



Interpolating sampled contours in 3-D: analyses of variability and bias

Paul A. Warren, Laurence T. Maloney^{*}, Michael S. Landy

Department of Psychology & Center for Neural Science, New York University, 6 Washington Place, New York, NY 10003, USA

Received 4 October 2001; received in revised form 4 June 2002

Abstract

In two experiments, we examined how observers interpolated the missing parts of sampled, planar contours in 3-D space. We varied (1) contour type (linear or parabolic), (2) orientation of the plane containing the contour and (3) the number of points on a sampled contour.

Interpolation performance was very accurate, comparable to results from Vernier tasks. Setting variability was highest along the line of sight and for the parabolic contour. Setting variability did *not* decrease with increasing number of points on either contour, suggesting that observers do not use all available, relevant information in this task.

© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Contour interpolation; Vernier acuity; Localization; Contour

1. Introduction

In everyday, cluttered scenes such as the one depicted in Fig. 1, nearer objects sometimes block our view of farther objects. Occlusions introduce two additional difficulties into the already difficult task of object recognition. First, the visual system must decide which visual fragments go together (the *grouping problem*), and, second, it must estimate the shape and position of the occluded parts of the objects, in particular their occluding contours (the *interpolation problem*).

Parts of surfaces and contours may be missing not only because they are occluded, but instead because the local visual information defining them is inadequate. Large parts of the occluding contour of a camouflaged animal may be simply undetectable in isolation. The wealth of illusory contour phenomena such as the Kanizsa figures (Kanizsa, 1979), suggest that, for a biological visual system, it is important to estimate an underlying boundary when information is incomplete.

The diagram in Fig. 2 serves to illustrate the complementary problems of grouping and interpolation. It is

likely that the visual system solves these two problems cooperatively, with the tentative results of surface and contour interpolation influencing grouping, and conversely (Gepshtein & Kubovy, 2000).

In this article, we are interested primarily in the interpolation of *sampled contours*, 1-D curves that are invisible except for a few discrete points (illustrated in Fig. 2). Several investigators have studied the grouping problem for contours, but few have focused on mechanisms of contour interpolation. A consequence of this imbalance is that we know very little about how human observers estimate the missing and degraded parts of contours.

An evident first question to consider is, what family of curves do human observers use in interpolating the gaps in the contours? There are infinitely many smooth curves that could be used to complete the missing parts of the contours in Fig. 1. Precisely which ones are available to the human visual system? If we asked human observers to complete a parabolic contour across a gap, would they choose completions that coincide with the parabola or would their responses be biased away from the parabola and toward a ‘human visual spline’?

A second question is how *reliably* can human observers estimate the parts of contours not visible in Fig. 1? Human performance in interpolating linear contours, as measured in Vernier tasks, is remarkably good and is

^{*} Corresponding author. Tel.: +1-212-998-7851; fax: +1-212-995-4349.

E-mail address: ltm@cns.nyu.edu (L.T. Maloney).

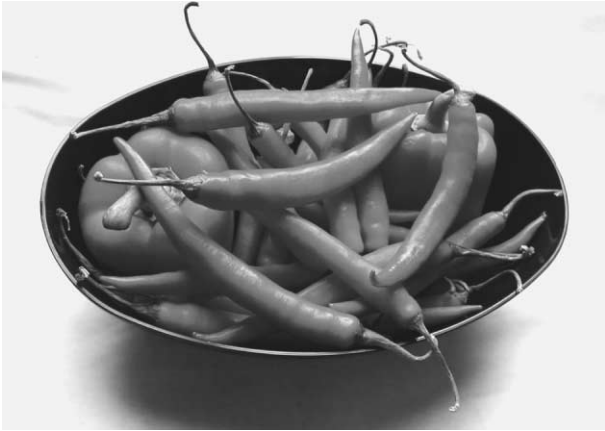


Fig. 1. A scene containing a number of interrupted contours. Accurate localization of the missing contours can potentially aid in segmenting objects, identifying them, and estimating their shapes.

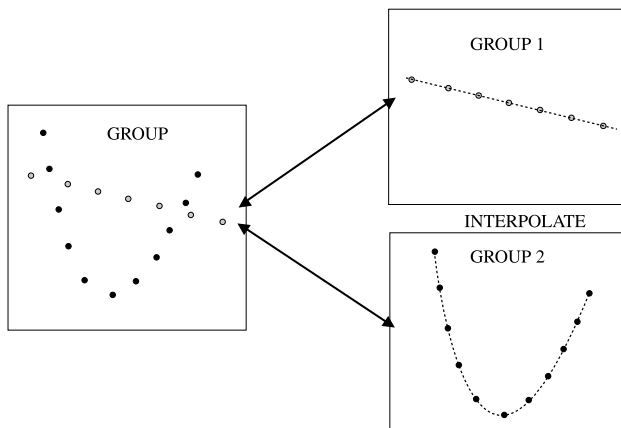


Fig. 2. A schematic representation of the grouping and interpolation problems. Both problems must be solved to obtain accurate form information.

affected primarily by the size of the gap to be interpolated. For sharply curved contours that span wide gaps, we might expect an additional increase in the variability of setting: the curve is simply more difficult to localize. We can ask how human performance degrades with curvature (*the cost of curvature*) with other factors held constant.

Third, the amount of available information about a contour presumably influences the reliability of human contour interpolation. If the visible parts of a contour are few, difficult to see, or distant from the point of interpolation, we would expect that interpolation would be less reliable than if they were many, well delineated, or near. In our experiments we vary the number of points visible on sampled contours and measure the change in reliability of interpolation.

Fourth, in normal viewing conditions, the visual system has to contend with contours (and missing con-

tour segments) that do not fall in a fronto-parallel plane. We would like to know how the visual system interpolates contours at arbitrary orientations in 3-D space.

The goal of the experiments reported here is to elucidate the four questions just posed.

1.1. Previous work

The grouping problem. The majority of studies investigating perceptual grouping attempt to provide a firm scientific base for the classical Gestalt principles of good continuation. Using oriented contour elements Field, Hayes, and Hess (1993) examined how several factors (including distance, path orientation and relative element orientation) affected an observer's ability to detect a contour embedded in a 'noise' background. They interpreted their results as evidence for the existence of an 'association field' comprising receptive fields linked together by inhibitory and excitatory connections. Pizlo, Salach-Golyska, and Rosenfeld (1997) also studied Gestalt grouping mechanisms by testing observer ability to detect curves, defined by dots, in a scene containing random background elements.

Polat and Sagi (1994) provided psychophysical evidence for the type of long-range lateral interactions defining the association field of Field et al. (1993). They have shown that contrast sensitivity for a central Gabor target can be affected depending on the orientation of two flanking targets relative to the global alignment of the three targets. Sensitivity was maximal when the angle between the oriented element and the line joining the three elements was zero—i.e. the Gabor elements were collinear ('end to end': Polat & Sagi, 1994).

Feldman (1997) proposed that the grouping of dots to form a curvilinear pattern is a probabilistic process in which contours are generated from 4-dot configurations. His data suggest an underlying local mechanism that assesses the 'regularity' of successive groups of four points, where 'regularity' is inversely related to curvature and is greatest for a contour whose curvature is zero (i.e. a straight line).

More recently, Geisler, Perry, Super, and Gallogly (2001) developed a model of contour element grouping based on measured scene statistics.

The models proposed to account for grouping can, in some cases, be extended to predict interpolation performance (see the discussion of Polat & Sagi's work above). However, these extensions are simply hypotheses. None of these studies directly measures which contour completion a human observer will select or the location of that contour.

The interpolation problem. Far fewer studies have measured interpolation performance directly. van Assen and Vos (1999) examined how observers' judgments of the collinearity of a point with two flanking reference points were affected by the presence of additional points

in the visual field. In some trials observers saw two additional ‘context points’ positioned such that, together with the reference points, they fell on four successive vertices of a regular polygon. When the points lay on a polygon with interior angle less than 30° , the context dots induced a bias in the collinearity judgement, as if the invisible line joining the two reference points had ‘bulged’ outward as a consequence of the presence of the context points. This result is quite striking—curvilinear interpolation occurred when as few as four points were presented in such a way that they could be interpreted as samples from a curvilinear contour.

Hon, Maloney, and Landy (1997) measured the contribution or influence of each point forming a sampled contour to the location of the contour. Observers interpolated parabolic sampled contours in a fronto-parallel plane by moving a displaced point until it lay on the contour. Another point was perturbed away from its base location on the contour. The *influence* of the perturbed point, in their terminology, is the ratio between the observer’s change in setting in response to the perturbation, and the size of the perturbation.¹ Their results suggest that the influence of a single point on interpolation decreases very rapidly with distance along the contour from the region of interpolation. Observers were able to interpolate a parabolic contour accurately using as few as four points other than the setting point. Influence was invariant under changes of orientation in the fronto-parallel plane and invariant under a scaling of the stimulus array by a factor of two. Hon et al.’s conclusions are consistent with Feldman’s 4-point hypothesis.

The above findings (concerning the number of points used in an interpolation task) are surprising given what we know about the relation between number of points and contour *detection* performance. Several authors (e.g. Braun, 1999; Kovacs & Julesz, 1993) report an improvement in performance as more elements are added up to a plateau at 12 elements. If we assume interpolation plays a part in successfully completing the tasks described in these studies then we might expect a similar improvement in our task as more points are added to the contour.²

¹ This definition coincides with the use of the term ‘influence’ in the robust statistics literature (Hampel, Ronchetti, Rousseeuw, & Stahel, 1986).

