Vision Research 153 (2018) 53–59

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

A preference for minimal deformation constrains the perceived depth of a stereokinetic stimulus

Yang Xing*, Zili Liu

Department of Psychology, University of California, Los Angeles, United States

A R T I C L E I N F O

Keywords:
Structure-from-motion
Stereokinetic
Maximal rigidity
Minimal deformation
Slow and smooth
Motion integration

A B S T R A C T

The current study examined whether the ‘slow and smooth’ hypothesis (Hildreth, 1984; Yuille & Grzywacz, 1989; Weiss, Simoncelli, & Adelson, 2002) could be extended to explaining a three-dimensional (3D) stereokinetic percept by specifying the smoothness term as a preference for minimal deformation. Stereokinetic stimuli are two-dimensional (2D) configurations that lead to 3D percepts when rotated in the image plane. In particular, a rotating ellipse with an eccentric dot gives rise to the percept of a cone with a defined height. In the current study, the spatial relationship between the ellipse and dot varied across trials in terms of the dot’s relative location and the aspect ratio of the ellipse. During each trial, participants (*n* = 8) adjusted the length of a 2D bar centered along the minor axis of the ellipse to indicate their perceived height of the cone. Upon rotation, the 2D bar was perceived to be perpendicular to the circular base of the cone. Our results were qualitatively and quantitatively consistent with the traditional hypothesis of minimum object change (Jansson & Johansson, 1973), which is also similar to the maximal rigidity assumption (Ullman, 1979). As the dot shifted from the major axis towards the minor axis of the ellipse, observers consistently reported an increasingly taller cone. The results illustrate the tendency of observers to perceive the apex of the cone at a height that minimized its 3D distance to the surface normal at the center of the circular base of the cone to reduce the relative motion between the dot and the base of the cone. The current study provides empirical evidence suggesting that, when presented with an ambiguous stereokinetic stimulus, the visual system prefers the interpretation that corresponds to a 3D percept that is slowest and maximally rigid.

1. Introduction

The perception of 3D structure from 2D visual motion remains a key issue in vision science. 3D reconstruction is intrinsically difficult because local measurements of a dynamic image on a 2D surface can correspond to an infinite number of object transformations in 3D space. In particular, the visual system encounters difficulty determining the exact location of each spatiotemporal point (x, y, t) at time t + Δt, formally known as the motion correspondence problem (Marr & Ullman, 1981). Thus, structure-from-motion (SFM) is an example of an ill-posed problem (Poggio, Torre, & Koch, 1985), which requires specific assumptions about the environment and the corresponding optic array to derive a unique solution. To gain insight into the constraints adopted by the visual system when integrating local motion signals, the current study compares human perception of a stereokinetic stimulus to a pre-defined optimal solution.

Stereokinetic stimuli are a class of SFM stimuli that have been employed experimentally to investigate how motion information supports depth perception in the absence of static pictorial cues. They are 2D images that result in non-veridical 2D and 3D percepts while rotated about an axis perpendicular to the image plane (Musatti, 1924; Wallach, Weisz, & Austin, 1956; Duncan, 1975; Shearer & Gould, 1999). For example, prolonged observation of a 2D rotating ellipse results in three sequential percepts: (1) percept I: the ellipse rotating rigidly in the image plane, (2) percept II: the ellipse deforming in the image plane, and eventually (3) percept III: a stereokinetic percept where the ellipse assumes the shape of a rigid 3D circular disk tilted relative to the image plane (Todorovic, 1993; Weiss & Adelson, 2000; Vezzani, Kramer, & Bressan, 2013).

If the ellipse includes a dot located on its minor axis (Fig. 1), a regular cone is perceived (Zanforlin, 1988). In contrast, when the dot is located on the major axis of the ellipse, observers report the dot sliding across a flat circular disk, tilted relative to the image plane. More generally, when the dot is not on the minor axis, the ellipse and dot configuration is seldom perceived as a rigid cone with an off-centered apex, but more frequently as constantly deforming. The perceived

---

* Corresponding author at: Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095-1563, United States.
E-mail address: yangxing92@ucla.edu (Y. Xing).

https://doi.org/10.1016/j.visres.2018.09.003
Received 5 January 2018; Received in revised form 31 August 2018; Accepted 1 September 2018

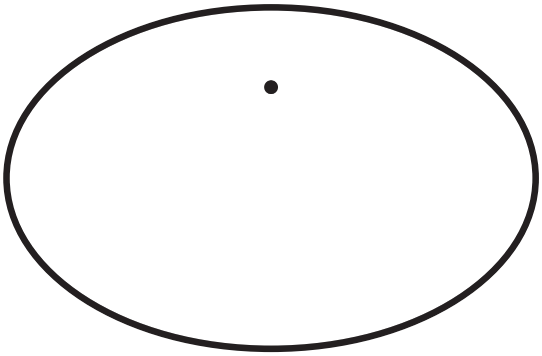


Fig. 1. Ellipse with an eccentric dot located on its minor axis. After extended viewing of this configuration rotating in the image plane, observers report perceiving a tilted regular cone of defined height.

height of the cone remains constant, but the cone appears to deform such that its apex appears to rotate around the surface normal at the center of the cone's circular base[1]. In the current study, we aimed to examine how the location of the eccentric dot influences the perceived height and deformation of the stereokinetic cone.

