolling pin" and "merry-go-round" responses; we shall refer to this aspect ratio as the *critical aspect ratio* (CAR). For each of the six conditions, four replications provided independent estimates of the CAR for each subject. The fitting procedure also estimated the semi-interquartile difference (SIQD), which is the change in the abscissa (log aspect ratio) required to change the response rate from 50% to 25% or 75%.

2.3. Smoothness estimates

Predictions were made using an adaptation of the computation proposed by Hildreth [8,9]. Hildreth defined a variational measure on possible two-dimensional velocity fields, and solved for the velocity field which minimized this quantity, calling this the "smoothest velocity field." Here we adopt a much simpler approach: instead of finding the minimum, we simply compute the "roughness" of each of the two rigid solutions, and assume that the visual system will prefer the interpretation having the lower value. No claim is made that either solution actually corresponds to a local minimum of the roughness function.

Following Hildreth, the quantity used to define "roughness" was

$$\mathbf{R}(\vec{v}) = \int \left| \frac{\partial v}{\partial s} \right|^2 \mathrm{d}s \tag{7}$$

This integral was approximated as a discrete sum:

$$\mathbf{R}(\vec{v}) = \sum_{i=1}^{N} \frac{|\partial v / \partial \theta|^2}{|\partial s / \partial \theta|} \mathrm{d}\theta \tag{8}$$

where θ is the parameter used to trace out the curves in equations (2) and (3) above, and dθ is equal to $\frac{2\pi}{N}$. This quantity was computed for each of the two possible interpretations; for rotations about a horizontal axis, all of the velocities are in the vertical direction:

$$\vec{v}_y = A_y \omega_y \cos(\omega_y \theta), \tag{9}$$

$$\frac{\partial \vec{v}_y}{\partial \theta} = -A_y \omega_y{}^2 \sin(\omega_y \theta), \tag{10}$$

For vertical axis rotation, all of the velocities are horizontal:

$$\vec{v}_x = A_x \omega_x \cos(\omega_x \theta + \phi_x), \tag{11}$$

$$\frac{\partial \vec{v}_x}{\partial \theta} = -A_x \omega_x{}^2 \sin(\omega_x \theta + \phi_x), \tag{12}$$

The remaining quantities needed to evaluate the sum in equation (8) are:

$$\frac{\partial s}{\partial \theta} = \left(\left(\frac{\partial x}{\partial \theta} \right)^2 + \left(\frac{\partial y}{\partial \theta} \right)^2 \right)^{1/2}, \tag{13}$$

$$\frac{\partial x}{\partial \theta} = A_x \omega_x \cos(\omega_x \theta + \phi_x), \tag{14}$$

$$\frac{\partial y}{\partial \theta} = A_y \omega_y \cos(\omega_y \theta). \tag{15}$$

Let us use the terms $\mathbf{R}_x$ and $\mathbf{R}_y$ to represent the roughnesses computed for motion about vertical and horizontal axes, respectively. These quantities depend on the amplitude factors A_x and A_y, the frequencies ω_x and ω_y, and the rotational phase ϕ_x. It turns out that the roughnesses vary as a function of rotational phase; this variation is shown in figure 5, where the log of the inverse smoothness is plotted as a function of phase. The phases at which the roughness measure attains a maximum correspond to rotational positions where the front and back sections of the generating curve project onto the same curve in the image, as occurs in the left-most and right-most panels of figure 2.

The measure of roughness defined in equation (8) is affected by the total arc length, which is a function of rotational phase. If we double *both* of the amplitude factors A_x and A_y, the roughnesses $\mathbf{R}_x$ and $\mathbf{R}_y$ also double (as does the total arc length). Thus we see that if we wish to have a roughness measure which depends only on shape and not on absolute size, we might obtain this by dividing by the total arc length L:

$$L = \int_0^{2\pi} \frac{\partial s}{\partial \theta} d\theta. \tag{16}$$

For the parameter values used to generate the data shown in figure 5, however, arc length variations as a function of phase are less than 5%, so the qualitative picture is not affected by this change.

Although the roughnesses vary as a function of phase, it can be seen from figure 5 that the *ratio* of the roughnesses is relatively constant, as indicated by the roughly constant vertical separation on the log ordinate. Log roughness ratio as a function of phase is plotted in figure 6 for different aspect ratios. Note that by taking the ratio of the roughnesses, we have divided out the effect of total arc length.