² It is possible that the observed improvement in detection performance can be explained by probability summation across detection of 4-point sub-contours within a 12-point contour. While we know that detection is better for 12-point contours than for 4-point contours, we cannot in fact, conclude the observers correctly segment all twelve points in the 12-point contour, just as we cannot assume anything about how observers localize the contour between the sampled points. These ‘conclusions’ go beyond the data available in detection experiments.

Our approach. In the experiments reported below, we examined the bias and variability of observer settings and used statistical methods to determine which of the contour points are important for the interpolation task. Since the vast majority of contours and surfaces in the visual scene are not confined to the fronto-parallel plane we extended the study to 3-D space. Several authors (Badcock & Schor, 1985; Blakemore, 1970; McKee, Levi, & Bowne, 1990; Ogle, 1953; Siderov & Harwerth, 1993) have noted that the location uncertainty inherent in the depth dimension is very different from that of the other dimensions. We therefore have no reason to anticipate that conclusions about interpolation performance in a fronto-parallel plane can be extended to interpolation of contours in space.

2. General methods

2.1. Apparatus

We used two Sony Trinitron Multiscan G500 monitors, positioned on either side of the observer, to display stimuli (Fig. 3). The two monitors formed part of a Wheatstone stereoscope: the image from the left monitor was projected to the observer’s left eye by a small half-silvered mirror placed at 45° to the observer’s Cyclopean line of sight. A second half-silvered mirror reflected the image of the second monitor to the observer’s right eye. The partial transparency of the mirrors facilitated spatial calibration of the monitors (described below) but played no other role in the experimental sessions. The optical distance from each eye to its corresponding monitor was approximately 68 cm. From this distance, the central region of each screen, used to display our stimuli, subtended approximately $18 \times 24^\circ$ of visual angle. The screens of the G500 monitors are close to physically flat. All stimuli were generated using

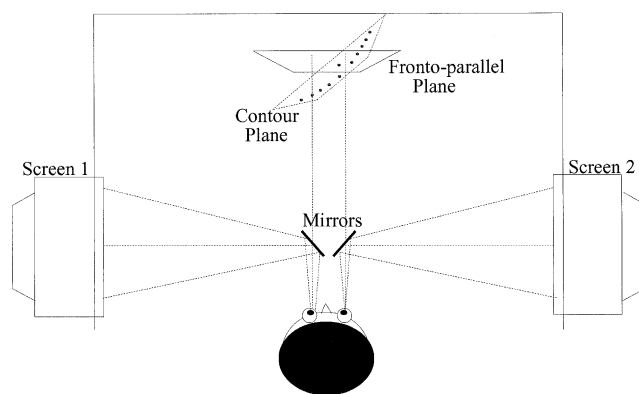


Fig. 3. Experimental apparatus. The observer was seated in a large-scale Wheatstone stereoscope.

Matlab® (The Mathworks Inc; Hanselman & Littlefield, 1997) on a Dell Precision 410 workstation running under the Linux operating system.

Observers were positioned in a chin rest and were asked to keep their heads still, although no head restraint was imposed. To minimize visual information in the scene that might have provided contextual cues, the apparatus was housed within a large box, the interior of which was covered in flocked black paper (Edmund Scientific), a highly light-absorbent surface. The observer could see only the points defining the stimulus, apparently floating in front of him or her.

We calibrated the apparatus spatially before each experimental session. Using only the left eye, the observer first viewed a 4×5 array of points on the left monitor optically superimposed on a reference target by one of the half-silvered mirrors. The reference target was a 4×5 array of points on a rigid, planar surface placed 68 cm in front of the observer. The observer moved the points on the left monitor until each coincided with the corresponding reference points. The observer repeated this alignment procedure with the right screen, using the right eye. These settings allowed us to estimate the mapping from screen coordinates on each monitor to the plane of the reference target separately. We used interpolation routines from the Numerical Recipes in C library (Press, Teukolsky, Vetterling, & Flannery, 1992) to compensate for any spatial distortions in the monitors or the alignment of the apparatus.

In the experiments described below, the task of the observer was to move a point in space until it appeared to lie on a contour sketched by other points. These points were 3-D Gaussian ‘blobs’ of light. These blobs were projected to the planes of the two CRT screens, and were rendered as 2-D Gaussians (with size inversely related to the distance from the observer to the point in 3-D). The Gaussians were created using gray levels from black, which coincided with the background to white at the highest pixel value of the CRT. The luminance values of the black background and the highest pixel value, at the center of the blob, were 0.3 and 114 cd/m² respectively.

In order to allow sub-pixel sized adjustments of the 2-D Gaussian blobs they were rendered as 11×11 pixel images with anti-aliasing as described by Georgeson, Freeman, and Scott-Samuel (1996). Using this method blobs could be centered on a location in a virtual 4×4 sub-pixel sized grid. Thus the physical spatial resolution was 1/4 of a pixel in the vertical and horizontal directions. Since each pixel is approximately $1.4' \times 1.4'$, this leads to a physical spatial resolution of approximately $0.35'$. This resolution proved to be adequate for the experiments reported here. We pre-computed blob images for each of the 16 sub-pixel positions, and for 10 different sizes of blob. The 10 different sizes represented

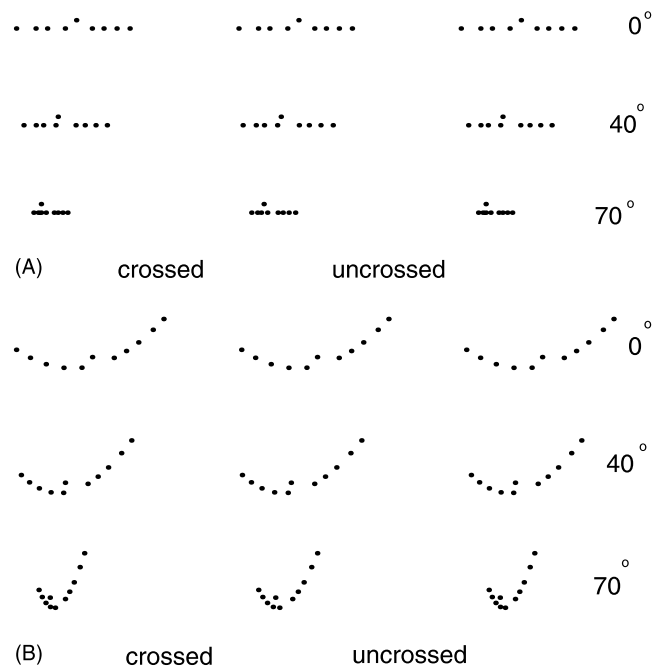


Fig. 4. (A) Stereo views of the linear segment stimuli used in Experiment 1. The left-hand pair is for crossed fusion, and the right-hand pair for uncrossed fusion. (B) Stereo views of the parabolic segment stimuli in the same format.

a gross approximation³ of the size cue due to perspective projection of the 3-D blobs into 2-D. Images were presented using gamma-corrected lookup tables. In the fronto-parallel plane condition, a blob on the contour subtended approximately $15'$.

Stimulus configurations. We used two different contour segments, linear and parabolic. The observer saw only *sampled points* constrained to lie on the segments, as illustrated in Fig. 4. For the parabolic segment, we selected 11 sample points: the two endpoints and nine more points whose positions were computed to be equally spaced in arc-length along the contour. For the linear segment nine sample points (two endpoints and 7 interior points) were similarly defined. We refer to such collections of points constrained to lie on an otherwise invisible contour as *sampled contours*. The number of points visible is one of the independent variables of interest and will be discussed later.

If the observer knew that the distance between successive points along the contour was always the same, he

³ The sizes were selected by choice of the standard deviation of the pre-computed Gaussians and ranged from 2.2 to 11 pixels. Since the Gaussians were truncated to an 11×11 grid, the Gaussian of size 11 was not as much as five times larger than the Gaussian of size 2.2. In this experiment, the total depth range used was approximately 70 ± 15 cm. Thus, the ratio between the largest size used and the smallest should be $85/55 = 1.55$. The ordering of stimulus size increased with virtual depth but, because of the truncation of the Gaussian on the 11×11 grid, our technique represented an exaggerated ordinal scale of size changes.

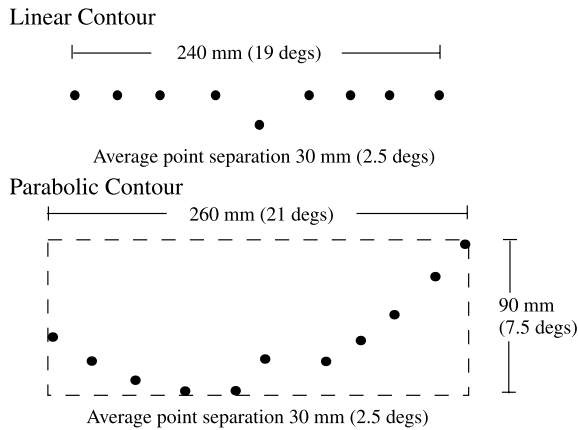


Fig. 5. A summary of the spatial dimensions of the stimuli and the corresponding visual angles when the stimuli are in the fronto-parallel plane.

or she might use spacing as a cue in interpolating the contour. We therefore ‘jittered’ the positions of the points along each of the contours. Specifically, each point was shifted along the contour by an amount selected at random from a Gaussian distribution with standard deviation equal to 5% of the contour length. Contour lengths were approximately 240 mm for the line and 300 mm for the parabola. Note that after jittering all points still fell precisely on their contour, but spacing between successive points was clearly non-uniform (Fig. 4). The locations of the points on the contour were not varied on a trial-by-trial basis but, instead, remained constant throughout the experiment.