Multiple interpretations associated with the rotating ellipse demonstrate that its 2D spatiotemporal sequence does not directly convey its true motion. When presented with ambiguous motion stimuli, it has been suggested that the visual system adopts a preference for *slow* and *smooth* motion percepts (Hildreth, 1984; Yuille & Grzywacz, 1989; Weiss, Simoncelli, & Adelson, 2002). The 'slow and smooth' hypothesis posits that, of all the potential velocity fields that can correspond to the observed 2D transformations, the visual system perceives the slowest and spatially smoothest velocity field. For example, a thin ellipse rotating in the image plane is perceived veridically, reflecting perception of the ellipse's true rotational motion. However, the perception of motion of a fat ellipse is ambiguous because segments of curvature along its *vertices* (end points of the major axis) are more similar to each other in comparison to a thin ellipse. This holds most true for the extreme case of an ellipse with aspect ratio of 1, which is a circle. Because of this ambiguity, there is greater potential for the visual system to be biased towards a percept that minimizes motion. Consequently, a fat ellipse is perceived as deforming, indicating a bias towards perceiving motion orthogonal to elements along the shape's contour during its rotation (Hildreth, 1984; Weiss et al., 2002).

Despite successfully explaining a wide range of 2D motions, the 'slow and smooth' hypothesis remains qualitative because there is no precise specification as to how smooth the resultant motion should be, or how the 'slow' and 'smooth' terms should be weighed relative to each other. For example, although rigid objects will give rise to smooth motion, smooth motion does not necessarily correspond to a rigid object in motion (Ullman & Yuille, 1987). As such, there is currently no principled manner for defining the ideal smoothness that the visual system assumes *a priori*.

The notion of a rigid object introduced in the example above has been emphasized by previously suggested heuristics. For instance, a specific proposal endorsing spatial smoothness of motion that predates the 'slow and smooth' hypothesis is the principle of minimal deformation (Jansson & Johansson, 1973). This principle states that the visual system prefers the percept that results in the minimal amount of shape change in 3D. For instance, the spatiotemporal sequence associated with a 2D planar square whose vertices gradually retreat towards its center has two potential interpretations: (1) a square shrinking in the image plane or (2) a square receding from the viewer. In accordance with the principle of minimal deformation, observers reported perceiving a square moving further away, indicating the visual system is biased towards a percept that preserves its size and shape (Johansson, 1964).

Relatedly, Ullman (1979, 1984) redefined the principle of minimal deformation as the principle of maximal rigidity. According to this rigidity assumption, the visual system prefers an interpretation that corresponds to a rigid object in motion (see also Kersten, Bülthoff, Schwartz, & Kurtz, 1992). In the case of a rotating ellipse with an eccentric dot, the 2D veridical percept is a rigid solution. However, unless the dot is located on the minor axis of the ellipse, observers report that the 3D cone is constantly deforming. As such, although the 2D distance between any two points remains constant during the configuration rotation, the perceived deformation suggests that the visual system utilizes preferences beyond rigidity. The compatibility of multiple rigid percepts with the same 2D spatiotemporal sequence suggests when inferring the 3D structure of an object from its 2D motion pattern, the visual system utilizes a preference for rigidity alongside additional constraints.

The evidence described above illustrates that both minimal motion assumptions and minimal shape change assumptions encounter difficulty accounting for stereokinetic phenomena, when applied independently. Assumptions focusing on minimizing velocity and relative velocity cannot distinguish between the stereokinetic percept (percept III) and the deforming ellipse (percept II), as both percepts are compatible with the same motion profile. Furthermore, minimal shape change assumptions cannot differentiate between the initial percept reported by observers (percept I) and the stereokinetic cone when the dot is located on the minor axis. Taken together, it appears that the perception of the stereokinetic cone involves heuristic processes, and the previously recorded successes of the 'slow and smooth' hypothesis and the rigidity assumption implies that the visual system utilizes both heuristics in some manner.

At balance, the 'slow and smooth' hypothesis remains qualitative due to the unconstrained nature of the smoothness term. In other words, there is no formal manner of specifying, *a priori*, how smooth the final percept should be. Most critically, the 'slow and smooth' hypothesis is restricted to 2D motion in its current form, while the stereokinetic percept is a 3D cone. How can the rigidity assumption and the 'slow and smooth' hypothesis be applied in conjunction to explaining 3D percepts?

