Recovering metric properties of objects through spatiotemporal interpolation ☆

Tandra Ghose [a,*], Janelle Liu [b], Philip J. Kellman [b]

[a] Department of Psychology, University of Kaiserslautern, Germany
[b] Department of Psychology, U.C. Los Angeles, USA

ABSTRACT

Spatiotemporal interpolation (STI) refers to perception of complete objects from fragmentary information across gaps in both space and time. It differs from static interpolation in that requirements for interpolation are not met in any static frame. It has been found that STI produced objective performance advantages in a shape discrimination paradigm for both illusory and occluded objects when contours met conditions of spatiotemporal relatability. Here we report psychophysical studies testing whether spatiotemporal interpolation allows recovery of metric properties of objects. Observers viewed virtual triangles specified only by sequential partial occlusions of background elements by their vertices (the STI condition) and made forced choice judgments of the object's size relative to a reference standard. We found that length could often be accurately recovered for conditions where fragments were relatable and formed illusory triangles. In the first control condition, three moving dots located at the vertices provided the same spatial and timing information as the virtual object in the STI condition but did not induce perception of interpolated contours or a coherent object. In the second control condition oriented line segments were added to the dots and mid-points between the dots in a way that did not induce perception of interpolated contours. Control stimuli did not lead to accurate size judgments. We conclude that spatiotemporal interpolation can produce representations, from fragmentary information, of metric properties in addition to shape.

1. Introduction

One of the important functions of visual perception is to provide descriptions of objects in the world, and thereby facilitate interaction with them. Many tasks require not only perceiving an object's shape but also metric properties such as its size. For example, even though a car and a miniature hot-wheels toy car may have the same shape, the way we interact with them is drastically different. Recovering size accurately is often fundamental in our interactions with the environment, for both behavioral tasks, e.g., reaching and grasping, as well as cognitive tasks, such as thinking and planning.

Obtaining useful object descriptions through vision is complicated by the fact that often only fragmentary information from the object reaches our eyes. Recent research has focused on

perhaps the most challenging version of fragmentation in visual information: cases where only parts of objects ever project to the eyes and these parts become available sequentially in time. Rather than being exotic phenomena unique to perception laboratories, these cases are quite common in a world in which objects and observers move, and in which occlusion is frequent. The question investigated in the current research is whether it is possible to recover characteristics of object size (e.g., the length of the object) when visual information does not specify the length in any momentary view. Actually, the question of metric perception from spatiotemporal interpolation examined here is even more demanding: Can we recover size of an object when that object never exists in any momentary view? Thus, an alternative title of this article could be: "What is the size of an object that is never present in the stimulus?"

In ordinary perception, most objects are partially visible due to occlusion by other opaque objects. For a moving observer, moreover, the pattern of occlusion changes with viewpoint; as a result, a full view of the object may never be present at any instance. Vision researchers have discovered a great deal about spatial interpolation processes that can produce perception of interpolated

contours and complete shape despite partial occlusion or in cases of camouflage, such as illusory contours and objects in laboratory settings (Fantoni & Gerbino, 2003; Field, Hayes, & Hess, 1993; Kanizsa, 1979; Kellman & Shipley, 1991; Lorenceau & Shiffrar, 1992, 1999; Petry & Meyer, 1987; Rubin, 2001; Tse, 1999; von der Heydt, Peterhans, & Baumgartner, 1984) and natural scenes (Elder & Goldberg, 2002; Geisler & Perry, 2009; Geisler et al., 2001).

Only more recently has research focused on the even more challenging case of *spatiotemporal interpolation*, in which the inputs to interpolation processes are obtained over time. Such cases are important in ordinary perception, as in interacting with objects in the environment; a moving observer frequently encounters changing patterns of occlusion, yet still may benefit from recovering the structure of objects and scenes. The role of motion has been explored for various aspects of perceptual grouping, including the influence of common motion of parts in unit formation (e.g., Johnson et al., 2003; Kellman, Gleitman, & Spelke, 1987; Kellman & Spelke, 1983), the role of motion information for perceiving illusory and occluded stimuli (Anderson & Barth, 1999; Anderson & Sinha, 1997; Anderson, O'Vari, & Barth, 2011; Bruno & Bertamini, 1990; Kellman & Cohen, 1984; Lorenceau & Shiffrar, 1992, 1999; Shipley & Cunningham, 2001), and the perception of boundaries and form in the absence of oriented edge inputs (Bruno, 2001; Cunningham, Shipley, & Kellman, 1998a, 1998b; Shipley & Kellman, 1990, 1992, 1993, 1994, 1997; Stanley & Rubin, 2005). Previous research shows that observers can recover and discriminate the shapes of dynamically occluded objects when their visible fragments satisfy certain spatial and temporal relations, formalized as *spatiotemporal relatability* (Palmer, Kellman, & Shipley, 2006). However, previous research has not shown whether metric properties can be effectively recovered by spatiotemporal interpolation for such dynamically occluded objects.

Metric properties can be recovered from visual units formed by either of the two different processes reported in literature, namely, "edge-insensitive" process (Kellman, 1996) or "edge-sensitive" process (Kellman & Cohen, 1984). For "edge-insensitive" process, motion information connects spatially separated visible parts through processes that depend on motion alone, as in the Gestalt principle of common fate. While for "edge-sensitive" processes unit formation is critically based on relations of oriented edges that lead to spatiotemporal interpolation if they satisfy the requirements for contour relatability (Kellman & Cohen, 1984; Palmer, Kellman, & Shipley, 2006). Length can be recovered from "edge-insensitive" processes by using motion information and the distance between spatial markers that appear at different time instances. For "edge-sensitive" processes, spatiotemporal interpolation based on edge relations is crucial. Edge-sensitive interpolation imposes greater constraints on the inputs, but as a result yields more in its outputs: Interpolation by edge-sensitive processes leads to detailed representations of interpolated boundary positions and clear shape. When object fragments are related by common motion alone (without relatable edges), unit formation may occur, but exact shape may be poorly specified.

The process of spatiotemporal interpolation (STI) relies both on spatial extraction of object features as well as an hypothesized representation, the *dynamic visual icon*, that not only preserves fragmentary information over time, but, over time, spatially updates the positions of previously seen fragments, allowing them to be appropriately connected to subsequently extracted object fragments (Palmer, Kellman, & Shipley, 2006). Because this mechanism of spatiotemporal interpolation includes within it a positional updating mechanism, it could provide the information needed for specifying metric properties in the object representations that result. The visual system may indeed know the size of an object that is never present in the stimulus, because despite its absence there, it exists in the dynamic visual icon.