The horizontal and vertical extents of the parabolic section in the contour plane were approximately 260 and 90 mm (21° and 7.5° visual angle in the fronto-parallel condition) respectively. The horizontal extent of the linear segment was 240 mm (19° visual angle in the fronto-parallel condition). The average linear separation of the points in the stimuli was about 25 mm for the line and 30 mm for the parabola when presented in the fronto-parallel plane. The dimensions of the stimuli are summarized in Fig. 5.

Clearly a rotation of the linear segment about a horizontal axis would lead to no change in stimulus. Thus, in Experiment 2 where such rotations are investigated, vertical linear contours were used instead. Also, to fit the contour in the screen it was reduced in scale by a factor of 0.75.

2.2. Procedure

In a single trial, observers saw a sampled planar contour in 3-D space. The contour was either a horizontal linear segment or a parabolic segment. An additional point (*AP*—the adjustable point) was positioned at random in 3-D space. The observer was instructed to

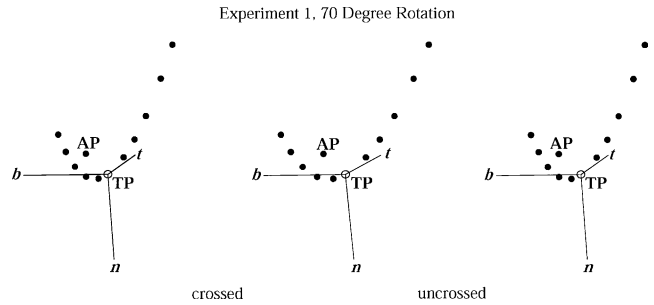


Fig. 6. Stereo views of the intrinsic coordinate system (t, n, b) used in this study. This coordinate system is centered on the true point TP. The unit vector t (‘tangent’) is tangent to the invisible contour at the point TP. The unit vector n (‘normal’) is orthogonal to t and in the plane of the contour, while the unit vector b (‘binormal’) is orthogonal to both t and n . The setting plane is determined by b and n and the adjustable point AP is confined to this plane. The contour plane contains t and n .

move the point until it lay on the perceived contour (method of adjustment). Movement of the AP was confined to a plane that we will refer to as the *setting plane*. For our choice of contours, the setting plane only intersected the contour once at a point we refer to as the *true point (TP)* and was always perpendicular to the contour at TP (Fig. 6). The observer’s task was, therefore, well defined. The invisible point TP was always in the middle of the series of visible sampled points defining the contour (Fig. 6).

Observers used six buttons throughout the course of the experiment. The observer used four of these to move the AP in the setting plane. Prior to the experiment we selected two direction vectors in the setting plane for each angle and curve condition. To each direction vector we assigned two of the four keys. Pressing one key of the pair moved the point one way along the vector, pressing the other moved it in the opposite direction. Observers quickly became comfortable with the mapping of keys to movement of the AP.

At the start of a trial the control program permitted ‘quick’ movement of the point—each key press displaced the point by approximately 0.5 mm in a direction in space defined by the experimental conditions. When the observer judged that the adjustable point AP was near the contour, they pressed a fifth key which allowed them to move the point with a greater precision (0.07 mm per key press) until they were satisfied with their setting. When the direction of motion was perpendicular to the Cyclopean line of sight and near the origin, the coarse and fine displacements corresponded to 2.5 and 0.35 min of visual angle (note the latter is the minimum spatial resolution allowed by our software). When the direction was along the line of sight, near the origin, they corresponded to 2.5’ and 0.35’ of disparity in each eye. A final press of the sixth key recorded the observer’s setting and triggered the next trial.

2.3. Rotations

In each of the experiments described below, the stimuli were rotated about an axis lying in the fronto-parallel plane that contained the true point TP (Fig. 4). In Experiment 1, the axis of rotation was vertical; In Experiment 2, it was horizontal.

2.4. Coordinate systems

Throughout this paper we will refer to absolute, intrinsic and setting plane coordinate systems. The *absolute coordinate system*, (X, Y, Z) , corresponds to a fixed frame of reference centered on the true point, TP. We chose the convention that this frame is left-handed with the X -, Y - and Z -directions corresponding to rightwards, downwards and towards the observer, respectively.

Where appropriate, absolute coordinate data will be presented in several different units. As well as in mm the X and Y coordinates will be given in units of visual angle. The Z component will be given in units of disparity relative to TP. In addition, the magnitude of the 3-D bias and standard deviation vectors will be reported as percentages of the average linear 3-D distance from the interpolation region to the nearest contour dots.

Intrinsic coordinates $(\mathbf{t}, \mathbf{n}, \mathbf{b})$ correspond to the left-handed triple defined by the tangent, normal, and binormal vectors to the contour at the true point TP (Fig. 6). The setting plane contains the vectors \mathbf{b} and \mathbf{n} and is perpendicular to \mathbf{t} . The contour plane⁴ contains the vectors \mathbf{t} and \mathbf{n} , and is perpendicular to \mathbf{b} . The *setting plane coordinate system* is a 2-D coordinate system with axes along \mathbf{b} and \mathbf{n} , the binormal and normal vectors of the intrinsic coordinate system. Any point on the setting plane can be expressed as a linear combination of these two vectors and it will prove convenient to describe observer performance in this coordinate frame.

Intrinsic coordinate biases and variabilities will be reported in mm and also as a percentage of the distance between the interpolation region and the nearest dots. Note that it does not make sense to report these data, which are invariant under rotation of the stimulus configurations, in units of visual angle.

3. Experiment 1—vertical rotation axis

In this experiment, we examine how human interpolation performance for linear and parabolic segments

⁴ Note that there are many planes containing the linear contour and, therefore, no uniquely defined *contour plane*. For convenience, we define the contour plane of the linear contour to be the same as that of the corresponding parabolic contour. The vector \mathbf{b} in the linear condition is chosen to be perpendicular to this contour plane and the vector \mathbf{n} is the vector in this contour plane perpendicular to the linear contour.

varies with the number of points used in defining a contour and with rotation of the contour plane about a vertical line. The dependent measures of interest are the observer's setting variability and the observer's setting bias.

3.1. Methods

On each trial, observers viewed a sampled contour, and adjusted the position of the AP within the setting plane so that it appeared to lie on the perceived contour. On each trial, the observer saw either a sampled parabolic contour consisting of 4, 6, 8 or 10 points or a sampled linear contour consisting of 2, 4, 6 or 8 points and a randomly displaced AP. The position of the points defining the contour was fixed over all trials. On trials with fewer than the maximum number of points displayed the most peripheral points were removed leaving only the central points. The factors in this experiment were *curve type* (parabola or line), *number of points* and *angle of rotation* of the contour plane (0° , 40° or 70°) about a vertical axis in the center of the display region.

In each session the observer saw each condition (curve type \times angle \times number of points) twice: each session, therefore, contained 48 settings. Observers completed a total of five sessions over a period of days with a different random ordering of the trials for each. Each observer contributed 240 settings in total. The parabola and line stimuli were interleaved so that the observer never saw the same curve type on two consecutive trials (to avoid the possibility that the observer would remember and repeat settings between successive trials with the same curve type). The order of presentation of the stimuli was otherwise random.

3.2. Observers

Two of the five observers were authors (PAW, LTM). The remaining three were not aware of the purpose of the experiment. Two had previously taken part in other visual psychophysics experiments.

3.3. Results

We first describe the results in terms of setting variability in absolute and intrinsic coordinates before describing biases in intrinsic coordinates. We also looked for biases in absolute coordinates (such as a consistent tendency to place the setting point to the right of the contour or away from the contour toward the observer). We found no such consistent pattern across conditions and observers and thus these biases are not reported. The variabilities are indications of the accuracy of observer settings. We will look for evidence for rotation-invariant biases in intrinsic coordinates. The presence of such biases would indicate that observers are, in effect,

interpolating a contour other than the linear or parabolic contour used to generate the stimuli.

Setting variability: absolute coordinates. The setting results for each observer, in absolute coordinates relative to TP, can be seen in Fig. 7. Since these data are in 3-D, we show two projections of the data: Fig. 7A is the X–Z projection and Fig. 7B is the Y–Z projection (the icons make the plotting conventions clear). Note that in the X–Z projection the rotation of the setting plane with the contour is evident. Note also that overall setting errors for the naïve observers (SDF, IM, JP) are higher

than those of the two authors (PAW, LTM). In spite of this, the vast majority of settings (80–100% depending on observer) are contained within a window of width ± 1.5 mm in the X- and Y-directions and ± 2 mm of depth ($\pm 7.5'$, $\pm 7.5'$, $\pm 1'$ disparity). Overall, there is more variability in the depth direction than in the fronto-parallel plane (note that the increased variability in the absolute depth dimension is not reflected in retinal coordinates where variability appears much smaller).

The magnitude of the 3-D standard deviation vector is approximately 10% of the average distance to the nearest contour points.

The average separation in the fronto-parallel plane between the interpolation region and the nearest neighboring dots flanking the AP is approximately 25 mm for the line and 30 mm for the parabola. Thus, performance comparable to vernier acuity studies requires setting errors to be in the region of 0.4–0.5 mm in the fronto-parallel plane (e.g. Klein & Levi, 1987). Note that the very largest of these errors are only a factor of 4 higher than would be expected from a comparable fronto-parallel, linear three dot Vernier task. Thus, in spite of rotating the stimulus in 3-D and adding curvature to the contour we can still claim that observers are able to interpolate with very high accuracy.

