Title
Understanding Spatiotemporal Boundary Formation: Processes, Models, and Scope

Author
Erlikhman, Gennady

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Understanding Spatiotemporal Boundary Formation: Processes, Models, and Scope

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Cognitive Psychology

by

Gennady Erlikhman

2014

ABSTRACT OF THE DISSERTATION

Understanding Spatiotemporal Boundary Formation: Processes, Models, and Scope

by

Gennady Erlikhman

Doctor of Philosophy in Cognitive Psychology

University of California, Los Angeles, 2014

Professor Philip J. Kellman, Chair

Often, differences in luminance, color, texture, and depth can help us determine object boundaries. However, when two surfaces have similar textures, as in the case of camouflage, or under dim lighting conditions, object segmentation can be difficult. In such cases, motion leading to the gradual accretion and deletion of texture information on a farther surface by a nearer one can be used to define the nearer object's boundary. It has been demonstrated that accretion and deletion is but one of a general class of texture element transformations that can give rise to the perception of illusory contours, global form, and global motion. This general process is called spatiotemporal boundary formation or SBF.

In the first chapter, I demonstrate two novel properties of SBF. First, SBF can be seen when element transformations are displacements in random directions. Second, global forms can be seen even when SBF-defined objects are rotating, expanding or contracting, accelerating, or smoothly deforming from frame to frame. I consider a two-stage model of SBF that can account for the perception of illusory contours and global form. In the first stage, oriented edge fragments are extracted locally from the sequential transformation of at least three elements in a small

spatiotemporal neighborhood. In the second stage, these fragments are integrated and missing regions are interpolated by the same processes that govern spatiotemporal interpolation between contrast-defined edges.

Chapter 2 tests the first stage of this model. I created a display in which small circular elements were arranged in a sawtooth pattern and disappear and reappear one at a time in sequence. The resulting percept was not of apparent motion, but of an illusory bar that occluded elements one at a time. Using both subjective and objective methods, I identified the spatial and temporal parameters under which SBF occurs. The experiments provide support for models of SBF that begin with extraction of local edge fragments and identify minimal conditions required for this process.

In the final chapter, I implemented the first stage of the SBF model and used it to predict edge orientations of SBF-defined edges. Model and human performance were compared in an orientation discrimination task as a function of element density, number of element transformation, and frame duration. The ideal observer model was able to perfectly predict edge orientation while human performance was suboptimal. I considered several constraints and sources of noise that could contribute to differences between human and ideal performance. In a second experiment, I measured the sensitivity to spatial and temporal display properties that may have acted as sources of noise. A model that incorporated these constraints and sources of noise was able to model human performance very closely with no additional free parameters. The behavioral and modeling work provide the first empirical evidence in support of the two-stage model of SBF.

The dissertation of Gennady Erlikhman is approved.

Hongjing Lu

Steven Nusinowitz

Dario Ringach

Philip J. Kellman, Committee Chair

University of California, Los Angeles

2014

To Allison,

Whose kindness, friendship, and love

Are the start of my day and the end,

Who holds me up and lifts me higher

Than I could ever do on my own.

To my family,

Stella, Jacob, and Boris,

For their loving support,

For giving me books, curiosity, and

A love for understanding how things work.

To my wife's family,

Julie, Steven, Jenna, and Natalie,

For their love, help, support, and delicious food.

Thank you.

Table of Contents

List of figures

target shape. Values closer to 100% indicate greater response precision. Data are shown for correct trials only.

Figure 1.9. Percentage of morph between target shape and another shape separated by whether the response came before or after the frame on which the pure target shape was presented (response time corrected, see text). Data are for correct trials only

Figure 1.10. Average shape identification accuracy from Experiments 1 (black, red, green, and blue lines) and 2 (purple line, "non-rigid") plotted as a function of element density. Also plotted are reproduced data from Experiment 3 from Shipley and Kellman (1994) in gray.

Figure 2.1. An example of three frames from a typical SBF display. The virtual object, a square, is indicated by the dashed line. Elements inside the object boundary are white, indicating that they share some surface property (e.g., color, orientation, shape), while those outside are black, having a different value for that property (e.g. red circles inside the boundary and green outside). As the virtual object moves, elements entering the boundary of the square become white and those exiting become black. Figure adapted from Figure 2 on p. 5 of Shipley T. F. & Kellman, P. J., (1994) Spatiotemporal boundary formation: Boundary, form, and motion perception from transformations of surface elements. *Journal of Experimental Psychology: General*, *123*(1), pp. 3-20.

Figure 2.2. Four apparent motion frames in which elements disappear one at a time in sequence. Consider the first two elements in the display as an apparent motion pair. Stimulus onset asynchrony (SOA) is the temporal interval between the appearance of the first element and the appearance of the second. Stimulus duration is the time one of the two elements is visible. ISI is negative because there are no frames when both elements are invisible.

Figure 2.3. A replotting of Figure 2.2, treating the white dot as a stimulus. SOA remains the same duration, but is now measured from the appearance of one dot until the appearance of the next. The stimulus duration is relabled as the flash duration, f, which is the duration that a white dot is present. The ISI is the temporal interval between the disappearance of a white dot and the appearance of the next dot.

Figure 2.4. The same display as in Figures 2.2 and 2.3. Element disappearances (changes to white) are triggered by the passing of a virtual bar, indicated by the dashed rectangle.

Figure 2.5. Illusory contour clarity ratings for Experiment 1, averaged across subjects. Ratings were on an integer scale of 1 to 5. Bars indicate the standard error of the mean. Inter-stimulus interval (ISI) corresponds to the duration that all elements were visible on the screen in between flashes. The seven curves represent different inter-element separations from 0.28 degrees (circles) to 1.13 degrees (inverted triangles).

Figure 2.6. Still image from Experiment 2. Elements of the sawtooth pattern disappeared one at a time. Subjects matched the perceived width of an illusory bar with one of seven possible choices shown above the sawtooth display.

Figure 2.7. Perceived illusory contour width as a function of true width averaged across all subjects. Width was computed from element flash duration and SOA (see text for details). Data were split for short (circles) and long (triangles) SOAs. Error bars are standard errors of the mean.

Figure 3.1. Depiction of a square "pseudosurface" moving over a field of circular black elements. All elements inside the square region are in one state and all those outside are in another. Each individual frame contains a collection of white circles in an amorphous group. As the square moves (frames 2 and 3), elements entering and exiting the region change states. The resulting percept is of a moving, colored region with crisply defined illusory contours. Figure from Erlikhman & Kellman (*submitted*).

Figure 3.2. A sequence of frames in which a moving edge successively transforms three elements (changing from black to white). (a). Three elements disappear, one at a time. $\mathbf{v}_{12}$ and $\mathbf{v}_{23}$ are transformation vectors defined by the spatial and temporal separation between elements. (b). Transformation vectors $\mathbf{v}_{12}$ and $\mathbf{v}_{23}$ can be combined to define the orientation of the moving edge. Figure from Shipley, T. F. & Kellman, P. J. (1997). Spatio-temporal Boundary Formation: the Role of Local Motion Signals in Boundary Perception. *Vision Research*, 37(10), 1281-1293.

Figure 3.3. Illustration of stimuli used in Experiment 1. An invisible, oriented bar moved laterally across a field of black elements on a white background. Whenever it passed the midpoint of an element, that element disappeared (became white; indicated by dashed circle in second panel) all at once. The element remained white for two frames and then reappeared (became black). The perception was of a moving, illusory, white bar.

Figure 3.4. Average orientation discrimination thresholds for three display conditions tested in Experiment 1. Thresholds are shown as a function of element density (a), number of events (b), and frame duration (c). Human performance data are shown in black. Bars indicate the 95% confidence intervals. Ideal observer performance is shown in blue.

Figure 3.5. An illustration of a trial in Experiment 2. Each row depicts one interval. A region of region of the display was cued in which element transformations would occur (first panel). An element within that region would disappear (second panel, indicated by dashed boundary) and reappear (third panel). A second element would disappear, also within the cued region (fourth panel). The two elements define a spatial, angular, and temporal value, which is compared to one in a subsequent interval (second row).

Figure 3.6. Data from Experiment 1 (black) replotted with model fits (blue) using noise parameter estimates from Experiment 2. Error bars are 95% confidence intervals. Model data reflect the average of 10 simulated experiment runs.

List of Tables

Acknowledgement

I would like to thank Phil Kellman for his guidance, wisdom, and inspiration over the years I have been fortunate enough to be his student. Many times over, he has shown by example the kind of scientist I strive to become. Thank you. I would also like to thank the Human Perception Lab lab members for helping to refine the ideas presented here, for subjecting themselves to countless pilot studies, and for their friendship in the time I have spent here.

The work presented in Chapter 1 is part of a paper in press that has been co-authored by Ying (Mac) Z. Xang and Philip Kellman. Mac has been a research assistant in the lab for two years and much of the work would not have been completed without his enthusiastic assistance. Chapter 2 is paper that has been submitted and is co-authored with Philip Kellman. Chapter 3 is a paper in preparation also co-authored with Philip Kellman. I wish to thank the invaluable discussions held with Hongjing Lu regarding computational modeling and general ideas. I would also like to acknowledge the help of Ariela Iancu and Maxwell Mansolf who assisted in collecting data for the experiments presented in that chapter.

VITA

2009 B.A. Philosophy and Cognitive Science
University of Pennsylvania
Philadelphia, PA
Honors Thesis in Cognitive Science: "Recall by Category"

2010 M.A. Psychology
University of California, Los Angeles
Los Angeles, CA
Specialization: Cognitive Psychology, Computational Cognition

PUBLICATIONS AND PRESENTATIONS

Erlikhman, G., & Kellman, P. J. Recovery of local edge fragments initiates spatiotemporal boundary formation. *Submitted*

Erlikhman, G., Xing, Y. Z., & Kellman, P. J. Non-rigid illusory contours and global shape transformations defined by spatiotemporal boundary formation. *In Press*

Keane, B., Erlikhman, G., Kastner, S., Paterno, D., & Silverstein, S. M. Multiple forms of contour grouping deficits in schizophrenia: What is the role of spatial frequency? *Under Review*

Kellman, P.J., Mnookin, J., Erlikhman, G., Garrigan, P., Ghose, T., Mettler, E., Charlton, D., & Dror, I. (2014). Predicting Fingerprint Match Difficulty Using Quantitative Image Measures. *PLoS ONE*, *9*(5)*:* e94617.

Kellman, P. J., Garrigan, P., & Erlikhman, G. (2013). Challenges in Understanding Visual Shape Perception and Representation: Bridging Subsymbolic and Symbolic Coding. In Sven Dickinson and Zygmunt Pizlo (Eds.), *Shape Perception in Human and Computer Vision: An Interdisciplinary Perspective*. 249-274.

Erlikhman, G., Keane, B., Mettler, E., Horowitz, T., & Kellman, P. J. (2013). Automatic Feature-Based Grouping During Multiple Object Tracking. *Journal of Experimental Psychology: Human Perception & Performance*, 39(6), 1625-1637

Polyn, S., Erlikhman, G., & Kahana, M. J. (2011), Semantic cuing and scale-insensitivity of recency and contiguity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 37(3), 766-775

Erlikhman, G., Caplovitz, G., & Kellman, P. J. (2014). Properties of spatiotemporal boundary formation. *VSS 2014*

Keane, B., Kastner, S., Paterno, D., Erlikhman, G., & Silverstein, S. M. (2014). Lateral interactions in schizophrenia: What is the role of spatial frequency? *VSS 2014*

Moses, R., Ghose, T., Erlikhman, G., & Kellman, P. J. (2014). Perceived occlusion velocity for fully visible and fragmented shapes. *VSS 2014*

Keane, B.P., Paterno, D., Suhail-Sindhu, T., Erlikhman, G., Kastner, S., & Silverstein, S.M. (2013). Kanizsa shape discrimination and contour integration deficits in schizophrenia: What is the role of spatial frequency? *Society for Research in Psychopathology*

Ghose, T., Erlikhman, G., Garrigan, P., Mnookin, J., Dror, I., Charleton, D., & Kellman, P.J. (2013). Perception, Image Processing and Fingerprint-Matching Expertise. *ECVP 2013*

Erlikhman, G., Ghose, T., Garrigan, P., Mnookin, J., Dror, I., Charleton, D., & Kellman, P.J. (2013). Fingerprint matching expertise and its determinants. *VSS 2013*

Caplovitz, G., Erlikhman, G., Lago, J., & Kellman, P. J. (2013). Neural Correlates of Spatiotemporal Boundary Formation (SBF). *VSS 2013*

Suhail-Sindhu T., Keane, B. P., Paterno, D., Erlikhman, G., Kastner, S., & Silverstein, S. M. (2013). Kanizsa shape discrimination and contour integration deficits in schizophrenia: What is the role of spatial frequency? *VSS 2013*

Erlikhman, G., Ghose, T., & Kellman, P.J. (2012) Contours and Surfaces Affect Stereoscopic Depth Perception in Dynamically Specified Displays. *VSS 2012*

Kellman, P.J., Erlikhman, G., Mansolf, M., Fillinich, R., & Iancu, A. (2012) Modeling Spatiotemporal Boundary Formation. *VSS 2012*

Mettler, E., Erlikhman, G., Keane, B., Horowitz, T., & Kellman, P.J. (2012) Further evidence for automatic, feature-based grouping in multiple object tracking. *VSS 2012*

Ghose, T., Erlikhman, G., & Kellman, P.J. (2011). Spatiotemporal Object Formation: contour vs. surface interpolation. *ECVP 2011*

Erlikhman, G., Ghose T., & Kellman, P.J. (2011). Spatio-Temporal Contour Interpolation in Four Dimensions. *VSS 2011*

Mettler E., Keane B. P., Erlikhman G., Horowitz T., & Kellman P.J. (2011). Automatic feature-based grouping during multiple object tracking. *VSS 2011*

Singh, G., Erlikhman, G., Ghose, T., & Liu, Z. (2011). Tilt aftereffects with orientations defined by motion or subjective contours. *VSS 2011*

Girshick, A., Burge, J., Erlikhman, G., & Banks, M. (2008). Prior expectations in slant perception: Has the visual system internalized natural scene geometry? *VSS 2008*

Chapter 1: Non-Rigid Illusory Contours and Global Shape Transformations Defined by Spatiotemporal Boundary Formation

Abstract

Spatiotemporal boundary formation (SBF) is the perception of form, global motion, and continuous boundaries from relations of discrete changes in local texture elements (Shipley & Kellman, 1994). In two experiments, small, circular texture elements underwent small displacements whenever an edge of an invisible (virtual) object passed over them. Unlike previous studies that examined only rigidly translating objects, we tested virtual objects whose properties changed continuously. Experiment 1 tested rigid objects that changed in orientation, scale, and velocity. Experiment 2 tested displays that transformed non-rigidly across a series of virtual object shapes. Results showed that robust SBF occurred for all of the rigid transformations tested, as well as for non-rigid virtual objects, producing the perception of continuously bounded, smoothly deforming shapes. These novel illusions involve perhaps the most extreme cases of visual perception of continuous boundaries and shape from minimal information. They show that SBF encompasses a wider range of illusory phenomena than previously understood, and they present substantial challenges for existing models of SBF.

Introduction

How do we perceive the boundaries of objects? This is, first of all, a question of what information is available in the optical input to the eyes. Often, objects differ from their backgrounds or other objects in surface characteristics; these differences produce discontinuities in luminance, color, or texture in their retinal projections. In ordinary environments (as opposed to pictures), there also tend to be depth discontinuities at object boundaries. These are manifest optically in stereoscopic disparities at boundaries as well as through changes in relative motion

of points along a boundary during object or observer motion.

In many situations, however, discontinuities in these stimulus properties are insufficient to reveal the complete boundaries of objects. Most pervasive are cases of occlusion, in which parts of an object's boundaries do not project to the eye due to a nearer, interposed object. Likewise, under conditions of camouflage, object surface properties may closely match properties of the background. Perception of complete objects in such cases depends on interpolation processes, as have been investigated in the perception of partially occluded and illusory contours and objects (e.g., Kanizsa, 1979; Michotte et al, 1954; Kellman, Garrigan & Shipley, 2005; for a review, see Shipley & Kellman, 2000). Experiments and models in this area have revealed a great deal about how the visual system goes beyond local visual information and uses spatial and temporal relations among physically specified edges to determine the occurrence and positions of interpolated edges.

These processes are used pervasively to overcome complex patterns of occlusion in in ordinary environments; yet perceiving object boundaries can be even more difficult. Suppose that *no* oriented edge fragments are visible. This can occur in camouflage, or more frequently, under dark viewing conditions, where a few sparse elements or features may be all that can be detected from the surfaces of objects or backgrounds. Gibson, Kaplan, Reynolds & Wheeler (1969) showed that even under such impoverished circumstances, objects with continuous boundaries can be perceived. Under conditions of relative motion of objects and observers, an object and its background undergo accretion and deletion of texture elements (Kaplan, 1969; Gibson, Kaplan, Reynolds & Wheeler, 1969). Accretion and deletion of even sparse texture elements on a farther surface by a nearer one can give rise to the perception of continuous boundaries, shape, and the relative depth of the two surfaces (Anderson & Cortese, 1989;

Braunstein, Andersen, & Riefer, 1982; Gibson, Kaplan, Reynolds, & Wheeler, 1969; Kaplan, 1969; Ono, Rogers, Ohmi, & Ono, 1989; Rogers & Graham, 1983; Yonas, Craton, & Thomspon, 1987).

Shipley & Kellman (1993, 1994, 1997) revisited accretion and deletion of texture and showed that it is just one example of transformations that can serve as the input to a more general process, which they called *spatiotemporal boundary formation* (SBF). They hypothesized that the crucial information for boundaries and shape in accretion and deletion is not the gradual covering or uncovering of texture elements, but the fact that those events are encoded as abrupt transformations (*spatiotemporal discontinuities*) in local visible elements. If this more general idea is correct, then transformations of other properties of local elements should also be capable of producing visual perception of continuous contours, shape, and relative motion. Their experiments revealed that discrete appearance and disappearance of texture elements, not just gradual covering or uncovering, produced SBF. Color change also produces SBF. Moreover, a whole range of ecologically bizarre transformations, including orientation change, position change (local element motion), and form change (of elements) also produce SBF. Figure 1.1 shows an example of SBF displays. All element transformations were unitary and discrete, meaning that they occurred instantaneously with no partial covering of the texture elements.

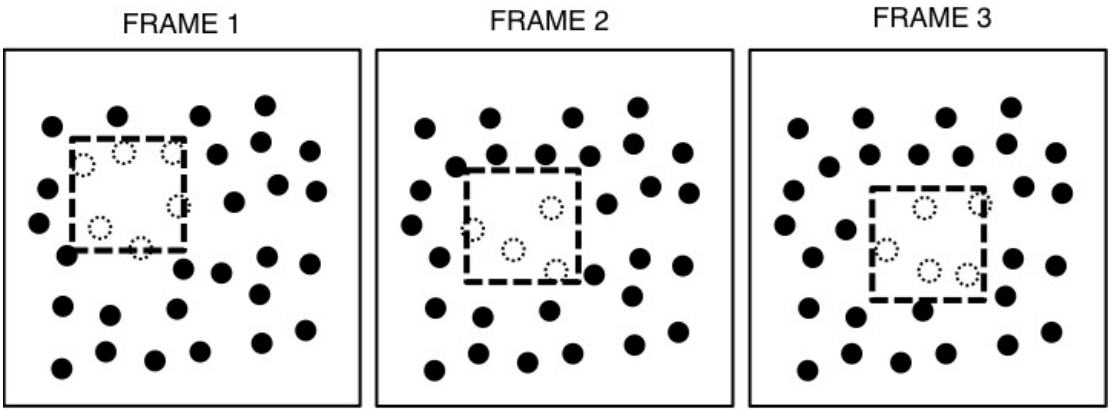

Figure 1.1. An example of three frames from a typical SBF display. The virtual object, a square, is indicated by the dashed line. Elements inside the object boundary are white, indicating that they share some surface property (e.g., color, orientation, shape), while those outside are black, having a different value for that property (e.g. red circles inside the boundary and green outside). As the virtual object moves, elements entering the boundary of the square become white and those exiting become black. Figure adapted from Figure 2 on p. 5 of Shipley T. F. & Kellman, P. J., (1994) Spatiotemporal boundary formation: Boundary, form, and motion perception from transformations of surface elements. *Journal of Experimental Psychology: General*, *123*(1), pp. 3-20.

As indicated in Figure 1.1, SBF occurs for both *unidirectional* and *bidirectional* transformations. In unidirectional displays, elements entering a specified virtual region all change their feature values in the same way. For example, in a unidirectional color change display, a virtual region moves among an array of white dots against a black background. Dots change from white to blue upon entering the virtual region and change back from blue to white upon exiting. Unidirectional color change displays have been extensively studied by Cicerone, Hoffman and colleagues, with an emphasis on perceived color spreading within such displays (Cicerone, Hoffman, Gowdy, & Kim, 1995; Cicerone & Hoffman, 1997; Fidopiastis et al., 2000; Miyahara & Cicerone, 1997; see also, Cicchini & Spillman, 2013; Gephstein & Kubovy, 2000). One feature of such displays is that in static views, a region corresponding to the virtual region (albeit with unclear boundaries) can be segregated from the background. In *bidirectional* displays, all texture elements are randomly assigned one of two feature values, so no such region is visible in any static view. Elements switch values when entering or exiting the virtual region. In a bidirectional color display, an array containing blue and white dots would be given, and when the virtual region passes over dots, the white dots turn blue and the blue dots turn white. Bidirectional displays also support SBF, producing the perception of continuous contours, complete shape, and global motion, but with no color spreading. The lack of uniform color across elements within the perceived shape's boundaries appears to prevent perceptual surface

formation. In SBF displays of this sort, ring-like objects with an empty interior region are seen. Besides homogeneous color, it appears that common motion of interior elements can also produce perception of a surface (Shipley & Kellman, 1994), as in classical accretion-deletion displays.

Ecology and SBF

Some transformations that produce SBF could arise from physical events involving real objects and surfaces. Ignoring for the moment the distinction between discrete and gradual element changes, both element disappearance and color change are ecologically natural events. As recognized in the pioneering work of Gibson, Kaplan, Reynolds & Wheeler (1969), relative motion between an opaque object and a textured background surface will lead to disappearance (covering) and reappearance of texture elements. Shipley & Kellman (1994) noticed that accretion-deletion displays are also consistent with an alternative, ecologically plausible interpretation. In a display with white visible elements and black surround, the white dots could be holes in a black surface, allowing visibility through the holes of a more distant white surface. The element change of disappearance in such a display could be occlusion of the white surface by a moving black object situated behind the front black surface and in front of the more distant white surface. This alternative physical arrangement, and the occasional appearance of SBF displays as containing an occluded object, has been discussed by Cicerone and Hoffman (1997) in the context of unidirectional color changes (color from motion). It corresponds to a similar parallel between ordinary partly occluded objects and illusory objects, which has been described previously (Kellman & Shipley, 1991). The reversibility of common illusory figures displays has been one of several bases for theories positing a common mechanism underlying illusory and occluded contour perception (Shipley & Kellman, 1992; see Kellman, Yin & Shipley, 1998 or

Kellman, Garrigan & Shipley, 2005 for more detailed discussion). Unidirectional color change displays (e.g., when elements change from white to blue upon entering the virtual region) are also ecologically plausible.[1] They could occur as modal completion if a blue, translucent filter, having the shape of the virtual region, passed in front of the array of elements. An ecologically plausible amodal version would occur under the circumstances described earlier for element disappearance, except that here a blue object, rather than a black one, is seen through holes.

In contrast to unidirectional color change, most element changes that support SBF could not arise ecologically from relative motion of an object and its background. For example, Shipley & Kellman (1994) showed that SBF is produced when an array consists of small horizontally and vertically oriented rectangles, and the element changes consisted of orientation changes, such that a moving virtual region cause vertical elements turn to horizontal and *vice versa*. Likewise, bidirectional color changes are not consistent with any surface moving over a background nor with an amodal version, in which a moving form is seen through apertures. Most relevant to the current studies is the fact that SBF can be produced from local element displacements (Shipley & Kellman, 1993, 1994). This is perhaps the most remarkable element transformation that evokes SBF, because the local motions used as the element transformation bear no relation to the global motion of the form that is seen. The form and global motion are resultants or spatial and temporal relations of element transformations, as in other SBF cases; the fact that the elements themselves move are incidental (see Movie 1.1 for an example). One might think that these local motions would degrade the information about shape in SBF displays, but this does not appear to be the case.

[1] Note that element disappearance is also an example of unidirectional color change; it is a special case in which the element changes to the background color.

Unlike kinetic or static illusory contours, or partly occluded contours, in SBF displays, no visible oriented contour fragments are required to produce contour and object perception. In fact, SBF occurs even in sparse displays, in which elements occupy as little as 1% or less of the total surface area. For these reasons, along with the fact that SBF occurs under many conditions that are ecologically bizarre, SBF might be considered the most extreme case of perceptual illusion in the construction of contours and objects. Perceived contours and shape are even more illusory than typical illusory contours, because so much is created from so little. Of course, as with most illusions, SBF may be seen as a byproduct of the mechanisms by which the visual system copes with the need to recover objects from information that is fragmented in space and time (Shipley & Kellman, 1994).

The aperture problem in SBF

In SBF, the only information available that can be used to recover moving contours are the positions and relative timing of abrupt element transformations. This presents a seemingly impossible version of the aperture problem. In the classic version (Wallach, 1935; Adelson & Movshon, 1982; Nakayama & Silverman, 1998a, 1998b; Shimojo, Silverman, & Nakayama, 1989), when an object's boundaries are seen through many small apertures, the visual system must determine the combined velocity structure of many spatially discrete, oriented contour segments whose global velocity is ambiguous. For each edge segment, its orientation and orthogonal motion velocity is available within the aperture. In SBF, the apertures are local elements that change discretely in some property. Unlike classic apertures, these changes by themselves do not produce any perception of a moving edge fragment. Moreover, a moving edge fragment in a classical aperture provides clear orientation information and constrains the directions of its movement to a 180 degree range. Individual element changes in SBF provide no

orientation information and no usable global motion information. Depending on the transformation used to produce SBF, there may indeed local orientation changes (when orientation change is used) or local motions (when element displacement is used), but these events not only provide zero information about a larger form and moving contours, they provide what would appear to be contradictory information. This more extreme version of the aperture problem in SBF has been referred to as the "point-aperture problem" (Prophet, Hoffman, & Cicerone, 2001).

One proposed solution relies on an intersection of constraints (Kellman, Erlikhman, Mansolf, Fillinich, & Iancu, 2013; Shipley & Kellman, 1994, 1997). Successive transformations of texture elements produce velocity signals that are determined by the spatial and temporal separation between transformation events. The velocity, orientation, and global motion direction of a region boundary are constrained by these signals. For example, consider several one-dimensional strips of evenly spaced texture elements at different orientations relative to the region boundary. Element transformations will be slowest along the strip that is orthogonal to the boundary, revealing the boundaries orientation. Given transformations along two strips (i.e., transformation of three non-collinear elements), both the velocity of the boundary and the local orientation can be recovered (see Shipley & Kellman, 1994, 1997 for details).