Given that local rigidity endorses interpretations that minimize relative motion between neighboring points, the smoothness term of the previously proposed 'slow and smooth' hypothesis can potentially be specified more precisely as a preference for rigidity. More specifically, the final 3D percept is perhaps not simply the slowest and spatially smoothest percept, but the slowest and minimally deforming (i.e., maximally rigid). In more intuitive terms, the visual system is biased towards perceiving a 3D structure that results in the minimal amount of motion and deformation by maintaining the 3D distances between any two points of its structure, as much as possible, throughout its spatiotemporal sequence.

In the current study, the perceived height of the stereokinetic cone was measured psychophysically to determine whether applying the two assumptions in 3D successfully explains human perception of a stereokinetic stimulus. As far as we know, the approach of applying the rigidity assumption and the 'slow and smooth' hypothesis simultaneously has never been applied previously in the literature.

Prior to applying this new hypothesis, three assumptions were postulated to govern perception of the stereokinetic cone. First, it was assumed that the motion of the ellipse is perceived as the motion of a 3D circular disk, tilted relative to the image plane. A rotating 3D circular disk produces a slower motion than that of a rigid ellipse or a deforming 2D ellipse (Yang, 2012). Furthermore, a rigid disk ensures that the resultant motion vector field is reasonably smooth. Relatedly, it was assumed that the perceived motion of the disk was the same

---

[1] Surface normal refers to the axis of spinning rotation extending through the circular disk's center. For the remainder of the paper, the circular disk's central axis of rotation will be referred to as surface normal.

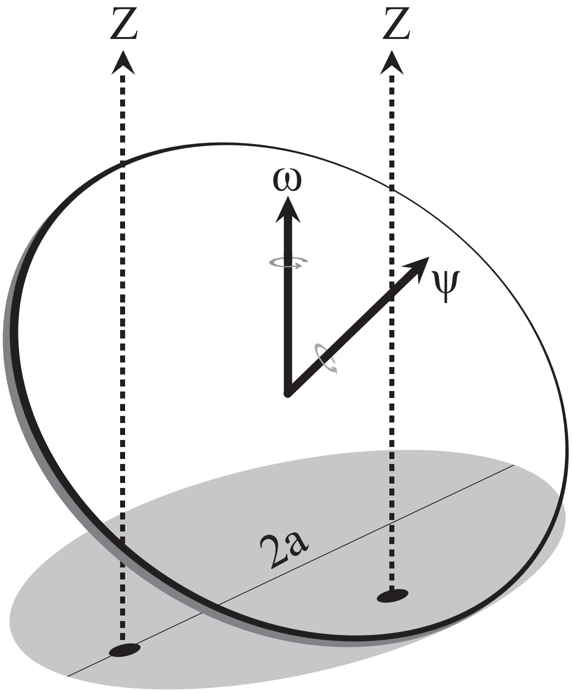


Fig. 2. Proposed hypothesis. The gray oval represents the projection of an ellipse on the image plane with eccentric dots on its major axis (labeled 2a) and minor axis. The presence of two dots is for illustrative purposes only. The configuration is presented as an ellipse with one eccentric dot within its interior. When the configuration is rotated about the Z-axis with angular velocity, ω, observers eventually perceive a 3D percept. The disk with a solid black contour represents the 3D circular base of the cone, which the visual system assumes to be responsible for the 2D ellipse projection. The black arrow labeled, ψ, represents the cone's central axis of rotation (i.e., the surface normal of the disk). The dashed arrow extending from each dot represents potential locations in 3D space where the depth of the dot can be perceived. When the dot is on the minor axis, the apex of the cone is perceived at the particular depth where the Z-line of the dot intersects with the surface normal of the circular disk. When the dot is positioned off the minor axis, its Z-line will never intersect with the surface normal of the disk. Thus, when the dot is on the major axis, its 3D distance to the surface normal is minimal when it is directly on the surface of the tilted disk.

regardless of whether or not the dot was present. Finally, we assumed that the dot's motion was unambiguous because there was no correspondence uncertainty in the x-y plane.

The first assumption above is motivated by previous psychophysical work by Rokers, Yuille, and Liu (2006). When a stereokinetic disk is perceived from the rotating ellipse, it exhibits two rotations: (1) ω around the Z-axis orthogonal to the image plane at the center of the stimulus and (2) ψ around the surface normal of the disk (see Fig. 2). When the frame of reference is on the eccentric dot, the dot will be stationary and no longer rotating around the Z-axis. The only motion left, with respect to the dot, is the disk's motion, ψ, which is the angular velocity of its spinning around the surface normal extending through its center. The relative motion between the disk and the dot depends on their relative distance in Z, because their (x, y) positions are determined already. This relative height from the base to the dot in the Z-dimension is undetermined and is the only free parameter in this problem. If the dot is located on the ellipse's minor axis, then the Z-line[2] passing through the dot will intercept the surface normal of the disk. If the Z-coordinate of the dot is chosen to be at this intercepting point, then there is no relative motion between the disk and the dot, since the dot defines the apex of a regular cone.