A phenomenon that has relations to STI, especially in the accumulation of shape information in some sort of storage buffer, is anorthoscopic perception. In anorthoscopic perception an object travels behind a narrow slit and researchers have reported that observers are able to perceive the entire shape of the object even though only small fraction of the object is visible at any given instant (Helmholtz,1962; Parks, 1965; Plateau, 1836; Zöllner, 1862). Hochberg (1968) showed observers a rotating plus-shaped figure behind a circular aperture in discrete frames. The observers could reliably detect overall form and distinguish between possible and impossible plus shaped 3D figure. Based on the performance of the observers, Hochberg argued for post-retinal visual storage that gets updated using incoming information in successive frames and leads to integration of visible regions into an accurate representation of shape. The perceived figure can be much bigger than the slit through which it is observed suggesting that shape information persists and is accumulated over time (e.g., Anstis & Atkinson, 1967).

However, there is a distinct difference between anorthoscopic perception and STI. In anorthoscopic displays, the entire figure is projected to the eyes over time as it moves behind a single aperture. In spatiotemporal interpolation large regions of the dynamically occluded objects are never projected to the eyes and the shape discrimination task requires interpolation across unspecified regions in the image. The figure is much larger vertically than any of the apertures and the apertures are offset and misaligned. These conditions ensure that shape perception occurs though integrating information across several narrow, misaligned apertures over time.

The goal of this research is to investigate whether reconstruction of dynamically occluded objects by spatiotemporal interpolation (STI) provides metric information and whether perception of metric properties when interpolation occurs (via edge-sensitive processes) is superior to a comparable control condition in which spatiotemporal interpolation is absent, and unit formation occurs via edge-insensitive processes or not at all. This was accomplished by comparing contour-sensitive STI displays with control displays that did not support STI but contained the possibility of common fate information (albeit common fate of elements those were not simultaneously present). The STI condition on the following website (http://www.sowi.uni-kl.de/fileadmin/wpsy/public/STI/STI_-Len.htm) is an example if edge-sensitive process, while the 3-dot controls, with and without oriented line segments, show grouping by edge-insensitive process. The control group also controlled for the possibility that cognitive strategies, along with perception of sequentially revealed endpoints and a sense of elapsed time, could allow estimation of the distance between endpoints. If spatiotemporal interpolation produces metric shape representations, it was predicted that the experimental condition would produce greater accuracy and/or precision of edge length than would be possible from estimation strategies or even potential common fate information available in a control group that did not support edge interpolation.

The research described in this paper draws upon work on visual unit formation in static and dynamically occluded scenes. To provide a context for the current experiment, we briefly review some aspects of visual unit formation below.

1.1. Types of visual completion

One of the influential observations of the Gestalt psychologists was that we are able to perceive parts of objects even when there are no local stimulus correlates (Koffka, 1935; Michotte, Thines, & Crabbe, 1964). Michotte, Thines, and Crabbe (1964) proposed the term "amodal" completion to describe perception of occluded areas in the absence of sensory attributes. For example, Fig. 1A shows a situation in which a white square on a black background
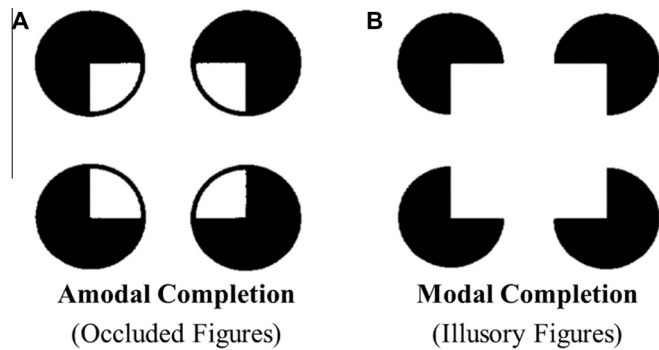
Fig. 1. Kanizsa squares. (A) Amodal completion, occluded object. (B) Modal completion, illusory figure.

is perceived amodally behind black circular windows on a white surface. Despite the perception of the square as connected behind the occluding surfaces, this representation of complete objects is amodal in that it does not contain local sensory information, such as perception of brightness or color in the occluded parts. Instead these are seen as situated behind other opaque objects. A different phenomenal experience occurs when perceptually completed forms are perceived in front of adjacent surfaces – what Michotte, Thines, and Crabbe (1964) called "modal" completion, and what are now more commonly referred to as illusory contours, surfaces, or objects. Fig. 1B shows a common example, deriving from the work of Kanizsa (1979, 1987). In the display, the form that is perceived has no real contours in the central regions marking its border with the surround. Despite this, the circles appear to be behind and occluded by the white square that is not explicitly drawn in the image.

Interpolation occurs spatiotemporally in kinetic illusory figures, rather than just spatially as in ordinary illusory figures (Bruno & Bertamini, 1990; Kellman & Cohen, 1984). In Fig. 2 sequential interruptions of the black circles on a white background caused by the movement of a triangle of the same color as the background over them give rise to a percept of unitary occluding figure. As in Kellman & Cohen's work, to be a kinetic illusory figure, rather than a conventional illusory figure, requires that inducing elements be configured so that no interpolated edges appear in any static frame. The phenomenon of kinetic illusory figures is that, despite this limitation, objects created by interpolation become clearly visible when the frames are played sequentially over time. The vari-

ous types of visual completion described here influenced the choice of the displays in the current experiment.
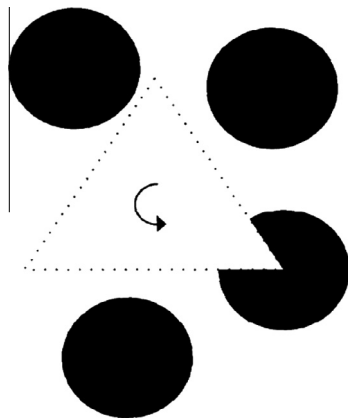
1.2. Processes specifying visual unit formation

Motion plays an important role in visual unit formation. Gestalt psychologists proposed the notion that units that move together get grouped together as a unit (the grouping principle of common fate). As we mentioned earlier, Kellman and Shipley (1991) distinguished between two processes of visual unit formation that depend on motion. *Primitive or edge-insensitive processes* specify unity only without indication location of particular boundaries. *Rich or edge-sensitive processes* specify both unity and form. The construction of detailed shape representations from edge-sensitive requires certain orientational and positional relations among the boundaries of the visible parts of the object while perception of unity alone from motion does not.