As in the case of the typical aperture problem, this model produces a coherent output only when several constraints are met. The texture element transformations are assumed to come from 1) a single, rigid entity that is 2) moving at a constant velocity. It is also assumed that the boundary can be decomposed into piecewise linear segments for which the orientation and velocity can be determined locally and independently. Such a model has been successfully

implemented for bar-like shapes whose boundaries have a single orientation and velocity along their length (Erlikhman et al. 2013).

Previous work has considered these constraints only in the case of unchanging shapes. Strong versions of the constraints underlying prior models would seem to imply that SBF should not occur for transforming shapes. For example, if a shape rotates, local edge orientations change continuously. If a shape scales, it also changes the local orientation of curved contours. If global motion of a shape accelerates, the assumption of constant velocity is violated. The models rely on the fact that fixed edge orientations and velocities produce specific spatiotemporal patterns of texture element transformations (which elements change and at what rate). If the pattern is constantly changing between element transformation events because edge orientation and velocity are changing, it would pose problems for current models in terms of recovering orientation and velocity of local edge fragments. In short, existing versions of SBF models, on the simplest account of their underlying assumptions, would work for a limited class of objects. Not coincidentally, these correspond to the objects that have been used in prior studies: rigid shapes moving with unchanging orientation and constant velocity.

Here we explore whether these limitations on SBF or its models may be arbitrary. In the real world, object motions are not limited to translation at constant velocity; objects rotate, accelerate, and scale (at least retinally). When objects rotate in depth, the retinal projection of their boundaries transform non-rigidly. In structure-from-motion (SFM) displays, we can readily see these as well as other non-rigid motions, such as the deformation of elastic objects or biological motion, even in sparse dot displays (e.g., Jain & Zaidi, 2011). Does SBF work with shapes of objects whose boundaries are changing in orientation or size be recovered? If illusory boundaries can be seen for SBF-defined objects that rotate, scale, transform non-rigidly, this may

force a reexamination of models of SBF. Perception of shape and illusory contours under these conditions would also provide the most spectacular versions of this class of visual illusion. Moving, deforming illusory contours would be seen between stationary texture elements in the absence of all local orientation and motion information.

We report two surprising visual illusions involving spatiotemporal boundary formation. In Experiment 1, SBF-defined illusory figures are seen that rotate, scale, and change velocity. Even though the displays contain only sparse texture elements such that no contour or shape information is available on any given frame, robust global form and motion is seen. In Experiment 2, observers were able to see non-rigid illusory contours produced by continuously deforming SBF-defined illusory figures. The displays demonstrate a new, easy way to create non-rigid illusory contours of arbitrary complexity.

Experiment 1

Experiment 1 used object transformations of rotation, scaling, and acceleration to test core assumptions about SBF. SBF is thought to arise from the integration of local motion signals across space and time. Shipley & Kellman (1997) distinguished form-precedes-motion models from motion-precedes-form models and found evidence for the latter in SBF. Their results suggested that pairs of discrete element changes proximate in space and time provide the input to SBF. If viewed in isolation, such pairs of element changes would produce perception of nearest neighbor apparent motion (Ullman, 1979), but in SBF, they do not result in perception of such motion (between two element changes). Rather, when two or more vectors produced by pairs of element changes are present within a certain spatiotemporal window, these are integrated to produce moving oriented contour fragments. At a higher level, perception of object shape and continuous boundaries in SBF appears to depend on spatiotemporal interpolation processes that

connect the edge fragments. Spatiotemporal contour interpolation, which has been studied in other contexts, relies on the updating of position information of contour segments that have disappeared based on a representation of their orientation and velocity. This persistence and positional updating of previously seen contour fragments in a temporary visual store allows such fragments to be integrated with contour segments that appear at a later time (Palmer, Shipley, & Kellman, 2006; Palmer & Kellman, 2014).

Existing models of SBF assume that local edge orientation and velocity are fixed within the integration window (Shipley & Kellman, 1994, 1997). Both the initial formation of edge fragments, and most, but not all, studies of spatiotemporal interpolation between edge fragments, have used contours with fixed orientations and velocities, and, as this would imply, rigid shapes of unchanging size and orientation. Experiment 1 tested whether SBF operates when these parameters change.

A secondary goal of the experiment was to determine whether element transformations consisting of motions in random directions could support SBF (see Figure 1.2). In previous work, consistent element motions (displacement in a uniform direction of all elements upon entering the virtual object) produced SBF (Shipley & Kellman, 1993, 1994). Preliminary work in our laboratory suggested that random element motion (consistent in extent but random in direction) could also support SBF, but no prior work has used these in SBF experiments.

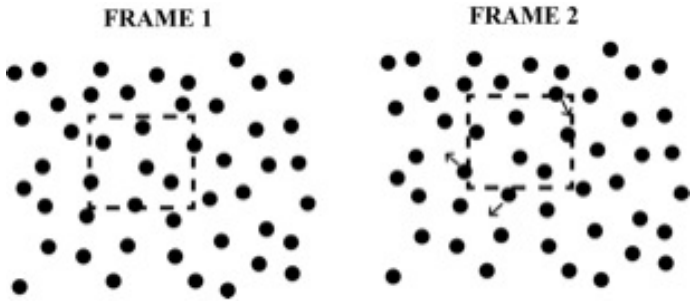

Figure 1.2. Example of the element transformation used in Experiments 1 and 2. The dashed region defines a "virtual object" which is not seen by the observer. As the virtual object moves,

elements that enter or leave the boundary of the virtual object are displaced in a random direction.

A virtue of using small, random element displacements as the inducing events in SBF displays is that no static view contains any information about a possible global shape in the display. Accurate shape perception in displays based on this transformation necessarily reflects the integrative SBF process that constructs global shape and continuous contours from series of element transformations integrated over time (Shipley & Kellman, 1994).

As in earlier research on SBF, we used a forced-choice shape identification paradigm. The paradigm is an objective performance method, in that there was an objectively correct answer (which virtual object was used in the display) on each trial. In the absence of global shape produced by SBF in displays such as those used here, consistently accurate shape perception is not possible (Shipley & Kellman, 1993).

Materials and Methods

Participants

Subjects were 16 undergraduate students (3 male, mean age: 19, range: 18-21) from the University of California, Los Angeles. All participants reported having normal or corrected-to-normal vision. Subjects received course credit for participating. Experiments were approved and conducted under the guidelines of the UCLA IRB. All subjects provided informed consent to participate.

Apparatus

All displays were created and displayed using the MATLAB programming language and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Stimuli were presented on a Viewsonic G250 CRT monitor, which was powered by a MacPro 4 with a 2.66 GHz Quad-Core Intel Xeon

processor and an NVidia GeForce GT120 graphics card. The monitor was set to a resolution of 1024x768 pixels and a refresh rate of 60 Hz.

Displays

Small red circles (diameter = 11.9 arcmin) were shown on a black background that filled the screen (40 cm x 30 cm; 25.06 deg x 18.92 deg). The total number of elements was either 200, 400, 600, or 1200. Elements were pseudo-randomly arranged by creating 100 equally sized regions and placing an equal number of elements in a random position within each region (see Shipley & Kellman, 1994). This minimized overlap between elements and ensured a nearly uniform distribution of elements across the display thereby also avoiding large, empty regions. The four element quantities corresponded to element densities of 0.42, 0.84, 1.27, and 2.53 elements per square degree of visual angle respectively. Elements covered 1.28%, 2.56%, 3.83%, or 7.67% of the pixels in the display area.

We defined ten virtual objects or "pseudosurfaces" similar to those used in Shipley and Kellman (1993)[2]. They are depicted in Figure 1.3. We will refer to these as virtual objects or virtual regions, while noting that they were referred to as "pseudosurfaces" in earlier work. Either label is intended to convey that the shapes are not physical entities; any static frame of the display is seen to contain only a field of undifferentiated texture elements. The shapes had varying degrees of symmetry and regularity. The virtual objects were on average 5.6 degrees of visual angle in height and width, within a range of 4.36 to 6.45 degrees in either dimension. When a virtual object came into contact with an element, the element was displaced by 10 pixels (14.9 arcmin) in a random direction (see Figure 1.2). The displacements were large enough to be

[2] In the original stimulus set, two of the "random" shapes (corresponding to the bottom row in Figure 1.2) were rotated or mirror reflections of one shape. Because shapes rotated in some conditions in this experiment, we generated new shapes that were complex and shared several features with other shapes, but were not confusable when rotated.

readily detectible (Shaffer & Wallach, 1966; Shipley & Kellman, 1993). When the element's original position was no longer within the boundary of the virtual objects, the element returned to that position. An element was defined as inside the virtual object if its center was on or inside of the virtual object boundary. On average, 1.74, 3.44, 5.27, and 10.51 elements transformed from frame-to-frame for each of the four element quantities respectively. An example of a scaling and rotating shape is shown in Movie 1.2.

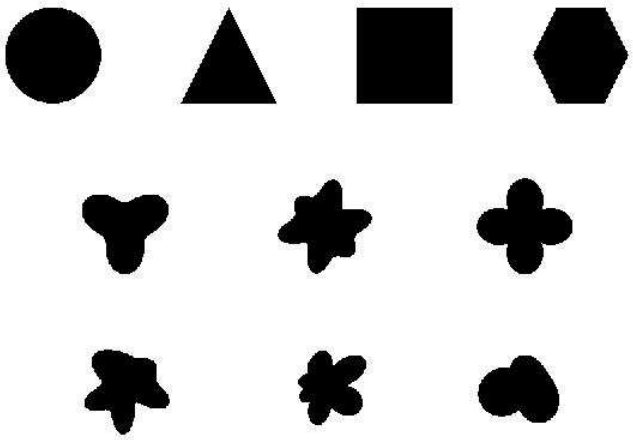

Figure 1.3. Ten shapes used in Experiments 1 and 2. The top four shapes are familiar, regular, and have multiple axes of symmetry. The second row contain shapes that are more unusual, but still symmetrical. The final row contains asymmetrical shapes. All shapes have approximately the same horizontal and vertical extent. They are modeled after the shapes used in Shipley & Kellman (1994). Throughout the text we refer to them as circle, triangle, square, hexagon, tri-leaf, butterfly, four-leaf, rand1, rand2, and rand3 starting from the top-left and going to the bottom-right.

Virtual objects traveled on a circular path centered on the middle of the screen, with a radius of 4.97 degrees of visual angle. The path was divided into 360 equidistant positions and the virtual object visited them sequentially. The virtual object traveled at a rate of four positions every frame (0.35 degrees per frame) and each frame was shown for 32.2 ms. A trial was complete when the virtual object made one complete circuit of the path. The starting position along the path was randomized across trials. A trial lasted 3 seconds.

As virtual objects traveled along the path, they underwent one of four possible transformations: scaling, rotation, rotation and scaling, or acceleration. In the scaling and rotation and scaling conditions, virtual objects increased or decreased in size at a rate of 1% per frame. The maximum size of a virtual object was 9.92 degrees and the minimum size was 2.49 degrees in any dimension. Upon reaching the size limit, scaling direction reversed. Initial scaling direction (shrinking or growing) was randomized across trials. If the virtual object was rotating, it rotated at a rate of 3 degrees per frame in the clockwise direction. Starting orientation of the shape was always upright. In the scaling and rotation condition, both of the transformations were applied simultaneously. In the acceleration condition, on each frame, there was a 30% probability of the velocity increasing, a 30% probability of the velocity decreasing and a 40% of the velocity remaining constant. Velocity changes affected which of the 360 positions along the path the virtual object would visit. Minimum velocity was the two positions per frame (compared to a base velocity of four) and maximum velocity was seven positions per frame. Constantly increasing velocity for the duration of the movie resulted in a final speed that too fast to follow and caused the trial to terminate very quickly.

Design

On each trial, participants performed a forced choice selection of the shape in the displays from among a fixed set of 10 alternatives. The four texture element quantities, the four shape transformation conditions, and the ten shapes were counterbalanced in a 4 x 4 x 10 design. Each trial was repeated twice, resulting in a total of 320 trials. Trial order was randomized. Prior to the experimental trials, there were 10 practice trials. Each practice trial had the highest density of elements and no shape transformation. Each of the ten shapes was shown once, in random order. The entire experiment lasted approximately 40 minutes.

Procedure

Subjects sat in a dark room at a distance of 90 cm from the computer monitor, with the only illumination coming from the monitor. They were given verbal and written instructions explaining that they were going to see a black screen with red dots in which an illusory shape would appear to move. Their task was to identify the shape that they had seen out of a set of ten possible shapes. Subjects then began the practice trials. At the start of each trial, a white fixation cross appeared in the middle of a black screen for 1 second. Then, the cross disappeared and the red texture elements were shown. The virtual object began to move as soon as the elements appeared. Once the object completed a full path around the screen, a new display with an image of the ten shapes was shown. Subjects made a response by clicking on one of the ten shapes with the mouse. A red, rectangular box appeared around the answer choice for 1.5 seconds to indicate the subject's response. For practice trials, feedback was provided by showing a green, rectangular, box around the correct choice. If the subject had selected the correct response, the green box surrounded the red one. In addition, the word "Correct" or "Incorrect" appeared in the top-left corner of the screen. Subjects had unlimited time to make a response. Once the practice trials were over, a message appeared on the screen, instructing subjects that the practice trials were over and that they would no longer receive any feedback.

Results

Mean accuracy data for Experiment 1 are shown in Figure 1.4. Highly accurate shape perception was possible under some of the conditions of the experiment, especially at the highest element density, and all conditions appeared to exceed chance accuracy. These observations were confirmed by the analyses. Accuracy data were collapsed across shapes and submitted to a 4 x 4 within-subjects ANOVA. There was a main effect of transformation type ($F(3,45)=90.18$,

p<0.001, η^2_p = 0.86), with highest accuracy for scaling shapes, followed by scaling and rotating shapes, rotating shapes, and accelerating shapes across most element quantities. There was a main effect of number of elements (F(3,45)=349.36, p<0.001, η^2_p = 0.959), with accuracy improving with an increasing number of elements. There was also a significant interaction (F(9,135)=2.70, p=0.006, η^2_p = 0.15).

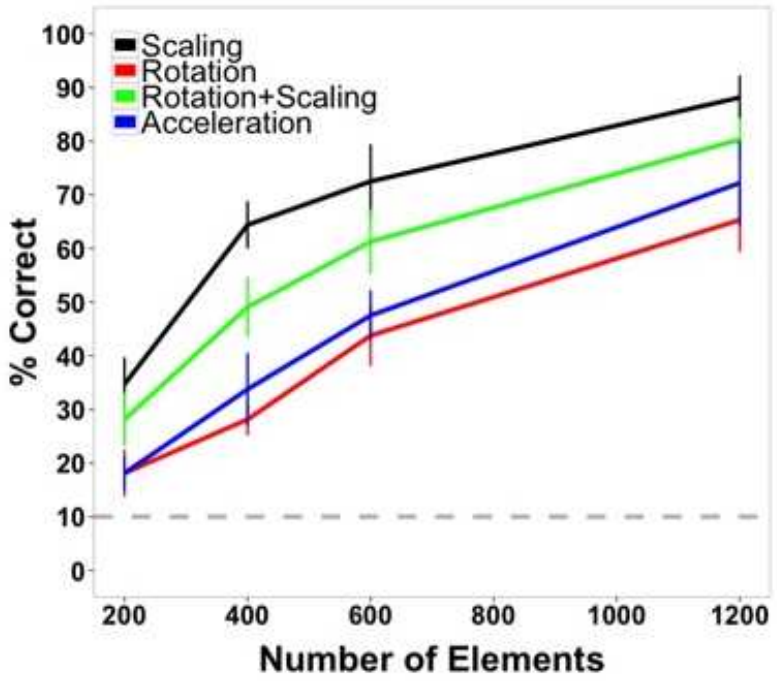

Figure 1.4. Average accuracy data from Experiment 1 as a function of number of texture elements in the display. Data are averaged across subjects and shapes. All transformation conditions were within-subject. Error bars indicate 95% confidence intervals. The gray, dashed line indicates chance performance (10%).

The highest accuracy was observed for the scaling condition with the largest element quantity (88.13%). Performance for this shape transformation at this number of elements was greater than all of the other transformation conditions (rotation: 65.31%, t(15)=6.80, p<0.0001 ; scaling + rotation: 80.31%, t(15)=2.74, p=0.015 ; acceleration: 72.19%, t(15)=5.08, p<0.001). For the lowest number of elements tested (200), performance in all conditions was above chance

(10%) (scaling: 34.69%, $t(15)=9.64$, $p<0.0001$; rotation: 18.12%, $t(15)=3.64$, $p<0.005$; rotation + scaling: 28.13%, $t(15)=7.15$, $p<0.0001$; acceleration: 18.12%, $t(15)=5.17$, $p<0.001$).

Discussion

The results of Experiment 1 show that boundaries and accurate shape can be perceived from SBF displays in which the virtual objects change orientation, scale and velocity. Moreover, these illusory figures were seen despite the transformations being displacements of individual texture elements in random directions, thereby producing incoherent local motion signals. Accurate perception of shape and the subjective appearance of continuous illusory contours bounding shapes illustrates the extreme nature of interpolation processes in SBF. Texture element transformations were spatiotemporally quite sparse in this study. For the 200, 400, 600 and 1200 element displays, there were on average about 1.6, 3.2, 5 and 10 element transformations per frame, respectively. These were spread along an average boundary length of the shapes of 17.6 degrees of visual angle. Shapes and illusory contours perceived in SBF, including the transforming shapes in this study, represent perhaps the most extreme illusion among illusory contours, in terms of spatial support. Even when a number of accumulated frames are considered together, there amount of total boundary specified by local stimulus information is a very small percentage. Displays perceived in SBF represent, in an important sense, the most perceived boundaries and shape perceived from the least stimulus input.

It is surprising that transforming virtual objects produced such robust SBF, given that changing orientation of edges in a neighborhood, changing stimulus velocity, and the changing size of virtual objects should all complicate recovery of the local edge fragments hypothesized to give rise to SBF. Transforming virtual objects create two confounding problems for an intersection of constraints solution to this aperture problem. First, in between successive

transformations of local texture elements, the edges that caused those transformations are changing not only in their position, but also in their orientation and velocity. Since several transformation events are needed constrain the orientation and velocity of an edge, it is unclear how the visual system can relate two or more events caused by, essentially, two different edges. Second, once local edge segments are recovered, they must be interpolated since they are recovered piecemeal, in different positions along the virtual object boundary and at different times. We return to these issues in the General Discussion.

Shape identification was affected predictably by the manipulation of element quantity, improving as a function of the number of elements. Performance was best for scaling and the combination of scaling and rotation shape transformations. Accuracy may have been better in those conditions because as objects become larger, more texture element transformation events occur along the virtual object boundary. For the highest density, there were, on average, 4.64 transformations per frame when a virtual object reached its smallest size (compared to an average 10.51 transformations per frame for rotating objects that did not change size) and 18.68 transformations per frame when the virtual object reached was largest. However, this increase in element changes scales directly with figure size, such that the number of element changes per unit of perimeter remains constant. Perhaps a more plausible account of improved performance with larger sized objects is that size may make differences between similar shapes larger and more discriminable. For example, at the highest density in the scaling condition, circles were never confused for hexagons or *vice versa*, but they were confused 17 times across all subjects at that element density when the objects were rotating without scaling.

The pattern of results in this experiment, with element position changes in random directions, was similar to experiments in which element transformations were position shifts in

only one or two directions and when virtual objects were rigid and not transforming (Shipley & Kellman, 1993, 1994). In those studies, performance also increased as a function of element quantity. Since different numbers of elements were used across studies, converting the independent variable to element density per degree of visual angle allows a standard metric for comparison. We take up these comparisons after considering the results of Experiment 2 below.

Experiment 2

The results of Experiment 1 suggest that shapes that rotate, scale, and accelerate can be accurately perceived in SBF. In Experiment 2, we further examined what kinds of global shape transformations are supported by SBF. Changes in orientation, scale, and velocity are rigid transformations of the virtual object. Perhaps non-rigid transformations can also be perceived. In these displays, virtual objects smoothly morphed from one of the ten shapes used in Experiment 1 into another. Morphing continued from shape to shape until all shapes were seen. Subjects were instructed to look for a target shape (say, the triangle) in the morphing sequence and to indicate when they saw that shape (see Movie 1.3).

If non-rigid illusory contours are seen in these displays, this presents a much more confounding problem for spatiotemporal interpolation. In addition to the difficulty in matching texture element transformation events with contours that are changing in position and orientation, the visual system must now deal with changes in contour curvature as the shape is morphing. Supposing that local edge segments can be somehow recovered even though the curvature of those segments changes in between transformation events, the segments must then be interpolated. While it has been demonstrated that contour fragments that change in orientation under occlusion can be interpolated with visible ones, it is not known whether contours

segments, real or illusory, can undergo changes in curvature while not visible and still be interpolated with other contour segments that are later revealed.

Materials and Methods

Participants

The participant group was composed of 12 University of California, Los Angeles undergraduate students (10 female, mean age = 22.75). All participants reported having either normal or corrected-to-normal vision. Participants were awarded course credit for their participation. Experiments were approved and conducted under the guidelines of the UCLA IRB. All subjects provided informed consent to participate.

Displays and Apparatus

Since the lowest element densities in Experiment 1 made shape identification difficult when the shape was not changing, higher element quantities were used to ensure that performance was not at floor. The three element quantities used were 529, 900, and 1600. In order to accommodate the larger number of elements on the screen, texture element diameter was reduced to 7 arcmin for a viewing distance of 134.5 cm. The element quantities corresponded to densities of 2.46, 4.18, and 7.43 elements per square degree of visual angle. Elements covered 2.62%, 4.47%, and 7.95% of the total display area.

The same shapes were used as in the first experiment. Average virtual object diameter was 4.45 degrees. The smallest size was 3.35 degrees and largest was 5.03 degrees. On average, there were 3.8, 6.47, and 11.47 element transformations per frame for each of the three element quantities respectively. As in Experiment 1, the virtual object traveled along a circular centered on the middle of the display. The radius of the path was 3.33 degrees. The path was divided into 120 equidistant positions. The distance between each position was 0.17 degrees. The virtual

object visited one position per frame. Each frame lasted for 33.2 ms. It took an object 4 seconds to make a full revolution. Starting position along the path was randomized across trials.

On each trial, the virtual object began as one of the ten shapes and smoothly morphed from one shape to another until it had become each of the ten shapes once. Shape morphing was performed by selecting 120 equally spaced points along the contour of each shape to use as reference points. A morphing algorithm generated 99 intermediate shapes between every pairing of shapes by creating matches between the nearest contour points of the two shapes and interpolating intermediate locations. In total, there were 90 such morphing sequences, one between each pair of shapes. The first and last steps of the morphing sequence were the original, un-morphed shapes. Each intermediate morphing step therefore reflected the relative proportion of the two shapes that were being morphed. For example, on the 31st step in the morphing sequence between shape A and B, the shape was 69% shape A and 31% shape B. The entire transformation sequence from one shape to another took approximately 3.3 seconds.

The transformation sequences on each trial involved nine transformations between the ten shapes. The order of shapes in the transformation sequence was randomized on each trial with the constraint that the first and last shapes could not be the target shape. Each trial lasted a maximum of 30 seconds. Each shape served as the target shape twice for each density, resulting in a total of 60 trials. Trial order was randomized. As in Experiment 1, there were 10 practice trials to help familiarize the participants with the task. Each of the ten shapes was the target for one of the practice trials. The highest density backgrounds were used for all practice trials. The entire experiment lasted approximately 30 minutes.

Design and Procedure

Participants were informed that the purpose of the study was to examine the perception of changing visual illusions. The stimulus was described as a morphing shape that would result from a pattern of flickering dots on the screen. At the beginning of each trial, the participant was presented with a target shape selected from one of the ten possible shapes. After a key press, the textured background appeared and the animation began. The participant was instructed to press a key when they believed the virtual object on the screen most closely resembled the target shape. The display was terminated immediately once the participant pressed the key. If no response was given during the course of the animation sequence, the trial was repeated (same target shape), but with a different shape transformation sequence. Subjects were instructed to try to make a response on the second or third viewing of a trial, and to avoid repeating a trial more often.

The first ten trials of the experiment were practice trials at the highest density. Each of the ten shapes was the target shape on one of the ten trials. Feedback was provided on the screen after every practice trial ("Correct" or "Incorrect"). Once the practice trials were over, the subject was informed via instructions on the screen that they would no longer receive feedback and that the number of texture elements would vary across trials.

Dependent Measures and Data Analysis

A response was scored as correct if it was made while the virtual object on the screen was a morph of 50% or more of the target shape. This occurred as one shape morphed into the target shape or as the target shape began morphing into another. Since each frame corresponded to a 1% morphing of the virtual object, the range within which a response was scored as correct was 50 frames on either side of the frame that contained a 100% morph of target shape.

The exact frame on which a response was recorded presumably includes time for response initiation and execution (i.e., response time). We applied a correction to account for the

delay between when a relevant perceptual event caused an observer to initiate a response and when a subsequent key press was recorded. For example, a response time correction that corresponded to 30 frames would mean that if an observer initiated a response when the virtual object was a 50% morph of the target shape, then the recorded response would occur 30 frames later, when the object was an 80% morph. Likewise, a recorded response when the object was a 60% morph of the target shape would actually correspond to a response initiation 30 frames earlier, when the object was only a 30% morph.

We defined the frame that contained the 100% morph of the target shape as the *target frame*, the frame on which a key press was recorded as the *response frame*, and the frame on which the response was initiated as the *decision frame*. The response time was defined as the difference between the response frame and the decision frame. Applying a response time correction shifted the center of the window within which a response was considered correct forward in time. With no correction, the window would be centered on the target frame and would span 50 frames on either side. A 30-frame correction would shift the window forward by 30 frames so that correct responses would be those response frames that occur between 20 frames before or 80 frames after the target frame.

We considered all integer response time corrections between 0 and 50 frames. For each correction, we determined the window within which responses were correct and computed the average accuracy across all subjects and conditions. The response time correction that resulted in highest average accuracy was 12-15 frames (all times in that range produced the same accuracy). Those numbers of frames corresponded to times of 398.4 - 498 ms, which roughly agree with response times from object priming studies (Vorberg, Mattler, Heinecke, Schmidt, & Schwarzbach 2003), recognition memory (Sternberg, 1969), and RSVP paradigms (Botella,

1992). The difference in average accuracy with and without the correction was less than 1% and all subsequent analyses were no different whether the correction was applied or not.

Results

Mean accuracy data for Experiment 2 are shown in Figure 1.5. It can be seen that successful shape identification occurred well above chance performance throughout, reaching very high accuracies at the highest element density. Accuracy data were collapsed across shapes and submitted to a one-way, within-subjects ANOVA. There was a significant main effect of density ($F(2,22)=19.38$, $p<0.001$, $\eta^2_p = 0.64$. Pairwise comparisons between the three densities revealed that accuracy at the highest density was greater than at the other two densities (high vs. medium: $t(11)=2.55$, $p=0.027$; high vs. low: $t(11)=5.43$, $p<0.001$) and that accuracy at the medium density was greater than at lowest density (medium vs. low: $t(11)=4.31$, $p=0.001$). These and the results that follow were the same for analyses without the response time correction.