In general, when the dot's position is not on the minor axis of the ellipse, the Z-line passing through the dot and the surface normal of the disk do not intersect. However, these two lines have a minimal distance between them (Fig. 2). If the dot's Z-coordinate is chosen such that it is closest to the surface normal of the disk, then the relative motion between the dot and the disk will be minimal. This relative motion exists because the dot is stationary in the current frame of reference, whereas the disk is spinning with an angular velocity ψ. We should reiterate that, here, we assume that the Z-position of the dot is independent of ψ, which is presumably determined completely by the 'slow and smooth' motion of a rigid disk (Rokers et al., 2006), independently of the dot.

If the visual system prefers a 3D percept that matches the proposed hypothesis above, then observers should report an increasingly taller cone as the eccentric dot progresses from the major axis towards the minor axis. In particular, the height of the cone when the dot is positioned on the major axis should be zero. Fig. 2 presents a schematic illustration of the ideal height of the stereokinetic cone. Fig. 7 provides clarification of how the height of the cone can be calculated.

2. Experiment: measuring the perceived height of the stereokinetic cone

In the current study, a perceptual task was utilized to assess whether a quantitative model involving a preference for slow motion and minimal deformation could account for the perceived height of the stereokinetic cone. Observers were presented with various configurations of the ellipse and the eccentric dot. The experimental paradigm involved systematically manipulating the angular location of the eccentric dot such that it would appear on the minor axis, major axis, and in between, across trials. In each trial, observers adjusted the length of a bar rotating simultaneously with the ellipse to indicate their perceived height of the cone. This bar was oriented along the minor axis of the ellipse in 2D, and was perceived to be perpendicular to the disk in 3D.

2.1. *Participants*

Eight observers participated in the study (5 females; age range: 18–29). One observer was author YZX. The remaining seven observers were six research assistants and one post-doctorate scholar who were naïve to the purpose of the study, but were trained psychophysical observers. The experiment was approved by the UCLA Institutional Review Board. Informed consent was obtained from all participants, who were treated in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki, 2013).

2.2. *Apparatus*

Stimuli were displayed on a 17-inch Dell Trinitron CRT with a resolution of 1280 × 960 pixels and a 60 Hz refresh rate. The viewing distance was 70 cm. Observers used a headrest to stabilize the position of their head. A viewing tube was appended to the computer screen to prevent any additional cues that could affect the perception of depth within the experiment. The computer screen (background luminance of 0.01 cd/m^2) provided the only light source in the room.

2.3. *Stimuli*

2.3.1. *Ellipse & eccentric dot*

The stimuli were developed and displayed using the MATLAB programming language and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Trial-to-trial, the spatial relationship between the ellipse and the eccentric dot was pre-programmed to vary in terms of the dot's location and the aspect ratio of the ellipse. There were four potential locations for the eccentric dot (0°: major axis, 30°, 60°, 90°: minor axis),

---

[2] Z-line and projection are synonymous. Both refer to the line of sight to the object when directly facing the stimulus.

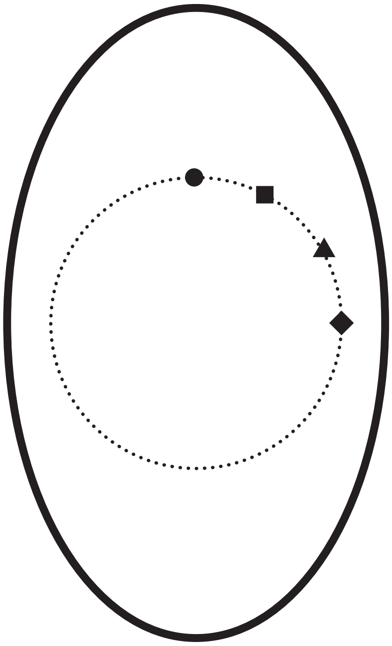


Fig. 3. Eccentric dot positions for ellipse of aspect ratio 0.6. The filled circle and diamond represent the major axis (0°) and minor axis (90°) positions, respectively. The square and triangle represent off-axis locations at 30° and 60°, respectively. The dot's distance to the center of the ellipse was held constant (77% of the semi-minor axis) across all angular locations for both aspect ratios.

two aspect ratios (0.6 and 0.8; major axis = 15.46 cm [~12.6° visual angle]), and finally, two rotation speeds of the stimulus (60° and 90°/s).