In the current study, the experimental STI condition was dependent on edge-sensitive processes with a form perceived via spatiotemporal interpolation while the spatial markers in control condition were grouped together simply due to the motion pattern. Thus the experimental and control conditions were chosen to investigate whether edge-insensitive processes are sufficient for the estimation of length or if form perception based on edge-sensitive interpolation is crucial.

1.3. Visual unit formation in static and dynamic scenes

The process of spatiotemporal interpolation (STI) leads to the formation of a visual unit when the stimulation is fragmentary across both space and time. Spatially, a single object may project separate fragments to discontinuous retinal areas. Temporally, such fragments may occur sequentially in time, when objects or observers move. These pervasive aspects of ordinary seeing require processes that interpolate across gaps in both space and time. Although there is an important complementary process of surface interpolation (e.g., Stanley & Rubin, 2005; Yin, Kellman, & Shipley, 1997), contour interpolation plays the key role in defining object shape (Kellman, Garrigan, & Shipley, 2005a; Kellman et al., 2005b).

The "sensitive" aspect of edge-sensitive interpolation is that it depends on certain constraints regarding the relative positions and orientations of edges given in the stimulus. A basic gating mechanism for interpolation is that inputs to the process are edges that end in contour intersections or junctions (Rubin, 2001; Shipley & Kellman, 1990). Junctions are points of tangent discontinuities – points where the contours have no unique orientation. (Visible contour junctions typically involve two edge orientations, although more are possible.)

For candidate edges (those leading into points of tangent discontinuity), the conditions that govern contour interpolation have been formally described in terms of spatial *relatability*.

Relatability is a notion related to the Gestalt idea of good continuation (Kellman et al., 2003, 2005b; Wertheimer, 1921). Relatability is the reciprocal to discontinuity, defined by the absence of corners or very sharp curves between the edges in consideration. Two edges are spatially relatable if their linear extensions intersect at an angle that is greater than or equal to 90 deg (Kellman & Shipley, 1991; for equations defining relatability in 2D and 3D, see Kellman et al., 2005b).

Palmer, Kellman, and Shipley (2006) proposed and tested an extension of spatial relatability to help understand spatiotemporal interpolation phenomena. The theory of spatiotemporal relatability (STR) explains dynamic object formation by the application of geometry of spatial relatability to currently seen parts and previously seen parts that were stored in a buffer – the dynamic visual icon. After a visible edge becomes occluded, it continues to be

Fig. 2. Schematic of a kinetic illusory figure. Sequential changes in black circles on a white background lead to the perception of a unitary central triangle of the background color moving infront of the circles (Kellman & Cohen, 1984).

represented as dynamic visual icon for some time and its position gets updated based on the velocity information obtained while it was visible. For being successfully represented as a dynamic visual icon the visual process must extract information about contour shape, motion (speed and direction), junctions, and boundary ownership of visible portions of the object. Experimental results are consistent with the existence of these features. When object fragments fulfill the characteristics for STR, evidence indicates that unified object representations are formed, and these representations confer advantages in object discrimination tasks (Palmer, Kellman, & Shipley, 2006).

1.4. Our study: recovery of metric properties (length) by STI

In the work reported here, we investigated whether recovery of form through STI is accompanied by metric representations, specifically representation of the length of an object's edge. We also investigated whether spatiotemporal relatability is essential for recovering metric properties of dynamically occluded objects or if, alternatively, an accurate estimate of length can be derived just by unit formation by motion as in edge-insensitive processes, assuming the latter processes can operate when separate dots and oriented line segments, that do not induce spatiotemporal interpolation, are revealed sequentially in time.

2. Methods and procedure

2.1. Participants

Fourteen undergraduate students of the University of California, Los Angeles (UCLA) and fourteen students from University of Kaiserslautern, Germany (Uni-KL), with normal or corrected-to-normal vision participated in this experiment. The observers were naïve to the purpose and nature of the experiment. They gave informed consent in accord with the policies of, UCLA and Uni-KL Committee for the Protection of Human Subjects, which approved the experimental protocol, in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

2.2. Apparatus

The displays were created and presented by a program written in MATLAB programming language (Mathworks Ltd.) using routines from the Psychophysics Toolbox (Brainard, 1997). Displays were presented on one of three 16″ × 12″ ViewSonic Graphic Series G225f computer monitors (UCLA), or a 16″ × 12″ Sun Microsystems CRT monitor (Uni-KL), each with a resolution of 1024 × 768 pixels and a refresh rate of 75 Hz. The observer sat 114 cm from the screen with his or her head stabilized using a chin-and-forehead rest. Participants responded by pressing one of two keys on the keyboard.

2.3. Units and conversion factors

For our experimental set-up, one degree of visual angle was equivalent to 53 pixels or 2 cm. The base-lengths of triangles used in our stimuli, which will be described below, were 6.0 cm on screen (3.0 deg or 160 pixels) and 4.5 cm (2.25 deg or 120 pixels). In this manuscript we will be using pixels (abbreviated as "px") as units to report the dimensions of our stimuli and results.

2.4. Stimuli

There were three conditions in this experiment, the spatiotemporal interpolation (STI) condition and two control conditions. Fig. 3-Bottom row shows the displays for the three conditions. In all three, elements were defined by sequential partial occlusion of black disks that were placed in the path of a white triangle oscillating on a white background. In the STI condition, the requirement for truly spatiotemporal interpolation was ensured by allowing no more than one vertex of the triangle to be visible at any moment. This design prevented observers from recovering the length of the triangle's base simply by spatial interpolation between two vertices. In the first control condition ("Control Condition 1"), dots were placed at the vertices of virtual triangles with the same dimension as the ones in the STI condition. Although the dots provided the same spatial and timing information as the triangle vertices in the STI condition, they did not induce perception of interpolated contours or a coherent object. In the second control condition ("Control Condition 2"), oriented line segments were added to the dots and mid-points between the dots in a way that did not induce perception of subjective triangle. These additional lines provided occluded stationary dots during the motion sequence as did the illusory triangle in the experimental condition. Moreover, they provided much more in the way of continuous, oriented reference information than was present in any other condition. If cognitive strategies based on sequentially available vertex information along with continuously available edge fragments allowed accurate size estimation, we expected this condition to produce the best accuracy. However, if illusory object formation through spatiotemporal interpolation provides metric information that is better than can be achieved through cognitive strategies, we expected this condition, along with the first control condition, to produce worse accuracy and or consistency of size estimation than in the spatiotemporal interpolation condition.