Setting variability: intrinsic coordinates. Recall (Fig. 6) that \mathbf{b} is the direction in the setting plane orthogonal to the contour plane and \mathbf{n} is the direction within the contour plane orthogonal to the curve at the true point, TP (it lies along the intersection of the contour plane and the setting plane). These directions rotate with the stimulus.

The pooled standard deviations across all observers were 0.78 mm (\mathbf{b}) and 0.11 mm (\mathbf{n}) for the line and 0.846 mm (\mathbf{b}) and 0.4196 mm (\mathbf{n}) for the parabola. The magnitudes of these standard deviation vectors were roughly 3.2% and 3.16% of the average distance to the nearest contour points. As expected, settings in the parabolic condition were less accurate than those in the line condition. Note that there is higher variability in the \mathbf{b} -direction for both curve types. Clearly then, for the parabola, setting variability is higher out of the plane of the curve than in the plane. We need to be careful in interpreting this result for the linear segment because of our arbitrary definition of the normal direction. However, it is safe to say that variability in the arbitrarily defined plane of the linear contour is lower than perpendicular to the plane. Total variance⁵ (measured as $\hat{\sigma}_b^2 + \hat{\sigma}_n^2$, the sum of the estimated variances in the \mathbf{b} - and \mathbf{n} -directions respectively) was calculated for each observer and averaged over the twelve conditions (four numbers of points and three angles). Between observers, values for the ratio of these numbers for the parabola

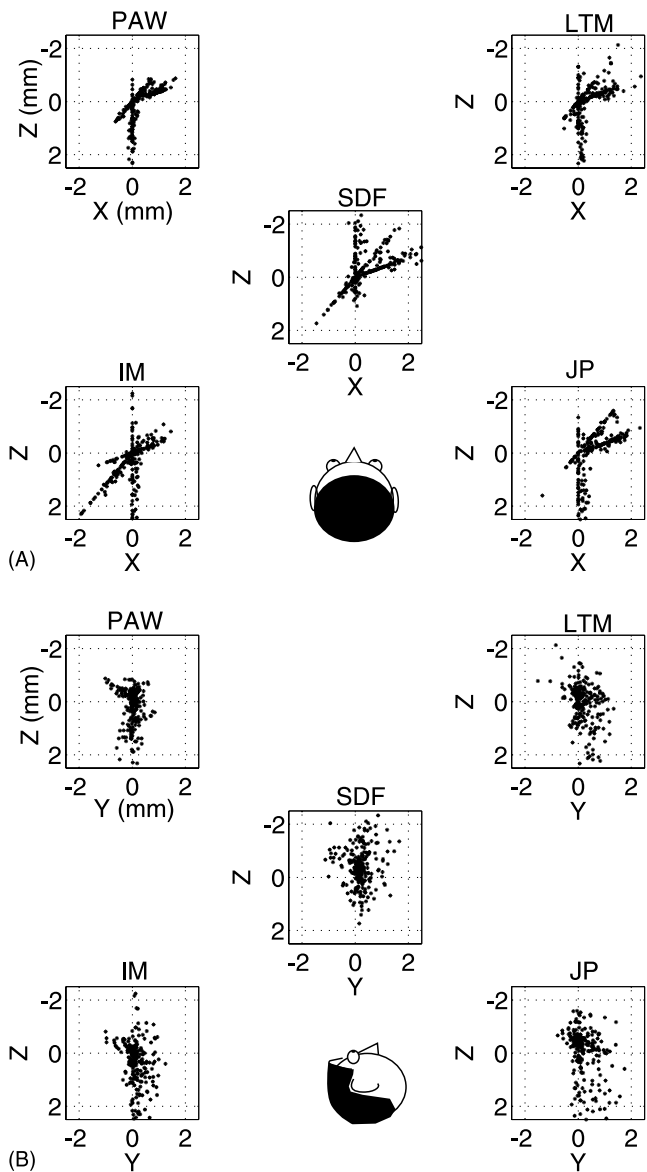


Fig. 7. Experiment 1: Combined 3-D setting errors for each observer in each condition. (A) View from above. (B) View from the side. At least 80% of data for each observer in each condition is contained in a window of size ± 1.5 mm in the X-direction, ± 1.5 mm in the Y-direction, and ± 2 mm in the Z-direction. The setting point is confined to a different plane in each angle condition. The influence of the planar constraints is evident in (A).

⁵ The total variance is the trace of the estimated $\mathbf{b} - \mathbf{n}$ covariance matrix and is rotation invariant.

and the line, ranged from 1.02 to 2.67 with an average of 1.62. Thus, on average, the estimated increase in total standard deviation of setting errors is the square root of this value—around 1.3. This 30% increase in variability comes from two sources. Firstly, the linear distance between the points nearest the interpolation region is slightly larger for the parabola than the line. Also there is clearly some additional error due to adding curvature to the task. Since the linear separation for the parabola is approximately 1.2 times that for the line, we estimate a lower bound for the *cost of curvature* to be around 10%. It is possible that the cost of curvature is higher than this value. For example, errors may increase with the length of the contour region across which interpolation occurs. In this case adding curvature interacts with the distance, since it necessarily lengthens the line between the points.

Setting bias: intrinsic coordinates. A rotation-invariant pattern of biases in this coordinate system would suggest that the ‘human visual spline’ does not coincide with the contour (linear or parabolic) that we used in generating the sampled contours. We look for rotation-invariant patterns of biases common to all observers, for the linear and the parabolic case separately.

When data from all observers were pooled and averaged in intrinsic coordinates over all number of points and angle conditions, the average setting biases were -0.1 mm (\mathbf{b}) and 0.08 mm (\mathbf{n}) for the line condition, and -0.22 mm (\mathbf{b}) and 0.24 mm (\mathbf{n}) for the parabolic condition. The magnitudes of these bias vectors were roughly 0.51% and 1.09% of the average distance to the nearest contour points. Thus, for the line condition, the average bias in the \mathbf{b} -direction (perpendicular to the contour plane) was roughly the thickness of a sheet of paper, the error in the \mathbf{n} (normal to the curve) direction even smaller. For the parabolic condition, the average biases were about twice as large. We performed four separate t -tests (separately for \mathbf{b} and \mathbf{n} , separately for the linear and parabolic cases) to test whether the measured biases were non-zero. We tested at the 0.05 level with a Bonferroni correction for multiple (4) tests. For the linear contour, we rejected the hypothesis that observers’ settings were unbiased in the \mathbf{b} -direction ($t_{540} = -3.28$; $p = 0.0011$) and in the \mathbf{n} -direction ($t_{540} = 16.59$; $p < 0.0001$). For the parabolic contours, we rejected the hypothesis in the \mathbf{b} -direction ($t_{540} = -6.33$; $p < 0.0001$) and in the \mathbf{n} -direction ($t_{540} = 13.99$; $p < 0.0001$). All of the rotation-invariant biases are significantly different from 0.

It is important to realize that, given the large number of settings made by each observer, we can potentially detect very small discrepancies between the observer’s settings and the invisible parabolic or linear contour. While the biases found are statistically significant, they are in physical units very small: in interpolating across a gap of approximately 50 or 60 mm width, the observers’

settings never stray more than a quarter of a millimeter from the linear or parabolic contour. We will return to this point in Section 5.

3.4. Nested hypothesis tests

We reformulated the hypotheses we sought to test in terms of a series of four nested models, each intended to account for the observed patterns of setting variability and setting bias found in the data. By comparing the adequacy of fit of these models, we can determine if, for example, the number of points contained in a contour affected human performance.

We fit all models to each observer’s data by maximum likelihood parameter estimation (see Mood, Graybill, & Boes, 1974, pp. 276–285) and all models share the assumption that observers’ settings in each condition can be treated as realizations of a bivariate Gaussian random variable in the setting plane coordinate system. The bivariate Gaussian has five parameters: two bias parameters corresponding to the mean of the distribution, two standard deviations and a parameter controlling correlation. We refer to the first two parameters collectively as ‘bias’, to the last three as ‘variability’. Before any further analyses were carried out, we plotted setting data (in the setting plane coordinate system) for all observers to assess whether there were deviations from the bivariate Gaussian model and found none.

Each model had a different number of parameters and specification of how these parameters are linked to the experimental factors. Our analysis will allow us to assess which parameters, and therefore which factors, are important in determining observer bias and variability. We emphasize that these ‘models’ are not intended as functional models of human interpolation mechanisms but instead as statistical descriptions of the data. As there are 24 conditions and five parameters summarizing the bivariate distribution that captures performance in each condition, we require at most 120 parameters to describe an observer’s performance (we refer to this minimally constraining model as Model 3, described below).

Nested hypothesis testing specifies a method for balancing the number of parameters used to fit a data set and the goodness of that fit. Essentially, we look for evidence that a model with fewer parameters can not be rejected as a worse fit to the data than a model with more parameters. Details of this method are provided in an appendix. We proceed with a description of our competing models.

Model 0. This is the most constrained model. Only six parameters were used to fit the entire data set for each observer. This model incorporates the assumption that the observed variability of observer’s settings is independent of experimental condition, and that observed variability and bias in settings could be modeled as the

result of an uncorrelated trivariate Gaussian error distribution in absolute coordinates. Thus, the six parameters correspond to the three absolute biases (center of the setting ellipse) and the three standard deviations (axis lengths of the setting ellipse in X , Y and Z coordinates). Since observer settings were confined to a plane we assumed that the bivariate setting plane distribution of observer settings represented a marginal projection of this trivariate distribution. Given the small number of parameters in this model and the large number of conditions, we expect this model to produce a relatively poor fit to the data. Rejecting this model simply tells us that at least some of the independent variables we varied in this experiment affected human performance.