Previous research has suggested that the dot's eccentricity influences the perceived height of the stereokinetic cone (Zanforlin, 1988). As such, we designed our stimuli such that the location of the dot was always equidistance from the center of rotation. The eccentric dot's distance from the center of the ellipse was always 77% of the semi-minor axis. See Fig. 3 for an illustration of the potential locations of the eccentric dot.

2.3.2. Rotating Bar

To measure the perceived height of the cone, we leveraged the ambiguity of a 2D bar rotating in the image plane. When oriented on the minor axis of the ellipse, the bar rotating about its center eventually gave rise to the percept of a tilted bar perpendicular to the base of the cone. Observers were instructed to adjust the bar such that one of its endpoints matched the perceived height of the cone. If the observer adjusted the length of the bar, both ends would either decrease or increase equivalently relative to the center of the bar. Please refer to Fig. 4 for a visual diagram of the depth probe.

2.4. Procedure

At the beginning of each trial, the ellipse and the eccentric dot were presented on the computer screen. Immediately afterwards, the configuration was rotated either clockwise or counterclockwise. If the observer perceived the stereokinetic cone, they were instructed to press the up arrow key to initiate the appearance of the rotating bar probe. The observer further adjusted the length of the rotating bar using the up and down arrow keys such that its tip matched the perceived height of the cone's apex. The up arrow key naturally increased the length of the rotating bar while the down arrow key decreased its length. The stimulus was presented until the observer pressed the spacebar key to proceed to the next trial.

Fig. 4. 2D bar probe. A bar positioned on the minor axis of an ellipse rotating about its center gives rise to a percept of a bar perpendicular to the circular base of the perceived cone. Observers used the up and down arrow keys to adjust the length of the bar. If adjusted, both ends of the bar would decrease or increase in length, equivalently, relative to its center.

2.5. Design

We manipulated the angular location of the eccentric dot (four levels: 0°, 30°, 60°, and 90° relative to the major axis), the aspect ratio of the ellipse (two levels: 0.6 and 0.8), and the rotation speed of the stimulus (two levels: 60°/second and 90°/second) in a 4 × 2 × 2 within subject's factorial design. The dependent variable of our experiment was the final length of the bar observers adjusted during each trial. In turn, the length of the bar was used to calculate the perceived height of the cone. Combining the various levels of the independent variables resulted in 16 different display types. Observers were presented with each display 13 times, resulting in an experiment consisting of 208 trials. On average, observers completed the experiment in 42 min.

3. Results

Participants' adjustments of the rotating bar probe were submitted to a repeated measures ANOVA as the dependent variable while angular location of the eccentric dot, aspect ratio, and rotation speed served as the independent variables. Because the analysis did not indicate a significant influence of the speed of rotation, observers' perceived height of the cone at the four different angular locations for both the 0.6 and 0.8 aspect ratios are plotted in centimeters in Fig. 5, after collapsing across rotation speed.

The analysis revealed a main effect of angle: $F(3,\ 21) = 485.51$, $p < .001$. The ideal height of the stereokinetic cone was predicted based on the assumption that observers prefer percepts that deform minimally. If that were the case, then observers should have reported a gradually taller cone as the eccentric dot shifted from the major axis toward the minor axis. The mean perceived height of the cone across both aspect ratios when the eccentric dot was positioned at 0° (major axis), 30°, 60°, and 90° (minor axis) locations are (1) 0.13 cm (cm), (2) 2.25 cm, (3) 4.89 cm, and (4) 6.45 cm, respectively. Thus, the perceived height of the cone increased systematically as the eccentric dot progressed towards the minor axis.

There was also a main effect of aspect ratio: $F(1,\ 7) = 194.31$, $p < .001$. More specifically, as the aspect ratio increased, observers

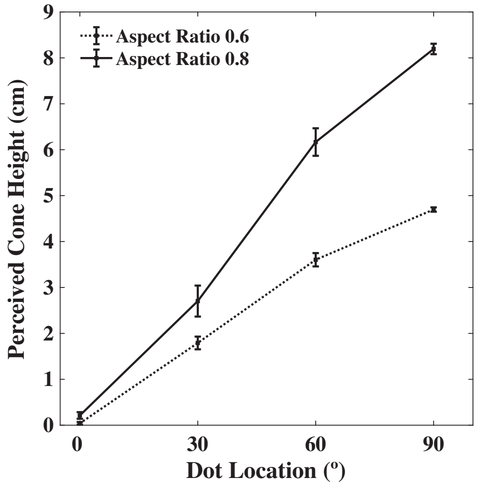


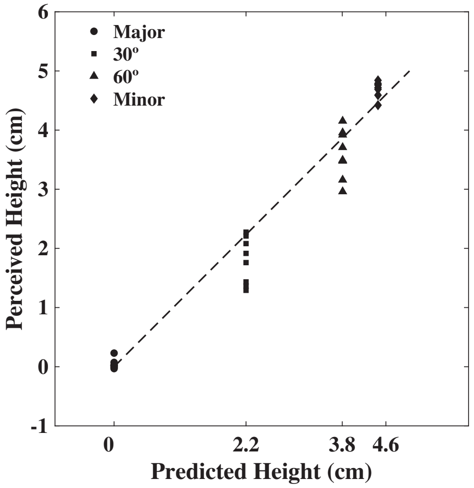
Fig. 5. Perceived cone height as a function of dot location for aspect ratios 0.6 and 0.8. Each data point indicates the mean perceived height (n = 8) at a specific dot location. Observers reported increasingly taller cones as the eccentric dot progressed towards the minor axis. Error bars signify standard error of the mean.