The illusory equilateral triangle was presented (Fig. 3-Top row) with two base-lengths of 120 px and 160 px respectively. The triangles were rotated by 15° from the horizontal direction of motion to ensure unambiguous motion information. (When a straight edge is oriented parallel to the direction of motion, its motion signal is ambiguous because the component of motion along its length produces no changes on the retina.) As the triangle translated across the screen, black inducing elements (disks with diameter of 60 px were sequentially occluded by vertices of the triangle. For each and every trial, the position of the three black inducing disks was randomly jittered by 15 px in vertical and horizontal direction about the center of the disks. Random jitter in the position of inducing disks was included so that the participants could not use the position of inducers to come up with an estimate of length of the base of the triangle. Additionally, in order to refresh the memory about the speed of triangle small dots (20 px in diameter) were placed in fixed positions in between the bigger black disks as shown in Fig. 3. It has been shown that the speed of recently seen, occluded or camouflaged object fragments is estimated very accurately with multiple refreshes (Palmer, Kellman, & Shipley, 2006), but across greater periods "occlusion velocity" is systematically slower than the velocity signal present when a fragment was visible (Palmer & Kellman, 2001, 2002, 2003). Two small dots (20 px in diameter, each) were placed close together on the triangle's base so as to identify which side was the target base. These dynamic displays created a clear representation of an illusory triangle despite the little information actually given. Taking into account the entire motion sequence, for the triangles with base-length 160 px and 120 px approximately 50% and 67% respectively of the triangle's complete border were ever specified by local stimulus information.

In the Control conditions, the dot stimuli contained the same black inducing elements, but the illusory triangle was specified
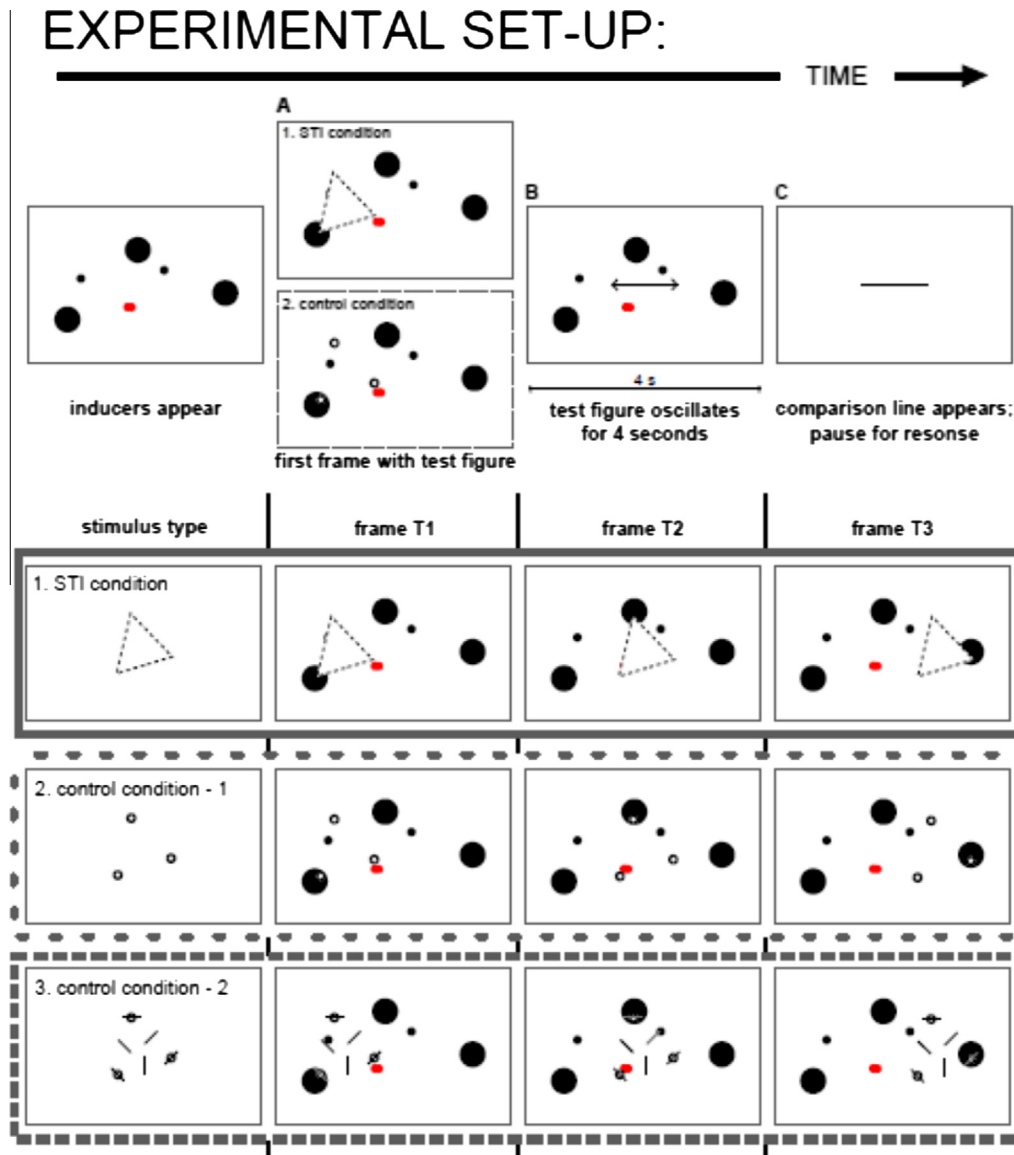
Fig. 3. Overview of displays and conditions. Top row: Phases of an experimental trial. First, the inducers appeared, followed by (A) either the STI stimulus or the control stimulus (B) The stimulus oscillated for 4 s before being replaced by (C) the comparison line and pause for response. Bottom row: Stills from the dynamic displays (1) the STI condition and (2) Control Condition 1 with dots placed at the vertices of the triangle. (3) Control Condition 2 with dots and oriented line segments that do not induce interpolation. For both stimulus type, the illusory triangle or the white dots translated horizontally and were revealed by sequential partial occlusion of the black inducers and different time frames. Only ONE vertex of the triangle was visible at any moment to ensure spatiotemporal (as opposed to simply spatial) interpolation. The task of the observer was to judge the length of the base of the "emergent triangle" that crossed the double dots. (This is schematic representation and dimensions of the stimuli do not match the ones used in the experiment. The white line segments and dots and a white background were used in the experiment. Outlines are being shown in black/gray to show the stimulus configuration in static frames).

by three white dots of 20 px in diameter located at each vertex of the underlying equilateral triangle (Fig. 3). The circumference of each of the white dot (62.8 px) was roughly equal to the maximum length of the parts of edges of the illusory triangle visible at any vertices (60 px). As in the STI condition, only one dot was visible at any instant. The presence of smooth circular dots instead of vertices with tangent discontinuities prevented contour interpolation based on the principles of spatiotemporal relatability. No definite form was perceived in the moving display; however, since the dots moved together, they could group together into a triangular visual unit as expected by the Gestalt grouping cue of common fate (assuming common fate can operate with elements not simultaneously visible). Phenomenologically, all participants reported that the three dots revealed a triangular configuration over the course of their motion sequence. In order to estimate the extent between

the dots at the base of the virtual triangle, the observer would have to use cognitive strategies based on speed and time instead of recovering the shape via spatiotemporal relatability and thereby the length of the edge.

