The Role of Temporal Structure in Human Vision

Randolph Blake
Department of Psychology, Vanderbilt University, Nashville, TN

Sang-Hun Lee
Department of Psychology, Seoul National University, Seoul, South Korea

Gestalt psychologists identified several stimulus properties thought to underlie visual grouping and figure/ground segmentation, and among those properties was common fate: the tendency to group together individual objects that move together in the same direction at the same speed. Recent years have witnessed an upsurge of interest in visual grouping based on other time-dependent sources of visual information, including synchronized changes in luminance, in motion direction, and in figure/ground relations. These various sources of temporal grouping information can be subsumed under the rubric temporal structure. *In this article, the authors review evidence bearing on the effectiveness of temporal structure in visual grouping. They start with an overview of evidence bearing on temporal acuity of human vision, covering studies dealing with temporal integration and temporal differentiation. They then summarize psychophysical studies dealing with figure/ground segregation based on temporal phase differences in deterministic and stochastic events. The authors conclude with a brief discussion of neurophysiological implications of these results.*

Key Words: visual grouping, temporal structure, common fate, temporal resolution, temporal integration, synchrony, figure/ground segmentation

Visual perception, unlike most other cognitive activities, seems automatic and effortless—open your eyes and the visual world immediately appears before you, populated with meaningful objects and events. In fact, however, the effortlessness of vision belies the complexity of the operations required to create our visual world—vision's ease is an illusion. To borrow a metaphor from Hoffman (1998), we are all visual geniuses who are naively unaware of our immense talents. Only when confronted with the challenging steps involved in seeing (Marr, 1982) do we appreciate that perception is a near miracle: vision somehow manages to assemble and collate data contained in the optical input to vision into recognizable, three-dimensional objects arrayed within a cluttered, three-dimensional space.

Why is vision inherently difficult? For one thing, the optical input is constantly changing as objects move about in the environment and as we ourselves move relative to those objects. The dynamic character of vision means that we must be able to recognize objects from multiple perspectives and from different viewing distances. For another, objects often appear in cluttered surroundings, which means that parts of objects often obscure, or occlude, one another. Figuring out which parts go with which objects—grouping, as it is called—represents another formidable challenge, as evidenced by the difficulty of dealing with occlusion in the case of machine vision. Visual perception is faced with other difficulties, too, some having to do with the variable lighting conditions illuminating the environment and others having to do with our inability to digest the entire visual scene in a single glimpse.

How does perception overcome these difficulties? Fortunately, the natural environment contains regularities that can be exploited by the processing machinery underlying visual perception. Some of these regularities arise from the nature of light and its interactions with the surfaces of opaque objects (e.g., "shadows are always attached to surfaces"). Others emerge from fundamental principles associated with the nature of matter (e.g., "two objects cannot occupy the same location at the

Authors' Note: We thank Sharon Guttman for comments on an earlier version of this article and David Bloom for help with preparing the manuscript. Supported by a NIH Grant (EY07760) to RB and a grant (M 103KV010021-04K2201-02140) from the Korea Institute of Science and Technology and Planning to SHL. Please address correspondence to Randolph Blake at Randolph.blake@vanderbilt.edu.

Behavioral and Cognitive Neuroscience Reviews
Volume 4 Number 1, March 2005 21-42
DOI: 10.1177/1534582305276839

same time"). Still others derive from geometrical regularities found in our natural environment, regularities that bias co-occurrence statistics associated with edges and boundaries (Geisler, Perry, Super, & Gallogly, 2001). These regularities can be—and evidently are—used by human visual systems to simplify the grouping of local features into global forms and to promote the segregation of those forms from their backgrounds. Indeed, grouping and figure/ground segregation have constituted two of visual perception's most enduring, widely studied problems. Dating back to the Gestalt psychologists, students of perception have compiled through the years a growing list of grouping principles that seem to capture key operations underlying human form perception. Most of those principles focus on relations among local "features" defined in terms of spatial discontinuities in luminance, color, and texture. These discontinuities constitute what can be termed *spatial structure*; together these various sources of spatial structure specify the locations of edges and borders within the visual scene. They provide the raw ingredients, so to speak, for the processes of grouping and figure/ground segregation. These forms of spatial structure, including two illustrated in Figure 1, can be construed as static sources of information in that their realization has no dependence on patterns of change over time. A "snapshot" of the visual scene is sufficient to portray contours specified by these spatial discontinuities.