perceived a taller cone. With ellipses of 0.6 and 0.8 aspect ratios, the mean perceived heights of the cone were 2.53 cm and 4.31 cm, respectively. However, the main effect of aspect ratio was expected, as the dot's position was always 77% of the ellipse's semi-minor axis. Increasing the aspect ratio of the ellipse naturally increased the distance between the dot and the center of the ellipse.

The only significant interaction within the study occurred between angle and aspect ratio: $F(1, 7) = 122.88$, $p < .001$. As can be seen in Fig. 5 the average difference in perceived height of the cone increases as the dot shifts from the major axis towards the minor axis. For instance, the average difference in perceived height increased from 0.17 cm when the dot was located on the major axis, to 3.5 cm when the dot was located on the minor axis.

Following the repeated measures ANOVA, a multilevel regression analysis was conducted to determine whether there was a linear correspondence between theoretical predictions and empirical data while accounting for correlated performance within subjects. The analysis was conducted using Hierarchical Linear and Nonlinear Modeling (HLM) software (Bryk, Raudenbush, & Congdon, 1996). Two separate regression analyses were conducted for aspect ratios 0.6 and 0.8 with observed height across all trials, regardless of rotation speed, as the outcome variable, and predicted height, a function of dot position and aspect ratio, as the predictor in the regression model. In both regression analyses, the slope and intercept were allowed to vary across participants.

(a) Aspect Ratio = 0.6



(b) Aspect Ratio = 0.8

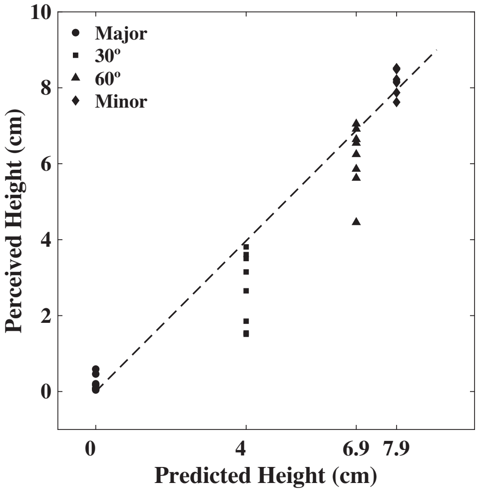


Fig. 6. Empirical results and theoretical predictions for aspect ratios (a) 0.6 and (b) 0.8. Predicted cone heights emphasizing minimal deformation are plotted on the abscissa. Empirical data are plotted along the ordinate axis. Individual data points reflect mean perceived height at a particular dot position for an observer. Across all participants, the perceived height of the cone increased, at a similar rate, as the eccentric dot progressed towards the minor axis. For both figures, an identity line was plotted to endorse visualization of the correspondence between theoretical predictions and empirical data.

For aspect ratio 0.6, the slope of the regression model was 1.00, $SE = 0.02$, $p = .87$[3], $R^2 = 0.83$, while the intercept was -0.13, $SE = 0.05$, $p = .05$ (without author YZX: $n = 7$; slope = 1.00, $SE = 0.02, p = .96, R^2 = 0.82$; intercept = $-0.09, SE = 0.05, p = .12$). For aspect ratio 0.8, the slope of the regression model was 0.98, $SE = 0.02$, $p = .32$, $R^2 = 0.74$, while the intercept was -0.28, $SE = 0.12$, $p = .06$ (without author YZX: $n = 7$; slope = 0.97, $SE = 0.02, p = .37, R^2 = 0.75$; intercept = $-0.21, SE = 0.11, p = .12$). The intercepts capturing each observer's performance did not significantly differ from each other, demonstrating that observers perceived a similar increase in height as the eccentric dot progressed from the major axis towards the minor axis. Taken together, the results of the multilevel regression suggest that the visual system appears to minimize the perceived deformation of the cone by interpreting the cone's apex at a height closest to the circular disk's central axis of rotation. Fig. 6 emphasizes the similarity between theoretical predictions and empirical data in two separate graphs corresponding to the two examined aspect ratios.