In Control Condition 2, oriented line segments were added to the dots and to the mid points between the dots. The lines were 27 px long subtending an angle of 0.5 deg. The orientation of the segments did not support relatability of edges of the triangle; however, it allowed for cognitive strategies that might be used to infer the size of as virtual object through vertex information, timing, and the presence of continuously available reference features (lines). If inference processes, not spatiotemporal interpolation, were used in the experimental condition, then participants might see one vertex at one time and somehow extrapolate the (hidden) movement of that vertex such that its updated position could be compared to

the position of a second vertex that becomes visible later. Such a strategy is equally available in both control groups (and the experimental condition). In addition, the second Control group provides continuously available reference lines which do not support spatiotemporal contour interpolation but do provide better information for relating vertex positions that are acquired sequentially in time. Also, these reference lines interact with the stationary dots in the displays in the same spatial and temporal patterns as in the experimental group.

2.5. Design

For the first experiment, the experimental design included three crossed within-subject variables: display-type (STI condition or Control Condition 1), target base-length (120 px and 160 px) and staircase direction (2up-1down or 2down-1up). The second experiment consisted of the control condition with dots and oriented line segments (Control Condition 2). The within-subject variables were target base-length (120 px and 160 px) and staircase direction (2up-1down or 2down-1up). Each observer completed 2 blocks of interleaved staircases in a given direction with a 5 min break in between the blocks. In each block, trials from each staircase were presented in random order.

2.6. Procedure

A schematic illustration of trial structure is shown in Fig. 3-Top row. The inducers, i.e., the three black disks and four black dots, appeared at the beginning of each trial. They were followed by the oscillating stimulus which was a white triangle in the STI condition or three white dots (with or without white oriented line segments) in the Control conditions. The oscillating stimulus completed two oscillations horizontally in 4 s at the speed of 10°/s. Immediately afterward, a black comparison line whose length was controlled by an adaptive staircase appeared on the screen. The observers made a forced choice of whether the comparison line was longer or shorter than the perceived base-length in the preceding display. The comparison line remained until the observer responded using either the "S" (shorter) or the "L" (longer) key. No feedback was given. The next trial was initiated by the observer's response.

For each display, the length of the comparison line was varied based on an adaptive staircase (Treutwein, 1995). Interleaved staircases were used by randomly intermixing different display sizes (120 px and 160 px). For the first experiment, the STI condition and the Control Condition 1 was interleaved within a block. One advantage of this format is that it tends to minimize the observer's memory for responses given to particular displays on previous trials. Data were collected for each experimental condition by using two different staircase procedures (2-up/1-down and 1-up/2-down). A 2-up/1-down staircase converged on the point of the underlying psychometric function where the observer was 29.3% likely to say that the comparison line was "longer" than the base-length of the triangle, and a 1-up/2-down staircase converged at the 70.7% likely "longer" point (for discussion of obtaining particular points on psychometric functions by using different types of staircases, see Derman, 1957; Falmagne, 1986; Levitt, 1970; Treutwein, 1995). For the 2-up/1-down staircases the comparison line remained at the same length until the observer made two consecutive "shorter" responses or one "longer" response. Two "shorter" responses caused the comparison line to increase in length on the next trial thus making the task easier. By contrast, one "shorter" response made the comparison line shorter, thus making the task more difficult. For the 1-up/2-down staircases the procedure was reversed. The comparison line changed in length for the next trial after one "shorter" response or two consec-

utive "longer" responses. The change in line length occurred in steps of 20 px for the first two reversals of direction in line-length changes. Next, the line changed in length by 10 px for next 8 reversals and finally by 5 px until the end of the staircase. The staircases converged after 16 reversals or a maximum of 100 trials. One staircase typically required approximately 5 min to complete. The initial value for the length of the comparison line was randomly selected from between 120 px and 160 px and differed for each staircase.

For the Control Condition 2 which was run after the first two conditions, the dots with oriented line segment were shown as interleaved 2-up/1-down and 1-up/2-down by randomly intermixing the different display sizes (120 px and 160 px).

2.7. Dependent measures and data analyses

The points of the underlying psychometric function that were sampled in the experiment were fitted with cumulative Gaussian using the PSIGNFIT toolbox. The PSIGNFIT toolbox implements the maximum-likelihood method described by Wichmann and Hill (2001) to fit psychometric functions to data points. The toolbox returned the threshold (the 50% point) and the slope at the mean for the best fitting cumulative Gaussian for the data points. The mean gives the measure of the point of subjective equality (PSE) or the measure of length at which the observer is 50% likely to make "longer" judgment. As in previous work (Guttman & Kellman, 2004), we took the mean to define the best estimate of the perceived value for the variable being measured. Perceptual values obtained in this way were compared to the physically specified length of the virtual object or separation of anchor points for a given condition. Also following earlier work (Guttman & Kellman, 2004), we took the *difference* between the two points on the psychometric function (the 29.3% and 71.7% likelihoods of a participant saying "longer") to define an *imprecision* measure for each stimulus. (We name the measure "imprecision" so that higher numbers for distance between the estimates correspond to higher values of the measure and linear relations between differences, such as the relations between distances of 40, 60, and 80 pixels are preserved in this measure. One could have the inverse – a precision measure going from 0 to ∞ – by taking the reciprocal.) The imprecision measure is directly related to the slope of the psychometric function: the greater the slope, the more precisely could the base-length to be estimated, whereas greater imprecision implies a shallower slope.

For the first experiment, data from 14 observers were analyzed by a 2×2 repeated measures ANOVA with display type and base-length as within-subject variables. Additionally, planned comparisons (*t*-tests) compared the accuracy and the imprecision results obtained with the STI condition and the Control Condition 1 for the two base-lengths. For comparing the STI condition and Control Condition 2, the data from two different groups of 14 observers were analyzed by 2×2 mixed ANOVA with repeated measures on base-lengths and display type as a between-subject variable. Additionally, independent samples *t*-tests were performed for comparing the data for the STI condition and Control Condition 2, and Control Conditions 1 and 2.