A single number characterizing the relative smoothness of the two interpretations was obtained by integrating the log roughness ratio across rotational phase. This was justified on the grounds that the ratio was relatively constant across phases, and because, although subjects' percepts are bistable, the transitions do not seem to be phase-locked with the rotation. Once this average ratio has been computed, we can estimate the predicted value of the critical aspect ratio by solving for the aspect ratio which yields a mean log roughness ratio of 0. Because the visual system might integrate on some other transformed representation, we should be prepared to accept an error on the order of the vertical variation of the curves in figure 6.

Figure 7 plots log roughness ratio as a function of log aspect ratio for $\omega_x=1$, $\omega_y=2$, and $\omega_x=2$, $\omega_y=3$. The points represent the mean of 256 different phases, sampled uniformly over the interval 0 to $\pi/4$. The sample phases were placed so as to straddle the phases at which the singularities occur, such as $\phi_x=0$. It may be observed that the points fall close to a straight line with a slope of 1/2. Linear regression was used to fit a line to the points; the log of the critical aspect ratio (CAR) was taken to be the X intercept from the regression equation. Critical aspect ratios were obtained in this way for a number of pairs of frequencies (ω_x,ω_y); log CAR is plotted against log frequency ratio ω_x/ω_y in figure 8, shown by the filled squares. Note that the points fall close to a line with a slope of -2. A weak explanation for this can be made from the fact that A_x appears in equation 10 with an exponent of 1, while ω_x appears with an exponent of 2. The graph figure 8 is symmetric: the positions of the points in the upper left quadrant are simply the positions of the points in the lower right quadrant reflected through the origin. In a practical sense, this means that the shape of the figure at the critical aspect ratio is unaffected if the entire figure is rotated 90 degrees.

2.4. Deviations from the smoothest velocity field

In the previous section, predictions have been made on the basis of which of the two rigid interpretations is "smoother," i.e., which rigid interpretation has the lower value of the inverse smoothness measure introduced by Hildreth. In this section, a slightly different approach is explored: we first solve for the smoothest velocity field (which does not correspond to rigid motion of a three dimensional figure), and then ask to which of the two possible rigid interpretations is it more similar.

The smoothest velocity field is obtained in the manner described by Hildreth. The expression for the roughness (equation 8 above) is differentiated with respect to the (unknown) tangential component of the velocity at each sample point, and then equated to zero. The result is a system of N linear equations in N unknowns which is easily solved using standard linear algebra. Figure 9 shows the resulting velocity field for a representative figure.

A measure of similarity between two velocity fields was formed by summing the squared vector differences between the rigid velocities and the corresponding velocity vectors from the smoothest field. This quantity was computed for each of the two rigid interpretations, at a number of different aspect ratios for each set of parameter values. When the log ratio of these two quantities is plotted versus log aspect ratio, a pattern of results similar to that seen in figure 7 is obtained. Linear regression was used to estimate the x intercept, i.e. the critical aspect ratio (CAR) for which the smoothest velocity field was equally different (using the integrated squared vector difference metric) from each of the two rigid solutions. The predicted CAR's generated by this method are shown by the open circles in figure 8, along with the predictions from the original method. Like the original predictions, these new predictions fall on a line with a slope close to -2, although this line is slightly steeper than that describing the original predictions.

3. RESULTS

The experimental results are shown in figure 10. The mean log CAR over the four replications is plotted as a function of log frequency ratio for the two subjects LR (triangles) and JBM (circles). The small filled squares show the corresponding predictions from figure 8. The numerical data used to generate figure 10 are given in table 1, along with the standard errors.

Several features of figure 10 are notable. First, although the data do deviate from the predictions, the overall slope of the data points is close to -2, like the prediction, giving qualitative support to our modified Hildreth model. Secondly, the deviations from the predictions are exclusively upwards from the predictions, indicating a bias in favor of the "merry-go-round" percept. It should be noted that the bias is evinced not simply by deviations from the predictions, but from the fact that pairs of points which correspond to two orientations of a given shape are not located symmetrically with respect to the origin. Any set of data possessing central symmetry would be evidence for isotropy, regardless of how unlike the prediction it might be. The prediction *does* show this symmetry, since there is no anisotropy built into the model.