As pointed out above, however, our visual world is highly dynamic: objects move within the visual scene, and observers themselves are chronically moving their eyes and heads as they view objects. Consequently, the retinal input to vision typically includes dynamic changes in the spatial structure of the retinal image. One might construe those temporal changes in visual signals as obstacles in the registration of meaningful visual information about objects, something biological vision must overcome. After all, we are taught to hold a camera still when taking photographs, for otherwise photographs of moving objects will likely be blurred. In the case of biological vision, however, visual changes over time might actually *enhance* perception, by contributing to grouping and segmentation of visual features. The Gestalt psychologists foresaw this possibility and formalized it as one of their principles, grouping by *common fate*. For them, common fate involved an ensemble of elements all moving in the same general direction at the same speed, relative to a background of other elements. Textbook examples of common fate include a flock of birds flying overhead and a marching band of musicians walking in lockstep. In recent years, however, the notion of common fate has been extended to dynamic stimuli other than motion, including unpredictable changes in the shapes and contrasts of a subset of contours. In modern parlance, these changes create what is called *temporal structure*. The focus of this article is the efficacy of temporal structure as a grouping agent. But for temporal structure to promote grouping, the dynamics portraying that structure must be reliably picked up and registered by the visual nervous system. Does this, in fact, occur? If so, to what extent do spatial structure and temporal structure interact to specify biologically relevant visual infor-

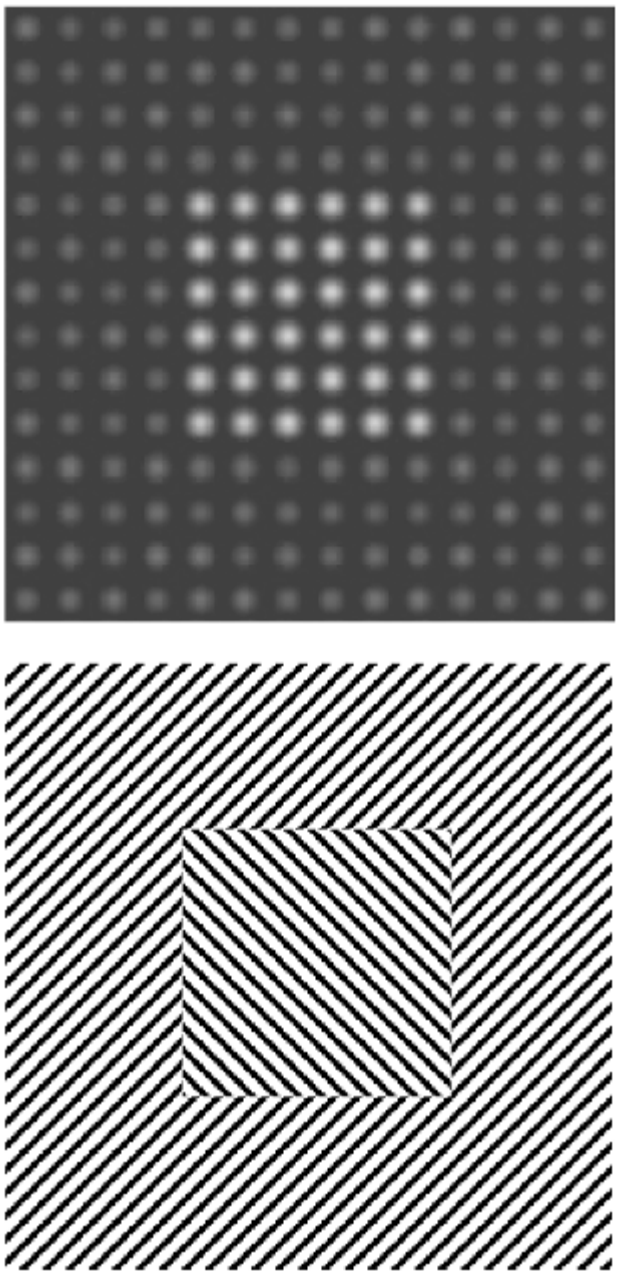

Figure 1: Two Examples of Static Sources of Visual Information Promoting Figure/Ground Segmentation.
NOTE: The top panel shows shape defined by luminance and the bottom panel shows shape defined by texture.

mation? These questions represent the central theme of this article.