3. Results

Participants' estimation of the base-length of triangles in this study is shown in Fig. 4. The two panels show the accuracy and imprecision measures for the STI and the Control conditions for the two base-lengths. The spatiotemporal interpolation condition (STI) produced a more accurate perception of length for both the 120 px and 160 px base-lengths. For base-length 120 px, the
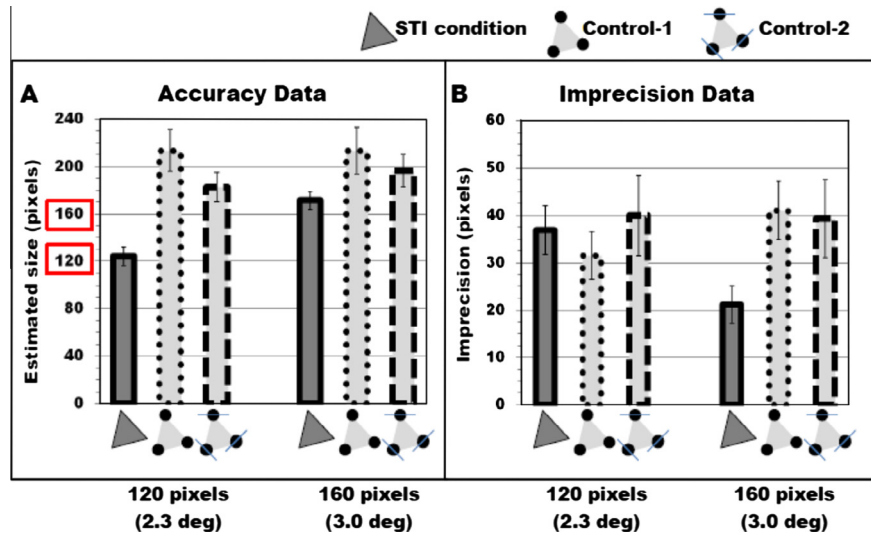
Fig. 4. Results of the experiment. Mean accuracy data (A) and precision data (B) are shown by condition and display size. Error bars indicate ±one standard error of the mean.

subjects perceived an average estimated length of 124.38 px (standard error of the mean 7.81 px) in the STI condition compared to 213.5 px (standard error of the mean 17.71 px) in the Control Condition 1 and 182.69 px (standard error of the mean 12.28 px) in the Control Condition 2. For base-length 160 px, the subjects perceived an average estimated length of 171.42 px (standard error of the mean 7.71 px) in the STI condition compared to 213.36 px (standard error of the mean 19.65 px) in the Control Condition 1 and 196.37 px (standard error of the mean 13.92 px) in the Control Condition 2. Qualitatively, the STI condition showed great sensitivity to the difference in sizes of the two triangle displays whereas the Control conditions showed little or none. The average accuracies in the STI condition corresponded to a 3.65% error in length estimation in the case of the smaller triangle and a 7.1% error for the larger triangle. In the Control Condition 1, average length estimations were higher by roughly an order of magnitude, off by 77.9% and 33.4% for the smaller and larger triangles respectively. In the Control Condition 2, average length estimations were higher by 52.2% and 22.7% for the smaller and larger triangles respectively.

For the first experiment with two display types (STI and Control Condition 1), these observations were confirmed by analyses using 2×2 repeated measures ANOVA. There were reliable main effects of base-length $\{F(1,13) = 17.02, p < .001, \eta_p^2 = 0.57\}$ and display-type $\{F(1,13) = 12.68, p < .005, \eta_p^2 = 0.49\}$ as well as a reliable base-length $\times$ display-type interactions $\{F(1,13) = 12.42, p < .005, \eta_p^2 = 0.49\}$. Planned comparisons indicated that the values differed significantly between the STI and the Control conditions both for 160 px $\{t_{13} = 2.1, p < .05\}$ and 120 px $\{t_{13} = 4.7, p < .001\}$. Planned comparisons further indicated that the values differed significantly between the 160 px and 120 px base-lengths for the STI condition $\{t_{13} = 7.1, p < .001\}$ but not for Control Condition 1 $\{t_{13} = 0.13, p = .99\}$.

For the second control experiment, a 2 (base-lengths: 120 px vs. 160 px) $\times$ 2 (STI vs. Control Condition 2) mixed ANOVA with repeated measures on base-lengths was conducted to compare the accuracy of length estimates in Control Condition 2 and the STI condition, respectively. A similar 2×2 mixed ANOVA analysis was performed for the two Control conditions one without and another with oriented line segments. There was a reliable interaction between condition and base lengths for the STI vs. Control Condition 2 $\{F(1,26) = 18.20, p < .001, \eta_p^2 = 0.41\}$, but not for Control Condition 1 vs. Control Condition 2 conditions $\{F(1,26) = 1.50, p = .23, \eta_p^2 = 0.055\}$. Independent samples t-test showed that the

STI Condition was reliably more accurate than Control Condition 2 for the 120 px display, $\{t_{26} = -4.0, p < .001\}$, but the difference was not reliable for 160 px $\{t_{26} = -1.57, p = .12\}$. There was no significant difference between the means for the two control conditions for either the 120 px display, $\{t_{26} = 1.43, p = .16\}$ nor the 160 px display, $\{t_{26} = 0.71, p = .49\}$. For Control Condition 2, a paired t-test was performed to compare the means for the two base-lengths. There was no reliable difference between means for the 160 px and 120 px displays in Control Condition 2 $\{t_{13} = 2.08, p = .06\}$.

The imprecision measure is an indicator to the sensitivity of the processes involved in the recovery of metric properties on spatio-temporal interpolation. Lower imprecision for STI condition compared to the Control condition would indicate that the availability of form from STI produces more precise and accurate results than those achieved through cognitive strategies. The imprecision measure for the base-length of 160 px was consistent with this prediction for STI vs. Control Condition 1. For base-length of 160 px, the imprecision for the STI condition, 21.14 px (standard error of the mean 3.88 px) was much lower than for Control Condition 1, 41.07 px (standard error of the mean 6.10 px). The planned t-test supported the findings that these values differed significantly $\{t_{13} = 4.1, p < .001\}$. However, the imprecision results for the STI and Control Condition 1 for base-length of 120 px did not differ reliably $\{t_{13} = -0.8, p = .41\}$. The imprecision values were 36.88 px (standard error of the mean 5.14 px) in the STI condition compared to an imprecision of 31.47 px (standard error of the mean 5.03 px) in Control Condition 1. In a 2×2 repeated measures ANOVA, there were no reliable main effects of base-length or $\{F(1,13) = .112, p = .74, \eta_p^2 = 0.009\}$ or display-type $\{F(1,13) = 3.9, p = .06, \eta_p^2 = 0.23\}$ but there was a reliable base-length $\times$ display-type interaction $\{F(1,13) = 10.98, p < .05, \eta_p^2 = 0.46\}$.