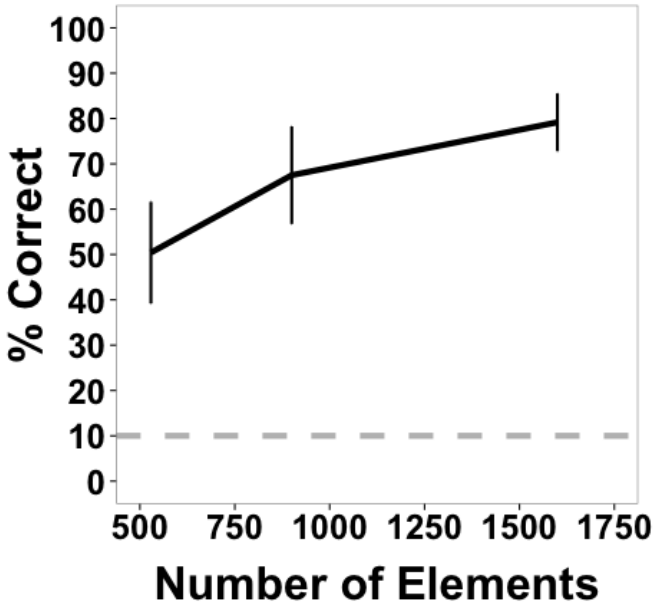

Figure 1.5. Average accuracy data from Experiment 2 as a function of number of texture elements in the display. Data are averaged across subjects and shapes. Error bars indicate 95% confidence intervals. The dotted gray bar indicates chance performance (10%).

Accuracy data were also examined separately for each shape across the three element quantities (Figure 1.6). Data were collapsed across subjects since each target shape was repeated only twice per subject. Low, medium, and high in the figure legend correspond to the three element quantities (529, 900, and 1600 elements). Identification accuracy was perfect for triangles and hexagons for the largest element quantities and exceeded 90% for squares, the quad-leaf shape, and the shape Rand3. Worst performance for any element quantity was for the shape Rand1 (25%). Worst performance at the largest quantity was for shape Rand2 (41.67%). Chance performance was 10%. Sensitivity (*d'*) for each shape is shown in Figure 1.7. False alarms were counted as those trials in which a subject responded with any shape other than the target shape. As with accuracy, sensitivity was computed from data collected from all subjects. Sensitivity was highest for triangles (4.65), squares (3.82), and hexagons (4.68) for the highest density, and was relatively high for circles (3.15), quad-leaf (3.11), and Rand3 (3.35). Sensitivity decreased with decreasing element density.

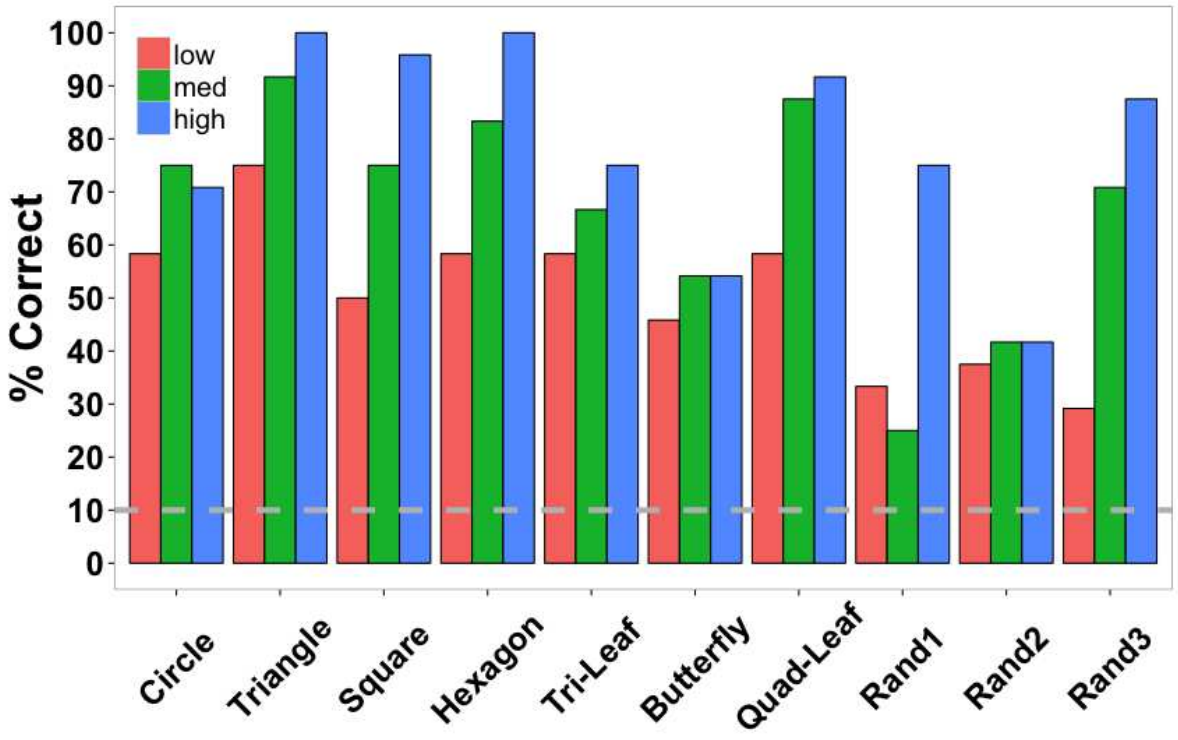

Figure 1.6. Shape identification accuracy in Experiment 2 separated by shape and element quantity (low = 529, medium = 900, and high = 1600 elements) and collapsed across subjects. The dashed gray line indicates chance performance. Shape names correspond to the shapes shown in Figure 1.2 starting at the top-left corner of the figure and proceeding left-to-right and top-to-bottom.

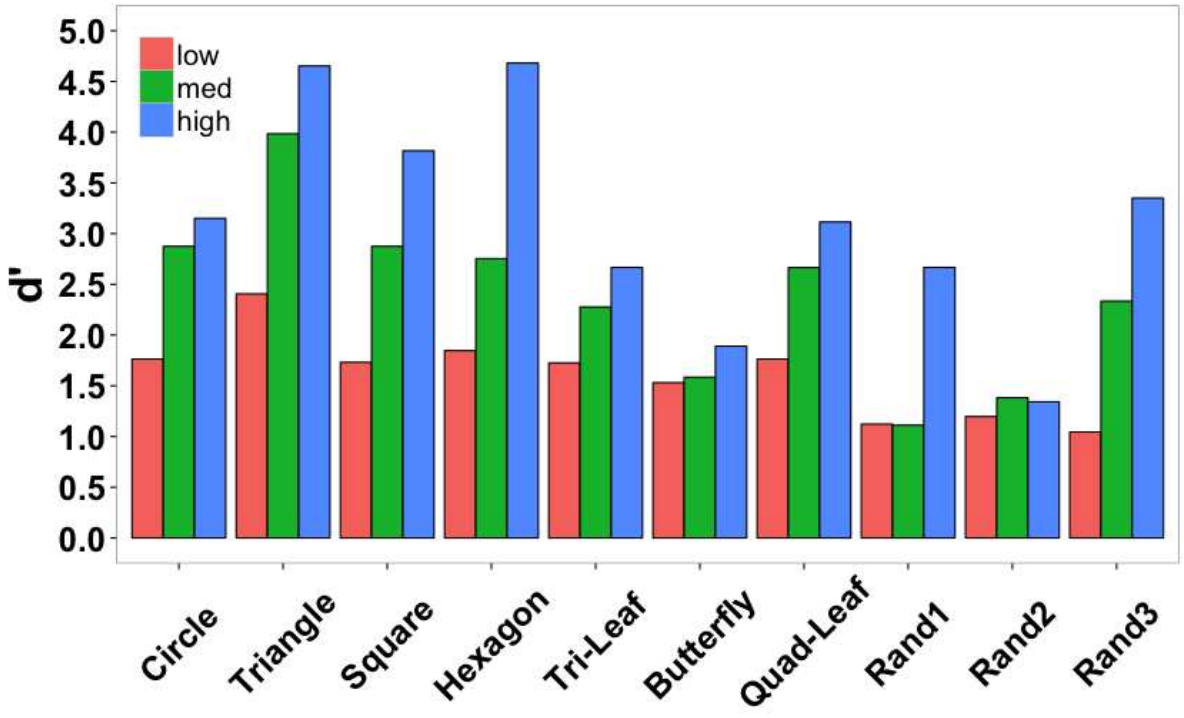

Figure 1.7. Shape identification sensitivity (*d'*) in Experiment 2 separated by shape and density (low = 529, medium = 900, and high = 1600 elements) and collapsed across subjects.

A secondary analysis examined the degree to which the virtual object on the screen resembled the target shape on the decision frame (response time corrected). Recall that subjects were instructed to respond as close as possible to the target frame (the frame containing the 100% morph of the target shape). Looking only at trials in which subjects made a correct response, the number of frames between the target frame and the decision frame is a measure of the extent to which the virtual object resembled the target shape. Because there were 100 frames between the target shape and the subsequent shape in the transformation sequence, a decision on the 16th frame after the target frame would indicate that the shape on the screen was an 84% (100-16) morph of the target shape. Likewise, a decision 16 frames before the target frame would also contain an 84% morph of the virtual object. Results were not significantly different if the response time correct was not applied.

Figure 1.8 shows the percentage of target shape on the decision frame (response frame - 15 frames) averaged across subjects as a function of element quantity. A one-way, within subjects ANOVA found a significant main effect of density ($F(2,22) = 5.65$, $p=0.010$, $\eta^2_p=0.34$). Post hoc, between-density comparisons revealed that the percentage of target shape on the decision frame for the highest density (84.96% target shape) was significantly greater than the percentage for the lowest density (79.31%; $t(11)=3.17$, $p=0.009$). No other differences were significant. As before, these results were the same when the response time correction was not applied.

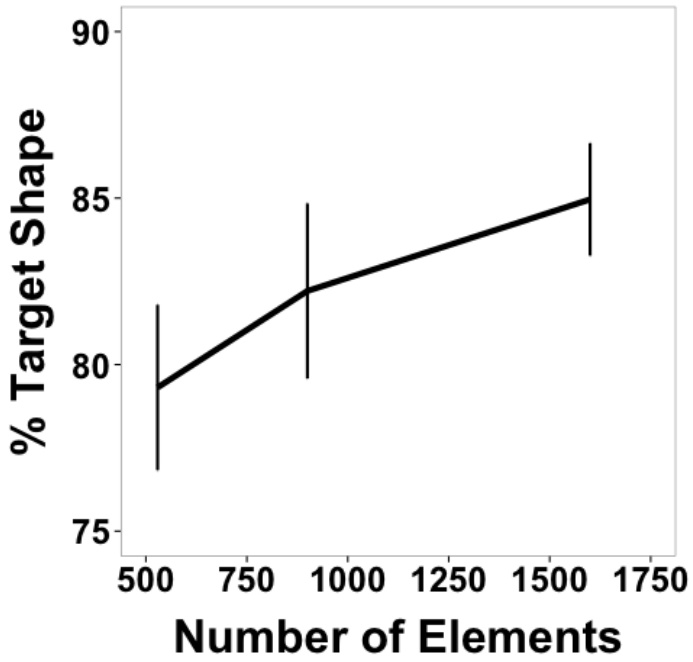

Figure 1.8. Percentage of morph between target shape and another shape when subjects initiated a response (response time corrected, see text) as a function of element quantity. Subjects were instructed to make a response when the figure on the screen matched as closely as possible the target shape. Values closer to 100% indicate greater response precision. Data are shown for correct trials only.

We further explored the data by distinguishing between decisions that came before the target frame and those that came after. The data are shown in Figure 1.9. A 2 x 3, within-subjects ANOVA found a significant main effect of decision time (before vs. after), ($F(1,10)=10.20$, $p=0.010$, $\eta^2_p=0.50$) and of element quantity ($F(2,20)=10.79$, $p=0.001$, $\eta^2_p=0.52$). There was also a significant interaction ($F(2,20)=4.52$, $p=0.024$, $\eta^2_p=0.31$). Post-hoc paired comparisons for percentage of target shape before and after the target frame revealed a difference for displays that contained the largest number of elements ($t(11)=4.17$, $p=0.002$). There were no significant differences for the two other element quantities.

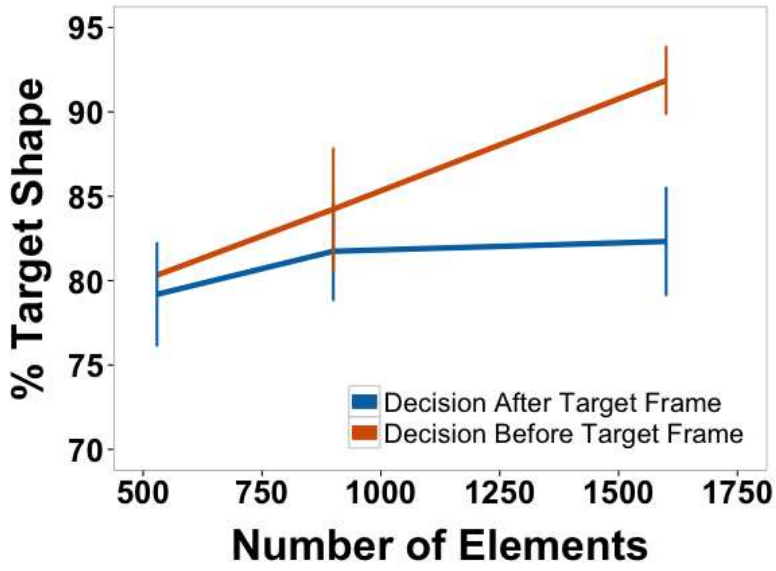

Figure 1.9. Percentage of morph between target shape and another shape separated by whether the response came before or after the frame on which the pure target shape was presented (response time corrected, see text). Data are for correct trials only.

Discussion

The results of Experiment 2 demonstrate that illusory contours can be accurately perceived in SBF displays even when those boundaries are smoothly deforming. As in Experiment 1, this has two implications for the visual processes involved in perceiving boundaries in these displays. First, local boundary segments can change not only in orientation and velocity, but also in curvature in between the texture element transformation events that define them. Second, interpolation between boundary segments occurs even when one segment continues to deform but is now invisible. Since events do not occur continuously along the entire boundary of the virtual object, they reveal only parts of the boundary at any given time. As transformation events reveal parts of the boundary, those newly visible regions interpolate with previously seen but now invisible ones. This process suggests a form of representation that encodes constructed edge fragments as well as their continuing trajectories and deformation, allowing such information to be preserved and updated for combination with later appearing information (c.f. Palmer, Kellman & Shipley, 2006; Palmer & Kellman, 2014) With the virtual

object morphing from frame-to-frame, the boundary is deforming non-rigidly. If the visual system encodes an orientation, velocity, and curvature of a boundary segment at one moment as fixed values, those features may not align with a segment recovered at a later time. We return to this possibility in the General Discussion.

Comparing the two experiments, the best performance (across all shapes) in Experiment 2 (79.58%) was within the range of best performances from the four conditions in Experiment 1 (65.31% - 88.13%). However, element density had to be doubled to 4.18 elements per square degree of visual angle before this level of performance was achieved. One reason for this difference could be because virtual objects were smaller in Experiment 2 (average diameter of 4.45 deg) than in Experiment 1 (diameter 5.6 deg). Because element size was smaller in Experiment 2, the total number of element transformations per frame was similar for the two largest densities in each experiment (12.85 in Experiment 1 and 11.47 in Experiment 2). Alternatively, a greater element density may have been needed in Experiment 2 to reach comparable performance because the task was harder. Responses were marked as correct only if they fell within 1.66 seconds of the target frame, whereas there was no response time limit in Experiment 1. In addition, some intermediate morphing stages may have appeared to be similar to other shapes. For example, morphing between a square and a circle may have resulted in intermediate morphs that resembled hexagons.

With the results of Experiment 2 in hand, we compared shape identification accuracy with the transforming and non-rigid virtual objects in Experiments 1 and 2 with shape identification accuracy in earlier work. Shipley & Kellman (1994, Experiment 3) used rigid, non-transforming shapes and local motion as the element transformation in a 10-AFC task. There were some differences from the present experiments. As mentioned earlier, we used a somewhat

revised set of figure choices here. Moreover, we used random directions of element motion, whereas the earlier study used consistent vertical displacements. The virtual objects used in the current experiment were also larger (4.45 degrees of visual of angle in diameter vs. 2 degrees). Although comparisons are inexact, they may be informative with regard to the primary purpose of the present work: to determine whether SBF occurs robustly for transforming shapes. The data are clear in showing the SBF occurs with transforming objects, but if SBF occurs from transforming objects but is notably weaker than in non-transforming shapes at comparable element densities, it would suggest that changing orientation, shape, or velocity do impact the recovery of shape in SBF.

Figure 1.10 plots the data from the two current experiments along with the earlier experiment with all conditions being displayed in terms of element density (elements/deg^2). As can be seen, performance at comparable densities for deforming shapes in Experiment 2 was comparable to that of rigid, non-transforming shapes in Experiment 3 of Shipley & Kellman (1994). The four densities used in Shipley and Kellman (1994) were 1.61, 3.21, 6.42, and 12.85 elements per degree of visual angle. (Performance was not significantly different for the two largest densities and density did not exceed 6.42 in the current experiment, so accuracy for only the first three densities is shown.) The densities used in the present Experiment 2 were 2.46, 4.18, and 7.43 elements per square degree of visual angle.

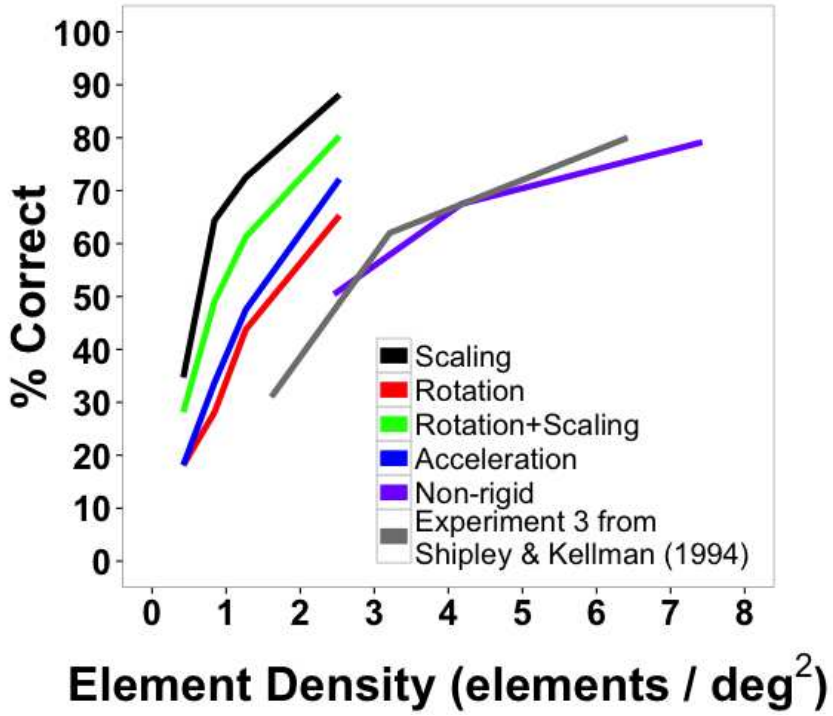

Figure 1.10. Average shape identification accuracy from Experiments 1 (black, red, green, and blue lines) and 2 (purple line, "non-rigid") plotted as a function of element density. Also plotted are reproduced data from Experiment 3 from Shipley and Kellman (1994) in gray.

Figure 1.10 also plots the results of the transforming, rigid shapes of Experiment 1 as a function of element density. Remarkably, all of these conditions produced *better* shape identification performance than occurred with non-transforming shapes in the earlier work. The densities used in our Experiment 1 were 0.42, 0.84, 1.27, and 2.53 elements per square degree of visual angle.

For all comparable element densities, accuracy was higher in the current experiment with transforming but rigid shapes than for non-transforming ones in earlier work. Even when densities were three times larger than those used in the current study, identification performance for non-transforming virtual objects reached 80%, while identification accuracy for scaling virtual objects reached 88%. This difference may be because virtual objects in the current study were more than twice as large (average diameter = 5.6 deg) as those used previously (2.0 deg). That larger shapes produce better shape identification is not entire intuitive. For displays with the

same density, the number of element changes in a unit of time per unit of perimeter remains constant for a large and small display of the same shape. Also, larger shapes would tend involve more of the retina outside of the fovea, with some attendant loss of visual acuity. It may, however, be the case that larger shapes make clearer a shape's parts and relations. We noted in the discussion of Experiment 1 that best performance observed in that study occurred in the scaling conditions, which included presentation of the largest shapes in the experiment. As suggested there, the exact reason for better performance with larger shapes in SBF is not entirely clear, but one plausible hypothesis is that larger visual angles allow better definition and discrimination of shape parts, resulting in improved discrimination.

It is clear from the data that identification accuracy varied depending on shape complexity and confusability (Figure 1.6). Sensitivity was greatest for triangles, squares, hexagons, quad-leafs, and shape Rand3 than for other shapes (Figure 1.7). Shape Rand2 appeared to be difficult to identify irrespective of the number of elements. For the lowest and intermediate element quantities, sensitivity was lowest for the shapes tri-leaf, butterfly, quad-leaf, Rand1 and Rand2. These shapes share in common regions of high curvature. Such regions require more proximal events to clearly specify the boundary. For a long, low curvature segment of a boundary, two edge fragments that are far apart would still be relatable (Kellman & Shipley, 1991). For high curvature regions, the segments would need to be from relatively nearby positions on the boundary to be relatable. Sparse texture displays would yield few recovered boundary segments.

In addition to improving accuracy, element density (or quantity, as these covaried in this study) was directly proportional to response precision. Subjects tended to respond on frames closer to the target frame (the one which contained the target shape) as texture element quantity

increased (Figure 1.8). Since the task instructions specified that subjects should respond as close as possible to the target frame, responses frames that contained shapes more closely morphed to the target shapes can be interpreted as more precise responses. Precision may have improved as a function of element quantity because subjects could more readily predict when the morphing sequence was approaching the target shape, or it could have improved because once the target frame was reached, subjects were quicker to identify the shape and respond. In order to distinguish between these two possibilities, data were split by whether responses came before or after the target frame (Figure 1.9). Responses after the target frame did not depend on element quantity. However, the more elements there were on the screen, the more precisely subjects could anticipate when the target frame was approaching.

General Discussion

Spatiotemporal boundary formation is known to produce form, continuous boundaries, and global motion from discrete transformations of sparse textural elements. The use of spatially separated, sequential changes in small, sparse elements to produce these perceptual outcomes comprises an amazing spatiotemporal integration capacity of the visual system. SBF occurs despite a lack of information that can define contours or surfaces in a single, static view or even across many frames. Within frames, in the displays used here, there are no differences between figure and ground in luminance, texture, or depth in any frame. Moreover, in terms of motion perception, every local stimulus event in the displays used here is fully accounted for as a single dot traveling a short distance in some direction. Indeed, viewed singly, that is the appearance of each element transformation. In terms of the literal stimulus events (discrete changes in local elements, with no local or momentary information for continuous boundaries or surfaces), SBF represents a startling illusion.

As with many illusions, SBF presents dual implications regarding the utility and function of perceptual processes. The events most likely to trigger SBF in real world situations are motions of objects that are poorly specified, either because of matching object and background surfaces (camouflage) or because an object is seen through multiple apertures. In these situations, SBF recovers whole objects accurately from minute bits of information spread across time. As Gibson, Kaplan et al. suggested in describing accretion and deletion of texture elements (the best known case of SBF), transformations produced by relative motion of objects and surfaces carry important information deeply rooted in the optics of opaque objects, depth order, and motion. In the ways we are most likely to encounter SBF in ordinary viewing environments, SBF is a highly ecological and sophisticated mechanism for detecting what is really occurring in the environment.

But as with the processes underlying many other illusions, SBF turns out to accomplish its ecologically relevant tasks by means of mechanisms that in other cases produce ecologically impertinent outcomes. Accretion and deletion is a fact about ecological optics, but when we ask how the visual system accesses that fact, it turns out to use discrete changes in local elements – virtually *any* detectable discrete changes. This is both more and less than the original idea that the visual system detects accretion and deletion. It is much more because virtually any element transformation can provide an input into SBF, even ecologically bizarre ones such as orientation change or local displacement of an element. It is *less* than accretion and deletion because elements need not be gradually covered nor must there be any array of texture elements that move together (as will always be present during relative motion in accretion and deletion displays).

When non-ecological element transformations are used, illusory contours and shapes are perceived that cannot arise from any known physics (apart from CRT displays and clever programmers). When for example, white and blue texture elements on a black background switch values upon entering or leaving a defined, moving, virtual region, the array of changes could not be caused by any moving translucent filter nor any movement of an object seen through apertures in an occluder. The fact of a bizarre illusion, here as in other illusion contexts, lays bare the functioning of the visual processes involved. The visual processing that apprehends objects passing in front of each other from sparse information also puts together illusory shapes from abrupt changes of other kinds, such as the objects formed from local, random direction element displacements in the experiments here.

In Experiment 1, we found that the orientations, sizes, and velocities of virtual object boundaries can change between successive transformation events and still be continuously seen. In Experiment 2, SBF was also found to support changes in boundary curvature, giving rise to robust percepts of non-rigid, illusory contours. Both experiments used displays in which boundaries were perceived without accompanying filling-in or surface completion suggesting that the two processes are separable and can be studied independently. The methods described can be readily adapted to generate dynamic, non-rigid, illusory contours with arbitrary form and complexity.

Implications for models of SBF

Perception of continuous boundaries and shape in SBF appears to depend on two processing stages. The first is to recover local edge segments from sparse texture transformation events and the second is to interpolate (connect) these segments to produce a representation of continuous contours and object shape. We consider each problem in turn.

The recovery of edges from transformation events is a difficult version of the aperture problem. Typically, local edge orientations and velocities are available in many small apertures, and the problem is to determine how they are connected and the global motion signal. In SBF, there is no local orientation information available since the apertures are points. The difficulty is compounded by the fact that in the displays used in these experiments, texture element transformations were element displacements in random directions, generating irrelevant and incoherent local apparent motion signals that were completely independent from the global motion of the virtual object. The relevant information for defining the virtual object boundary was solely the position and timing of transformation events. Despite these difficulties, it is possible to solve this point-aperture problem by assuming that contour segments of the virtual object boundary are rigid, moving at a constant velocity and not changing their orientation.

Experiments 1 and 2 demonstrated that contours can change in these properties and still support the perception of global shape and motion. Does this invalidate existing models? The answer is that models may need to be modified but that the underlying concepts may survive. Theoretically, a local edge orientation in SBF can be recovered from three non-collinear element transformations in some local neighborhood (Shipley & Kellman, 1997). The aperture problem in SBF may get solved many times, relatively quickly, and in relatively small regions. Thus, an object may not have to be rigid or otherwise unchanging for initial edge segments to be constructed.

If the texture is sufficiently dense, the aperture problem can be solved multiple times in a small spatiotemporal window, resulting in several oriented edge representations over time. These will be small illusory contour segments. Straight or curved apparent motion might then be seen between successively recovered illusory segments that are proximal in space and time. In effect,

once the aperture problem is solved for a local segment, the problem becomes a matter of detecting correspondences between sequentially recovered segments. There is no reason to suspect that this would be any different for real or illusory contours: whatever the solution to the correspondence problem that allows the matching of real, contours across success frames can be applied to rotating illusory contours in SBF.

Difficulty will arise when texture displays are very sparse. In order to solve the aperture problem, multiple transformation events are needed; if the contour transforms too much between the events, then the solution might not be correct. This could explain why SBF deteriorates with decreasing element density.