In this article, we address the following questions: (a) How accurately and reliably does the visual system register temporal structure defining visual events? (b) Can the visual system derive spatial structure solely on the basis of temporal structure without discontinuities in static properties? and (c) What properties of temporal structure are critical for the visual system to group or segregate visual components? From the outset, we wish to stress the distinction between temporal structure in the stimulus domain and temporal structure in the neural domain. The former refers to time-varying changes in the optical input to vision created by events in the world; the latter refers to temporal patterning in the trains of action potentials within ensembles of neurons. Much recent debate has centered on the existence and possible functional significance of synchronized neural activity (Engel & Singer, 2001), including the role of synchronization in visual feature binding (Treisman, 1999). Temporal structure among neural discharges could be evoked by external, stimulus-driven temporal structure but may also arise from internally generated, dynamic interactions among neurons. A review of the growing literature on neural synchronization is beyond the scope of our article; our focus is on temporal structure contained in the optical input to vision, structure that may or may not evoke synchronized activity within the visual nervous system. Of course, evidence that external temporal structure can indeed promote perceptual grouping ("binding" as some would term it), although not definitive, would be encouraging to advocates of the neural synchrony viewpoint.

With that disclaimer in place, we are ready to start with a brief overview of temporal resolution in human vision, to set the stage for considering temporal structure and grouping.

TEMPORAL RESOLUTION IN HUMAN VISION

Repeating a point made above, the optical input to vision often contains rich temporal structure, produced by object motion, sudden illumination changes, and observer-produced head and eye movements. Can this dynamic temporal structure, which can be noisy and unpredictable, effectively contribute to spatial grouping and segmentation of visual features? Before attempting to answer this question, it is important to ask whether temporal structure is reliably registered by the visual nervous system. Does human vision have the requisite temporal resolution to exploit temporal structure contained in the optical input? Several lines of evidence, reviewed in the following paragraphs, suggest an affirmative answer to this question.

We shall begin by distinguishing between two seemingly conflicting demands faced by vision when confronted with dynamic visual events. On the one hand, those events may need to be integrated into a unified perceptual representation, a requirement we can term *temporal integration*. There are many instances where the optical input to vision is temporarily interrupted (e.g., during eye blinks), yet we seamlessly piece together visual signals over time to maintain perceptual continuity. Temporal integration can also enhance visual sensitivity by summing weak neural signals over time (a capacity embodied in Bloch's law). Yet in other situations, effective vision requires segregating visual events occurring closely in time, a requirement we can term *temporal differentiation*. When one object briefly and rapidly passes in front of another, we certainly don't want to incorporate the images of the two objects into some nonexistent hybrid object, and this requires being able to differentiate visual events occurring closely in time. Likewise, we rely on temporal differentiation every time we read messages rapidly flashed on at the same location on a video monitor. So, then, how does vision achieve these two seemingly incompatible demands? Let's consider each in turn.

Temporal Integration

The vision literature describes a number of phenomena that bear on the question of temporal integration. Here we shall summarize just a few, starting with *visual masking*.

Figure 2a schematically illustrates the stimulus conditions defining backward masking. A "target" is briefly presented and followed closely in time by "mask." Under appropriate spatio-temporal conditions, the trailing mask can markedly reduce the visibility of the target stimulus even though the two visual events do not overlap in time (Breitmeyer & Ganz, 1976; Kahneman, 1968). Accounts of masking typically assume that the visual system retains an iconic representation of the target that is susceptible to temporal integration with the mask.[1] Given this view, one should be able to estimate the integration time by varying the interval between target and mask to reveal the interval necessary to preclude the target's reduced visibility. A number of studies point to a critical interval in the range 80-120 msec, with the specific value depending on spatial frequency and retinal location (Breitmeyer, 1978; Breitmeyer & Ganz, 1976; Corfield, Frosdick, & Campbell, 1978; Meyer & Maguire, 1977).