Model 1. This model assumes that the setting variability and bias in each condition do not vary with number of points. For each section of data, i.e. a curve type (line or parabola) and an angle (0, 40, 70), we combine setting data over all numbers of points and estimate the bias and variability parameters of the maximum likelihood bivariate Gaussian in each of six setting plane conditions (2 curve types \times 3 angles). This leads to a model with $5 \times 6 = 30$ parameters. If we fail to reject this model, then we fail to reject the hypothesis that neither biases nor variability change with the number of points. If we reject this model then it is of immediate interest to determine whether it is bias or variability or both that vary with number of points.

Model 2. This model is similar to Model 1 in that the variance structure of the settings is independent of the number of points. However, each level of the number of points condition is allowed a separate pair of bias parameters in the setting plane. We now permit separate bias parameters for each of the 24 separate conditions but require that the variability parameters be independent of the number of points. This model has 2 (bias parameters) \times 24 (conditions) bias parameters and 3 (variability parameters) \times 6 (condition sets) variance parameters, a total of 66 parameters.

If we reject Model 1 in favor of Model 2, then we reject the hypothesis that setting biases are not affected by number of points. If, after rejection of Model 1, we reject Model 2 in favor of Model 3, then we conclude that setting variability varies with number of points. As noted in the introduction, it is plausible that setting variability would decrease with increasing number of points and such a decrease is consistent with rejecting Model 2.

Model 3. This model is minimally constrained in the sense that subjects' setting performance in each condition is estimated separately from that of any other condition: performance in one condition does not constrain performance in any other. This model has 5 (parameters) \times 24 (conditions) = 120 parameters and is the 'top' of a hierarchy of nested models—Model 0 is nested within (i.e., is a special case of) Model 1, which

is nested within Model 2, which, in turn, is nested within Model 3. The fit of any model lower in the hierarchy can be duplicated in a higher-level model by some choice of parameter settings.

In summary, we will ascend the hierarchy of models, beginning with Model 0, until we find a point where we cannot reject a model in favor of its immediate superior in the hierarchy. We will first test whether Model 1 produces a better fit to the data than Model 0. This test is essentially asking whether setting errors can simply be explained as coming from a trivariate Gaussian distribution in absolute coordinates or whether there is a significant effect due to the different conditions. If so, we will next test whether Model 2 is significantly better than Model 1. This second test asks whether the number of points factor is important in determining observer bias and variability. We then tested whether Model 3 produced a significantly better fit to the data than Model 2. These latter two tests were calculated separately for the parabola and the line. In total this could lead to as many as 25 separate tests—thus, we used a Bonferroni correction and rejected the null hypothesis (Model j is not significantly better than Model i , with $j > i$) if the p -value of our log-likelihood ratio was smaller than $0.05/25 = 0.002$.

3.5. Results of nested hypothesis tests

The findings of the nested hypothesis tests are summarized in Fig. 8. The hierarchy of four models is represented by four stacked, labeled rectangles, one 'stack' for each observer and curve type. We began at the lowest model (Model 0) and tested each model against its successor. If we cannot reject a model, we stop at that level of the hierarchy. The points where we stopped in each hierarchy are emphasized by a shaded rectangle. The likelihood ratio statistic Δ corresponding to each 'step' up the hierarchy is shown as well as the corresponding p -value. The first instance of a non-significant p -value, indicating failure to reject the null hypothesis, in the hierarchy is ringed. If the p -value is less than 0.0001 then a value of '***' is reported.

For all but the test between Model 0 and Model 1 the log-likelihood was calculated separately for the parabolic and the linear conditions so that we could better assess the source of any significant difference in the fits of the two models to the data. For the test between Model 0 and Model 1 we compare the fits to the entire data set because Model 0 has no separate curve type parameters.

The first test compared the fit of Model 0 with Model 1. Our results led us to reject Model 0 in favor of Model 1. Next we tested whether Model 1 captured enough of the characteristics of the data for us to allow us to prefer Model 1 over Model 2. For two observers (JP & LTM) Model 1 was not rejected for both the

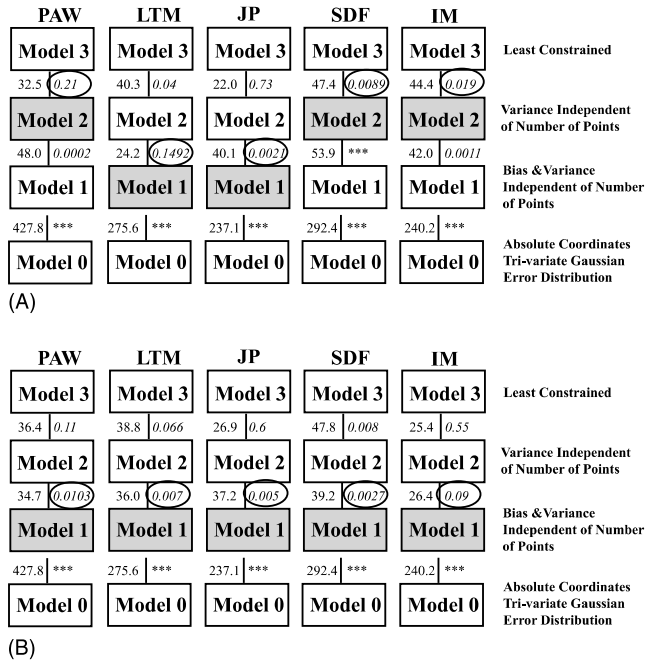


Fig. 8. Experiment 1: The results of the nested hypothesis tests. For each subject and each contour type, the hierarchy of four models is represented by four stacked, labeled rectangles. We begin at the lowest model (Model 0) and test each model against its successor. If we cannot reject a model, we stop at that level of the hierarchy. The points where we stopped in each hierarchy are denoted by a shaded box. The numbers between each pair of boxes are the likelihood ratio decision statistic Δ and the p -value corresponding to the hypothesis that the data are consistent with the lower model. (A) The linear condition. (B) The parabolic condition.

linear and parabolic curve types. For all other observers Model 1 was accepted for the parabolic contour only.

In spite of rejecting Model 2 in favor of Model 1 in most cases we have proceeded to test whether Model 2 is preferred over Model 3 (the least constrained model) for all subjects. This test was performed to assess whether Model 2 and preferred models lower in the hierarchy were good summaries of the data. In all cases Model 2 could not be rejected in favor of Model 3, which had a parameter for every variable in the experiment. Consequently Model 2 and, for some subjects, Model 1, are parsimonious descriptions of the data.

Recall that the difference between Models 1 & 2 is based solely on extra parameters to model the variation of the biases with the number of points. Thus for all observers in all conditions we conclude that the variance structure is not affected by the number of points variable. This result is consistent with the perturbation analysis of influence performed by Hon et al. (1997) who found a steep decrease in influence away from the AP. That is, additional points beyond those required to specify a curve were not used for interpolation.

To illustrate this result, we plot the parabolic contour settings of a single observer for a single angle condition (0°) over all four numbers of points (Fig. 9). If an ob-

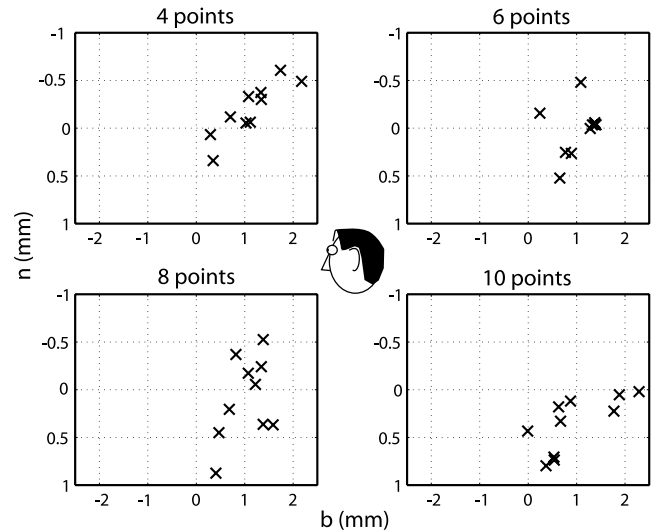


Fig. 9. Experiment 1: Data for the parabolic contour for one subject (PAW) in a single angle condition (0°) for each number of points. Note that there is no tendency for the distributions of points to become less variable as the number of points increases.

server were using all available information we would expect a decrease in the spread of errors as the number of points increased.⁶ In fact, the settings for this observer show no such trend. This pattern is evident for all observers and for all angle conditions. We return to this point in the discussion.

4. Experiment 2—horizontal rotation axis

4.1. Methods

Experiment 2 was identical to Experiment 1 in all respects except for the axis of rotation of the stimulus configurations, which was now a horizontal line, perpendicular to the line of sight, in the middle of the display area. In choosing a new rotation axis we are able to better disassociate intrinsic and absolute coordinates. Note that in Experiment 1 the n -direction was close to the Y -direction for all conditions. In this experiment n and Y differ markedly in rotated conditions.

In addition, in place of the horizontal linear stimulus of Experiment 1 a vertical line was used and scaled by a factor of 0.75 to fit in the display area. We expect to find a similar pattern of results in this experiment as those of Experiment 1. This will show that our findings are independent of rotation axis.

Observers. The observers were the same as in Experiment 1. Two of the five observers were authors (PAW, LTM). The remaining three were unaware of the pur-

⁶ We have no prediction for changes in the bias (the mean of the distributions of settings). Only the overall spread of the distributions is under discussion here.

poses of the experiment. All observers had normal or corrected-to-normal vision.