Once the aperture problem is solved locally in order to recover a segment of the boundary, these boundary segments must be interpolated to produce a representation of the global shape. Since element transformation events are spatiotemporally sparse, boundary segments are recovered piecemeal, in different regions and times. This leads to the second level of processing in SBF: interpolation connecting basic contour fragments that have been formed. Because these do not appear simultaneously or in register spatially, the visual system therefore needs a way of encoding recovered segments, and storing and updating their representations to be interpolated with segments that are recovered at a later time. Such spatiotemporal interpolation has been found with real edge fragments in rigidly translating (Palmer, Shipley, & Kellman, 2006; Palmer & Kellman, 2014) and rotating, luminance-defined edges (Kellman & Cohen, 1984), but not yet for illusory contours and not for non-rigidly deforming shapes.

According to models of spatiotemporal interpolation, when a part of an object becomes occluded, a visual, iconic representation of that surface continuous to persist for a brief time (Palmer, Shipley, & Kellman, 2006). That icon is an encoding of the position, orientation, and

velocity of the surface contours. If another part of the object is revealed (appears from occlusion), the visual system interpolates the relatable contours of the visual icon with those of the newly revealed object part. Interpolation is possible because the representation of the position of the occluded segment (the visual icon) is updated under occlusion (for a short time).

The visual system faces the same problem in SBF displays: since the aperture problem is solved locally for different areas along the virtual object boundary, edges are not recovered all at once. It is as if parts of the boundary become disoccluded whenever the problem is solved, and are occluded otherwise. The visual system must then interpolate between recovered edge segments that are visible only for short periods of time. One possibility is that the representation of occluded edges is very flexible and capable of both first and second order deformations. For rotating shapes, for example, when transformations along one part of the virtual object boundary reveal an edge segment, the representation of the position and orientation of that segment continue to change even when there are no further transformations to support its perception. When the aperture problem is then solved again in a nearby position, the resulting segment is interpolated with the shifted and rotated representation of the past segment if the two are relatable. A second possibility is that the representation of the segment remains fixed in terms of orientation and curvature at the moment of occlusion. A snapshot is taken, and it can only be minimally manipulated. When the next segment is recovered, the two segments must fall within the range of relatability (Kellman & Shipley, 1991) in order to be interpolated. Further studies are needed to distinguish between these two possibilities.

The present studies show that SBF encompasses a wider range of illusory phenomena than previously realized. Scaling and rotating, even accelerating rigid shapes can be recovered in SBF. Even more remarkable, deforming shapes can be perceived, and recognition of a shape is

possible even when it is part of a rapidly changing series of shapes. These phenomena clearly expand the envelope beyond what previous models anticipate or explain. Although we sketched an outline of how more advanced models might encompass these perceptual illusions, the current results raise more questions than they answer, and further research will be required to achieve a detailed understanding of these amazing phenomena in which the visual system does so much with so little.

References

Adelson, E. H., & Movshon, J. T. (1982). Phenomenal coherence of moving visual patterns. *Nature, 300*, pp. 523-525.

Andersen, G. J., & Cortese, J. M. (1989). 2-D contour perception from kinematic occlusion. *Perception and Psychophysics, 46*, pp.49-55.

Anstis, S. M. & Atkinson, J. (1967). Distortions in moving figures viewed through a stationary slit. *The American Journal of Psychology, 80*, pp.572-585.

Aydin, M., Herzog, M. H., & Öğmen, H. (2008). Perceived speed differences explain apparent compression in split viewing. *Vision Research, 48*(15), pp. 1603-1612.

Barraza, J. F., & Chen, V. J. (2006). Vernier acuity of illusory contours defined by motion. *Journal of Vision, 6*, 923-932.

Botella, J. (1992). Target-specific and target-categorized conditions in RSVP tasks as reflected by detection time. *Bulletin of the Psychonomic Society, 30*(3), p.197-200.

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

Braunstein, M. L., Andersen, G. J., & Riefer, D. M. (1982). The use of occlusion to resolve ambiguity in parallel projections. *Perception & Psychophysics, 31*, pp. 261-267.

Bruno, N., & Bertamini, M. (1990). Identifying contours from occlusion events. *Perception & Psychophysics, 48*, pp. 331-342.

Bruno, N., & Gerbino, W. (1991). Illusory figures based on local kinematics. *Perception, 20*, pp. 259-274.

Chen, V. J., & Cicerone, C. M. (2002). Subjective color from apparent motion. *Journal of Vision, 2*(6), pp. 424-437.

Chen, V. J., & Cicerone, C. M. (2002). Depth from subjective color and apparent motion. *Vision Research, 42*, pp. 2131-2135.

Cicchini, M., & Spillmann, L. (2013). Neon color spreading in dynamic displays: Temporal factors. *Journal of Vision, 13*(12), 1-9.

Cicerone, C. M., & Hoffman, D. D. (1997). Color from motion: Dichoptic activation and a possible role in breaking camouflage. *Perception, 26*, pp. 1367-1380.

Cicerone, C. M., Hoffman, D. D., Gowdy, P. D., & Kim, J. S. (1995). The perception of color from motion. *Perception & Psychophysics, 57*, pp. 761-777.

Cooke, T., Cunningham, D. W., & Bulthoff, H. H. (2004). The perceptual influence of spatiotemporal noise on the reconstruction of shape from dynamic occlusion. *Lecture Notes in Computer Science, 3175*, pp. 407-414.

Cunningham, D. W., Shipley, T. F., & Kellman, P. J. (1998). The dynamic specification of surfaces and boundaries. *Perception, 27*(4), pp. 403-415.
Cunningham, D. W., Shipley, T. F., & Kellman, P. J. (1998). Interactions between spatial and spatiotemporal information in spatiotemporal boundary formation. *Perception & Psychophysics*, *60*(5), pp. 839-851.
De Weerd, P., Vandenbussche, E., & Orban, G. A. (1993). Occlusion cues contribute to orientation judgments of occlusion-defined contours. *Perception & Psychophysics*, *54*(6), pp. 706-715.
Fidopiastis, C., Hoffman, D. D., Prophet, W. D., & Singh, M. (2000). Constructing surfaces and contours in displays of color from motion: The role of nearest neighbors and maximal disks. *Perception, 29*, pp. 567-580.
Gepshtein, S., & Kubovy, M. (2000). The emergence of visual objects in space-time. *Proceedings of the National Academy of Sciences*, *97*(14), pp. 8186-8191.
Gibson, J. J., Kaplan, G. A., Reynolds, H. N., & Wheeler, K. (1969). The change from visible to invisible: A study of optical transitions. *Perception and Psychophysics*, *3*, pp. 113-116.
Jain, A. & Zaidi, Q. (2011). Discerning nonrigid 3D shapes from motion cues. *Proceedings of the Natural Academy of Sciences*, *108*(4), pp. 1663-1668.
Kaplan, G. A. (1969). Kinetic disruption of optical texture: The perception of depth at an edge. *Perception & Psychophysics*, *6*, pp. 193-198.
Kellman, P. J., & Cohen, M. H. (1984). Kinetic subjective contours. *Perception & Psychophysics*, *35*(3), pp. 237-244.
Kellman, P. J., Erlikhman, G., Mansolf, M., Fillinich, R., & Iancu, A. (2012). Modeling spatiotemporal boundary formation. *Journal of Vision*, *12*(9). [Abstract.]
Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, *23*, 141-221.
Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), pp. 2729-2737.
Larsson, J., Heeger, D. J., & Landy, M. S. (2010). Orientation selectivity of motion-boundary responses in human visual cortex. *Journal of Neurophysiology*, *194*(6), pp. 2940-2950.
Lorenceau, J., & Boucart, M. (1995). Effects of a Static Textured Background on Motion Integration. *Vision Research*, *35*(16), pp. 2303-2314.
Lui, L. L., Dobiecki, A. E., Bourne, J. A., & Rosa, M. G. P. (2012). Breaking camouflage: responses of neurons in the middle temporal area to stimuli defined by coherent motion. *European Journal of Neuroscience*, *36*(1), pp. 2063-2076.
Mateeff, S., Popov, D., & Hohnsbein, J. (1993). Multi-aperture viewing: perception of figures through apertures. *Vision Research*, *33*(17), pp. 2563-2567.
Miyahara, E., & Cicerone, C. M. (1997). Color from motion: separate contributions of chromaticity and luminance. *Perception*, *26*(11), pp. 1381-1396.
Mysore, S. G., Vogels, R., Raiguel, S. E. & Orban, G. A (2008). Shape selectivity for camouflage-breaking dynamic stimuli in dorsal V4 neurons. *Cerebral Cortex*, *18*(6), pp.1429-1443.
Nakayama, K., & Silverman, G. H. (1988a). The aperture problem I: Perception of non-rigidity and motion direction in translating sinusoidal lines. *Vision Research*, *28*, pp. 739-746.
Nakayama, K, & Silverman, G. H. (1998b). The aperture problem II: Spatial integration of velocity information along contours. *Vision Research*, *28*, pp. 747-753.s

Nawrot, M., Shannon, E., & Rizzo, M. (1996). The relative efficacy of cues for two-dimensional shape perception. *Vision Research*, *36*(8), pp. 1141-1152.
Ono, H., Rogers, B. J., Ohmi, M., & Ono, M. (1989). Dynamic occlusion and motion parallax in depth perception. *Perception*, *17*, pp. 255-266.
Palmer, E. M., & Kellman, P. J. (2014). The aperture capture illusion: Misperceived forms in dynamic occlusion displays. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), pp. 502-524.
Palmer, E. M., Kellman, P. J., & Shipley, T. F. (2006). A theory of dynamic occluded and illusory object perception. *Journal of Experimental Psychology: General*, 135(4), 513-541.
Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437-442.
Poom, L. (2001a). Visual inter-attribute contour completion. *Perception*, *30*, pp. 855-865.
Poom, L. (2001b). Visual summation of luminance lines and illusory contours induced by pictorial, motion, and disparity cues. *Vision Resarch*, *41*(28), pp. 3805-3816.
Prazdny, K. (1986). Illusory contours from inducers defined solely by spatiotemporal correlation. *Perception & Psychophysics*, *39*, 175-178.
Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. http://www.palamedestoolbox.org
Prophet, W. D., Hoffman, D. D., & Cicerone, C. M. (2001). "Contours from apparent motion: A computational theory" in *From fragment to objects: Segmentation and grouping in vision*. P. Kellman and T. Shipley (Eds), Amsterdam: Elsevier Science Press, pp. 509-530.
Rogers, B. J., & Graham, M.E. (1983). Dynamic occlusion in the perception of depth structure. *Perception*, *12*, A15.
Rovden, C. S., Baker, J. F., & Allman, J. (1988). Perceptions of depth elicited by occluded and shearing motions of random dots. *Perception*, *17*, pp. 289-269.
Sáry, G., Vogels, R., & Orban, G. A. (1994). Orientation discrimination of motion-defined gratings. *Vision Research*, *34*(10), pp. 1331-1334.
Schmid, A. M. (2008). The processing of feature discontinuities for different cue types in primary visual cortex. *Brain Research, 1238*(31), pp. 59-74.
Shaffer, O., & Wallach, H. (1966). Extent-of-motion thresholds under subject0relative and object-relative conditions. *Perception & Psychophysics*, *1*, 447-451.
Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research, 29*(5), pp. 619-626.
Shipley, T. F., & Kellman, P. J. (1993). Optical tearing in spatiotemporal boundary formation: When do local element motions produce boundaries, form, and global motion? *Spatial Vision, 7*(3), pp. 323-339.
Shipley, T. F., & Kellman, P. J. (1994). Saptiotemporal boundary formation: Boundary, form, and motion perception from transformations of surface elements. *Journal of Experimental Psychology: General*, *123*, pp. 3-20.
Shipley, T. F., & Kellman, P. J. (1997). Spatio-temporal boundary formation: the role of local motion signals in boundary perception. *Vision Research*, *27*(10), pp. 1281-1293.
Sinha, P. (2001). Role of motion integration in contour perception. *Vision Research, 41*(6), pp. 705-710.

Stappers, P. J. (1989). Forms can be recognized from dynamic occlusion alone. *Perceptual and Motor Skills*, *68*, pp.243-251.
Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, *57*(4), p. 421-457.
Todd, J. T. (1984). The perception of three-dimensional structure from rigid and nonrigid motion. *Perception & Psychophysics*, *36*(2), pp. 97-103.
Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences, USA*, *100*(10), pp.6275-7280.
Wallach, H. (1935). Ueber visuell wahrgenommene Bewegungsrichtung. *Psychologische Forschung*, *20*, pp. 325-380.
Yonas, A., Craton, L. G., & Thompson, W. B. (1987). Relative motion: Kinetic information for the order of depth at an edge. *Perception & Psychophysics*, *41*, pp. 53-59.

Chapter 2: Recovery of Local Edge Fragments Initiates Spatiotemporal Boundary Formation

Abstract

Spatiotemporal boundary formation (SBF) is the perception of illusory boundaries, global form, and global motion from spatially and temporally sparse transformations of texture elements (Shipley & Kellman, 1993a, 1994). It has been theorized that the positions and times of texture element transformations can be used to extract the orientation of local edge fragments, which form the basic shape units in SBF. To test this theory, we created a novel display consisting of a sawtooth arrangement of circular elements that disappeared and reappeared sequentially, one at a time. Within the appropriate display settings, the resulting percept was not of apparent motion between elements, but of a larger oriented edge fragment that traveled laterally across the display. Experiment 1 identified the spatial and temporal intervals within which SBF occurred using a contour clarity rating task. Experiment 2 extended and refined the temporal limits using an objective performance task, in which the perceived widths of moving bars in SBF could be compared to the virtual objects used to generate the visible element changes. The two experiments, using perceptual reports and accuracy in size perception, converged in revealing highly constrained spatial and temporal parameters under which SBF occurs. The experiments provide clear support for models of SBF that begin with extraction of local edge fragments and identify minimal conditions required for this process.

Introduction

Spatiotemporal boundary formation (SBF) is the perception of continuous contours, global form, and global motion from the sequential transformation of sparse texture elements

(Shipley & Kellman, 1993a, 1994). SBF is perhaps the most extreme case in which the visual system constructs contours and objects from fragmentary input, as it requires no oriented edge fragments and produces complete perceived boundaries with little stimulus support. For example, the concept of support ratio (the ratio of illusory or occluded edge length to total edge length) has been shown to predict interpolation strength of edge fragments for occluded and illusory contours (Banton & Levi, 1992; Shipley & Kellman, 1992). Roughly speaking, robust contour interpolation occurs with support ratios of .5 or greater, and noticeable interpolation may still be present at support ratios of .2 or .3. In typical examples of SBF, with widely spaced, small background elements, support ratio would be very close to zero, yet robust perception of continuous contours and clear overall shape are present (Shipley & Kellman, 1994). Functionally, SBF may exist as a visual mechanism for apprehending objects under conditions of minimal information, as when the surface properties of two objects are similar or difficult to discriminate, or when an object is viewed under complex occlusion situations (e.g., through foliage), or under dim viewing conditions when surface features are difficult to resolve.

How are shapes seen in SBF? One hypothesis is that SBF first defines local edge fragments from transformations of texture elements (e.g. occlusion) that occur closely together in space and time. These edges fragments, once created, connect across gaps to form concrete objects. Shipley & Kellman (1994, 1997) have proposed a model of how edge orientation can be computed from the positions of elements, the distances between them, and the temporal interval between their transformations. If the model is accurate, it should be possible to create displays in which only a single, illusory edge fragment is seen, and its perception should be constrained by the spatial and temporal intervals between texture element transformations. In this paper, we describe a novel display that comprises or approaches the minimal conditions for the occurrence

of SBF. The display can be parametrically varied to find the range of spatiotemporal texture element transformation intervals that support SBF under these minimal conditions. The present experiments provide more precise information relevant to modeling SBF than has previously been available, and they also allow for a more ready comparison of SBF to other, well-characterized visual phenomena in which elements transform successively, such as apparent motion.

Surprisingly, relatively little is known about spatiotemporal constraints on SBF. Prior work has exclusively focused on two-dimensional virtual shapes that travel in an array of randomly arranged elements (Shipley & Kellman, 1993a, 1993b, 1994, 1997; Cunningham, Shipley & Kellman, 1998; Fidopiastis et al. 2000). As a result, the spatial and temporal properties of the displays, in particular, inter-element distance and the time between element transformations are highly variable depending on both the shape and the distribution of elements. In a typical display, small, uniform texture elements, usually circles, are distributed randomly in a large area (Figure 2.1). A virtual object is specified that moves across the texture field. The object is virtual in the sense that its boundaries are not defined by luminance differences with the background. As the virtual object moves, elements that were previously inside the boundary may now be outside and *vice versa*. For ease of description, we refer to this as elements exiting or entering the virtual object region, even though it is usually the region that is moving while the elements are stationary. Elements that move across the boundary change in one of their properties. For example, a red element inside the object boundary may become red when the boundary moves and no longer contains that element. This produces a pattern of element transformations along the boundary of the object as it moves across the display. Surprisingly, clear, continuous illusory contours are seen that correspond to the virtual object's boundary.

These boundaries are seen even if fewer than one element on average transforms per frame and when elements cover only a small portion (e.g., 1.9 %) of the entire display area (Shipley & Kellman, 1994). When transformations are color changes, as in the example above, color spreading is also seen within the boundary of the object and the entire region may be perceived as a transparent surface moving in front of a textured background. This phenomenon has been studied separately under the name “color from motion” or “dynamic color spreading” (Cicerone, Hoffman, Gowdy, & Kim, 1995; Cicerone & Hoffman, 1997; Miyahara & Cicerone, 1997). However, illusory contours can be seen in the absence of color spreading. Shipley & Kellman (1994) showed that a wide range of local element transformations can support SBF, including changes in element position, shape, and orientation. A number of transformations they studied, including bidirectional color and orientation changes, and local element motion, provide no information in any static frame about an approximate region in which a form may reside.

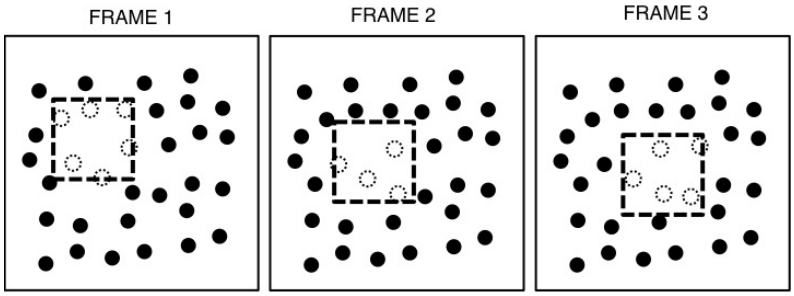

Figure 2.1. An example of three frames from a typical SBF display. The virtual object, a square, is indicated by the dashed line. Elements inside the object boundary are white, indicating that they share some surface property (e.g., color, orientation, shape), while those outside are black, having a different value for that property (e.g. red circles inside the boundary and green outside). As the virtual object moves, elements entering the boundary of the square become white and those exiting become black. Figure adapted from Figure 2 on p. 5 of Shipley T. F. & Kellman, P. J., (1994) Spatiotemporal boundary formation: Boundary, form, and motion perception from transformations of surface elements. *Journal of Experimental Psychology: General*, *123*(1), pp. 3-20.

The perception of illusory contours in these displays is known to depend on several factors. First, element transformations must be encoded as discontinuities or disruptions in the regular pattern of the textured background (Shipley & Kellman, 1993a). If changes are detectable, but small (e.g., displacement over a short distance), then no illusory shape is seen. At frame durations longer than 165 ms, illusory contours, global form, and global motion are no longer seen (Shipley & Kellman, 1994). One instead sees apparent motion between nearby transforming elements. However, because texture element transformations may occur in any region of the boundary from frame to frame, it is unclear from this result what the temporal integration limits are for local edge fragments (i.e., for spatially proximal texture elements that whose transformations define a local contour). Third, SBF also degrades with texture density. However, manipulating texture density confounds spatial and temporal distances between transformation events as well as the total number of transformations that occur. Density also interacts with contour complexity. Spatially sparse texture will result in fewer interactions with the virtual object border. Regions of high curvature that fall in between places on the border that interacted with an element have no way of being recovered. A study in which the arrangement of texture elements was not random, also found that boundaries were more clearly seen for higher densities in which there were smaller gaps between elements (Fidopiastis et al. 2000).

SBF displays based on random arrangements of texture elements and regular or arbitrary two-dimensional virtual objects produce perception of continuous contours completely bounding a region, and clearly defining a shape. However, these displays make it difficult to isolate the effects of inter-element distance and inter-transformation time and to evaluate models of the process. We wondered what are the minimal conditions to elicit the key phenomenon in SBF – the construction of a contour. To that end, we created novel displays in which elements were

arranged in a sawtooth pattern and disappeared one at a time sequentially. This allowed us to independently manipulate the absolute spatial and temporal distances between element transformation events, to restrict the number of events that occurred per frame, and enabled us to identify the spatiotemporal limits within which integration occurs that leads to SBF.

Experiment 1

A display was created in which a sawtooth arrangement of dots was presented, and one dot at a time flashed off and then on again in sequence (Figure 2.2). What would we expect to perceive in such a display? For many spatial and temporal intervals, the laws governing apparent motion should cause us to see a single entity (a white dot or blob) moving along the sawtooth pattern (Wertheimer, 1912; Korte, 1915; Ekroll, Faul, & Golz, 2008). Indeed, this is exactly what is seen in Movie 1. In this movie, all elements are visible for fourteen frames (140 ms); one element disappears for four frames (40 ms); all elements are again visible for fourteen frames (140 ms), etc. These settings are well within the range when apparent or phi-motion should be seen (Wertheimer, 1912; Steinman, Pizlo, & Pizlo, 2000; Ekoll, Faul, & Golz, 2008).

In a two-element apparent motion display, the temporal interval between the appearance of one element and the appearance of the next is the stimulus onset asynchrony (SOA), the time that an element is visible is the stimulus duration, and the interval during which no elements are visible is the interstimulus interval (ISI). Considering the first two elements in Figure 2.2, the SOA is the interval between the appearance of the first element and the appearance of the second (180 ms). The stimulus duration is length of time an element is visible before it disappears (140 ms + 40 ms + 140 ms = 320 ms). Finally, since the elements are never invisible simultaneously, ISI is negative (-140 ms), indicating that the stimulus duration is longer than the SOA. The absolute value of the ISI therefore corresponds to the interval when all elements are visible.

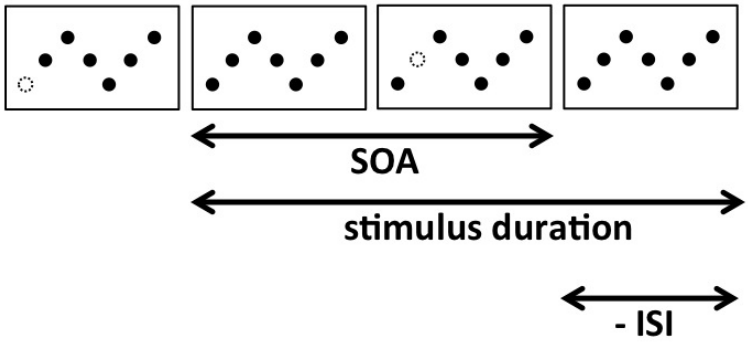

Figure 2.2. Four apparent motion frames in which elements disappear one at a time in sequence. Consider the first two elements in the display as an apparent motion pair. Stimulus onset asynchrony (SOA) is the temporal interval between the appearance of the first element and the appearance of the second. Stimulus duration is the time one of the two elements is visible. ISI is negative because there are no frames when both elements are invisible.

A simpler way to characterize the display in terms of these intervals is to consider the disappearance of a black dot to be the appearance of a white dot in that position. In that case, the appearance of a white, invisible element at each element position is the stimulus relative to which all time intervals are measured (see Figure 2.3). Consequently, the duration that an element is occluded becomes the stimulus or flash duration (f), the temporal interval between the appearances of white elements (flashes) becomes the ISI, and the total time between the disappearance of one element and the disappearance of the next is the SOA. This parametrization is more intuitive in that it avoids negative values for ISI and captures the idea that an element transformation event is the stimulus of interest.

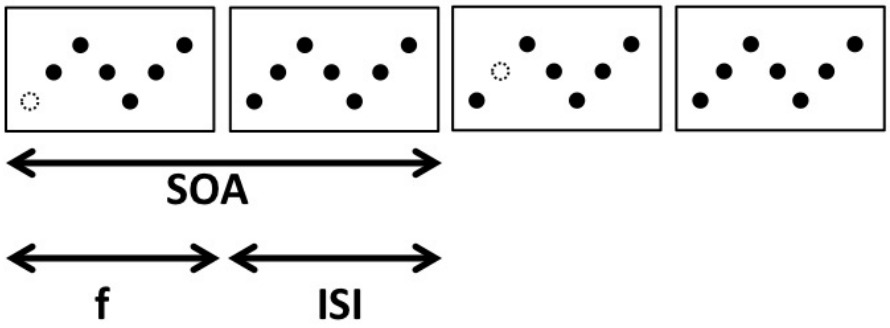

Figure 2.3. A replotting of Figure 2.2, treating the white dot as a stimulus. SOA remains the same duration, but is now measured from the appearance of one dot until the appearance of the next. The stimulus duration is relabled as the flash duration, f, which is the duration that a white

dot is present. The ISI is the temporal interval between the disappearance of a white dot and the appearance of the next dot.

We developed this display, however, not to examine local apparent motion, but to investigate the minimum conditions for SBF. As is typical in SBF, the local apparent motion correspondences (e.g., Ullman, 1979) are not what is seen; instead, groups of element changes support perception of a larger moving form or edge. Figure 2.4 shows how this display is in fact a good testbed for minimum conditions in SBF. Although the display has a perfectly valid physical description as a moving (white) dot that successively occludes each element (Figures 2.2 and 2.3), there is another description that characterizes this display as an SBF stimulus: a virtual bar moves through the space such that it causes a change in each element it touches (Figure 2.4). Movie 2.2 shows this effect with a vertical virtual edge and the element change of disappearance (and reappearance after the bar has passed). The only difference between Movie 2.2 and Movie 2.1 is that the temporal interval that all elements were visible has been shortened in Movie 2.2, from fourteen to two frames (140 to 20 ms). This change corresponds to shortening both the SOA and ISI to 40 and -20 ms respectively, while keeping stimulus duration fixed. In the alternative formulation of the timing parameters, these values correspond to an SOA of 40 ms and an ISI of 20 ms.

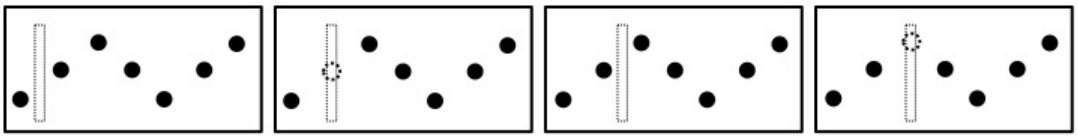

Figure 2.4. The same display as in Figures 2.2 and 2.3. Element disappearances (changes to white) are triggered by the passing of a virtual bar, indicated by the dashed rectangle.