Although the deviations from the predictions shown in figure 10 look rather modest, it should be noted that the ordinate is a logarithmic scale, so small deviations correspond to profound differences in shape. In order to assess the significance of the anisotropy, however, it is necessary to compare these deviations with the range of aspects ratios for which the percept is bistable. This is indicated by the transition zone of the psychometric function shown in figure 4, which is typical for all subjects. In most cases the size of the anisotropy effect (sum of the logs of the CAR's for corresponding conditions) is larger than the width of this this transition zone, indicating that there are stimuli which are consistently perceived in merry-go-round motion regardless of the orientation in which they are presented.

4. DISCUSSION

4.1. The vertical-horizontal illusion

The *vertical-horizontal illusion* (VHI) refers to the fact that a vertical line will appear longer than a horizontal line of the same physical length. The details of this much-studied illusion are summarized well by Robinson [14]. Observers similarly overestimate the vertical component of motion in obliquely moving targets [15]. This suggests a simple explanation of the anisotropy observed in the data: namely, that the visual input is subjected to an affine distortion prior to the analysis of motion. A deformation of the image consistent with the VHI *would* produce an anisotropy of the correct sign for the subjects shown in figure 10; if the visual input were stretched in the vertical dimension by a factor α and compressed in the horizontal dimension by the same factor, then the lines in figures 8 and 10 would simply be shifted horizontally by an amount $2\log(\alpha)$. We can estimate the amount of deformation needed to account for the subjects data by calculating the horizontal shift necessary to bring the regression lines for the predicted and observed results in figure 10. When this is done, an aspect ratio factor of 1.20 (0.079 log units) is obtained for subject JBM, and a factor of 1.21 (0.0825 log units) for subject LR.

4.2. Anisotropy in 2-D apparent motion correspondence

Gengerelli [16] has demonstrated an anisotropy of two-dimensional apparent motion correspondence with the stimulus illustrated in figure 11.a, sometimes referred to as a "bistable quartet." The stimulus consists of two pairs of luminous dots which are flashed in alternation. Each pair is located at opposite corners of a rectangle. When the pairs are alternated in time, several percepts are possible: The two dots which are visible at any given time may be seen to oscillate in either a horizontal or vertical direction. It would also be physically consistent for the dots to be seen in circulating motion around the perimeter of the figure, but this is rarely observed.

In this stimulus the aspect ratio of the figure affects the perceived direction of motion. When the horizontal separation is very small compared to the vertical separation, it is more likely that horizontal motion will be seen. It is possible to measure a psychometric function relating aspect ratio to the proportion of the time that horizontal motion is seen. The inflection point of this psychometric function corresponds to a critical aspect ratio for this task, i.e. the aspect ratio for which horizontal and vertical motions are equally likely to be perceived. With an aspect ratio of unity, the subjects showed a preference for *vertical* correspondence when the figure was fixated centrally. Note that this is the *opposite* of what would be predicted if motion correspondence were determined simply by proximity *after* a deformation consistent with the vertical-horizontal illusion.

Gengerelli [16] found that this bias disappeared when the display was fixated eccentricly, so that the entire display fell within a single cortical hemifield, and concluded that the bias resulted from a preference to make correspondences within a cortical hemifield. Ramachandran, Cronin-Golomb and Myers [17] performed a similar study on commissurotomy or "split brain" patients, and found that although the bias was exaggerated, the stimulus was still quite ambiguous, suggesting that the perception of apparent motion across the vertical midline was easily mediated by subcortical structures.

These results pose a puzzle with respect to the results of the present experiment: if there is, for whatever reason, a preference for vertical correspondences in ambiguous motion displays, then we would expect to see a preference for "rolling pin" rotation in the ambiguous KDE figures used in the present study. This is opposite to the bias observed in the present study.

If, as is commonly supposed, computation of two-dimensional optic flow is a necessary precursor to the computation of structure from motion, then any biases inherent in the two-dimensional process should be reflected in the responses of the three dimensional system. If there are no biases in the two-dimensional motion process, then there cannot be any biases in the three dimensional process *if* the ambiguity must be resolved at the two-dimensional stage.

An alternative possibility is that two-dimensional ambiguities are *not* resolved prior to computation of 3-D structure. An architecture which would permit this is a distributed representation in which all possible two-dimensional velocities are represented; the perceived direction would usually correspond to the most active unit, but less active units could still pass their signals to higher levels. Even if a bias existed at an early stage, it would be possible for a different (stronger) bias at a later stage to dominate the resolution of the ambiguity. Distributed models for solving the aperture problem have been proposed by Heeger [18], Sereno [19], and Perrone [20]. Simoncelli *et al.* [21] have proposed a form of this architecture based on probabilities which includes stages for three-dimensional representation.