However, the imprecision results for Control Condition 2 for base-length of 120 px was 39.9 px (standard error of the mean 8.4 px) and 160 px was 39.3 px (standard error of the mean 8.3 px). There was no reliable difference between the imprecision measures for the two base-lengths. Independent sample t-test showed that the imprecision results did not differ reliably from the STI and Control Condition 1.

4. Discussion

The experiments sought to investigate the role played by spatiotemporal interpolation in recovering metric properties such as

length of visual units formed from information that is fragmentary across both space and time. The data from the experiments showed that for conditions where a form was perceived by the virtue of spatiotemporal interpolation, the length of the perceived triangle was accurately estimated. Average error for length estimates were in the 3–7% range for the STI condition. Errors of length estimation were far higher in the Control Condition 1, despite equivalent spatial and temporal markers that were displayed. In fact, in both Control displays, subjects gave very similar size estimates for the 120 px and 160 px displays (213.5 px and 213.36 px, respectively). In a second control experiment the addition of oriented line segments that do not support interpolation did not lead accurate estimation of length and produced estimates that did not differ significantly from Control Condition 1. Thus, cognitive inferences about length based on position and timing, as in the dots control condition, with or without oriented line segments, were much less accurate. This pattern of results is consistent with the notion that estimation of metric properties such as length is more accurate when form is recovered due to STI than when length has to be estimated through cognitive strategies. The inferior performance with the Control stimuli may be attributed to cognitive strategies relying on spatial cues provided by the distance between the inducers rather than actually recovering the length. The predicted STI effects for accuracy were seen in both display sizes. This pattern of results suggests a special role of perceptual object formation from spatiotemporal interpolation in producing representations of functionally important object properties.

In the imprecision data, we found a reliable condition difference for large displays but not for small displays. The greatest precision of estimates was found in the STI condition with the largest display. In that case, a difference of 21 pixels was found, corresponding to about 24 arcmin of visual angle. This means that, on average, a change of 12 arcmin in the comparison stimulus from the estimated perceived length was enough to markedly affect length comparison, such that a 12 arcmin decrease produced approximately 29% judgments of "longer" and a similar increase raised the proportion of "longer" responses to about 71%. For the smaller STI display, the average range for imprecision was 37 pixels, corresponding to about 0.71 deg (42.7 arcmin) of visual angle. Thus, a shift of 21.35 arcmin in either direction from the approximately veridical perceived length was enough to markedly change proportions of "longer" responses in the STI condition. Although the precision of length perception as measured here was clearly better for the larger display, both displays showed that perceived length was quite constrained and that performance became easier as comparison stimuli moved further from the actual base-length of the virtual object.

The imprecision data in the Control conditions are somewhat harder to interpret. For Control Condition 1, for the smaller display, in particular, imprecision was as low (numerically lower, in fact, although not statistically significant) as for the corresponding STI display. However, we note that this means something quite different in this Control condition, as the estimated length in this condition is quite discrepant from the length specified by spatiotemporal information in the stimulus. In fact, the limits of imprecision (between the 29.4% and 71.6% proportions of "longer" responses) in neither case in Control Condition 1 encompassed the true (virtual) value of length. For Control Condition 1, the smaller display, combining the estimated perceived length of 213.5 px with the measured imprecision of 41 px deg produces a range of 182–245 px for perceived length for a display that was actually 120 px in length, and the similar estimated perceived length and imprecision for the larger display gives 172.3–254.3 px for a display that was actually 160 px in length. Besides failing to encompass the veridical size of the virtual length, these data are notably insensitive to the difference between the stimulus size

conditions. We interpret the comparable precision of responding in this Control condition across both displays, and the lack of variation in responses to the two displays, as indicating a somewhat stable strategy by subjects and across subjects, but not one that had much to do with length perception.

The imprecision results are even harder to interpret for Control Condition 2. For Control Condition 2, the smaller display, combining the estimated perceived length of 182.7 px with the measured imprecision of 39.9 px produces a range of 143–223 px for perceived length for a display that was actually 120 px in length. However, the estimated perceived length and imprecision for the 160 px display gives 156–236 px that does encompass the true (virtual) value of the length. Based on the similar imprecision values obtained for the two sizes in Control Condition 2, we interpret that the presence of oriented line elements leads to cognitive strategies that is stable across displays of different sizes. However, the errors in length estimation compared to the true (virtual) length shows that these cognitive strategies do not help much in length estimation if the oriented line elements do not induce visual unit formation via spatiotemporal relatability.

Taken together, these results indicate that metric perception can arise from formation of objects based on spatiotemporal interpolation. Length was accurately perceived under conditions known to support object formation by spatiotemporal interpolation. Perception of such objects requires satisfaction of certain constraints in the edge-sensitive interpolation process, mainly the spatial and temporal relations described as spatiotemporal relatability (Palmer, Kellman, & Shipley, 2006). When clear spatial and temporal reference points are provided that do not lead to perception of illusory figures through the edge-sensitive process (although these may still be related by common fate or an edge-insensitive grouping process), perception of object characteristics was inaccurate and not at all comparable to that under STI condition. Cognitive strategies for recovering length from the same basic spatial and temporal data do not approach the accuracy of spatiotemporal interpolation. The present data provide further evidence of sophisticated routines that the visual system uses to decode objects and scenes despite inputs that can be fragmentary in both space and time. The remarkable capacity of STI to accumulate fragments across space and time and produce perception of coherent objects even allows us to perceive the sizes of objects that, in an important sense, never really exist in the stimulus.

Based on this study we conclude that spatiotemporal contour interpolation not only produces perception of shape but furnishes representations with reasonably accurate metric properties of dynamically perceived objects. Even though an object constructed through spatiotemporal interpolation does not exist at any moment in the stimulus, its size and shape appear to be accurately perceived.

References

Anderson, B. L., & Barth, H. C. (1999). Motion-based mechanisms of illusory contour synthesis. *Neuron, 24*, 433–441.