Why is this a minimal SBF display? A collinear array of dots would be simpler. However, an important characteristic of SBF is that it does not occur for collinear arrays, and models of SBF become degenerate under this condition. The reasons involve an especially

difficult aperture problem that occurs in SBF (Shipley & Kellman, 1994, 1997; Prophet, Hoffman, & Cicerone, 2001). A simple summary is that SBF specifies an edge orientation as well as its motion from the spacing and timing of element changes. In a collinear array, the timing of element transformations will be identical for all edges regardless of virtual edge orientation. Only by adding a non-collinear element can orientation be determined from relative timing of flashes (Shipley & Kellman, 1997).

Virtual object properties can easily be manipulated simply by changing the temporal features of these displays. For example, it is possible to create displays with moving bars of any orientation simply by changing the relative timing of element changes. An oriented edge will result in variable delays between element changes, with shorter intervals between elements that lie on paths parallel to the bar and longer intervals between changes of elements that lie on paths perpendicular to the bar. An example can be seen in Movie 3. In Experiment 2, we examine how changing the duration that an element is invisible can affect the perceived width of the illusory bar.

The goal of Experiment 1 was to identify the spatial and temporal spacing parameters of texture element transformation events that support the perception of SBF in the displays described above. If there is a spatial integration limit, then the perception of illusory contours should deteriorate as distance increases for a fixed temporal interval between transformations. If there is a temporal integration limit, then, even for small inter-element distances, no illusory contours would be seen provided a sufficiently long interval between element transformations. The advantage of the current displays is that these parameters can be easily controlled and manipulated independently simply by changing inter-element spacing and change duration (i.e. occlusion or flash duration). Using a vertical bar for a virtual object and a sawtooth pattern with

fixed inter-element distances ensured that the distance and timing between each *event* (not just between each element) was constant within a display.

In Experiment 1, we used a subjective rating method. This method allowed us to determine as directly as possible the conditions under which participants actually perceived a moving, oriented edge. It also allowed us to sample a wide variety of spacing and timing parameters. We collected subjective contour clarity ratings for a variety of spacing and timing parameters and found that there is indeed an orderly pattern of spatial and temporal limits within which an illusory bar was visible. In Experiment 2, we followed up these results using an objective performance method, both to validate the perceptual reports of Experiment 1 as well as to explore when SBF under minimal conditions has measurable effects on performance on a task with an objectively correct answer.

Method

Participants

Subjects were eight research assistants or graduate students (one of whom was one of the authors, GE) who volunteered for the study (5 female; mean age: 25.38). All subjects reported having normal or corrected-to-normal vision. All subjects except for the author were naïve to the purposes of the study.

Design

Two properties of the displays were manipulated independently: inter-element distance (seven levels) and the ISI, the interval between the reappearance of one element and the disappearance of the next (13 levels). All combinations of the 7 inter-element distances and 13 temporal intervals were tested, resulting in 91 unique displays. Each display was presented 5

times, for a total of 455 trials. Subjects provided contour clarity ratings on a scale of 1-5 for each display.

Stimulus and Apparatus

All displays were created and displayed using the MATLAB programming language and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Stimuli were presented on a Viewsonic G250 CRT monitor, which was powered by a MacPro 4 with a 2.66 GHz Quad-Core Intel Xeon processor and an NVidia GeForce GT120 graphics card. The monitor was set to a resolution of 1024x768 pixels and a refresh rate of 100 Hz.

Displays consisted of black dots (diameter = 7.9 arcmin) on a white background (13.42 x 10.08 degrees of visual angle) arranged in a sawtooth pattern with five dots per cycle. The dots were centered on the screen vertically and spanned its entire width. The vertical and horizontal distances between each pair of points (measured from their centers) were equal and varied from 0.2 to 0.8 degrees of visual angle in steps of 0.1. These values correspond to Euclidean distances between element centers of 0.28, 0.42, 0.57, 0.71, 0.85, 0.99, and 1.13 degrees respectively. The number of dots needed to span the width of the screen varied with separation: 68, 45, 34, 27, 23, 20, and 17, for each of the 7 inter-element separations.

Dots flashed (i.e., disappeared and then reappeared) in sequence, beginning with the left-most dot and continuing, one at a time, from left-to-right. Once the right-most dot flashed, the pattern reversed, with flashes going from right-to-left. Displays lasted until the subject made a response. The flash duration (i.e., time each dot was invisible) was 20 ms. The inter-stimulus interval, or ISI was defined as the time between the reappearance of one dot and the disappearance of the next. ISI varied from 0 to 120 ms in steps of 10 ms. An ISI of 0 describes the case in which once an element reappears, the next one disappears on the same frame, i.e.,

there are no frames on which all elements are simultaneously visible. The ISI can therefore be thought of as the duration of the interval during which all elements are visible. Displays were made with all combinations of inter-element distances and ISIs. Trial order was randomized.

Procedure

Subjects sat at a distance of 170 cm from the computer monitor with their heads stabilized by a chin rest. The only illumination in the room came from the monitor. Subjects were instructed that they would see several movies containing a sawtooth pattern of black of dots on a white screen and in which one dot would flash at a time. They were told that if they tracked the flashing pattern laterally, they would sometimes see a moving illusory bar and that their task was to judge the clarity of the bar. They were then shown an example movie in which a bar was easily discernable and told that this movie corresponded to a high clarity rating (a five on a 1-5 scale). They were then shown a movie in which no bar was seen as an example of a low clarity rating stimulus (a one). Subjects were instructed to track the flashing sequence and not maintain fixation. Subjects provided a clarity rating on a scale from 1 to 5, 1 indicating poor illusory contour clarity and 5 indicated maximal contour clarity. Text at the top of every display reminded subjects that 1 indicated poorest clarity and 5 best clarity. Displays lasted until a response was made. After providing a rating, a blank, white screen was shown for 500 ms and then the next trial began immediately. A break was provided every 50 trials. The entire experiment lasted 30-40 minutes.

Results

Figure 2.5. Illusory contour clarity ratings for Experiment 1, averaged across subjects. Ratings were on an integer scale of 1 to 5. Bars indicate the standard error of the mean. Inter-stimulus interval (ISI) corresponds to the duration that all elements were visible on the screen in between flashes. The seven curves represent different inter-element separations from 0.28 degrees (circles) to 1.13 degrees (inverted triangles).

The results are displayed in Figure 2.5. Contour clarity ratings were averaged across subjects and are displayed as separate curves for each inter-element separation as a function of ISI. The data were submitted to a 7x13 within-subjects ANOVA. There was a significant main effect of inter-element separation ($F(6,42)=30.69$, $p<0.001$, $\eta_p^2=0.81$, Greenhouse-Geisser corrected), ISI ($F(12,84)=78.04$, $p<0.001$, $\eta_p^2=0.92$, Greenhouse-Geisser corrected), and a significant interaction ($F(72,504)=17.71$, $p<0.001$, $\eta_p^2=0.72$, Greenhouse-Geisser corrected).

Clarity ratings approached 1 (lowest clarity) for ISIs of 60-80 ms for all inter-element separations. Beyond 80 ms, ratings were flat at 1 for all separations. For ISIs between 0 and 60 ms, ratings decreased as a function of distance, with low average ratings (< 2) for inter-element distances greater than 1 degree (diamond and inverted triangle symbols in Figure 2.5) regardless of ISI. For most distances, ratings decreased as a function of ISI. However, for several of the

element separations such as 0.71 and 0.85 degrees (star and square symbols in Figure 2.5), it appears as though ratings first increased from 0 to 20 ISI before decreasing. *Post hoc* paired *t*-tests found no difference between ratings at 0 ISI and 20 ISI for either element separation.

Discussion

The experiment revealed spatial and temporal integration limits for SBF. Transformation events had to occur within 1 degree of each other and within 60-80 ms in order to produce a percept of an illusory bar. Illusory contour strength was graded, decreasing gradually as a function of both inter-element separation and ISI.

Past experiments with SBF had found that illusory shape perception degrades as SOA increases up to 165 ms, beyond which only apparent motion between elements is seen with no corresponding global form or illusory contours (Shipley & Kellman, 1993b, 1994). Shipley & Kellman (1994) showed that performance on an objective shape identification task improved monotonically with the number of frames shown within a temporal window of about 165 ms and that integration of information did not occur beyond that temporal interval. However, these findings are not directly comparable to those considered here. First, the earlier work varied SOA but, unlike the present experiment, used a fixed ISI of zero. Second, the 165 ms limit encompasses a number of processes that must occur in order to see complete shapes: the extraction of local edge orientations, the recovery of a sufficient number of such fragments to provide information around an entire 2D shape, and some persistence of these fragments over time to enable spatiotemporal interpolation (Palmer, Kellman & Shipley, 2006). Only the first process is relevant to the current study. Third, although some elements changed every frame as the object moved (on average, 5.44 elements at the highest densities tested in their study), the event positions may have corresponded to different contour fragments. It is unclear how much

time was needed for three non-collinear elements to transform in any local region of the object boundary.

The present experiment allowed us to examine the dynamics of the creation of a single oriented edge fragment. The sawtooth displays are somewhat unique in allowing measurements of the conditions that specifically restrict the extraction of edges in SBF. Further work is needed to determine how many edge fragments are used to construct a complete object representation and how those edges are integrated.

It is important to note that the results of Experiment 1 also support the general idea that the most basic process in SBF is extraction of local oriented edge fragments. Possible theories of SBF that could be based on defining 2D regions and progressively gaining evidence about their shapes (Prophet et al., 2001) would not apply to these displays (see also Shipley & Kellman, 1997).

Experiment 2

In Experiment 1, the flash duration was held constant while ISI varied. For a fixed flash duration, increasing the ISI also increases the SOA by an identical amount (SOA = flash duration + ISI). It is therefore impossible to determine whether the limiting factor is ISI or SOA. It is also unknown whether flash duration has an independent effect. Experiment 2 was designed to decide whether SOA or ISI is the temporal constraint on SBF.

It was found that illusory contours were not visible for ISIs greater than 60-80 ms and SOAs greater than 80-100 ms with a fixed flash duration of 20 ms. For this experiment, SOA was held constant at 80 ms while ISI varied from 0 to 60 ms. If SBF is determined by SOA, then illusory contours should be very weak or not seen at all as was found in Experiment 1. On the

other hand, if SBF is determined by ISI, then illusory contours should be seen for all but the longest ISI.

These temporal properties of the sawtooth stimulus affect more than just contour clarity. Because we began from an apparent motion display, it is easy to forget that flash duration, ISI, and SOA are actually determined by the properties of the virtual bar, namely its width and velocity. It is possible to look at these displays from a functional vantage point, in terms of a mechanism allowing the visual system to recover edges and objects from sparse information (Shipley & Kellman, 1994). The bar *causes* elements to disappear and reappear as it passes over them. It is possible to write an expression relating bar width and velocity to these temporal parameters. Assuming that the bar is traveling at a constant velocity in the rightward direction, its speed can be determined from the time it takes to travel between two elements, i.e., the time between the disappearance of one element and the disappearance of the next, or SOA.

$$v = \frac{h}{SOA} \tag{1}$$

Here, v is the object velocity and h is the horizontal separation between elements. Since SOA is defined as the sum of ISI and flash duration, any of these temporal terms can be substituted by a combination of the other two. The time that an element is occluded (the flash duration) is the time it takes for the bar to completely pass by the element. Given that elements disappeared and reappeared discretely with no gradual covering or uncovering, we treated the moment of element disappearance and reappearance as the point when the bar's edge reached (disappearance) or passed (reappearance) the midpoint of the element. The distance that the bar travels during the flash duration is therefore its width:

$$w = v * f \tag{2}$$

Substituting *v* from equation (1) into equation (2) yields an expression for bar width solely in terms of the spatial and temporal properties of the display:

$$w = \frac{h * f}{SOA} \tag{3}$$

Note that these calculations are only possible because the virtual bar is vertical and translating laterally across the display. As a result, ISI, SOA, and frame duration are the same for all elements in the display.

Since displays are created by setting the temporal parameters, it is possible to determine the objective width and velocity of the virtual bar used to generate the displays. If SBF serves to recover edge and object information from sets of sparse element changes, then an illusory bar, when perceived, might have a perceived width corresponding to that of the virtual object. Changes in the temporal parameters might have predictable effects on the properties of the perceived bar. For example, holding the horizontal separation between elements and SOA constant while increasing flash duration, would increase bar width while holding bar velocity constant. The forward reasoning is more intuitive: given a constant velocity, wider bars will occlude elements longer, resulting in a longer flash duration.

If the visual system interprets the sequence of changes as caused by a moving bar, it should be possible to affect the perceived bar properties simply by changing temporal settings. This affords the opportunity to create an objective measure of SBF in sawtooth displays. If an illusory bar is seen, observers should be able to accurately estimate bar width. This estimated value can be compared to the objective width as determined by the temporal settings. If observers are seeing an illusory bar, then the perceived and objective widths should be in agreement. This should be possible, of course, only under the conditions where SBF occurs, and a bar is actually seen. Experiment 1 showed clear limits in the spatial and temporal parameters

that produce perception of a bar. Only under such conditions – where SBF produces perception of the object – would we expect participants to be able to recover accurately its width (c.f. Ghose, Liu, & Kellman, 2014).

This observation motivated the control condition. It is possible that observers could perform the task solely by estimating the time that an element was invisible and inferring that this corresponded to a wider bar. Although unlikely because the subjects were naïve with respect to how displays were constructed, we tested them in a control condition with an SOA of 160 ms and the same flash durations as for the 80 ms SOA. This resulted in ISIs that were greater than 80 ms, a value that Experiment 1 showed did not support SBF, regardless of whether the temporal integration window is constrained by ISI or SOA. If subjects were using a strategy based on matching flash duration to bar width, then they should perform equally well in the experimental and control conditions, as it would not matter whether timing parameters did or did not support SBF. In contrast, if subjects really were seeing the illusory bar and were basing their responses on its properties in the experimental condition, then we would expect poor performance, and possibly chance performance, in this control condition.

Method

Participants

Participants were four student volunteers and one of the authors (GE) (females: 3, mean age: 22). All reported having normal or corrected-to-normal vision. With the exception of the author, all subjects were naïve to the purposes of the experiment.

Design

Two properties of the displays used in Experiment 1 were manipulated independently: stimulus onset asynchrony (SOA), the temporal interval between the disappearance of one element and the disappearance of the next, and flash duration, the temporal interval during which an element was invisible. We designated two sets of stimulus parameters as the experimental and control conditions, respectively. Displays with short SOAs (80 ms) were used as the experimental condition and displays with long SOAs (160 ms) comprised the control condition. Seven flash durations were used from 20 to 80 ms in steps of 10 ms. This resulted in ISIs (SOA - flash duration) of 0-60 ms for the short SOA and 80-140 ms for the long SOA. All combinations of SOAs and flash durations were used to create 14 unique displays. Each display was presented 20 times for a total of 280 trials. Equation (3) was evaluated for all SOAs and flash durations to compute the objective width of an illusory bar that would have produced those temporal intervals. Subjects performed a seven-alternative forced choice task by matching the perceived width of the illusory bar in the display to one of seven possible choices for each SOA.

Stimulus and Apparatus

Apparatus, stimuli, and procedure were similar to that of Experiment 1 unless otherwise noted. Horizontal and vertical spacing between elements was fixed at 0.4 degrees, for which illusory contours were relatively clear in Experiment 1. Using equations (1) and (3), bar width and velocity were computed for each SOA and flash duration. The bar velocity for the short and long SOAs was 5 deg/s and 2.5 deg/s respectively. The bar widths were 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, and 0.4 degrees for each of the flash durations respectively for the short SOA, and 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, and 0.2 degrees for the long SOA. We refer to these computed widths as the true bar widths. Above the elements, seven black horizontal lines were drawn with widths corresponding to the seven true bar widths for each SOA. Lines were ordered from

shortest to longest from left to right across the display (see Figure 2.6). Each line had a number above it corresponding to an answer choice (1-7), with 1 above the shortest line and 7 above the longest. The answer choices were always present during a trial. All combinations of the SOAs and flash durations were tested. Trial order was randomized for each subject.

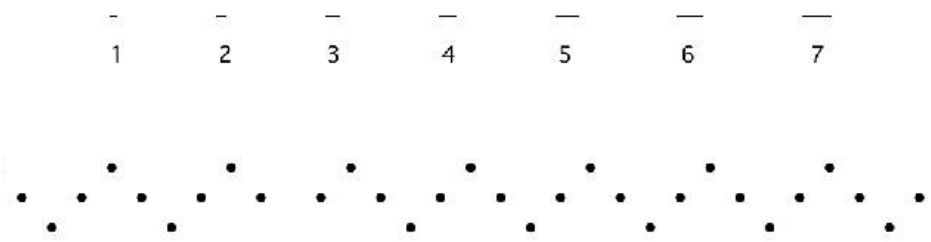

Figure 2.6. Still image from Experiment 2. Elements of the sawtooth pattern disappeared one at a time. Subjects matched the perceived width of an illusory bar with one of seven possible choices shown above the sawtooth display.

Procedure

Subjects sat 170 cm away from the monitor with their heads stabilized with a chinrest. Subjects were informed that they would be judging the width of an illusory bar and comparing it to one of several choices. They were also told that sometimes the bar would be difficult to see, and that they should make their best guess from amongst the answer choices if that was the case. They were then shown an example display with a readily seen illusory bar of intermediate width. Once subjects confirmed that they could see the bar, they initiated the experiment with a keypress. Trials began immediately. A trial would last until a subject made a response by pressing one of the numbered keys 1-7 that corresponded to the seven response choices. A break was provided every 50 trials. The experiment lasted for 20-30 minutes.

Results and Discussion

The results are shown in Figure 2.7. Average perceived bar width is shown for the experimental condition (short SOA, blue points) and control condition (long SOAs, red points) as a function of true bar width. The dashed, black line indicates when perceived width matches true width. The data were averaged across subjects for each condition and fit with a linear model. The fitting results are shown as dashed, gray lines in Figure 2.7. A linear model fit the data revealed a positive relationship between perceived and true bar width in the experimental condition: perceived width = 0.746 * true width + 0.025, $R^2 = 0.997$. Evaluated individually, all subjects showed monotonically increasing bar width estimates for increasing virtual bar widths (slopes ranging from 0.63 to 0.85). In contrast, there was no relationship between responses and true bar width in the control condition: perceived width = 0.065 * true width + 0.080, $R^2 = 0.518$. The slope of the regression line fit to the averaged data was not significantly different from zero ($p = 0.07$). Individual subject slopes were in the range of -0.01 to 0.19.

Figure 2.7. Perceived illusory contour width as a function of true width averaged across all subjects. Width was computed from element flash duration and SOA (see text for details). Data

were split for short (circles) and long (triangles) SOAs. Error bars are standard errors of the mean.

In Experiment 1, for the same inter-element spacing of 0.4 degrees and SOA of 80 ms (frame duration = 20 ms, ISI = 60 ms), contour clarity ratings were very low (1.6). Here, for the same SOA, but for ISIs between 0 and 60 ms, subjects were not only able to see illusory contours, but were also able to discriminate between different bar widths that differed only by 0.05 degrees. Within this set of tested display parameters, the property that determines the temporal integration window for SBF appears to be ISI rather than SOA. It is possible that there is an additional constraint imposed by SOA. Holding ISI constant and increasing flash duration would produce long SOAs that would correspond to a very wide, slowly moving bars.

In the control (long SOA) condition, subjects gave similar responses regardless of specified bar width. In this condition, both the SOA and ISIs were outside the range that produced high contour clarity ratings in Experiment 1. We therefore expected that subjects would not see the illusory figures in these displays. If subjects were performing the task simply by matching flash duration to bar width without seeing any illusory figures, then they should have been able to do so for both the short and long SOAs since the same flash durations were used with both. However, subjects only matched bar width for the short SOA, when ISIs were 60 ms or less. At least two subjects in the control condition selected the same bar width on nearly every control trial (one consistently selected the thinnest bar width, and the other consistently selected an intermediate width). The data suggest that there was no impression of bar width in the control displays, and subjects resorted to guessing strategies.

Even if the displays produced different percepts, it is possible that subjects used the widths of the response choices as a cue. Because frame durations were the same for both SOAs, the bars for the longer SOA were half the width of those for the shorter SOA. As a result, the

difference between bar widths was also halved from 0.05 degrees to 0.025 degrees. No subject reported noticing the different response choices or a relationship between response choice and phenomenology. Poor performance for long SOAs may have also occurred because shorter bar widths are more difficult to see or because the smaller differences between bar widths made the task more difficult. There are several reasons why this was likely not the case. First, bar widths of 0.1, 0.15, and 0.2 degrees were tested for both SOAs and width estimates were closer to true widths in for shorter SOAs (e.g., bars of a 0.2 degree width were rated on average as having a width of 0.17 degrees in the short SOA condition, and a width of 0.09 degrees in the long SOA condition). Second, bias was smallest and accuracy was highest for narrower bars for the short SOA. Narrow bars in the short SOA condition were therefore perceived more clearly. If narrower bars were more accurately perceived, then responses should have been more accurate for the long SOA condition, in which bar widths were have the size of those used in the short SOA, but this was not observed.

In the experimental condition, for wider bars, there was a tendency to underestimate bar width. Average perceived width for the widest bars (0.32 deg) was 80% of true width (0.4 deg). It is unclear from this experiment whether this bias is particular to the estimation of metric properties of illusory objects or whether it would also exist if subjects had to perform the task with real objects.

An interesting observation is that subjects were able to discriminate between bars whose widths differed by less than half of an element's diameter (0.13 degrees). In particular, perceived width was significantly different for bars with true widths of 0.1 and 0.15 degrees ($t(4)=3.21$, $p=0.03$, Cohen's $D=0.23$). This suggests that SBF allows edges to be recovered at sub-element resolution. Note that all element changes in SBF are discrete (e.g., a dot was either fully present

or fully invisible on any frame, with no gradual covering or uncovering). No studies as yet have systematically investigated the effect of element size, and this could be a question for future research.

General Discussion

In this paper, we set out to identify the minimal conditions for spatiotemporal boundary formation (SBF). We sought the simplest conditions that might produce perception of a single oriented illusory edge fragment from the sequential transformation of texture elements. We hypothesized that these minimum conditions for SBF might be met or approximated by a novel display in which circles arranged in a sawtooth pattern disappeared and reappeared one at a time (Figure 2.2). Such a display would ordinarily produce perception of apparent motion between elements (Movie 2.1), but with different timing parameters, it produced SBF -- perception of a laterally translating, vertically oriented, illusory bar (Movie 2.2). The display structure allowed for precise and independent manipulation of the temporal and spatial intervals between element transformation events. As a result, values for those parameters that led to one or the other percept could be identified.

Using subjective ratings of illusory contour clarity in Experiment 1, we found that the perception of a moving illusory contour decreased gradually as the separation between elements increased up to one degree of visual angle (Figure 2.5). For displays containing elements that were arranged farther apart illusory contours were no longer seen. Illusory contour clarity was also constrained by ISI, the temporal interval between element transformations. Clarity ratings gradually decreased with increasing ISI up to 60-80 ms. Displays in which ISIs were greater than 80 ms appeared to be ineffective in producing SBF.

In Experiment 1, for all ISIs tested, the corresponding SOAs – the time between the transformation of one element and the next – were 20 ms longer than the ISIs. It was therefore unclear whether the limiting temporal factor was a 60-80 ms ISI or an 80-100 ms SOA. Experiment 2 was designed to distinguish between these two properties. The temporal and spatial properties of the displays were used to compute the width of an illusory bar that would have led to texture element transformations at those rates. A wider bar, for example, would take longer to pass over an element, resulting in a longer flash duration (time that the element is white or invisible). In an objective performance task, subjects matched perceived illusory bar width with the objective width computed from display parameters. When tested with an SOA of 80 ms, which yielded poor contour clarity ratings in Experiment 1, and varying ISIs from 0 to 60 ms, subjects were very accurate in estimating bar width (Figure 2.7). The results suggest that the temporal constraints at this SOA are strongly influenced by ISI, since an SOA that produced poor contour clarity ratings in Experiment 1 was able to yield accurate estimates of illusory bar width, as long as ISIs were short. A control condition with higher SOAs and ISIs verified that subjects were performing the task based on perceived illusory bar width, and not directly from temporal properties of the display that may have served as a response cue. The control condition yielded no evidence of SBF, as indexed by participants' inability to detect bar width accurately.

An important feature of these experiments is that they bring to bear both subjective and objective methods for evaluating illusory contour perception. Each has potential advantages and disadvantages. SBF is a perceptual phenomenon: what observers see matters. But whereas we assume that perceptual reports convey information about what is seen, they also potentially reflect many other factors, including variations in criteria or use of scales by participants, their understanding of instructions, and possibly their hypotheses about what the experimenter is

looking for. Objective paradigms, on the other hand, in which participants' performance can be compared to an objectively correct answer, may more readily avoid some criterion issues, and can be more revealing about underlying mechanisms, but only if the task really depends on the relevant perceptual representations. If there are other strategies for succeeding at a task, all bets are off. In the present work, the concordance of the spatial and temporal parameters that produce clear perception of contours in Experiment 1 with the parameters that allow successful performance in Experiment 2 strongly supports the idea that the methods are converging on contour perception through SBF.

Taken together, the experiments reveal several novel properties of SBF. SBF can arise in very simple displays with few elements arranged in a regular pattern and does not require the formation of a closed contour: A single edge fragment can be seen. Whether apparent motion or an illusory figure is seen is determined by the spatial and temporal intervals between texture element transformations (Experiment 1). These intervals are the spatial and temporal integration limits within which oriented edges can be extracted. The dimensions of the shape that is perceived can be determined by the spatial and temporal properties of the elements with which it interacts (Experiment 2).

How are illusory contours and shapes seen in SBF?

In these displays, illusory edge orientation and velocity were recovered from sequential transformations of elements. The spatial and temporal parameters governing the displays determined whether the percept was of apparent motion (Movie 2.1), a vertically oriented, moving illusory bar (Movie 2.2), or a tilted illusory bar (Movie 2.3 – here, the timing between element transformations varied). These results provide direct support for a class of models of SBF first proposed by Shipley and Kellman (1994, 1997) over two decades ago. They proved

formally that three, non-collinear texture element transformations where sufficient to define the local orientation of a contour fragment, and they theorized that two pairs of element transformations in some local region (involving as few as three element transformations) might be the mechanism producing perception of an oriented edge fragment and its current motion. On this view, locally generated edge fragments are the basic constituents of shape perception in SBF. Here, we have shown that SBF can indeed generate such local edge fragments from small numbers of element transformations, provided these events occur within particular spatial and temporal windows.

These findings offer new insight about the perception of contours and objects from sparse texture element changes. Most previous work had exclusively looked at 2D virtual shapes in SBF (e.g., Shipley & Kellman, 1993; Cicerone, Gowdy & Kim, 1995; Cooke, Cunningham & Bulthoff, 2004; but see Chambeaud, Martin, & Barraza, 2014). The confirmation that SBF can occur for short edge fragments indicates that such fragments may indeed be recovered without more global shape information and that such fragments are the likely basic units in SBF. The clear spatial and temporal constraints on this process suggest modular mechanisms. In a separate paper, we consider how this computation is implemented neurally and suggest that filters for recovering motion energy from moving contrast-defined edges (Adelson & Bergen, 1985; van Santen & Sperling, 1984) may also function as edge filters when edge orientation is not given by static stimulus properties. On this hypothesis, motion energy filters have a dual function, one that is not readily discernible when stimulus orientation is explicitly given by contrast: they are also edge filters. Moreover, on this hypothesis, SBF is not an esoteric perceptual illusion, but is actually more indicative of basic processes implemented across the visual field for extraction of edges, motion, and the relationship of these. Of course, many details of this conjectured edge-

motion duality in basic visual filtering remain to be worked out, including how a variety of aperture and cross-scale ambiguity problems get resolved.