4.3. Ecological considerations

We have considered several possibilities for the site at which the bias is introduced; we have not, however, said anything about *why* the bias might be present, or if it has any functional significance. The possibility has been mentioned that the bias is a direct result of an early warping of the visual field by the vertical-horizontal illusion, and has no intrinsic significance. If, on the other hand, the bias is restricted to the interpretation of three dimensional objects and scenes, then we are faced with the intriguing possibility that the bias arises due to some aspect of the three dimensional world. According to signal detection theory an observer's criterion is influenced by the *a priori* probability of a given stimulus [22]; in the present case, it would be sensible to expect a bias similar to that observed if "merry-go-round" motion is in fact more prevalent in the environment.

Does such an ecological imbalance exist? If so, it is likely to arise from *observer self-motion,* as opposed to the motion of other objects. We note that there is usually a rotational component to the relative motion between objects and a moving observer, the axis of which depends on the relation between the object and the direction of motion. For example, when an ambulant observer moves through the forest, the tree trunks at eye level have a small component of merry-go-round motion in addition to a large translational component. Similarly, when the observer surmounts a fallen log, the log has a component of rolling pin motion. Perhaps the bias developed in creatures living under open skies, so that they never had any fallen logs overhead, i.e. that portion of the superior visual field which would have produced a component of rolling pin motion was devoid of pattern. This argument seems somewhat contrived, however, and requires many assumptions about both the nature of the environment and the behavior of the observer with regard to locomotion and eye movements. If one assumes that an ambulant observer tends to look in the direction of motion, then one would expect that the types of motion encountered would be different for the different regions of the visual field; in particular, this might produce a bias towards rolling-pin motion in the inferior visual field. An investigation of the dependence of the observed bias on position in the visual field would be an interesting topic of future research.

5. Conclusions

In spite of the fact that Hildreth's theory does not predict the anisotropy seen in this study, the fact that a modified version of the theory predicts the correct dependence on frequency ratio (i.e., the correct slope in figure 10) is strong support for the theory. The theory can easily be made to predict the biases if an affine transformation consistent with the vertical-horizontal illusion is assumed to precede the motion analysis. Reported anisotropies in two-dimensional motion correspondence, however, are inconsistent with this view. One possibility is that the phenomena studied by Ramachandran *et al.* involve completely different mechanisms subject to their own distinct biases. An alternative explanation is that the observed bias is introduced at a level involving three-dimensional representation; an implication of this hypothesis is that the two-dimensional aperture problem is not resolved independently of three-dimensional interpretation.

6. ACKNOWLEDGMENTS

The bulk of the experimental data were collected while the author was a graduate student in the Department of Psychology at the University of California at San Diego and supported by NIH grant EY-01711 (awarded to Professor Donald I. A. MacLeod, whose personal support is also gratefully acknowledged) and an IBM graduate fellowship. Computational modeling and preparation of the manuscript was performed at NASA Ames research Center, where the work was supported by NASA RTOP 506-47. The initial implementation of the Hildreth algorithm was based on a Mathematica program provided by Maggie Shiffrar and Al Ahumada. Thanks to John Perrone for donating some subject hours.

7. REFERENCES

1. W.R. Miles, "Movement interpretation of the silhouette of a revolving fan," *Amer. J. Psychol.,* **43**, 392-405, 1931.

2. H. Wallach and D. O'Connel, "The kinetic depth effect," *J. Exp. Psychol.,* **45**, 205-217, 1953.

3. S. Ullman, *The Interpretation of Visual Motion,* MIT Press, Cambridge, Mass, 1979.

4. S. Ullman, "Maximizing rigidity: the incremental recovery of 3-D structure from rigid and rubbery motions," *Perception,* **13**, 255-274, 1984.

5. H. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image," *Proc. R. Soc. Lond. B,* **208**, 385-397, 1980.

6. D. D. Hoffman, "Inferring local surface orientation from motion fields," *J. Opt. Soc. Am.,* **72**, 888-892, 1982.

7. D.J. Heeger, and A. Jepson, "Visual perception of three-dimensional motion," *Neural Computation,* **2**, 127-135, 1990.

8. E. Hildreth, "The computation of the velocity field," Proc. R. Soc. Lond. B,

GET VOL & PAGES, 1984.