4.2. Results

Setting variability: absolute coordinates. The magnitudes of errors for this experiment in absolute coordinates were similar to those of Experiment 1 (Fig. 10). Once again at least 80–100% of errors were contained within a window of $\pm 1.5 \times \pm 1.5 \pm 2$ mm ($\pm 7.5'$, $\pm 7.5'$, $\pm 1'$ disparity). Thus, the magnitude of the standard

deviations vector is roughly 10% of average distance to the nearest contour points. Note that once again performance is better in the fronto-parallel plane than in depth. This suggests that, on the whole, observer performance was of comparable accuracy to Experiment 1 in spite of the different rotation axis.

Setting variability: intrinsic coordinates. The pooled intrinsic coordinate standard deviations across all observers were 0.75 mm (**b**) and 0.1 mm (**n**) for the line and 0.63 mm (**b**) and 0.68 mm (**n**) for the parabola. The magnitudes of these standard deviation vectors were roughly 4.03% and 3.09% of the average distance to the nearest contour points. If we assume that variability scales with separation of the points we would expect smaller standard deviations for this line than that in Experiment 1 because the contour was scaled by a factor of 0.75 to fit in the screen in its new vertical orientation. In fact, linear contour standard deviations are similar in both experiments. This is at odds with the results of Vernier acuity studies (e.g. Klein & Levi, 1987) and a pilot study we carried out to test the effect of scaling on variability (see Section 5). In this pilot study the stimulus of Experiment 1 was simply scaled by a factor of 0.5—the standard deviations decreased accordingly. We suggest that the failure to find a scaling of variability in the results of Experiment 2 reflect the fact that the linear contour had not been seen before in this orientation and thus subjects were inexperienced.

Calculation of the cost of curvature parameter in this experiment relies upon the assumption that we can scale the variances of the linear stimulus to make them comparable to the parabolic variability. Thus, due to the issues raised above we will not report the cost of curvature in this experiment.

Setting bias: intrinsic coordinates. When data from all observers were pooled and averaged over each number of points and angle condition, setting biases for the line were 0.16 mm (**b**) and 0.03 mm (**n**). For the parabola the average biases were 0.05 mm (**b**) and 0.55 mm (**n**). The magnitudes of these bias vectors were roughly 0.87% and 1.84% of the average distance to the nearest contour points.

As in Experiment 1, we performed four separate *t*-tests (separately for **b** and **n**, separately for the linear and parabolic cases) to test whether the measured biases were non-zero. We tested at the 0.05 level with a Bonferroni correction for multiple (4) tests. For the linear contour, we rejected the hypothesis that observers' settings were unbiased in the **b**-direction ($t_{540} = 5.36$; $p < 0.0001$) and in the **n**-direction ($t_{540} = 8.4$; $p < 0.0001$). For the parabolic contours, we failed to reject the hypothesis in the **b**-direction ($t_{540} = 2.07$; $p = 0.04$) but did reject it in the **n**-direction ($t_{540} = 19.93$; $p < 0.0001$). Again, the observers' rotation-invariant constant errors were very small but (with the exception of the **b**-direction for the parabola) highly significant.

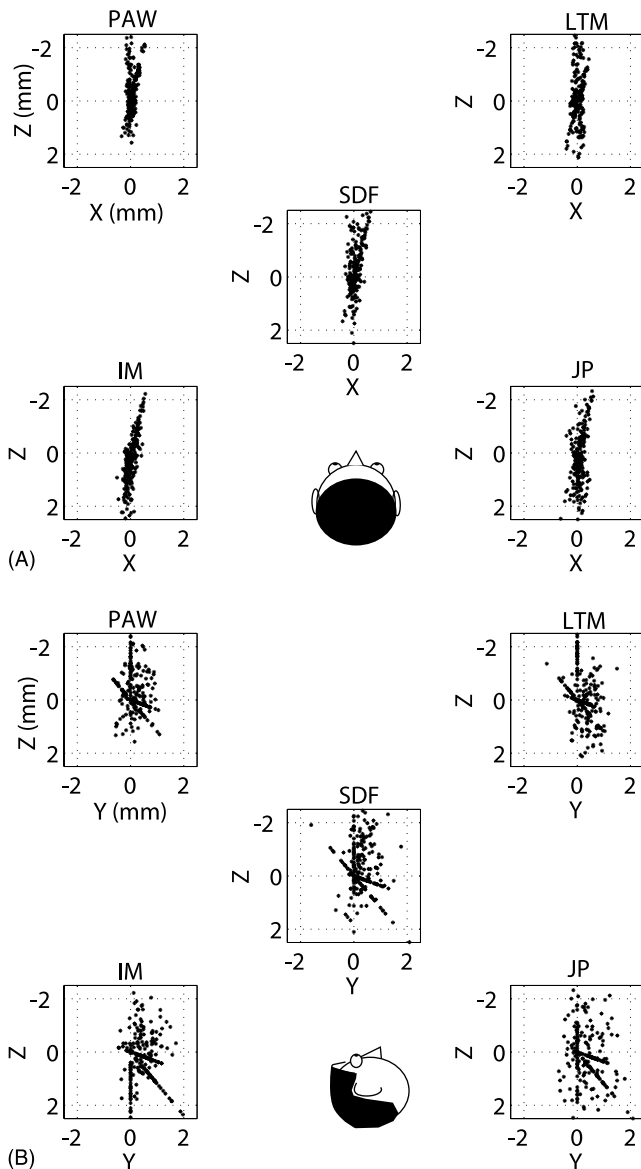


Fig. 10. Experiment 2: Combined 3-D setting errors for each observer in each condition. (A) View from above. (B) View from the side. At least 80% of data in each observer and in each condition is contained in a window of size ± 1.5 mm in the *X*-direction, ± 1.5 mm in the *Y*-direction, and ± 2 mm in the *Z*-direction. The setting point is confined to a different plane in each angle condition.

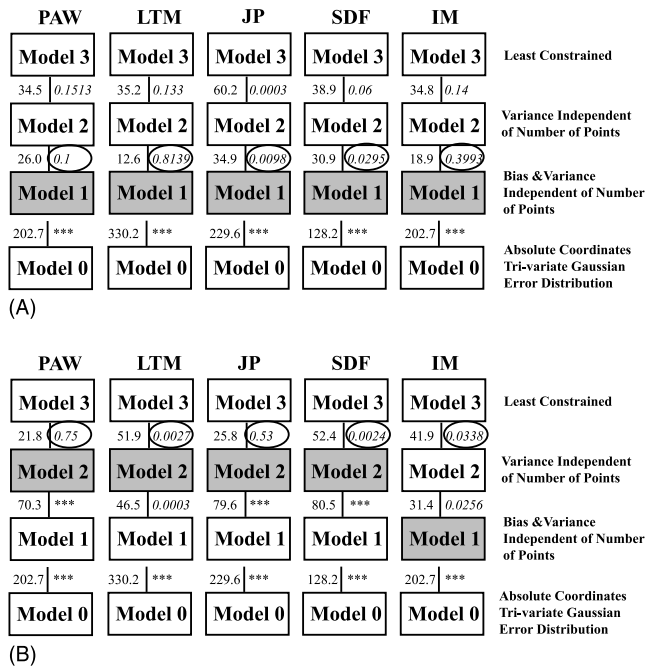


Fig. 11. Experiment 2: The results of the nested hypothesis tests in the same format as Fig. 8.

Note that there were individual differences in biases (not shown) between observers and that the pattern of *b*-direction biases in the 0° parabolic contour condition was different from the same condition in Experiment 1 in spite of the stimuli and task being identical. We take this as evidence that there is no systematic interaction of setting bias with angle.

4.3. Results of the nested hypothesis tests

Fig. 11 shows that we were forced to reject Model 0 in favor of Model 1 in all cases. However, we can not reject Model 1 in favor of Model 2 for any observers in the linear condition and one observer in the parabolic condition. For the four remaining observers in the parabola condition we could not reject Model 2 in favor of the least constrained Model 3. These results are in line with Experiment 1 and imply that for all observers the variance structure of their settings is independent of the number of points variable. Note that the concerns raised above regarding the fact that the linear contour variability was unexpectedly high are not important in this analysis because parabolic and linear conditions are analyzed separately.

5. General discussion

Many previous studies have examined human ability to solve the grouping problem. We address a second, complementary problem: how do observers, given the

grouped fragments of a contour, estimate the missing parts of the contour? Our stimuli were *sampled contours*, invisible except for a small number of visible points. Observers were asked to move a point so that it fell on the invisible contour. In this study, the contours were not confined to the fronto-parallel plane but could appear in arbitrary orientations in 3-D space.

We have shown that the human observer is able to integrate sparse information across large regions of 3-D space and use that information to infer accurately the position of an invisible contour. Observers interpolate with accuracy comparable to that of hyperacuity, in spite of the curvilinear nature of the parabolic contour and the rotation of these contours into the third dimension. Given the large distances between the points forming our sampled contour it is remarkable that observers are able to infer a parabolic segment so accurately.

Settings were, on the whole, more uncertain in the depth direction. As expected, variability is greater for the parabolic condition than for the linear condition, for all observers. The estimated cost of curvature for our choice of stimuli (Experiment 1) may have been as low as 10%.

In different conditions, we rotated the stimulus configuration around two different axes, a vertical axis in Experiment 1, and a horizontal axis in Experiment 2. We looked for rotation-invariant biases in observers' settings for both the linear and the parabolic contours. The presence of such biases would suggest that the 'human visual spline' is effectively fitting different classes of curves to the sampled contours than we used in generating the stimuli. While it is perhaps implausible that the observers were fitting non-linear contours to the linear sampled contours, it was distinctly possible that the observers might choose to use a non-parabolic contour (such as a circular segment fit to a subset of three points) in interpolating the parabolic sampled contours.