The spatial integration limits found in Experiment 1 constrain the maximum length of a single edge fragment. However, the experiments did not identify the minimal length that can be perceived. From causal observation, the height of the illusory bars in Experiment 1 was constrained by the height of the sawtooth pattern (i.e., the vertical separation between the element at a peak and trough). The smallest height of the sawtooth tested was 0.4 degrees of visual angle. In Experiment 2, illusory bar widths were resolvable at a sub-element level (smaller than 0.13 degrees); however, bar width may have more to do with the perception of an illusory surface than minimum length. It is also worth noting that minimum length may be constrained by element size: since three non-collinear elements are needed to be able to determine edge orientation, perhaps edge length is determine by the minimum distance that one of the elements needs to be shifted in order to appear as not being collinear with the other two. This distance may depend on element size. Other experiments have used SBF shapes with curved contours (e.g., Shipley & Kellman, 1994) and their illusory boundaries appear smooth, suggesting that minimum edge length is likely smaller than the values tested here. Further work is therefore needed to determine the minimum length of an illusory fragment that can be extracted.

Once a fragment is extracted, it is related to other fragments and the missing boundary regions between fragments is interpolated. The orientations and velocities of several fragments can be used to determine the global motion direction of the set of fragments (Shipley & Kellman, 1998). Contour interpolation of illusory edge fragments may be determined by the same geometric constraints of relatability that govern interpolation of real contours (Kellman & Shipley, 1991). Because several events need to occur before an edge can be extracted, these

fragments will never be available all at once. Rather, the visual system needs a method for maintaining a persisting representation of a fragment once extracted and updating its position relative to other fragments extracted at a later time. A model has been proposed for how such spatiotemporal interpolation may occur between visible contour fragments and occluded contour fragments that had been visible at a prior time (Palmer, Kellman, & Shipley, 2006). Global form in SBF would be constructed in the following manner: (1) local edge fragments recovered from sequences of element transformations within a small spatiotemporal window, (2) relatable contour fragments are interpolated to produce a coherent boundary, (3) the global motion of the completed object is recovered from the individual motions of the edge fragments.

Perceptual postdiction

An interesting property of the sawtooth displays is that when the illusory bar is seen, both its trailing and leading edge are visible. Given that elements disappear quantally and one at a time, this is actually quite surprising. Consider the displays used in Experiment 2. In order to determine bar width, an element must first reappear, since the width determines the duration that an element is invisible or occluded. However, the bar's trailing edge is seen continuously, even while elements are still invisible and therefore before the width can be determined. This is a form of "postdiction" by which the visual system constructs a surface representation *after* the offset of a stimulus (Choi & Scholl, 2006; Kawabe 2011). Postdiction always occurs in apparent motion, in which an observer sees intermediate states when none exist. These intermediate states can be positions, as when an object appears to move between two locations, or can be more complex, as in the perception of intermediate shapes when a square and triangle are flashed alternatively (Kolers & von Grünau, 1976). Postdiction also occurs on an element-by-element scale in SBF in

that while no apparent motion is seen between elements, the magnitude of a velocity vector between two elements can only be determined after the second element has disappeared.

It is additionally surprising that no apparent motion is seen since the displays, especially when ISIs are 0, which corresponds more closely to a regular apparent motion stimulus in which an element disappears on the same frame as the preceding element reappears. Correspondence models of apparent motion and first-order motion detectors would both predict that motion should be seen between elements (e.g., Ullman, 1979). However, a number of studies have found that form perception can alter or suppress motion perception (Bruno & Gerbino, 1991; Lorenceau & Alais, 2001; Petersik & McDill, 1981; Ramachandran & Anstis, 1986). The integration of local motion signals into a boundary may prevent them from being seen (Shipley & Kellman, 1997). In particular, when element transformation events can be interpreted as occlusions, apparent motion is suppressed (Ekroll & Borzikowsky, 2010; Holcombe, 2003; Sigman & Rock, 1974). Outside of the integration range, however, the percept reverts to inter-element apparent motion. Since at least two non-collinear signals are needed to define an edge, inter-element apparent between the first two elements (the first motion signal), is suppressed only after the third element disappears and a second motion signal is generated. Although it has been theorized that three transformations are sufficient to determine the orientation of an edge that caused those transformations, subsequent studies are needed to determine the minimum number of successive element required.

Conclusion

Using both subjective and objective methods, the experiments reported here provide strong evidence that the illusory global forms seen in spatiotemporal boundary formation begin with a more local process of extracting oriented edge fragments from small sets of local element

transformations. This process is constrained by both clear spatial and temporal properties, suggesting modular mechanisms for local edge recovery from sparse stimulus information. It is possible that this computation may be performed by detectors that have been previously characterized as motion energy filters; in the absence of oriented edge information in static views, these or related filters may be used to extract edge orientation. Further research will be needed to investigate whether this conjecture will prove to be useful in understanding the relation of SBF to basic visual mechanisms.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, A, 2*(2), pp. 284-299.

Banton, T., & Levi, D. M. (1992). The perceived strength of illusory contours. *Perception & Psychophysics, 52*(6), pp. 676-684.

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

Bruno, N., & Bertamini, M. (1990). Identifying contours from occlusion events. *Perception & Psychophysics, 48*, pp. 331-342.

Bruno, N., & Gerbino, W. (1991). Illusory figures based on local kinematics. *Perception, 20*, pp. 259-274.

Chambeaud, J.G., Martin, A. & Barraza, J.F. (2014). Effect of disparity on the perception of motion-defined contours. *Psychology & Neuroscience, 7,* 2, 91 − 10.

Choi, H, & Scholl, B. J. (2006). Perceiving causality after the fact: Postdiction in the temporal dynamics of causal perception. *Perception, 35*, pp. 385-399.

Cicerone, C. M., & Hoffman, D. D. (1997). Color from motion: Dichoptic activation and a possible role in breaking camouflage. *Perception, 26*, pp. 1367-1380.

Cicerone, C. M., Hoffman, D. D., Gowdy, P. D., & Kim, J. S. (1995). The perception of color from motion. *Perception & Psychophysics, 57,* pp. 761-777.

Cooke, T., Cunningham, D.W., & Bulthoff, H. (2004). The perceptual influence of spatiotemporal noise on the reconstruction of shape from dynamic occlusion. In Rasmussen, C.E., Bulthoff, H.H., Giese, M.A. & Scholkopf, B. (Eds.). *Pattern recognition: Proceedings of the 26th DAGM Symposium*. Tubingen, Germany, August/September, 2004.

Cunningham, D. W., Shipley, T. F., & Kellman, P. J. (1998). Interactions between spatial and spatiotemporal information in spatiotemporal boundary formation. *Perception & Psychophysics, 60*(5), pp. 839-851.

Ekroll, V., & Borzikowsky, C. (2010). The role of occlusion cues in apparent motion. *Perception, 39,* pp. 1606-1623.

Ekroll, V., Faul, F., & Golz, J. (2008). Classification of apparent motion percepts based on temporal factors. *Journal of Vision, 8*(4), pp. 1-22.

Fidopiastis, C., Hoffman, D. D., Prophet, W. D., & Singh, M. (2000). Constructing surfaces and contours in displays of color from motion: The role of nearest neighbors and maximal disks. *Perception, 29*, pp. 567-580.

Ghose, T., Liu, J., & Kellman, P. J. (2014). Recovering metric properties of objects through spatiotemporal interpolation. *Vision Research, 102*, pp. 80-88.

Holcombe, A. O. (2003). Occlusion cues resolve sudden onsets into morphing line motion, disocclusion, and sudden materialization. *Journal of Vision, 3*, pp. 562-572.

Kawabe, T. (2011). Nonretinotopic processing is related to postdictive size modulation in apparent motion. *Attention, Perception, & Psychophysics, 73*, pp. 1522-1531.

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23*, pp. 141-221.

Kolers, P. A. & von Grünau, M. (1976). Shape and color in apparent motion. *Vision Research, 16*, pp. 329-335.

Korte, A. (1915). Kinematoskopishe Untersuchungen. *Zeitschrift für Psychologie mit Zeitschrift für angewandte Psychologie, 72*, pp. 194-296.

Lorenceau, J., & Alais, D. (2001). Form constraints in motion binding. *Nature Neuroscience, 4*(7), pp. 745-751.

Miyahara, E., & Cicerone, C. M. (1997). Color from motion: separate contributions of chromaticity and luminance. *Perception, 26*(11), pp. 1381-1396.

Palmer, E. M., Kellman, P. J., & Shipley, T. F. (2006). A theory of dynamic occluded and illusory contour formation. *Journal of Experimental Psychology: General, 135*(4), pp. 513-541.

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

Petersik, J. T., & McDill, M. (1981). A new bistable motion illusion based upon "kinetic optical occlusion". *Perception, 10*, pp. 563-572.

Prophet, W. D., Hoffman, D. D., & Cicerone, C. M. (2001). "Contours from apparent motion: A computational theory" in *From fragment to objects: Segmentation and grouping in vision*. P. Kellman and T. Shipley (Eds), Amsterdam: Elsevier Science Press, pp. 509-530.

Ramachandran, V. S., & Anstis, S. (1986). Figure-ground segregation modulates apparent motion. *Vision Research, 26*(12), pp. 1969-1975.

Shipley, T. F., & Kellman, P. J. (1993a). Optical tearing in spatiotemporal boundary formation: When do local element motions produce boundaries, form, and global motion? *Spatial Vision, 7*(3), pp. 323-339.

Shipley, T. F., & Kellman, P. J. (1993b). Spatiotemporal boundary formation: Temporal integration is confined to a 150 msec window. *Investigative Ophtalmology and Visual Science Supplement, 34*(4), pp. 1082.

Shipley, T. F., & Kellman, P. J. (1994). Saptiotemporal boundary formation: Boundary, form, and motion perception from transformations of surface elements. *Journal of Experimental Psychology: General, 123*, pp. 3-20.

Shipley, T. F., & Kellman, P. J. (1997). Spatio-temporal boundary formation: the role of local motion signals in boundary perception. *Vision Research, 27*(10), pp. 1281-1293.

Sigman, E., & Rock, I. (1974). Stroboscopic movement based on perceptual intelligence. *Perception, 3*, pp. 9-28.

Steinman, R. M., Pizlo, Z., & Pizlo, F. J. (2000). Phi is not beta, and why Wertheimer's discovery launched the Gestalt revolution. *Vision Research, 40*, pp. 2257-2274.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge: MIT Press.
van Santen, J. P. H., & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America, A, 1*(5), pp. 451-473.
Wertheimer, M. (1912). Experimentelle Studien über das Segen von Bewegung. *Zeitschrift für Psychologie, 61*, pp. 161-265.

Chapter 3: Modeling Spatiotemporal Boundary Formation

Abstract

Transformations of widely separated visible elements, such as appearance/disappearance, color change, orientation change or motion, can give rise to robust percepts of continuous contours, shape, and global motion. Shipley and Kellman (1994) found that the well-known example of accretion and deletion of texture is only one of many element transformation types produces these effects and called the more general process *spatiotemporal boundary formation* (SBF). Efforts to model SBF have included formal proofs that orientations and motion direction of local edge fragments could be recovered from small sets of element changes (Shipley & Kellman, 1994, 1997), but little work has examined SBF in simplified situations, and no models have taken into account noise in human detection of the basic inputs to SBF. Accordingly, no model has been able to predict accurately human SBF performance. Here, we measured orientation discrimination thresholds for thin, oriented edges as a function of element density, display duration, and frame duration. Thresholds decreased with increasing density and display duration and increased as frame duration increased. We implemented an ideal observer version of the Shipley & Kellman (1997) model, and it exceeded human performance, predicting perfectly edge orientation on a trial-by-trial basis solely from element positions and the times of their transformations. In a second experiment, we measured human precision in detecting inputs to the model (distance, angular separation, and time between the transformations of pairs of elements). A model that added encoding imprecision for these parameters to the estimated from experiment 2 to the ideal observer closely fit human data from experiment 1 with no free parameters.

Introduction

The boundaries of objects can indicated by many cues, some of the most common being discontinuities in luminance contrast, color, depth (Julesz, 1971) or texture (Kaplan, 1969; Julesz, 1975) at the object boundary. Other cues can come from dynamic information as objects or observers move in the world, such as a motion parallax and the accretion and deletion of texture (Braunstein, Andersen, & Riefer, 1982; Gibson, Kaplan, Reynolds, & Wheeler, 1969; Kaplan, 1969; Ono, Rogers, Ohmi, & Ono, 1989; Rogers & Graham, 1983; Yonas, Craton, & Thomspon, 1987). The boundaries of a textured surface may be undetectable if it is in front of a similarly textured surface. Once one of the surfaces begins to move, however, the elements of the farther surface are gradually occluded and revealed at the boundaries of the nearer surface. Such occlusions can result in the perception of illusory boundaries, surfaces, and global motion (Andersen & Cortese, 1989; Cicerone & Hoffman, 1997; Cicerone, Hoffman, Gowdy, & Kim, 1995; Cunningham, Shipley, & Kellman, 1998a, 1998b; Fidopiastis, Hoffman, Prophet, & Singh, 2000; Prazdny, 1986; Shipley & Kellman, 1993, 1994, 1997; Rovden, Baker, & Allman, 1988).

As elements are accreted or deleted, they provide oriented contour information about the occluding boundary. Shipley and Kellman (1993, 1994) suggested that accretion and deletion of texture is a special case of a more general process. Sets of changes in some property of visible elements can produce perception of continuous illusory boundaries, global form, and global motion, a process they called *spatiotemporal boundary formation* (SBF). A wide variety of element transformations can produce similar perceptions of boundaries and surfaces, including discrete (i.e., all-at-once) element disappearance, as well as changes in orientation, shape, color, or position (Shipley & Kellman, 1993, 1994).

How are global shapes seen in SBF? It has been proposed that shape in SBF depends on two processing stages. First, information from element changes in certain small neighborhoods

produces local edge fragments having specific orientations. Second, these edge fragments connect across gaps according to well-known interpolation processes that operate in the perception of illusory and occluded contours (Kellman & Shipley, 1991; Kalar et al, 2010; Palmer, Kellman & Shipley, 2006; Shipley & Kellman, 1994; Erlikhman & Kellman, submitted). Whereas the second stage involves processes that are well-understood, the first has remained mysterious. Although Shipley & Kellman (1994) described mathematically minimum conditions for SBF (three non-collinear element transformations), little empirical research has examined SBF with single edges and relatively few elements. Virtually all previous studies of SBF have used closed objects with smooth contours as stimuli (although see Barraza & Chen, 2006). Recently, we demonstrated that oriented, illusory edge fragments can be recovered from extremely minimal displays in which elements are arranged in a jagged wave or sawtooth pattern and disappear and reappear sequentially, one at a time (Erlikhman & Kellman, *submitted*). These results support the two-level theory of SBF, specifically in implicating a process that recovers local oriented edge fragments. These local edge fragments are likely the basic units from which larger shapes are constructed in SBF. In this paper, we sought to develop a model of how such edges may be extracted. First, we describe how an SBF display is constructed and discuss several properties. Next, we consider and test an ideal observer model of how edges can be extracted in SBF displays. In the first experiment, we measured orientation discrimination thresholds for SBF-defined edges across a variety of display properties. Human sensitivity relative to the ideal observer model was suboptimal. We consider several constraints and noise factors that could have affected human performance. In a second experiment, we empirically measured noise to low-level features of the displays, such as element separation. A model that incorporated the constraints and the spatial and temporal noise parameters measured in the second experiment

was able to accurately predict human performance in the first experiment across all tested display conditions.

Spatiotemporal Boundary Formation (SBF)

Figure 3.1 shows an example of an SBF display. The dotted line in Figure 3.1 defines the boundary of a virtual object. The elements are always stationary and the virtual object moves across the display. As the object movies, elements that move across the boundary change in some property such as color. The change is instantaneous and discrete (i.e., not gradual). The percept is of a moving figure with crisp illusory boundaries. In uni-directional transformations (Figure 3.1), elements inside the boundary of the object have the same value along some feature dimension (e.g., color) and those outside have a different one. In bi-directional transformations, elements are randomly assigned one of two values and switch to the other upon entering or exiting the boundary of the moving object. The clarity of the illusory contour boundary depends on element density, luminance differences between elements, the velocity of the occluding surface, and frame duration (Andersen & Cortese, 1989; Shipley & Kellman, 1994; Cicerone et al., 1995).

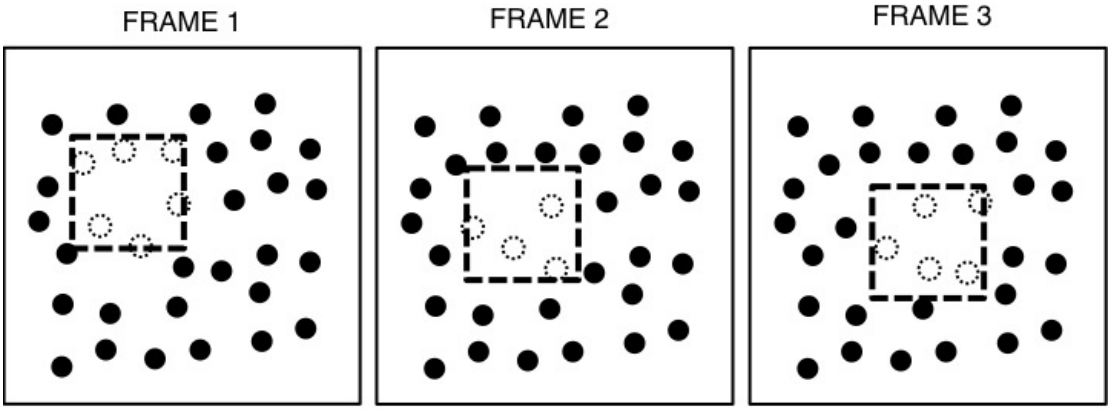

Figure 3.1. Depiction of a square "pseudosurface" moving over a field of circular black elements. All elements inside the square region are in one state and all those outside are in

another. Each individual frame contains a collection of white circles in an amorphous group. As the square moves (frames 2 and 3), elements entering and exiting the region change states. The resulting percept is of a moving, colored region with crisply defined illusory contours. Figure from Erlikhman & Kellman (*submitted*).

There are many other cases in which the visual system is able to recover structure from sparse information. For example, gradual, dynamic occlusion of one surface by another can be used to create "kinetic illusory contours" (Andersen & Cortese, 1989; Bruno & Bertamni, 1990; Bruno & Gerbino, 1991; Kellman & Cohen, 1984; Stappers, 1989). In these displays, spatially disparate portions of a figure become visible one at a time as it moves in front of another surface. Regions of the figure that are never visible are interpolated and an illusory surface is perceived, despite many of the object's regions never being visible simultaneously. Similarly, in multi-aperture or slit-viewing displays, portions of an object's surface are gradually revealed as the object moves behind many small apertures (Anstis & Atkinson 1967; Aydin, Herzon, & Ogmen 2008; Mateeff, Popov, & Hohnsbein, 1993; Palmer, Shipley, & Kellman, 2006). In such displays, information about the object's surfaces and boundaries are also only available sporadically, must be represented when occluded, and related to other, visible regions when they become available.

Spatiotemporal boundary formation is an even more extreme case in terms of the paucity of available information for the construction of edges and surfaces. In SBF, the position of object boundaries is only revealed by their interaction with elements. Because elements transform all at once and are not gradually occluded, there is no oriented contour information at those locations. This creates a very difficult kind of aperture problem, what has been referred to as a "point aperture problem", in which *both* the orientation and velocity of an edge are indeterminate (Shipley & Kellman, 1994, 1997; Prophet, Hoffman & Cicerone, 2001). In the traditional aperture problem, the task is to construct a velocity field from many local signals recovered from

the motion of oriented edge fragments. In the point aperture problem, there are no oriented edge fragments. The visual system must simultaneously recover *both* the orientation and the velocity of edge from spatially and temporally sparse and discrete element transformations. Next, we consider one proposed solution to this problem.

Modeling SBF

A model of SBF has been proposed for how local, oriented edge fragments can be extracted from the sequential transformation of elements (Shipley & Kellman, 1994, 1997). These illusory fragments may then be related and the regions between interpolated by the same processes that govern contour grouping and interpolation for real contours (i.e., relatability) resulting in a representation of a completed object contour (Kellman & Shipley, 1991; Palmer, Kellman, & Shipley, 2006). Erlikhman and Kellman (*submitted*) have provided evidence for the first stage of this process and have shown that single edges can indeed be recovered.

A geometric proof of the point aperture problem is possible given the positions and times of three, non-collinear element transformations, the orientation of an edge that caused those transformations can be computed assuming a constant edge velocity and orientation (Shipley & Kellman, 1997). An intuition for the proof appears in Figure 3.2. Figure 3.2a depicts a sequence of element transformations for three elements (labeled 1, 2, and 3) caused by a moving edge. When two elements transform (in this case, disappear and reappear) in succession, a transformation vector, $\mathbf{v}_{12}$, is formed between them. The magnitude of the vector is determined by the spatial and temporal separation of the transformations. We use the term “transformation vector” in lieu of motion vector to emphasize that apparent motion is *not* seen between individual elements during SBF. The transformation of a third element defines a second transformation vector, $\mathbf{v}_{23}$, between the second and third elements. If the tails of these two

transformation vectors are placed on the same point (Figure 3.2b), then the orientation of the vector connecting their heads ($\mathbf{v}_{12}$ - $\mathbf{v}_{23}$) has the orientation of the illusory edge, provided that the edge was moving at a constant velocity and had a constant orientation between transformation events.

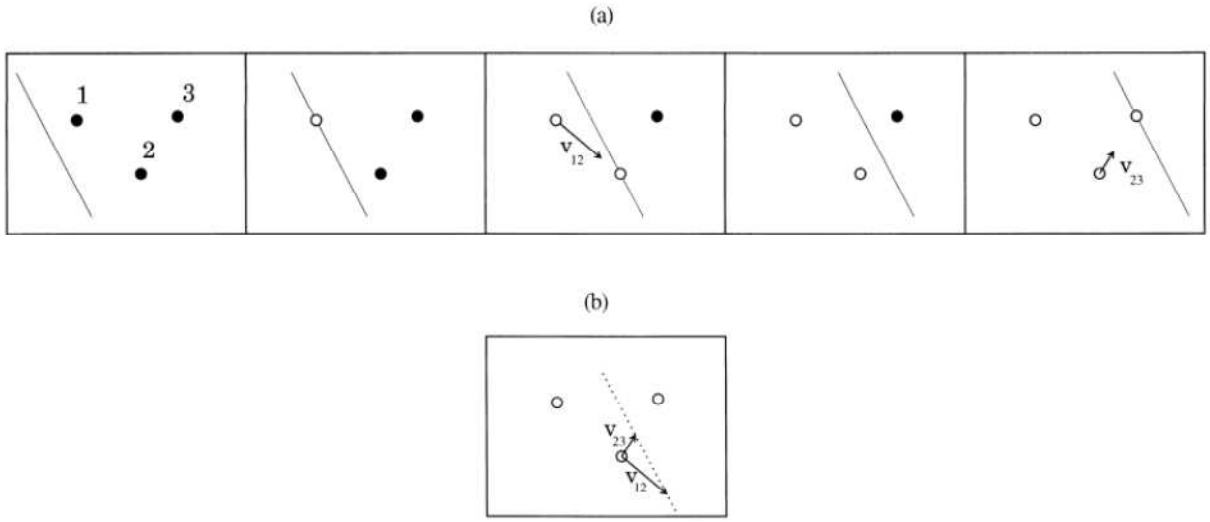

Figure 3.2. A sequence of frames in which a moving edge successively transforms three elements (changing from black to white). (a). Three elements disappear, one at a time. $\mathbf{v}_{12}$ and $\mathbf{v}_{23}$ are transformation vectors defined by the spatial and temporal separation between elements. (b). Transformation vectors $\mathbf{v}_{12}$ and $\mathbf{v}_{23}$ can be combined to define the orientation of the moving edge. Figure from Shipley, T. F. & Kellman, P. J. (1997). Spatio-temporal Boundary Formation: the Role of Local Motion Signals in Boundary Perception. *Vision Research*, 37(10), 1281-1293.

The orientation of the illusory edge, θ, can be expressed with the following equation:

$$\theta = \tan^{-1}\left(\frac{v_{23} * \sin\varphi_{23} - v_{12} * \sin\varphi_{12}}{v_{23} * \cos\varphi_{23} - v_{12} * \cos\varphi_{12}}\right) \qquad (1)$$

Where φ_{ij} is the angle formed between a horizontal line passing through element *i* and a line connecting element *i* to element *j*, and v_{ij} is the magnitude of the transformation vector between the two elements. v_{ij} can be computed from the distance between two elements (A_{ij}) and the time between the transformations of those elements (ΔT_{ij}):

$$v_{ij} = \frac{A_{ij}}{\Delta T_{ij}} \qquad (2)$$

One interesting difference between this solution and the solution to the classical aperture problem in motion perception (Nakayama & Silverman, 1988a, 1998b; Shimojo, Silverman & Nakayama, 1989) is that the aperture problem only requires two frames from a motion sequence and two local velocity estimates from differently oriented edges along different regions of the object boundary. Edge orientations are given by contrast information. Here, only a single edge is being recovered, so all element transformation events are occurring along the length of an edge. Moreover, both the local edge orientation, as well as the edge's motion, must be recovered from discrete element changes. At least three events, or three frames if one event occurs per frame, are needed in order to recover the edge orientation. Once edge orientation is recovered, the motion direction of the edge is itself ambiguous, i.e., there is an aperture problem for the recovered edge segment (Shipley & Kellman, 1994). This aperture problem for the global motion of the object can be solved if several edge segments along the object boundary are recovered. One of the surprising features of SBF is that the object boundary is continuously seen even though edge segments are recovered sporadically. In typical considerations of the aperture problem, contrast-defined edges are always visible, and it is only a matter of integrating local motion signals that can be extracted from any pair of frames.

The model makes several assumptions. First, at least three transformation events are needed. The orientation of the illusory edge is ambiguous for any two events because there are an infinite number of combinations of edge orientations and velocities that could cause the same temporal interval between two element transformations. Second, the three elements cannot be collinear. If they were, then $\varphi_{23} = \varphi_{12}$ and substituting into equation 1 leads to the equality $\theta = \varphi$, that is, that the estimated edge orientation is the same as the orientation of the line connecting all

three elements. Correct edge orientation cannot be recovered from a collinear triplet of elements for the same reason that orientation cannot be recovered from the transformations of only two elements, namely that many combinations of edge orientations and velocities could have caused the transformations. The collinearity restriction applies only to *successive* element transformations. Elements can be arranged in a regular grid and SBF would still occur (Fidopiastis et al., 2000), as long as the elements that transformed in sequence were not collinear. Third, the orientation and velocity of the moving edge are constant between transformation events. If velocity or orientation change between the first and second or second and third elements in a triplet, then the triplet can be treated as two independent pairs of elements, and the same problem in determining orientation for a pair of elements arises.