9. E. Hildreth, "Computations underlying the measurement of visual motion," *Art. Intell.*, **23**, 309-355, 1984.

10. C.O. Weber, "Apparent movement in Lissajou figures," *Am. J. Psychol.,* **42**, 647-649, 1930.

11. B.R. Philip and V.R. Fisichelli, "Effect of speed of rotation and complexity of pattern on the reversals of apparent movement in Lissajou figures," *Am. J. Psychol.*, **58**, 530-539, 1945.

12. V.R. Fisichelli, "Effect of rotational axis and dimensional variations on the reversals of apparent movement in Lissajous figures," *Am. J. Psychol.,* **59**, 669-675, 1946.

13. J.B. Mulligan and D.I.A. MacLeod, "Reciprocity between luminance and dot density in the perception of brightness," *Vision Res.*, **28**, 503-519, 1988.

14. J.O. Robinson, *The Psychology of Visual Illusion,* Hutchinson & Co., London, 1972.

15. R.B. Post, and M. Chaderjian, "Perceived path of oblique motion: horizontal-vertical and stimulus-orientation effects," *Perception*, **16**, 23-28, 1987.

16. J.A. Gengerelli, "Apparent movement in relation to homonymous and heteronymous stimulation of the cerebral hemispheres," *J. Exp. Psychol.*, **38**, 592-599, 1948.

17. V.S. Ramachandran, A. Cronin-Golomb, and J.J. Myers, "Perception of apparent motion by commissurotomy patients," *Nature (Lond.),* **320**, 358-359, 1986.

18. D.J. Heeger, "Model for the extraction of image flow", *J. Opt. Soc. Am. A,* **4**, 1455-1471, 1987.

19. M.E. Sereno, "Implementing stages of motion analysis in neural networks", Proc. 9th Ann. Conf. Cog. Sci. Soc., 405-416, 1987.

20. J.A. Perrone, "Simple technique for optical flow estimation," *J. Opt. Soc. Am. A*, **7**, 264-278, 1990.

21. E.P. Simoncelli, D.J. Heeger, and E.H. Adelson, "Perception of 3D motion in the presence of uncertainty," *Invest. Ophth. Vis. Sci., suppl.,* **31**(4), 173, 1990.

22. D.M. Green, and J.A. Swets, *Signal Detection Theory and Psychophysics,* Wiley, NY, 1966.

8. FIGURE LEGENDS

Figure 1: Illustration of the three dimensional ambiguity of Lissajous figures. The top row depicts a random waveform (shown on the left), which is "rolled" into a cylinder. Several views of the cylinder are shown. In the second row, the random waveform is replaced by two cycles of a sinusoid. When this curve is rolled into a cylinder and viewed from the side (far right), the parametric equations describing the projected curve are symmetric in x and y (with the exception of the frequency parameters), implying that the projected curve could equally well lie on a horizontally oriented cylinder. The curve which generates the same projection when rolled into a horizontal cylinder is shown in the bottom row. Although only a single phase is shown, this ambiguity remains when the cylinders are rotated.

Figure 2: Illustration showing a series of possible stimulus frames together with oblique views of the two possible generating shapes. The upper row shows slightly oblique views of the corresponding vertical cylinder inscribed with the saddle-shaped figure, while the lower row depicts views of a horizontal cylinder inscribed with the pretzel-shaped figure. Note that the cylinder in the upper row rotates through 90 degrees from left to right, while the cylinder in the lower row rotates through a full 180 degrees.

Figure 3: Diagram showing the velocities associated with the two possible interpretations of an animated Lissajous figure. On the left are shown the velocities corresponding to perceived rotation about a vertical axis ("merry-go-round" motion). This corresponds to the saddle-shaped figure shown in the middle row of figure 1. Note that the velocities of the two limbs which intersect in the middle of the figure have opposite directions. On the right are shown the velocities corresponding to rotation about a horizontal axis ("rolling pin" motion); this corresponds to the pretzel-shaped figure shown in the bottom row of figure 1. Note that where the curve intersects itself in the center the velocities match, since the intersection in the figure corresponds to an actual three-dimensional intersection in the projected figure.

Figure 4: Typical data from a single run of the experiment. The abscissa represents the log of the aspect ratio A_x/A_y, while the vertical axis represents the proportion of responses indicating "rolling pin" motion seen. The open circles are for the condition $\omega_x=2, \omega_y=1$, while the filled circles are for the dual condition $\omega_x=1, \omega_y=2$. Each curve represents 50 judgments collected with a single staircase. Raw data such as these were fit with a cumulative normal to estimate the critical aspect ratio at which the two percepts were equally likely. For these data, the fitting procedure produced estimates of the critical aspect ratio of -0.30 and 0.64. The inequality of the absolute values of these numbers is evi-

dence of anisotropy or bias.