Following the multilevel regression, we observed differences in the average prediction bias across the four dot locations. To examine these differences, we conducted post-hoc two-tailed t-tests with a Bonferroni correction ($\alpha = 0.00625$) for the average prediction bias, comparing each group to a null hypothesis with no prediction error. All tests were performed at both levels of aspect ratio. For aspect ratio 0.6, when the dot was located on the minor axis, perceived height was significantly biased in the positive direction $+0.23$ cm, $t(7) = 4.87$, $SE = 0.05$, $p = .0018$. There was no statistically significant bias in the remaining conditions.

Additionally, we observed that the variance in task performance across individuals varied as a function of the dot's position. As can be seen in Fig. 6, observers demonstrated significantly larger variance in task performance when the eccentric dot was presented at an off-axis

---

[3] For slopes, H_0: slope = 1.

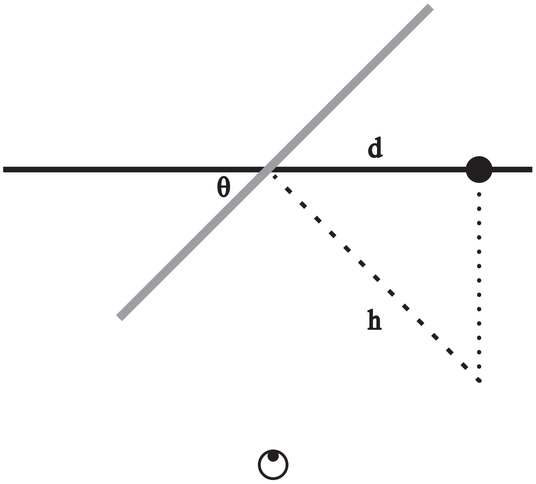


Fig. 7. Top-down view of the intersection of the circular disk's surface normal and the projection of the eccentric dot. The solid black dot represents the 2D projection of the dot on the monitor. The gray line represents the tilted stereokinetic disk. h indicates the perceived height of the cone. θ represents the tilt of the circular disk. d denotes the distance of the eccentric dot to the major axis.

location (i.e., when the dot was located 30° or 60° relative to the major axis). To examine these differences, we conducted post-hoc F tests with a Bonferroni Correction ($\alpha = 0.025$) comparing the on-axis positions (i.e., when the dot was on the major axis or minor axis) to the off-axis locations at both levels of aspect ratio. For both aspect ratios, the variance of individuals' performance in the off-axis dot locations was significantly greater than variance in performance across the on-axis positions (for aspect ratio $b/a = 0.6$, $F(7,7) = 11.03$, $p = .0052$; for $b/a = 0.8$, $F(7,7) = 9.00$, $p = .0096$). These F ratios can be interpreted directly as the ratio of the variance of these two conditions. For an aspect ratio of 0.6, the variability in performance in the off-axis conditions was 11.03 times as high as in the on-axis conditions, while for an aspect ratio of 0.8, the ratio of these variances was 9.00.

This can be explained as follows. Recall that the 2D bar probe was oriented alongside the minor axis of the rotating ellipse, and was perceived as the surface normal of the 3D disk. Therefore, if the dot is perceived as the apex of a regular cone when it is positioned on the minor axis, our theory predicts that one tip of the bar will necessarily intersect with the dot. This intersection provides a visual reference, simplifying the task for observers. In a similar vein, when the dot was located on the major axis, observers reported little to no height; thus, it was unnecessary to increase the bar's length from its minimal size. Both of these extreme conditions (i.e., on-axis locations [0° and 90° relative to the major axis]) provided explicit visual reference, making these particular conditions easier than when the dot was presented at an off-axis location, where its depth could only be matched by the height of the bar in the perceived 3D space.

4. Discussion

In the current study, we quantified the perceived height of the stereokinetic cone to evaluate whether the 'slow and smooth' hypothesis (Hildreth, 1984; Yuille & Grzywacz, 1989; Weiss et al., 2002) could accurately predict the perceived height of various stereokinetic configurations, when defining the smoothness term as a preference for minimal deformation (Jansson & Johansson, 1973; Ullman, 1979, 1984; Braunstein & Andersen, 1984). We assumed that the ellipse was perceived as a circular disk to ensure a slow and rigid solution. Based on that assumption and the observation that the perceived cone is deforming unless the dot is located on the minor axis, we suggest that the visual system allows for minimal deformation to occur to minimize relative motion between the apex of the cone and its base.

The height of the stereokinetic cone that corresponds to the slowest and maximally rigid percept has its apex situated closest to the axis of rotation of its circular base to minimize relative motion between the configuration's constituent elements. When the eccentric dot is located on the minor axis, the height of the cone is at its zenith and is simply dictated by the intersection of the dot's projection line and the surface normal at the center of the tilted circular disk. However, when the dot is positioned on the major axis, its projection will never intersect with the surface normal of the tilted circular disk and the eccentric dot's 3D distance to the axis of rotation is shortest when directly on the surface of the disk. Therefore, the 3D percept was predicted to be flat. The current study provides supporting evidence for the proposed hypothesis as observers ubiquitously reported decreases in the perceived height of the cone as the eccentric dot progressed from the minor axis of the ellipse towards the major axis position. The multilevel regression comparing the relationship between the eccentric dot's location and the perceived height of the cone indicated by observers' adjustment of the 2D bar probe confirms the current hypothesis.