Several findings support the hypothesis that motion-like signals, or vectors relating pairs of element transformations elements serve as the input to an edge extraction process. We refer to these signals as motion-like because actual local motion between element transformations are not perceived in SBF. If element transformations are color changes, the perception of illusory boundaries can be disrupted by the addition of spurious flickering or moving elements in other regions of the display (Shipley & Kellman, 1997). Continuously moving elements should be readily distinguishable from the stationary ones that that change color at the object boundary (when the object boundary passes across then), and might not be expected to have an effect on boundary formation. However, these spurious transformations greatly reduced contour clarity, suggesting that additional motion signals interfere with those used to construct edges. Contour clarity has also been found to depend on the relative contrast of elements. In particular, if element transformations are isoluminant color changes, illusory contour perception is greatly reduced (Cicerone, et al. 1995; Miyahara & Cicerone, 1997). First-order motion perception is

also known to be poor under isoluminance (Cropper 2005; Cropper & Derrington, 1994; Derrington & Henning, 1993).

Experiment 1

Despite mounting behavioral evidence in support of a basic edge-extraction process, using vectors relating element transformations, as an initial step to SBF, no working model that could take an SBF display as input and produce a local edge orientation as output has previously been implemented or tested. Doing so would require displays that focus more specifically on construction of a single orientation, whereas most prior work on SBF has used 2D shapes. As mentioned earlier, perception of shape and continuous illusory boundaries in such displays probably involves two stages of processing, the construction of local edge fragments *and* the connections of those fragments via spatiotemporal interpolation processes that connected oriented edge fragments across gaps (Palmer et al., 2006). Here, we focus on the first stage to evaluate models of it. We generated a display in which the SBF-defined shape was a single, thin, oriented bar that moved across a field of black, circular elements (Figure 3.3). Whenever the bar passed the midpoint of an element, that element disappeared (became white) all at once and remained invisible (white) for two frames at which point it reappeared (cf. Shipley & Kellman, 1993, 1994, 1997). This resulted in a sequence of element transformations as the bar moved across the display. The sequence could then be broken down into triplets of events from which bar orientation could be computed using equation 1. All that is needed to determine edge orientation are the relative element positions and the temporal interval between their transformations.

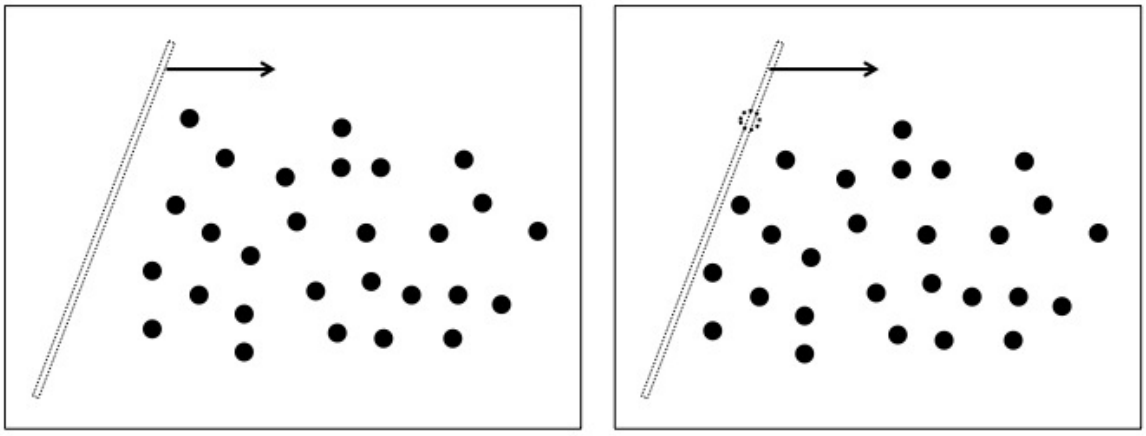

Figure 3.3. Illustration of stimuli used in Experiment 1. An invisible, oriented bar moved laterally across a field of black elements on a white background. Whenever it passed the midpoint of an element, that element disappeared (became white; indicated by dashed circle in second panel) all at once. The element remained white for two frames and then reappeared (became black). The perception was of a moving, illusory, white bar.

Human orientation discrimination thresholds were measured as a function of several display parameters: element density, number of element transformation events, and frame duration. These have previously been shown to affect the perception of illusory contours in SBF (Shipley & Kellman, 1994). If the model is correct, it should be able to accurately model performance under a variety of display settings and be affected by the same properties that affect human performance. The model described in equation 1 was used to predict edge orientation on a trial by trial basis in simulated experimental trials. This enabled the computation of an orientation discrimination threshold for the model, which was directly compared to human data. We discuss the properties of the model after presenting the behavioral results.

Method

Design

A between-subjects design was used to test the effects of three display properties on orientation discrimination of SBF-defined edges. Displays varied in element density (number of

elements per square region), number of transformation events, or frame duration. Each group of subjects was exposed to only one of the three display manipulations. All subjects performed an edge orientation discrimination task, judging whether an SBF-defined edge was tilted clockwise or counterclockwise away from vertical. Orientation sensitivity was measured for six densities, six element quantities, and three frame durations.

Participants

Subjects were 45 students from the University of California, Los Angeles, split into groups of 15 for each of the three experimental conditions. Subjects were compensated with course credit for participating. All reported having normal or corrected-to-normal vision. All subjects were naïve to the purposes of the experiments.

Apparatus

Stimuli were created and displayed using the MATLAB programming language and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Stimuli were presented on a Viewsonic G250 CRT monitor, which was powered by a MacPro 4 with a 2.66 GHz Quad-Core Intel Xeon processor and an NVidia GeForce GT120 graphics card. The monitor was set to a resolution of 1024x768 pixels and a refresh rate of 60 Hz.

Stimulus

Displays contained black, circular elements with a diameter of 10 pixels (0.25 degrees of visual angle) on a white background. The elements were placed within a 614.4 pixel by 614.4 pixel region (15.19° x 15.19°) centered on the computer monitor. The elements were pseudo-randomly arranged by dividing the display area into a grid of equally sized regions and placing a single element at a random position within each region. This placement method ensured that

there were no large areas in the display that lacked elements and also prevented their overlap while preserving a somewhat uniform distribution over the entire display (cf. Shipley & Kellman, 1993, 1994).

A one-pixel-wide bar was specified that spanned the height of the display. On each frame, the bar moved laterally 5 pixels (0.125 deg/frame, 7.5 deg/sec). Whenever the bar passed the midpoint of an element, that element disappeared (became white) for two frames (33.2 ms) and the bar paused. After two frames, the element reappeared (became black) and the bar continued moving. Elements appeared and disappeared discretely without gradual occlusion. The resulting percept was of a horizontally translating, illusory bar. Whether the bar started on the left side of the display and moved right or *vice versa* was randomized across trials. The movie lasted until the bar reached the opposite end of the screen. The bar only traveled across the screen once, so each element transformed only one time. A new arrangement of elements was generated for every trial.

On each trial, the bar was tilted clockwise or counterclockwise with respect to the vertical. The degree of tilt was set by an adaptive staircase procedure (Psi method (Kontsevish & Tyler, 1999) implemented in the Palamades Toolbox (Prins & Kingdom, 2009)) that was used to find the 75% orientation discrimination threshold. Whether the bar was rotated clockwise or counterclockwise was randomized across trials.

Three properties of the displays were manipulated, element density (number of elements in the display area), number of transformation events, and frame duration. One property was manipulated at a time, resulting in three experimental conditions. In the density condition, element density was varied by drawing 9, 16, 25, 36, 49, or 64 elements in the display area. These quantities corresponded to densities of 0.04, 0.07, 0.11, 0.16, 0.21, and 0.28 elements per

squared degree visual angle. A separate staircase was used for each density to determine orientation discrimination thresholds. The six staircases were interleaved and terminated after 50 trials.

In the event number condition, element density was held constant at one element per 0.28 square degrees of visual angle (the highest density in the density condition). Each display contained 64 elements. The trial lasted until the illusory bar came into contact with 9, 16, 25, 36, 49, or 64 elements. Starting horizontal position and motion direction of the bar were randomized with the constraint that there would be enough elements in the direction of motion that would allow for the required number of element contacts. As with density, six interleaved staircases were used to determine orientation discrimination thresholds for each element quantity.

In the temporal condition, 64 elements were placed with the highest density used from the density and event conditions. Frame durations were 16.7, 33.3, or 66.7 ms. Subjects were allowed to make a response at any point during the trial because at long frame durations the movies lasted for a long time. Three interleaved staircases were used, one for each frame duration. The shortest frame duration was the same frame duration that was used in the other conditions. As such, there was one display type that was identical in across all three conditions (64 elements, 64 events, 16.7 ms frame duration).

Procedure

Subjects sat in a dark room at a distance of 89.5 cm from the monitor. The only illumination came from the monitor. Subjects were instructed that they would be making orientation judgments about slanted edges and were shown examples of real edges that were tilted clockwise and counterclockwise. After each stimulus presentation, a response screen appeared asking whether the line was tilted clockwise or counterclockwise. Subjects made a

response by pressing a key on the keyboard. Before beginning experimental trials, subjects first performed 10 practice trials at the highest element density and quantity. Feedback was provided after each practice trial. Once complete, subjects were told that they would receive no further feedback and that in the rest of the experiment certain aspects of the displays would change such as the total number of elements, the display duration, or the speed of the illusory line. Short rest breaks were provided throughout the experiment every 100 trials.

Results and discussion

Orientation discrimination thresholds for each of the three conditions are shown in Figure 3.4 (black lines). The 75% correct orientation discrimination thresholds were computed for each subject for each condition and averaged across subjects. In the density condition, thresholds decreased as a function of density with the highest threshold of 19.08° for the lowest density and 3.17° for the highest. In the element quantity condition, displays with 36 or more element transformations had similar thresholds: 3.54, 3.30, and 3.36° for 36, 49, and 64 transformations respectively. Displays with 16 or 25 element transformations had slightly higher thresholds at 5.20° and 4.20° respectively. Displays with only nine element transformations had the highest thresholds, 8.8°. In the frame duration condition, average thresholds were similar for the two fastest durations, 3.90° and 3.66° respectively. However, sensitivity was worse for the longer, 66.7 ms frame duration, at 5.70°. Displays with the highest density, largest number of element transformations, and shortest frame duration (64 elements, 64 events, frame duration = 16.7 ms) appeared in all three conditions. Thresholds were between 3 and 4 degrees across all conditions.

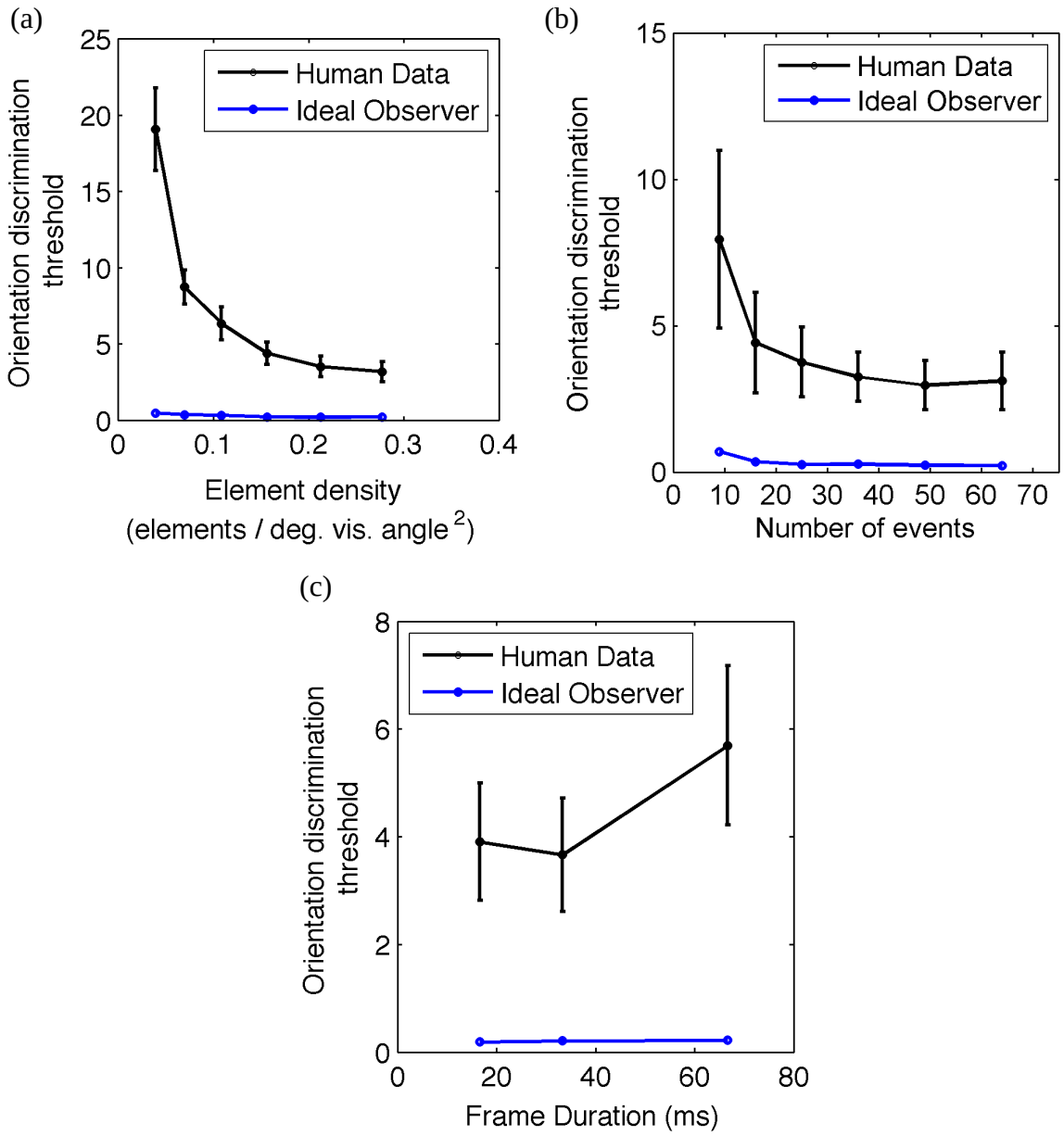

Figure 3.4. Average orientation discrimination thresholds for three display conditions tested in Experiment 1. Thresholds are shown as a function of element density (a), number of events (b), and frame duration (c). Human performance data are shown in black. Bars indicate the 95% confidence intervals. Ideal observer performance is shown in blue.

For each condition, data were submitted to a within-subject, one-way ANOVA to test for the effect of the manipulated display property. Increasing density (Fig. 4(a), black line), decreased thresholds (Mauchly's test: $\chi^2(14) = 43.85$ $p < 0.001$, Greenhouse-Geisser $\varepsilon = 0.37$,

$F(1.86, 26.06) = 87.77$, $MSE = 16.65$, $p < 0.001$, $\eta^2_p = 0.86$). Similarly, increasing the number of element transformations (Fig. 4(b), black line) decreased thresholds (Mauchly's test: $\chi^2(14) = 77.7$, $p < 0.001$, Greenhouse-Geisser $\varepsilon = 0.36$, $F(1.81, 27.13) = 10.08$, $MSE = 19.4$, $p=0.001$, $\eta^2_p = 0.40$). Increasing the inter-frame interval increased thresholds ($F(2, 28) = 7.78$, $MSE = 2.37$, $p = 0.002$, $\eta^2_p = 0.36$) (Fig. 4(c), black lines).

We performed post hoc, Bonferroni corrected, two-tailed, paired *t*-tests comparing the thresholds from the worst condition to the rest of the conditions. Performance in the density condition was worse in the lowest density condition compared to all others (all *p*s < 0.001). The threshold for the smallest number of transformations (9) was larger than for all other numbers except for 16 elements (all *p*s < 0.01). In frame duration condition, there was no significant difference between the two shortest frame durations. However, sensitivity for the longest duration was worse than either of the shorter delays (16.7 vs. 33.3 ms: $t(14)=3.88$, $p=0.0017$; 33.3 ms vs. 66.7 ms $t(14)=3.42$, $p=0.0042$). Performance for displays that had identical features for all three conditions was not significantly different across conditions (*p*s>0.05).

The results replicated previous findings that element density and frame duration affect shape perception in SBF (Shipley & Kellman, 1994). It has also been shown that the virtual object may be a single edge or thin bar if the elements are arranged in a regular pattern with equal spacing and inter-element transformation times (Erlikhman & Kellman, *submitted*). Here, we demonstrate that single edges can be recovered even when elements are arranged randomly with varying transformation times. These findings lend support to the notion that edges are indeed extracted as a first step in constructing complex shape representations in SBF. One interesting finding was that in the event quantity condition, performance continued to increase as a function of the number of events up to approximately 25 events, after which performance

leveled off. An earlier study of SBF also found that performance in a shape identification task improved as a function of number of frames shown, but only up to five frames, after which performance was constant (Shipley & Kellman, 1993). The number of events needed to reach best performance might depend on the difficulty of the task, complexity of the shape, and distance that the object travels per frame.

Ideal Observer Model

The ideal observer model described by equation 1 was used to predict bar orientation on a trial-by-trial basis for each of the conditions in Experiment 1. On each trial, the relative distances (A_{ij}), angular relationships (φ_{ij}), and timing (ΔT_{ij}) of element transformations were recorded for all elements directly from the displays by simulating the motion of an illusory edge. The sequence of events was divided into triplets and edge orientation was computed for each triplet. From n elements each of which transformed once, n-2 triplets were created, each triplet providing an estimate of bar orientation, θ. All elements except for the first and last appeared in multiple triplets. The median of the orientation estimates in a single trial was used to generate a "clockwise" or "counterclockwise" response. If the median was 90°, one of the two responses was chosen randomly. The responses were then submitted to the same staircase procedure for each of the experimental conditions and display settings. This resulted in estimates of orientation discrimination sensitivity for the model that could be directly compared to human sensitivity measures. Importantly, while the model output was an orientation, the comparison to human data was at the level of discrimination thresholds.

The model was able to predict edge orientation very accurately, producing orientation discrimination thresholds below one degree for all densities (blue line, Figure 3.4). However, in examining individual orientation estimates derived from a triplet, there was some deviation from

true orientation. Average orientation estimate error was 1.85° per triplet. When the median was taken across all orientation estimates computed from all triplets in a single trial, error was 0.37, 0.28, 0.24, 0.20, 0.17, and 0.14° for the six element densities from smallest to largest respectively. Error was reduced for higher density displays because there were more triplets that contributed to the final estimate. The model's performance may have been imperfect because the bar advanced in discrete steps of 5 pixels every 16.7 ms (i.e., every frame). This introduced error in the amount of time between element transformation events, which could only be in multiples of the frame rate. To test this explanation, a separate set of simulations was run for which the velocity of the bar was used to compute the time when an element should have transformed. Using these "true" times, average orientation estimate error was less than 0.1° per triplet. The model is therefore able, in principle, to perfectly determine edge orientation from three element transformation events. In all subsequent modeling, the timing correction was not applied because true bar velocity cannot be known *a priori*. Even without the correction, however, the model's median orientation estimates differed little (less than 0.5°) from true orientation, and the final model thresholds were well below those of human observers, even for the highest density (3.17°).

It is possible that simultaneous element transformation events could have affected both human and model performance. Simultaneous events could be used to perform the task perfectly: because the stimulus was a thin bar, simultaneous events could only have occurred if the orientation of the edge was the same as the angle between the event positions. Knowing the angle would therefore be sufficient to determine whether the edge was oriented clockwise or counterclockwise. Such a strategy would be particular to the displays used in this experiment. If the contour of the virtual object was curved or composed of more than a single edge, a straight

line connecting the positions of simultaneous events on different parts of the curve or on different edges would not correspond to the shape's contour. Nevertheless, if observers discovered that they could use simultaneous events to do the task, then one might have expected performance to have been near-perfect, especially for higher densities which had the highest frequencies of simultaneous events on a per-trial basis. On average, there were 0.21, 0.74, 1.93, 4.06, 7.56, and 12.73 simultaneous events per trial for each of the six element densities respectively. However, even at the highest density, average human thresholds were around 3.5°. If two simultaneous events over the course of a trial were sufficient to perform the task, we may have expected better performance. Furthermore, because the virtual edge spanned the height of the display, simultaneous events were often far apart. The average distance between simultaneous events for the highest density was 3.38°. Given that simultaneous events lasted for only 33.2 ms and that the elements were small, it would have been difficult for observers to detect them at all.

As a further check, we also performed a control experiment (not reported here) in which only a single element disappeared on every frame. Subjects reported seeing an illusory edge and orientation discrimination thresholds were very similar to those found in Experiment 1. The model was also able to perfectly predict edge orientation in these displays. Human and model performance therefore did not depend on the presence of simultaneous transformation events to drive performance.

Experiment 2

Experiment 1 indicated that observers could very accurately discriminate between edge orientations of illusory edges defined by SBF and that their sensitivity depended on the spatial and temporal properties of the displays. Human performance, however, was far worse than the ideal observer model, especially for low density displays and displays with few transformation

events. One limitation on human performance could be spatial and temporal integration limits beyond which events cannot be combined to recover edge orientation. In sparse displays, edges may not be formed because elements are far apart from one another and the temporal intervals between their transformations are long. Similar integration limits exist, for example, in apparent motion, where the perception of motion between two alternatively flashing elements is constrained by the inter-element distance and the element flash timing (Wertheimer, 1912; Korte 1915). Under minimal conditions, when few elements transform one at a time, perception of illusory contours is strongly degraded for inter-element distances greater than one degree of visual angle or inter-stimulus intervals greater than 80 ms (Erlikhman & Kellman, *submitted*). For dense displays with random arrangement of elements and large, extended objects similar to those used in Experiment 1, perception of SBF improves with the number of frames that can be fit into a 165 ms temporal window, with additional frames adding little or no additional benefit. In the event quantity condition in Experiment 1, performance improved with increasing number of events up to 25-36, beyond which there was no added benefit to sensitivity. This may reflect a maximum threshold on the number of events that can be usefully integrated. Additionally, both Experiment 1 and prior work demonstrate a gradual reduction in SBF perception as a function of the display's spatial and temporal properties, suggesting an effect of noise. Inaccurate representations of the distance between elements, for example, would affect the computation of the inter-element velocity signal and ultimately the estimate of the bar's orientation.

There are several possible low-level sources of noise in these displays. Accurate representation of the edges position and velocity requires precise estimates of element positions and transformation times. The ideal observer model showed that even slight deviations from correct temporal values could cause a 1.8° error in the orientation estimate. Additional sources of

noise arise from imprecise measurement of inter-element position or the angular separation between elements. Temporal, spatial, and angular distances are the three variables in equation 1 used to derive edge orientation. Experiment 2 was designed to empirically measure sensitivity to these quantities in SBF-like displays. Noisy measurement of low-level stimulus properties has previously been used to account for error in visual speed perception (Hürlimann, Kiper, & Carandini, 2002; Stocker & Simoncellli, 2006) and cue reliability of spatial and orientation signals in biological motion (Thurman & Lu, 2014). These low-level sources of noise may result in mis-estimation of edge orientation, resulting in poorer sensitivity. Including them in the model might be able to account for the suboptimal human performance.

Method

Participants

Subjects were 3 volunteers from the University of California, Los Angeles, and one of the authors, GE. All reported having normal or corrected-to-normal vision. Two of the subjects were experienced psychophysical observers.

Design

A method of constant stimuli design was used to measure sensitivity to spatial, temporal, and angular separation between pairs of circular elements flashed successively for various distances, temporal intervals, and angles. Subjects performed a two-interval, forced choice task in which they selected the interval that contained the flashed pair of elements that were either closest together in space, closest together in time, or which formed the smallest angle relative to horizontal. Seven spatial distances from 20 to 140 pixels (0.5 to 3.5 degrees of visual angle) were used as references. Each reference was compared to 10 other distances to obtain points on a psychometric function. A similar procedure was used for six reference temporal intervals (50 to

300 ms) and five reference angular differences (15 to 75º). For each reference, a cumulative Gaussian function was fit to the data to obtain mean and standard deviation estimates. Spatial, temporal, and angular sensitivity data were collected between subjects.

Stimuli

The apparatus was the same as that for Experiment 1. Stimuli consisted of a background array of 400 randomly placed white, circular elements (diameter = 0.25º) on a black background and two pairs of target elements which were identical to the background elements. All elements appeared within a 13.69º by 13.69º area centered on the middle of the screen. Each trial was composed of two intervals. In each interval, the elements of one pair flashed (disappeared) for 50 ms and then reappeared one at a time. The positions and times of flashes defined the spatial, temporal, and angular relationship between elements in a pair. One pair of elements defined a reference distance, temporal interval, or angle, and the other a corresponding comparison value.

In the spatial task, seven reference distances were used: 20, 40, 60, 80, 100, 120, and 140 pixels (0.50, 1.00, 1.49, 1.99, 2.49, 2.98, and 3.48 degrees of visual angle, respectively). These were the Euclidean distances between elements in one of the target pairs. Comparison distances for the other pair were offset by between -50 and 50 pixels (-1.24º to 1.24º) from the reference distance. Ten comparison distances were used for each reference to cover a range of values to along the psychometric function. Whether target elements in the first and second interval were separated by the reference distance and comparison distance respectively or vice versa was randomized across trials. Each element pair was centered on a random position within the display area. The angle formed between the elements in a pair and the horizontal was randomized across trials, but was the same for both pairs of elements within a trial.

A similar pairing of reference and comparison was used for temporal and angular separations. For the temporal task, six reference durations were used: 50, 100, 150, 200, 250, and 300 ms. Ten comparison durations were used for each reference, with offsets in the range of -180 to 180 ms. For this task only, the monitor refresh rate was set to 100 Hz to allow intervals to occur in steps of 10 ms. For angular separation, five reference angles were used from 15º to 75º in steps of 15º. Ten comparison angles were used for each reference, with offsets in the range of -22º to 22º. For both temporal and angular tasks, elements within a pair were 3.75º apart. Element pairs in the temporal task appeared at random angular positions across trials, but the formed the same angle within a trial (across pairs). Each reference-comparison pair for each task was tested 20 times. Trial order was randomized.

Before target elements in a pair flashed, a red outline of a square (7.45º by 7.45º) centered on the elements appeared for 300 ms. Pilot work had found that without this attentional cue, observers often missed the disappearance of one or both elements in a target pair. This attentional cue occurred at the beginning of each interval, before the first element of an element pair flashed. Even with the attentional cue, it was sometimes difficult to detect an element flash. In order to prevent guessing in such cases, subjects were allowed to press a key to repeat a trial. The same reference and comparison values were used, but a new display was generated with background and target elements appearing in new positions and with the interval order randomized.