Figure 5: Variation of the logarithms of the roughnesses $\mathbf{R}_x$ and $\mathbf{R}_y$ as a function of rotational phase, which is expressed as a fraction of 2π (one complete rotation). The plot shows the log of roughness computed at 256 uniformly spaced values of rotational phase ϕ_x. The values of the parameters were $A_x = A_y = 1$, $\omega_x = 1$, $\omega_y = 2$, $N = 1024$. The upper curve is the roughness computed for the "rolling pin" interpretation, $\mathbf{R}_y$, while the lower curve is the roughness computed for the "merry-go-round" interpretation, $\mathbf{R}_x$. The separation of the curves indicates a strong preference of the model for the "merry-go-round" interpretation for this set of parameter values.

Figure 6: Variation of log roughness ratio as a function of rotational phase. The values of the parameters were $\omega_x = 1$, $\omega_y = 2$, $N = 1024$. The three curves represent three different values of the aspect ratio A_x / A_y; from the upper curve to the lower the aspect ratios were 4.0, 1.0, and 0.25.

Figure 7: Log roughness ratio integrated over rotational phase as a function of log aspect ratio. Log aspect ratio was sampled uniformly in 20 steps from -1 to 1. The values of the parameters used to generate the lower line (filled squares) were $\omega_x = 1$, $\omega_y = 2$, $N = 1024$, while the upper line (triangles) represents $\omega_x = 2$, $\omega_y = 3$.

Figure 8: Log critical aspect ratio as a function of log frequency ratio ω_x/ω_y. The points sampled from the abscissa correspond to the following ordered pairs (ω_x, ω_y): (1,5), (1,4), (1,3), (2,5), (1,2), (3,5), (2,3), (3,4), (4,5), (5,4), (4,3), (3,2), (5,3), (2,1), (5,2), (3,1), (4,1), (5,1). Filled squares indicate predictions based on the roughness of the rigid interpretations; open circles indicate predictions based on difference between the rigid interpretations and the smoothest velocity field.

Figure 9: The smoothest velocity field, computed using Hildreth's algorithm.

Figure 10: Log critical aspect ratio versus log frequency ratio for three subjects are plotted together with model predictions. The model predictions are indicated by the small squares which lie on the straight line with slope approximately equal to -2; filled triangles indicate data for subject LR, filled circles subject JBM, large open squares subject JAP. Frequency/aspect ratio combinations in the upper right quadrant are seen primarily in "rolling pin" motion, while those in the lower left quadrant are seen primarily in "merry-go-round" motion. The negatively sloped lines indicate the boundary in the parameter space between these two regimes. The fact that pairs of data points representing rotated stimuli are not located symmetrically with respect to the origin indicates the anisotropy, which for these data favors the "merry-go-round" interpretation.

Figure 11: a) Stimulus configuration used by Gengerelli [16], Ramachandran *et al.* [17]. The open circles represent dots present at time t_1, which are replaced by dots at the positions shown by the filled circles at time t_2. When this sequence is presented cyclically, the percept is usually of a pair of dots in oscillatory motion, either side-to-side or up-down. b) The perceived direction of motion can be biased by changing the aspect ratio of the figure.

subject	ω_x	ω_y	log(CAR)	SEM
JBM	1	2	0.667	0.048
JBM	2	1	-0.385	0.084
JBM	1	3	1.030	0.020
JBM	3	1	-0.752	0.066
JBM	2	3	0.680	0.052
JBM	3	2	-0.294	0.032
LR	1	2	0.657	0.040
LR	2	1	-0.185	0.022
LR	1	3	0.913	0.020
LR	3	1	-0.547	0.064
LR	2	3	0.436	0.020
LR	3	2	-0.283	0.062
JAP	1	2	0.594	0.011
JAP	2	1	-0.201	0.003
JAP	1	3	0.869	0.007
JAP	3	1	-0.673	0.004
JAP	2	3	0.502	0.005

JAP	3	2	-0.318	0.007

Table 1: Raw data used to generate the graph in figure 9. For subjects JBM and LR, the mean log of the critical aspect ratio (CAR) was computed over four replications of each of the six frequency pairs, while three replications were used for subject JAP. The fifth column shows the standard error for each mean.