Anderson, B. L., O'Vari, J., & Barth, H. (2011). Non-Bayesian contour synthesis. *Current Biology, 21*(6), 492–496.

Anderson, B. L., & Sinha, P. (1997). Reciprocal interactions between occlusion and motion computations. *Proceedings of the National Academy of Sciences, 94*(7), 3477–3480.

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

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

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

Bruno, N. (2001). Breathing illusions and boundary formation in spacetime. In T. F. Shipley & P. J. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision* (pp. 531–556). Amsterdam: Elsevier.

Cunningham, D. W., Shipley, T. F., & Kellman, P. J. (1998a). The dynamic specification of surfaces and boundaries. *Perception, 27*, 403–415.

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

Derman, C. (1957). Non-parametric up-and-down experimentation. *Annals of Mathematical Statistics, 28*, 795–797.

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision, 2*(4).

Falmagne, J. C. (1986). Psychophysical measurement and theory. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.). *Handbook of perception and human performance: Sensory processes and perception* (Vol. I, pp. 1–66). New York: John Wiley & Sons.

Fantoni, C., & Gerbino, W. (2003). Contour interpolation by vector-field combination. *Journal of Vision, 3*, 281–303.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Research, 33*, 173–193.

Geisler, W. S., & Perry, J. S. (2009). Contour statistics in natural images: Grouping across occlusions. *Visual Neuroscience, 26*(01), 109–121.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41*, 711–724.

Guttman, S. E., & Kellman, P. J. (2004). Contour interpolation revealed by a dot localization paradigm. *Vision Research, 44*, 1799–1815.

Helmholtz, H. von (1962). *Treatise on physiological optics* (Vol. 3). New York: Dover (Original work published 1867).

Hochberg, J. (1968). In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception* (pp. 309–331). New York: Holt, Rinehart & Winston.

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Cheshire, A. (2003). Infants' perception of object trajectories. *Child Development, 74*, 94–108.

Kanizsa, G. (1979). *Organization in vision*. New York: Praeger.

Kanizsa, G. (1987). Quasi-perceptual margins in homogeneously stimulated fields. In G. E. Meyer, & S. Petry (Eds.), W. Gerbino (Trans.), *The perception of illusory contours* (pp. 40–49). New York: Springer-Verlag, Inc. (Reprinted from *Rivista di Psicologia*, 1955, *49*, pp. 7–30).

Kellman, P. J. (1996). The origins of object perception. In R. Gelman & T. Au (Eds.). *Handbook of perception and cognition: Perceptual and cognitive development* (Vol. 8). Academic Press.

Kellman, P. J., & Cohen, M. H. (1984). Kinetic subjective contours. *Perception and Psychophysics, 35*, 237–244.

Kellman, P. J., Garrigan, P., Kalar, D., & Shipley, T. F. (2003). Good continuation and relatability: Related but distinct principles. *Journal of Vision, 3*(9), 120.

Kellman, P. J., Garrigan, P., & Shipley, T. F. (2005a). Object interpolation in three dimensions. *Psychological Review, 112*(3), 586–609.

Kellman, P. J., Garrigan, P., Yin, C., Shipley, T., & Machado, L. (2005b). 3D interpolation in object perception: Evidence from an objective performance paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 31*(3), 558–583.

Kellman, P. J., Gleitman, H., & Spelke, E. S. (1987). Object and observer motion in the perception of objects by infants. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 586–593.

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

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive psychology, 15*(4), 483–524.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace & World.

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 33*, 467–476.

Lorenceau, J., & Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Research, 32*, 263–273.

Lorenceau, J., & Shiffrar, M. (1999). The linkage of visual motion signals. *Visual Cognition, 6*, 431–460.

Michotte, A., Thines, G., & Crabbe, G. (1964). *Les compliments amodaux des structures perceptives (Amodal complements of perceptual structures)*. Oxford, England: Publications U. Louvain.

Palmer, E. M., & Kellman, P. J. (2001). The aperture capture effect: Misperceived forms in dynamic occlusion displays. *Journal of Vision, 1*, 381a (Abstract).

Palmer, E. M., & Kellman, P. J. (2002). Underestimation of velocity after occlusion causes the aperture–capture illusion. *Journal of Vision, 2*, 477a (Abstract).

Palmer, E. M., & Kellman, P. J. (2003). (Mis)perception of motion and form after occlusion: Anorthoscopic perception revisited. *Journal of Vision, 3*, 251a (Abstract).

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.

Parks, T. E. (1965). Post-retinal visual storage. *American Journal of Psychology, 78*, 145–147.

Petry, S. E., & Meyer, G. E. (1987). *The perception of illusory contours*. New York: Springer-Verlag Publishing.

Plateau, J. A. F. (1836). Anorthoscop. *Bulletin de l'Académie de Bruxelles, 3*, 364.

Rubin, N. (2001). The role of junctions in surface completion and contour matching. *Perception, 30*, 339–366.

Shipley, T. F., & Cunningham, D. W. (2001). Perception of occluding and occluded objects over time: Spatiotemporal segmentation and unit formation. In T. F. Shipley & P. J. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision* (pp. 557–585). Amsterdam: Elsevier.

Shipley, T. F., & Kellman, P. J. (1990). The role of discontinuities in the perception of subjective figures. *Perception and Psychophysics, 48*, 259–270.

Shipley, T. F., & Kellman, P. J. (1992). Perception of partly occluded objects and illusory figures: Evidence for an identity hypothesis. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 106–120.

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*, 323–339.

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*, 3–20.

Shipley, T. F., & Kellman, P. J. (1997). Spatio-temporal boundary formation: The role of local motion signals in boundary perception. *Vision Research, 37*, 1281–1293.

Stanley, D. A., & Rubin, N. (2005). Rapid detection of salient regions: Evidence from apparent motion. *Journal of Vision, 5*(9).

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research, 35*, 2503–2522.

Tse, P. U. (1999). Volume completion. *Cognitive Psychology, 39*, 37–68.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron response. *Science, 224*, 1260–1262.

Wertheimer, Max (1921). Untersuchungen zur Lehre von der Gestalt, I: Prinzipielle Bemerkungen. *Psychologische Forschung, 1*, 47–58.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception and Psychophysics, 63*(8), 1293–1313.

Yin, C., Kellman, P. J., & Shipley, T. F. (1997). Surface completion complements boundary interpolation in the visual integration of partly occluded objects. *Perception, 26*, 1459–1479.

Zöllner, F. (1862). Uber eine neue Art anorthoskopischer Zerrbilder (On a new type of anorthoscopic distorted image). *Annalen der Physic, 117*, 477–484.