We found statistically significant rotation-invariant biases (constant errors) in observers' mean interpolation settings for both the linear and the parabolic contour. However, the physical magnitude of these biases were less than 0.55 mm in all directions and with one exception (the *n*-direction for the parabolic contour in Experiment 2) were all less than a quarter of a millimeter, slightly more than the thickness of two sheets of paper. Again, recall that observers are interpolating across a gap of around 50–60 mm. The rotation-invariant biases were much smaller than observers' setting uncertainties: it is plausible that observers could not detect the presence or absence of their own constant errors.

We expected that, as the number of points defining a sample contour increased, then the observer's setting variability would decrease. We originally intended to see whether a difference in setting variability in the linear

and parabolic cases could be canceled by increasing the number of points on the parabolic contours. Our results preclude this possibility. Nested hypothesis tests reveal that the goodness of fit of maximum likelihood Gaussian models to the data did *not* improve with the addition of parameters for the setting variability for the different numbers of points. We cannot directly conclude from our study that observers do not use all the information available to them, but we can state that their setting variability did not benefit from it. The results of Feldman (1997) and Hon et al. (1997) suggest that, at least in the fronto-parallel plane, observers do not use all the points in interpolation. Our results are consistent with theirs and extend them to 3-D: we have no reason to claim that observers use more than four points at a time when interpolating.

We emphasize that this pattern of results is not due to a simple floor effect, i.e. the reason that positioning variability did not decrease with number of points was simply that it was already as low as it could get for such a task. As noted above, the largest of our setting errors are four times larger than those found in comparable Vernier studies in the fronto-parallel plane (Klein & Levi, 1987) and in depth (McKee et al., 1990). In Klein and Levi (1987), observers fixated a central horizontal line target while two additional co-linear lines were briefly presented at eccentricities of up to 10° of visual angle. Observers were asked to rate the alignment of the three lines. Performance in this vernier acuity study produced a Weber fraction of around 0.01. Corresponding values for our task (intrinsic coordinate total standard deviation scaled relative to the average distance from the interpolation region to the nearest contour points) are in excess of 0.03 for both the line and parabola. Thus, the Weber fractions corresponding to our data are roughly three times as large as previous studies have reported. Clearly then, subject performance could have improved. We return to this point in the section below where we show that the human interpolator is non-ideal.

It is surprising that observers do not seem to gain a significant advantage from the extra points. One possible explanation is that visual interpolation mechanisms are confined to fixed regions of the retinas roughly centered on the foveas of the two eyes. A first difficulty with such an explanation is that the linear sampled contour with four points occupied almost the same region of the retinas as the parabolic sampled contour with four points, yet observers seemed to use only the inner two points of the linear sampled contours while they must have used more than two points in curvilinear interpolation.

This objection could be answered by postulating that there are distinct mechanisms for linear and non-linear interpolation and, for our choice of experimental conditions, the two-point linear configuration just happened

to fall within the linear interpolation area and the four-point parabolic configuration just happened to fall within the curvilinear interpolation area. This explanation can readily be tested by a replication of Experiments 1 and 2 with the stimulus configurations for both linear and parabolic sampled contours reduced in scale by a factor of two. Such a scaling would double the number of points that fall within the hypothetical linear and curvilinear areas. We would expect to see a decrease in the variability of linear interpolation between two and four points (now that four points fall within the linear interpolation area) and a similar decrease in the variability of parabolic interpolation between four and eight points (now that eight points fall within the curvilinear interpolation area).

Using a different experimental method, Hon et al. (1997) measured the influence of each visible point on subject's settings in an interpolation task similar to ours. Influence, as they defined it, was a measure of how much each visible point affected interpolation settings. They found that the influence measures for parabolic interpolation were invariant when stimulus size was scaled by a factor of two. In particular, points that had little influence did not increase their influence when moved closer to the point of fixation, an outcome inconsistent with the claim that there are fixed retinal areas for linear and curvilinear interpolation.

Hon et al. reported influence, not setting variability, and did not vary number of points in the contour. Accordingly, we performed a simple control experiment in which one observer (PAW) repeated one of the linear contour conditions for Experiment 1 (the 0° angle condition), but with the entire stimulus configuration reduced in scale by a factor of two. The results (Fig. 12) are unequivocal. If there were fixed interpolation areas, then after reduction in scale by a factor of two, we would expect to see a reduction in setting variability between two and four points (while we saw no such reduction in Experiment 1). Instead, the distributions of settings in the replicated conditions (Fig. 12B) are roughly scaled copies of the corresponding settings in the original data (Fig. 12A). These results, together with those of Hon et al. (1997), lead us to conclude that the lack of a reduction in setting variability between two and four points for the linear sampled contour, and between four and six points for the parabolic sampled contours cannot be explained by interpolation mechanisms confined to particular retinal regions.

A second possible explanation is that the additional points beyond the minimum needed to define the curve do not, in fact, carry very much information. Suppose, for example, that an observer is fixating the setting point and attempting to interpolate a linear sampled contour. The observer's performance is limited by the accuracy with which the visible points can be localized and this accuracy is limited, first of all, by the retinal eccentricity

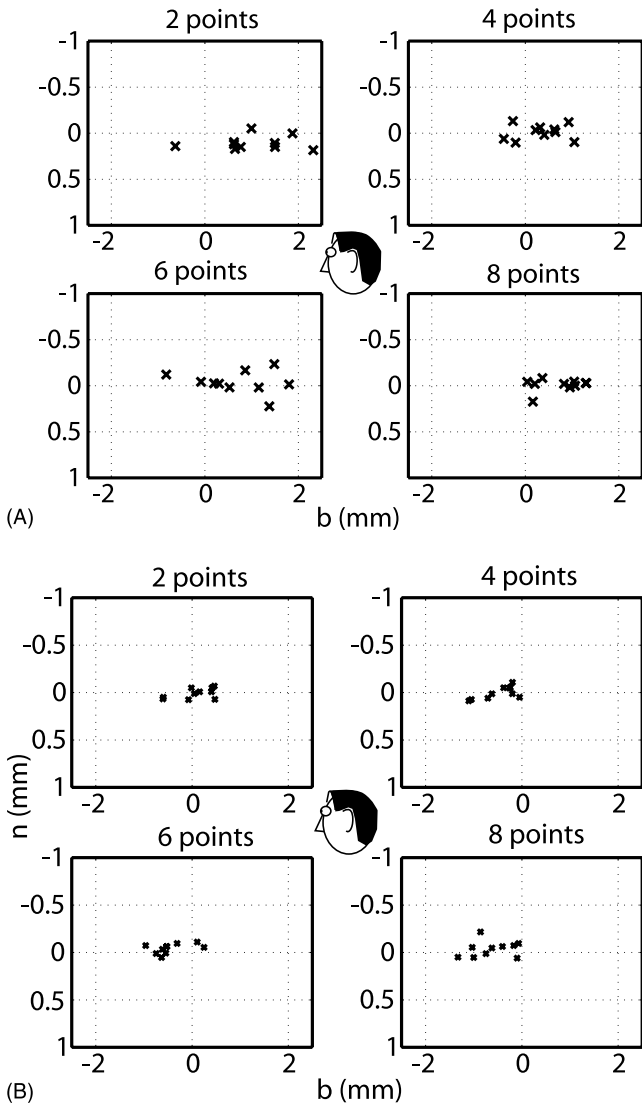


Fig. 12. (A) The results of Experiment 1 (0° rotation case, the linear contour) for observer PAW. (B) A replication of Experiment 1 (0° rotation case, the linear contour) for observer PAW but with the stimulus configuration (all interpoint differences) reduced in scale by a factor of 2. Note that the effect of scaling the stimuli leads to results that are scaled by roughly the same factor in agreement with the results of Hon et al. (1997).

of each point and its disparity. If we assume that the positional uncertainty associated with each visible point can be modeled as a Gaussian random variable and the uncertainties for distinct points are independent, then we can readily derive the maximum likelihood estimator for the intersection of a line with the setting plane. An ‘ideal interpolator’ would then move the setting point to this estimated point of intersection of line and setting plane. It can be shown that, in estimating the setting point, the ideal interpolator makes use of all visible contour points, with weights inversely proportional to the variance of each point in each direction in space. The ideal interpolator would not only make use of every

visible contour point in interpolating the contour, but would also gain some benefit (as measured by a decrease in variability) from each additional point placed on the contour. The amount of benefit to be expected in going from two to four to six or more points is entirely determined by the positional uncertainty associated with the added points.

An ideal interpolator. We could test the above hypothesis by simulating an ideal interpolator’s performance in our experiment, using estimates of human positional uncertainty obtained from the literature. However, data for relative and absolute 3-D location uncertainty away from fixation are rather scarce. Past measurements have focused on the fronto-parallel plane containing the fixation point and the Cyclopean line-of-sight. Accordingly, we limited ourselves to an analysis of linear interpolation in the fixation plane, one of the conditions in our experiment.

Localization uncertainty data in the literature are usually reported as threshold acuity. We assume these threshold values are a good approximation to positional variability. We also assume isotropic uncertainty in the *X*- and *Y*-directions, which depends only on eccentricity. Vernier acuity away from fixation is usually modeled as a linear function of eccentricity (e.g. Beard, Levi, & Klein, 1997; Klein & Levi, 1987). Accordingly, we assume uncertainty in the *X*- and *Y*-directions is determined by a linear function, with slope equal to 1% in line with these studies. The intercept was determined by considering hyperacuity thresholds in the literature.