In sum, based on the comparison between observers' bar adjustments and theoretical predictions, *a priori* assumptions of motion slowness and minimal deformation appear to accurately account for how the location of the dot influences the perceived height and deformation of the stereokinetic cone. Thus, the visual system does not perceive a cone simply because it is a 3D solution. Instead, the current study provides empirical evidence demonstrating that, when the visual system is confronted by ambiguous 2D motion signals, it selects the 3D structure that produces the slowest total motion and the minimal amount of deformation. In the case of the rotating ellipse and dot configuration, the ideal percept is a 3D cone.

Acknowledgments

We thank Gennady Erlikhman, Maxwell Mansolf, James Kubricht, Mark Straccia, and Michell McCool for their valuable comments and suggestions.

Conflict of Interest statement

The authors declare that there were no extant commercial or financial relationships that could be interpreted as a potential conflict of interest while the reported study was conducted.

Appendix A:. Ideal height of the stereokinetic cone

An ellipse can be defined as having a major and minor axis. However, in the case of a circle, its major and minor axes are equivalent (i.e., one diameter). Because the length of the major axis is never perceived to change, the tilted circular disk's diameter is directly derived from the principal ellipse's major axis. The ideal height of the cone can be calculated using the following equation:

$$h = \frac{d}{\sin(\theta)}$$

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.visres.2018.09.003.

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Braunstein, M. L., & Andersen, G. J. (1984). A counterexample to the rigidity assumption in the visual perception of structure from motion. *Perception, 13*(2), 213–217.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. (1996). *HLM 4 for windows [computer software]*. Chicago, IL: Scientific Software International Inc.

Duncan, F. (1975). Kinetic art: On my psychokinematic objects. *Leonardo, 8*, 97–101.

Hildreth, E. C. (1984). *The measurement of visual motion.* Cambridge, MA: MIT Press.

Jansson, G., & Johansson, G. (1973). Visual perception of bending motion. *Perception, 2*(3), 321–326.

Johansson, G. (1964). Perception of motion and changing form. *Scandinavian Journal of Psychology, 2*, 171–208.

Kersten, D., Bülthoff, H. H., Schwartz, B. L., & Kurtz, K. J. (1992). Interaction between transparency and structure from motion. *Neural Computation, 4*(4), 573–589.

Marr, D., & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B: Biological Sciences, 211*(1183), 151–180.

Musatti, C. (1924). Sui fenomeni stereocinetici. *Archivio Italiano di Psicologia, 3*, 105–120.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature, 317*, 314–319.

Rokers, B., Yuille, A., & Liu, Z. (2006). The perceived motion of a stereokinetic stimulus. *Vision Research, 46*, 2375–2387.

Shearer, R. R., & Gould, S. J. (1999). Of two minds and one nature. *Science, 286*, 1093–1094.

Todorovic, D. (1993). Analysis of two- and three-dimensional rigid and nonrigid motions in the stereokinetic effect. *Journal of the Optical Society of America A, 10*, 804–826.

Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London B: Biological Sciences, 203*(1153), 405–426.

Ullman, S. (1984). Rigidity and misperceived motion. *Perception, 13*, 219–220.

Ullman, S., & Yuille, A. (1987). *Rigidity and smoothness of motion (no. AI-M-989).* Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Vezzani, S., Kramer, P., & Bressan, P. (2013). Stereokinetic effect, kinetic depth effect, and structure from motion. In J. Wagemans (Ed.). *Oxford handbook of perceptual organization*. Oxford, UK: Oxford University Press.

Wallach, H., Weisz, A., & Austin, P. (1956). Circles and derived figures in rotation. *American Journal of Psychology, 69*, 48–59.

Weiss, Y., & Adelson, E. H. (2000). Adventures with gelatinous ellipses – constraints on models of human motion analysis. *Vision Research, 29*, 543–566.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*(6), 598–604.

Yang, X. (2012). *Experimental and computational studies on human visual perception of structure from motion and natural scenes.*(Ph.D. thesis). Los Angeles: Department of Psychology, University of California.

Yuille, A. L., & Grzywacz, N. M. (1989). A mathematical analysis of the motion coherence theory. *International Journal of Computer Vision, 3*, 155–175.

Zanforlin, M. (1988). The height of a stereokinetic cone: A quantitative determination of a 3-D effect from 2-D moving patterns without a "rigidity assumption". *Psychological Research Psychologische Forschung, 50*, 162–172.