Procedure

Subjects sat at a distance of 89.5 cm from the monitor and had their heads stabilized by a chin-rest. Subjects were given verbal instructions that they would be making discrimination judgments between the spatial, temporal, or angular distances defined by the flashing of two dots

in a field of dots. A trial began with all elements, background and target, displayed on the screen. After 300 ms, the outline of a red, square appeared centered on the first target element pair. This marked the beginning of the first interval. The square remained on the screen for 300 ms and then disappeared. After a further 300 ms, the first element of the first element pair disappeared for 50 ms and reappeared. The second element in the pair then disappeared for 50 ms. In the temporal duration task, a pause was inserted after the reappearance of the first element and before the disappearance of the second. This pause defined the temporal interval about which subjects made a judgment. Once the second element reappeared, all elements remained on the screen for another 300 ms, at which point the second interval began. An attentional square was again shown for 300 ms and the second pair of elements flashed one at a time. After the last target element reappeared, the display remained on the screen for another 300 ms and was then replaced by a blank, black screen. White text instructed subjects to make a response by pressing one of two keys on the keyboard to indicate whether the first or second interval contained the pair of target elements that were farthest apart (spatial task), that flashed furthest apart in time (temporal task), or that formed the smallest angle with the horizontal (angular task). If subjects missed one or more target element flashes, they were instructed to press a third key to repeat a trial. Subjects were explicitly instructed not to repeat trials in which they were unsure of the answer, but saw all four target element flashes. Subjects were given a break every 100 trials. An illustration of a trial sequence is shown in Figure 3.5. The three noise conditions were run independently in separate sessions. Each session lasted approximately one and a half hours.

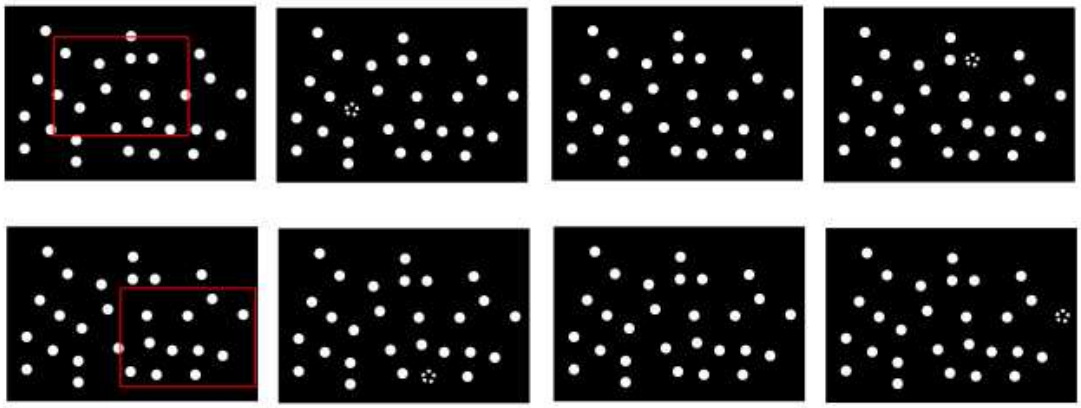

Figure 3.5. An illustration of a trial in Experiment 2. Each row depicts one interval. A region of region of the display was cued in which element transformations would occur (first panel). An element within that region would disappear (second panel, indicated by dashed boundary) and reappear (third panel). A second element would disappear, also within the cued region (fourth panel). The two elements define a spatial, angular, and temporal value, which is compared to one in a subsequent interval (second row).

Results and Discussion

For each reference-comparison pair the percentage of largest distance, longest duration, and smallest angle responses was computed for each of the three tasks respectively. Cumulative normal distributions were fit to the data for each reference and for each subject separately using a non-linear least squares procedure and the mean and standard deviation of the functions were estimated. The standard deviation estimates appear in Tables 3.1-3.3 for all subjects in each of the three tasks.

Observer	**0.5 deg**	**1 deg**	**1.5 deg**	**2 deg**	**2.5 deg**	**3 deg**	**3.5 deg**
GE	0.16	0.18	0.24	0.26	0.44	0.33	0.47
RO	0.27	0.35	0.35	0.32	0.42	0.41	0.79
SC	0.20	0.20	0.19	0.29	0.25	0.37	0.35
YX	0.75	0.54	0.68	0.96	0.88	1.00	1.58
Avg.	0.35	0.32	0.36	0.46	0.50	0.53	0.80

Table 3.1. Standard deviation estimates for each subject in the spatial task for each of the seven reference distances.

Observer	**50 ms**	**100 ms**	**150 ms**	**200 ms**	**250 ms**	**300 ms**
GE	68.54	133.45	99.80	91.59	107.02	86.23
RO	48.28	27.8014	50.83	51.53	69.63	62.15
SC	116.63	61.89	67.92	72.83	96.19	120.48

YX	182.43	181.26	147.96	192.33	137.94	153.60
Avg.	103.97	101.10	91.63	102.07	102.70	105.61

Table 3.2. Standard deviation estimates for each subject in the temporal task for each of the six reference temporal durations.

Observer	**15º**	**30º**	**45º**	**60º**	**75º**
GE	6.96	7.94	12.85	9.27	7.87
RO	7.89	13.51	9.88	14.31	10.23
SC	6.14	6.41	8.17	12.13	6.610
YX	5.45	14.84	20.58	19.33	13.46
Avg.	6.61	10.67	12.87	13.76	9.54

Table 3.3. Standard deviation estimates for each subject in the angular task for each of the five reference distances.

Typically, empirical noise estimates are used to model one individual's performance at a time because sensitivity varies from subject to subject. This can be seen, for example, in subject YX's standard deviation estimates for distance sensitivity (Table 3.1), which are two to three times larger than those of the other subjects across all reference distances. However, as a first step, we sought to test as simple and general a model as possible. Standard deviation estimates were therefore averaged across all references and all subjects resulting in average estimates of 0.47 degrees of visual angle, 101.18 ms, and 10.69º for spatial, temporal, and angular separation respectively. On the one hand, these average quantities may have overestimated the amount of noise and glossed over subtle differences as a function of reference value (e.g., increasing variability with increasing inter-element distance). On the other hand, average values may be more appropriate to apply to the data from Experiment 1, in which subjects were naïve observers who were not specifically instructed to pay attention to inter-event properties. Instead of examining subject-specific fits, we therefore used the average data from Experiment 2 to attempt to predict performance from a completely different group of subjects from Experiment 1.

Model Results

The ideal observer model introduced in Experiment 1 was modified by adding two constraints and three sources of noise. The first constraint was on the number of integrated elements from which the final orientation estimate was derived. In the event quantity condition, additional elements beyond 25-36 did not affect sensitivity. We therefore restricted the number of elements to be integrated to 25. A sequence of 25 consecutive element transformations was sampled for each trial, and orientation estimates were derived only from triplets within that set of elements. The second constraint was temporal – triplets containing inter-event times greater than 165 ms were excluded from the final set from which the average orientation was computed. Previous work using similar displays found that perception of SBF was greatly reduced beyond this limit (Shipley & Kellman, 1994).

The average noise parameters estimated in Experiment 2 were applied by including additive noise to the spatial (A), temporal (ΔT), and angular (φ) inter-element properties as indicated in equation 1. It was assumed that noise was normally distributed with a mean of zero and a standard deviation given by the noise estimates derived in Experiment 2. For spatial and temporal noise, truncated normal distributions were used to ensure that the sampled values were non-negative. Each condition in Experiment 1 was simulated ten times and the average of the threshold estimates from the ten simulations are shown in Figure 3.6.

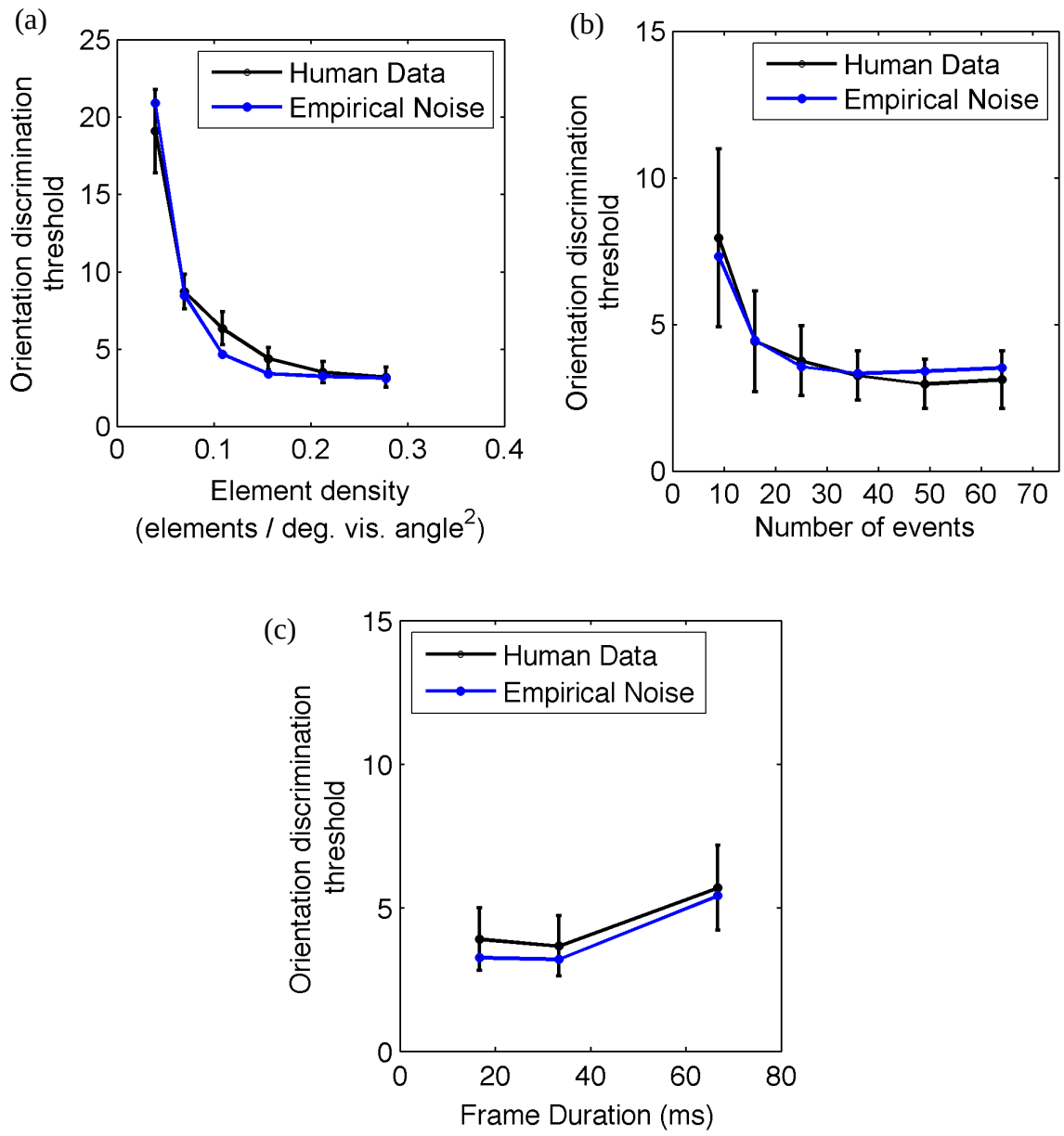

Figure 3.6. Data from Experiment 1 (black) replotted with model fits (blue) using noise parameter estimates from Experiment 2. Error bars are 95% confidence intervals. Model data reflect the average of 10 simulated experiment runs.

Model performance was measured by computing the root mean squared error between the human data and the model fit. The overall fits were highly accurate across all conditions: element density RMSE = 3.42; event quantity RMSE = 0.89; frame duration RMSE = 0.83. As a comparison, for the density condition, we examined models that did not include the element

quantity and temporal constraints, as well as models that only included those constraints and added no noise. Both constraint-only and noise-only models provided worse fits to the data: constraint-only RMSE = 11.18; noise-only RMSE = 13.15.

Increasing noise as a function of inter-element separation in Experiment 2 suggests that spatial noise may be multiplicative in nature rather than additive. A role for multiplicative computation has been suggested for looming signals (Gabbiani, Krapp, Koch, & Laurent, 2002), contrast-gain control (Albrecht & Geisler, 1991; Määttänen & Koenderink, 1991), and orientation selectivity (Beaudot & Mullen, 2005). To test whether multiplicative noise might better capture human performance, the distance data from Experiment 2 were log-transformed refit with a cumulative normal, and the standard deviation for each distance for each subject was computed. The average across all references and subjects was 0.30. The model was then rerun and compared to human data from Experiment 1. The multiplicative noise model was able to fit the density condition slightly better than the additive noise model (RMSE = 3.11), but was worse for the number of events (RMSE = 1.74) and frame duration (RMSE = 2.15) conditions.

General Discussion

The current study provides new evidence in support of a model of spatiotemporal boundary formation (SBF) that extracts local, oriented edge fragments by solving the point aperture problem. We have recently demonstrated behaviorally that such edge fragments can indeed be recovered in minimal displays in which elements disappear and reappear one at a time, suggesting that they serve as the basic units from which global shapes are constructed in SBF (Erlikhman & Kellman, *submitted*). Although a model had been proposed for how such fragments may be recovered solely from the positions of elements and times of their transformations (i.e., by solving the point aperture problem), the model had not been implemented or applied to behavioral data (Shipley & Kellman, 1994, 1997). Here, we present

the first evidence that not only can such a model determine edge orientation given accurate input, but that it can also very precisely model human performance across a variety of conditions using only a few constraints and sources of noise. Importantly, this was possible without any parameter fitting. In addition, although the model made predictions about edge orientation on a trial-by-trial basis, the final comparison to human data was performed on orientation discrimination thresholds that were generated by a staircase procedure. It is all the more impressive that it was possible to fit these data with a mechanistic model.

An ideal observer model demonstrated that edge orientation could be unambiguously determined from three, non-collinear element transformations. In Experiment 1, special displays were created in which only a single illusory edge was visible. It was possible to model human sensitivity to the edge's orientation across a variety of display conditions known to affect perception of SBF by introducing two integration constraints and noise to the estimates of element positions and event timings. The two constraints were: (1) element transformations had to occur within 165 ms of each other, a cutoff beyond which illusory contours are not seen (Shipley & Kellman, 1994), and (2) orientation estimates were computed from a subset of 25 events. In Experiment 1, orientation sensitivity was similar for displays containing 25 events or more. The amount of variability or noise in measuring inter-element distance, angular separation between elements, and the temporal interval between events (the three variables in model) were estimated on a separate group of subjects (Experiment 2) and their averages were used to predict performance in Experiment 1.

There are several limitations to the model. First, it is assumed that the velocity and orientation of the illusory edge is constant. We have recently shown that SBF supports a wide range of transformations of the pseudosurface including scaling, rotation, acceleration, and non-

rigid transformation (Erlikhman, Xing, & Kellman, *in press*). Second, the model only gives the orientation for a single edge. In displays with 2D shapes or curved edges, the model would need some spatial parameter that limits the integration of element transformations to small neighborhoods to allow for the extraction of edge information along all portions of the contour. For example, if the illusory figure is a circle, then there must be some way of keeping separate element transformations on opposite sides of the circle. Future work is needed to address these concerns.

For a moving object, two frames from a motion sequence are sufficient to solve an aperture problem that occurs locally for each contour (e.g., Weiss, Simoncelli, & Adelson, 2002). Such a solution is impossible for the point aperture problem, particularly in SBF displays in which edges can be seen even when only a single element transforms from frame to frame (Erlikhman & Kellman, *submitted*). If a biological process instantiates the computation performed by the model, it must be integrating element transformation events over an extended region of space and period of time. This brings to mind a class of motion-energy models with spatiotemporal filters that are thought to reflect neural properties and which can detect moving, contrast-defined edges over time (Adelson & Bergen, 1985; Challinor & Mather, 2010; van Santen & Sperling, 1984). However, such models predict nearest-neighbor apparent motion between transforming elements. An alternative possibility is that a set of large, oriented motion filters that capture the transformation of several elements may be used to determine edge orientation. Evidence for the existence of such filters have been found in primates in V1 (Maracar, Raiguel, Xiao, & Orban, 2000; Schmid, 2008), V2 (Lu et al, 2010; Chen et al., 2014) and MT (Marcar & Cowey, 1992; Marcar, et al., 1995). If an explanation in terms of these detectors is possible, then SBF is not simply an esoteric visual illusion, but is actually the result

of a fundamental visual process involved in the extraction of edges, motion, and their interaction. We are currently exploring the possibility of linking models of SBF to the known properties of oriented motion energy filters.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perceptions of motion. *Journal of the Optical Society of America, A*, *2*(2), pp. 284-299.

Ahumada Jr., A. J. (2002). Classification image weights and internal noise level estimation. *Journal of Vision, 2*, pp. 121-131.

Albrecht, D. G., & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in visual cortex.

Andersen, G. J., & Cortese, J. M. (1989). 2-D contour perception from kinematic occlusion. *Perception and Psychophysics*, *46*, pp.49-55.

Anstis, S. M. & Atkinson, J. (1967). Distortions in moving figures viewed through a stationary slit. *The American Journal of Psychology*, *80*, pp.572-585.

Aydin, M., Herzog, M. H., & Öğmen, H. (2008). Perceived speed differences explain apparent compression in split viewing. *Vision Research*, *48*(15), pp. 1603-1612.

Barraza, J. F., & Chen, V. J. (2006). Vernier acuity of illusory contours defined by motion. *Journal of Vision, 6*, 923-932.

Beaudot, W. H. A., & Mullen, K. T. (2005). Orientation selectivity in luminance and color vision assessed using 2-d band-pass filtered spatial noise. *Vision Research*, *45*(6), pp. 687-696.

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

Braunstein, M. L., Andersen, G. J., & Riefer, D. M. (1982). The use of occlusion to resolve ambiguity in parallel projections. *Perception & Psychophysics*, *31*, pp. 261-267.

Bruno, N., & Bertamini, M. (1990). Identifying contours from occlusion events. *Perception & Psychophysics*, *48*, pp. 331-342.

Bruno, N., & Gerbino, W. (1991). Illusory figures based on local kinematics. *Perception*, *20*, pp. 259-274.

Challinor, K. L., & Mather, G. (2010). A motion-energy model predicts the direction discrimination and MAE duration of two-stroke apparent motion at high and low retinal luminance. *Vision Research*, *50*, pp. 1109-1116.

Chen, V. J., & Cicerone, C. M. (2002a). Subjective color from apparent motion. *Journal of Vision, 2*, pp. 424-437.

Chen, V. J., & Cicerone, C. M. (2002b). Depth from subjective color and apparent motion. *Vision Research, 42*, pp. 2131-2135.

Chen, M., Li, P., Zhu, S., Han, C., Xu, H., Fang, Y., Hu, J., Roe, A. W., & Lu, H. D. (2014). An orientation map for motion boundaries in V2. *Cerebral Cortex*, bhu235.

Cicerone, C. M., & Hoffman, D. D. (1997). Color from motion: Dichoptic activation and a possible role in breaking camouflage. *Perception, 26*, pp. 1367-1380.

Cicerone, C. M., Hoffman, D. D., Gowdy, P. D., & Kim, J. S. (1995). The perception of color from motion. *Perception & Psychophysics*, *57*, pp. 761-777.

Cooke, T., Cunningham, D. W., & Bulthoff, H. H. (2004). The perceptual influence of spatiotemporal noise on the reconstruction of shape from dynamic occlusion. *Lecture Notes in Computer Science, 3175*, pp. 407-414.

Cropper, S. J. (2005). The detection of motion in chromatic stimuli: first-order and second-order spatial structure. *Vision Research, 45,* pp. 865-880.
Cropper, S. J., & Derrington, A. M. (1994). Motion of chromatic stimuli: First-order or second-order? *Vision Research*, *34*(1), pp. 49-58.
Cunningham, D. W., Shipley, T. F., & Kellman, P. J. (1998). The dynamic specification of surfaces and boundaries. *Perception*, *27*(4), pp. 403-415.
Cunningham, D. W., Shipley, T. F., & Kellman, P. J. (1998). Interactions between spatial and spatiotemporal information in spatiotemporal boundary formation. *Perception & Psychophysics*, *60*(5), pp. 839-851.
De Weerd, P., Vandenbussche, E., & Orban, G. A. (1993). Occlusion cues contribute to orientation judgments of occlusion-defined contours. *Perception & Psychophysics*, *54*(6), pp. 706-715.
Derrington, A. M., & Henning, G. B. (1993). Detecting and discriminating the direction of motion of luminance and colour gratings. *Vision Research*, *33*(5/6), pp. 799-811.
Erlikhman, G., & Kellman, P. J. Recovery of local edge fragments initiates spatiotemporal boundary formation. *Submitted.*
Erlikhman, G., Xing, Y. Z., & Kellman, P. J. Non-rigid illusory contours and global shape transformations defined by spatiotemporal boundary formation. *In Press.*
Fidopiastis, C., Hoffman, D. D., Prophet, W. D., & Singh, M. (2000). Constructing surfaces and contours in displays of color from motion: The role of nearest neighbors and maximal disks. *Perception, 29*, pp. 567-580.
Gabbiani, F., Krapp, H. G., Koch, C., & Laurent G. (2002). Multiplicative computation in a visual neuron sensitive to looming. *Nature, 420*, pp. 320-324.
Gibson, J. J., Kaplan, G. A., Reynolds, H. N., & Wheeler, K. (1969). The change from visible to invisible: A study of optical transitions. *Perception and Psychophysics, 3i, pp. 113-116.*
Hürlimann, F., Kiper, D. C., & Carandini, M. (2002). Testing the Bayesian model of perceived speed. *Vision Research, 42*, pp. 2253-2257.
Julesz, B. (1971). *Foundations of Cyclopean Perception*. University of Chicago Press: Chicago.
Julesz, B. (1975). Experiments in the visual perception of texture. *Scientific American, 232*, pp. 34-43.
Kalar, D. J., Garrigan, P., Wickens, T. D., Hilger, J. D., & Kellman, P. J. (2010). A unified model of illusory and occluded contour interpolation. *Vision Research*, *50*(3), pp. 284-299.
Kaplan, G. A. (1969). Kinetic disruption of optical texture: The perception of depth at an edge. *Perception & Psychophysics*, *6*, pp. 193-198.
Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), pp. 2729-2737.
Korte, A. (1915). Kinematoskopishe Untersuchungen. *Zeitschrift für Psychologie mit Zeitschrift für angewandte Psychologie, 72,* pp. 194-296.
Larsson, J., Heeger, D. J., & Landy, M. S. (2010). Orientation selectivity of motion-boundary responses in human visual cortex. *Journal of Neurophysiology*, *194*(6), pp. 2940-2950.
Lu, H. D., Chen, G., Tanigawa, H., & Roe, A. W. (2010). A motion direction map in macaque V2. *Neuron, 68*, pp. 1002-1013.
Lui, L. L., Dobiecki, A. E., Bourne, J. A., & Rosa, M. G. P. (2012). Breaking camouflage: responses of neurons in the middle temporal area to stimuli defined by coherent motion. *European Journal of Neuroscience*, *36*(1), pp. 2063-2076.

Mateeff, S., Popov, D., & Hohnsbein, J. (1993). Multi-aperture viewing: perception of figures through apertures. *Vision Research*, *33*(17), pp. 2563-2567.
Määttänen, L. M., & Koenderink (1991). Contrast adaptation and contrast gain control. *Experimental Brain Research*, *87*, pp. 205-212.
Marcar, V. L., & Cowey, A. (1992). The effect of removing superior temporal cortical motion areas in the macaque monkey: II. Motion discrimination using random dot displays. *European Journal of Neuroscience*, *4*, pp. 1228-1238.
Marcar, V. L., Raiguel, S. E., Xiao, D., & Orban, G. A. (2000). Processing of kinetically defined boundaries of the macaque monkey. *Journal of Neurophysiology*, *84*(6), pp. 2786-2798.
Marcar, V. L., Xiao, D. K., Raiguel, S. E., Maes, H., & Orban, G. A. (1995). Processing of kinetically defined boundaries in the cortical motion area MT of the macaque monkey. *Journal of Neurophysiology*, *74*, pp. 1258-1270.
Mysore, S. G., Vogels, R., Raiguel, S. E. & Orban, G. A (2008). Shape selectivity for camouflage-breaking dynamic stimuli in dorsal V4 neurons. *Cerebral Cortex*, *18*(6), pp. 1429-1443.
Nakayama, K., & Silverman, G. H. (1988a). The aperture problem—I. Perception of nonrigidity and motion direction in translating sinusoidal lines. *Vision Research*, *28*(6), pp. 739-746.
Nakayama, K., & Silverman, G. H. (1988b). The aperture problem—II. Spatial integration of velocity information along contours. *Vision Research*, *28*(6), pp. 747-753.
Nawrot, M., Shannon, E., & Rizzo, M. (1996). The relative efficacy of cues for two-dimensional shape perception. *Vision Research*, *36*(8), pp. 1141-1152.
Ono, H., Rogers, B. J., Ohmi, M., & Ono, M. (1989). Dynamic occlusion and motion parallax in depth perception. *Perception*, *17*, pp. 255-266.
Palmer, E. M., Kellman, P. J., & Shipley, T. F. (2006). A theory of dynamic occluded and illusory object perception. *Journal of Experimental Psychology: General*, 135(4), 513-541.
Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437-442.
Prazdny, K. (1986). Illusory contours from inducers defined solely by spatiotemporal correlation. *Perception & Psychophysics*, *39*, 175-178.
Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. http://www.palamedestoolbox.org
Prophet, W. D., Hoffman, D. D., & Cicerone, C. M. (2001). "Contours from apparent motion: A computational theory" in *From fragment to objects: Segmentation and grouping in vision*. P. Kellman and T. Shipley (Eds), Amsterdam: Elsevier Science Press, pp. 509-530.
Rogers, B. J., & Graham, M.E. (1983). Dynamic occlusion in the perception of depth structure. *Perception*, *12*, A15.
Rovden, C. S., Baker, J. F., & Allman, J. (1988). Perceptions of depth elicited by occluded and shearing motions of random dots. *Perception*, *17*, pp. 289-269.
Sary, G., Vogels, R., & Orban, G. A. (1994). Orientation discrimination of motion-defined gratings. *Vision Research*, *34*(10), pp. 1331-1334.
Schmid, A. M. (2008). The processing of feature discontinuities for different cue types in primary visual cortex. *Brain Research*, *1238*(31), pp. 59-74.
Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision research*, *29*(5), pp. 619-626.

Shipley, T. F., & Kellman, P. J. (1993). Optical tearing in spatiotemporal boundary formation: When do local element motions produce boundaries, form, and global motion? *Spatial Vision, 7*(3), pp. 323-339.
Shipley, T. F., & Kellman, P. J. (1994). Saptiotemporal boundary formation: Boundary, form, and motion perception from transformations of surface elements. *Journal of Experimental Psychology: General*, *123*, pp. 3-20.
Shipley, T. F., & Kellman, P. J. (1997). Spatio-temporal boundary formation: the role of local motion signals in boundary perception. *Vision Research*, *27*(10), pp. 1281-1293.
Sinha, P. (2001). Role of motion integration in contour perception. *Vision Research*, *41*(6), pp. 705-710.
Stappers, P. J. (1989). Forms can be recognized from dynamic occlusion alone. *Perceptual and Motor Skills*, *68*, pp.243-251.
Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), pp. 578-585.
Thurman, S. M., & Lu, H. (2014). Bayesian integration of position and orientation cues in perception of biological and non-biological forms. *Frontiers in Human Neuroscience, 8*:91. doi: 10.3389/fnhum.2014.00091.
van Santen, J. P. H., & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America, A, 1*, pp. 451-473.
Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), pp. 598-604.
Wertheimer, M. (1912). Experimentelle Studien über das Segen von Bewegung. *Zeitschrift für Psychologie*, *61*, pp. 161-265.
Yonas, A., Craton, L. G., & Thompson, W. B. (1987). Relative motion: Kinetic information for the order of depth at an edge. *Perception & Psychophysics*, *41*, pp. 53-59.