Turning now to the *Z*-direction, for small disparities depth acuity thresholds are thought to be close to linearly dependent on disparity (McKee et al., 1990). In our simulation the disparity away from the fixation plane will always be zero (the stimulus is in the fronto-parallel plane). Clearly though, we would expect some degradation of Vernier acuity as location becomes more eccentric from fixation. The data of Blakemore (1970) for the depth acuity as a function of eccentricity allow us to approximate this relationship.

With an estimate of location uncertainty in the *X*-, *Y*- and *Z*-directions we can now simulate an ideal observer. On a given trial this observer performs weighted least squares regression on data points coming from a number of the uncertainty distributions surrounding each of the points in the stimulus. We define the ideal interpolator’s setting to be the intersection of the regression line with the setting plane. We can then use Monte Carlo techniques to estimate the variability of these settings as a function of the number of points.

In Fig. 13, we report the results of such an analysis for the linear sampled contour of Experiment 1, using the model described above to estimate positional uncertainty. The horizontal axis is the number of points defining the linear sampled contour and the vertical axis is the estimated total variance of the ideal interpolator’s

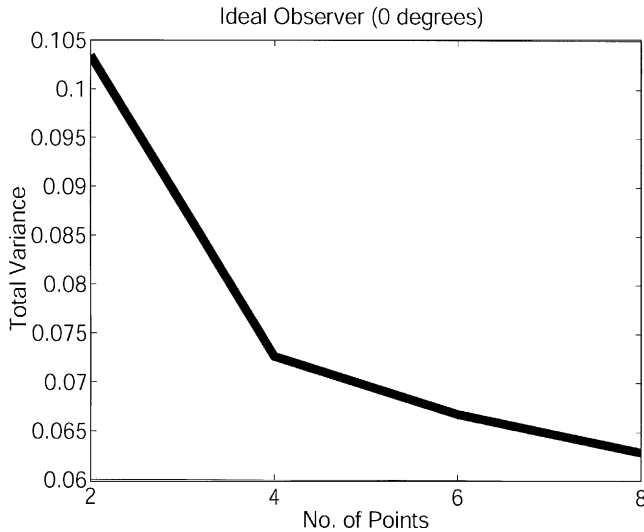


Fig. 13. Ideal interpolator accuracy. The vertical axis is the estimated total variance in setting of a maximum likelihood estimator (an ‘ideal observer’) designed to interpolate linear segments similar to those used in Experiments 1 and 2. The horizontal axis is the number of points on the contour available to the ideal observer. The ideal observer makes use of all of the available points on the contour and (correctly) weights each point according to a model of positional uncertainty as a function of distance from fixation. The uncertainty model is based on previous results in the psychophysical literature, as described in the text. The setting variability of the ideal observer declines markedly with increasing number of points, in contrast to the measured variability of the human observers in Experiments 1 and 2.

settings (recall that total variance is the sum of the variances in the b - and n -direction in the setting plane). Each total variance data point is calculated from a distribution of 10,000 interpolation trials. The total variance in the four-point case is a factor of 0.7 smaller than in the two-point case. Therefore the additional information not used by a human interpolator leads to a 30% reduction in total variance for the ideal interpolator.

In summary, our results suggest that the human visual system is a non-ideal interpolator: only a subset of available information is used and the potential contribution of the neglected information is certainly not negligible.

Robust interpolation. Robust statistics (Hampel et al., 1986) concerns how optimal statistical estimators are affected by small deviations from their underlying assumptions. The ideal Gaussian interpolator (like most Gaussian estimators) may be a poor choice for a less than ideal visual system since small failures in estimating positional variability can lead to a sharp degradation in performance. The ideal interpolator described above must have full knowledge of the position uncertainty associated with every point in visual space. For points distant from fixation, this uncertainty is large, and the ideal interpolator weights the contribution of such points by very small weights (the inverse of the variance). Accordingly, a robust visual interpolation algo-

rithm that does not make use of all available points may outperform an ideal interpolator that is not robust to failures of its assumptions. Unfortunately, a rigorous test of this hypothesis would require more basic psychophysical data regarding positional uncertainty in 3-D space than is currently available in the literature.

The human visual spline. An alternative possibility is that we have stumbled across a basic structural principle constraining human visual interpolation of curves. Under this account, the human visual system, in localizing a sample contour, does so by computing piecewise approximations to sections of the contour containing four points at a time. Our results lead us to suggest that, in interpolating parabolic contours, the piecewise approximation curves that we measure are remarkably close to parabolic. Only further research can establish if this four-point support hypothesis is a general property of human visual interpolation and whether the ‘the human visual’ spline will continue to be a piecewise parabolic interpolation for other classes of smooth, sampled contours such as cubics.

Acknowledgements

All three authors were supported in part by grant EY08266 from the National Institute of Health. LTM and MSL were supported in part by Human Frontiers Science Program grant RG0109/1999-B. We thank Sergei Gepshtein for comments on an earlier draft of the manuscript.

Appendix A

Here we describe how nested hypothesis testing allows us to select the ‘best’ of our models. We want to strike a balance between number of parameters and ‘goodness of fit’ (see Burnham & Anderson, 1998; Zucchini, 2000). As our measure of goodness of fit we use likelihood and we compare models by comparing their likelihood. Let θ_i to be the set of parameters which maximize the likelihood of the data \mathbf{d} in Model i and let $L_i(\theta_i, \mathbf{d})$ be the corresponding maximized likelihood. For example, if we wish to compare Model 1 with Model 2 we compute the likelihood ratio

$$\rho_{12} = \frac{L_2(\theta_2, \mathbf{d})}{L_1(\theta_1, \mathbf{d})}. \quad (\text{A.1})$$

If Model 2 is significantly better than Model 1 we would expect this ratio to be high. If not, we would expect a value close to 1 (recall that Model 1 is nested within Model 2 and so $\rho_{12} \geq 1$). We will reject Model 1 in favor of Model 2 if the likelihood ratio is high enough and the remaining issue to address is, what cutoff value

should we use? How high must the likelihood ratio be before we reject the more constrained model?

The following result provides a basis for setting this cutoff. Under the null hypothesis that Model 1 is correct, the *likelihood ratio statistic* $\Delta = 2 \log \rho_{12}$ is asymptotically distributed as a χ^2 random variable, with degrees of freedom equal to the difference in the number of parameters between the models (Mood et al., 1974, p. 441). We can test this null hypothesis by comparing the value of the likelihood statistic Δ to cutoffs taken from an ordinary χ^2 table.

References

- Badcock, D. R., & Schor, C. M. (1985). Depth increment detection functions for individual spatial channels. *Journal of the Optical Society of America A*, 2, 1211–1215.
- Beard, B. L., Levi, D. M., & Klein, S. A. (1997). Vernier acuity with non-simultaneous targets: the cortical magnification factor estimated by psychophysics. *Vision Research*, 37, 325–346.
- Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *Journal of Physiology, London*, 211, 599–622.
- Braun, J. (1999). On the detection of salient contours. *Spatial Vision*, 12, 211–225.
- Burnham, K. P., & Anderson, D. R. (1998). *Model selection and inference: a practical information-theoretic approach*. New York: Springer.
- Feldman, J. (1997). Curvilinearity covariance, and regularity in perceptual groups. *Vision Research*, 37, 2835–2848.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local “association field”. *Vision Research*, 33, 173–193.
- Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41, 711–724.
- Georgeson, M. A., Freeman, T. C., & Scott-Samuel, N. E. (1996). Subpixel accuracy: psychophysical validation of an algorithm for fine positioning and movement of dots on visual displays. *Vision Research*, 36, 605–612.
- Gepshtein, S., & Kubovy, M. (2000). The emergence of visual objects in space-time. *Proceedings of the National Academy of Science, USA*, 14, 8186–8191.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics; the approach based on influence functions*. New York: Wiley.
- Hanselman, D. C., & Littlefield, B. R. (1997). *Mastering MATLAB 5: a comprehensive tutorial and reference*. Prentice-Hall.
- Hon, A. K., Maloney, L. T., & Landy, M. S. (1997). The influence function for visual interpolation. In B. E. Rogowitz & T. N. Pappas (Eds.), *Proceedings of the SPIE* (3016, *Human vision and electronic engineering II* pp. 409–419).
- Kanizsa, G. (1979). *Organization in vision*. New York: Praeger.
- Klein, S. A., & Levi, D. M. (1987). Position sense of the peripheral retina. *Journal of the Optical Society of America A*, 4, 1543–1553.
- Kovacs, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure ground segmentation. *Proceedings of the National Academy of Science, USA*, 90, 7495–7497.
- McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research*, 30, 1736–1779.
- Mood, A., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). New York: McGraw-Hill.
- Ogle, K. N. (1953). Precision and validity of stereoscopic depth perception from double images. *Journal of the Optical Society of America*, 43, 906–913.
- Pizlo, Z., Salach-Golyska, M., & Rosenfeld, A. (1997). Curve detection in a noisy image. *Vision Research*, 37, 1217–1241.
- Polat, U., & Sagi, D. (1994). The architecture of perceptual spatial interactions. *Vision Research*, 34, 73–78.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C: the art of scientific computing*. Cambridge: Cambridge University Press.
- Siderov, J., & Harwerth, R. S. (1993). Precision of stereoscopic depth perception from double images. *Vision Research*, 33, 1553–1560.
- van Assen, M. A., & Vos, G. (1999). Evidence for curvilinear interpolation from dot alignment judgements. *Vision Research*, 39, 4378–4392.
- Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology*, 44, 41–61.