In the case of masking, we infer something about temporal integration from the deterioration in target visibility dependent on the asynchrony between presentation of mask and target. A complementary strategy, and one that arguably taps temporal integration more directly, is

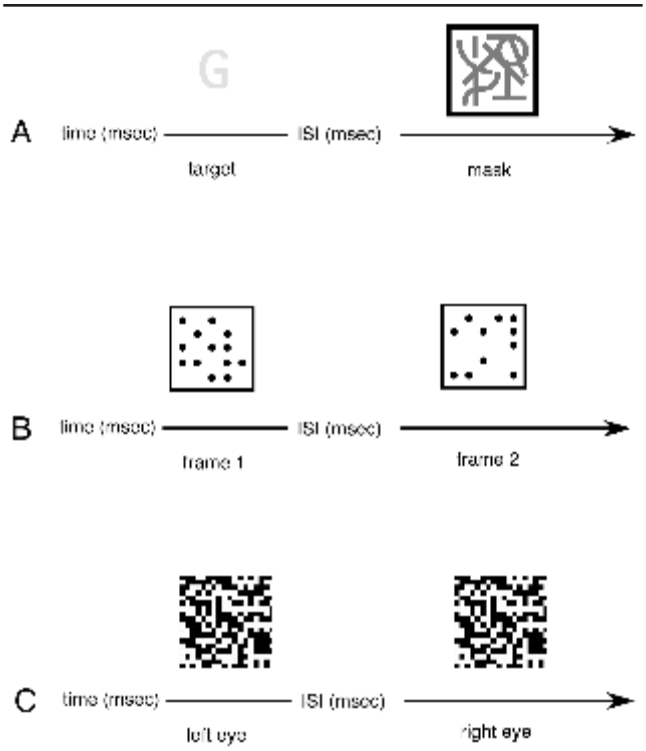

Figure 2: Schematics of Procedures for Estimating Temporal Integration.

NOTE: (A) Visual backward masking, where a "target" element (star in this example) is followed closely in time by a "masker" that can render the target invisible when the interval separating the two (interstimulus interval: ISI) is brief. (B) Form-part integration task, where two complementary parts of a component figure are presented sequentially, with a brief interval (ISI) separating the two. Observers make a perceptual judgment that requires stimulus information from both components. In this example (modeled after Hogben & Di Lollo, 1974), one "cell" of a 5 × 5 matrix of black circles is empty—the cell with the missing circle is conspicuous when the two half-images are presented with brief ISIs. (C) Stereopsis from random-dot stereograms, with the two half-images presented sequentially to left and right eyes, with a brief ISI between presentations.

the *form-part integration task* (Eriksen & Collins, 1967; Hogben & Di Lollo, 1974). Figure 2b shows the essence of this procedure. An original visual pattern is divided into small, complementary components distributed between two individual images. These "half" images are then presented successively, with the interval between images (ISI) varied over trials. The observer's task is to perform a perceptual judgment based on the original pattern. This task can only be accomplished by integration of the components into a global, complete figure; the requisite stimulus information cannot be ascertained from either half-image alone. Observers are able to perform at above chance levels when the half-images are sequentially presented with ISIs up to about 120 msec (Dixon & Di Lollo, 1994; Eriksen & Collins, 1967; Hogben & Di Lollo, 1974, 1985). This outcome implies that the neural signals associated with the half-image components greatly outlast the duration of the visual stimuli themselves, with the visual system integrating those signals to recreate a usable representation of the composite.

Another way to tap into temporal integration is to utilize sequential dichoptic (i.e., separate stimulation of left and right eyes) presentation of two half-images that together yield stereoscopic depth perception (see Figure 2c). For this purpose, random-dot stereograms are particularly useful, for the separate half-images contain no hint of the shape of the region defined by retinal disparity (Julesz, 1971). To recognize this shape, neural signals associated with the two monocular images must be combined binocularly, with the disparity between those images then specifying the shape of the region seen in depth. We know, of course, that observers can easily extract stereoscopic depth from a random dot stereogram when the two monocular images are presented simultaneously. What happens, however, when those two images are presented sequentially, not simultaneously? Several studies have shown that the visual system can extract stereoscopic depth from temporally separated half-images of a random-dot stereogram, so long as the interval separating the two half-images does not exceed 50-70 msec (Julesz & White, 1969; Ross & Hogben, 1974). Binocular integration time using pictures containing recognizable monocular forms can be stretched a little beyond 100 msec (Efron, 1957; Ogle, 1963).

The three examples described above represent just a few of the many visual phenomena all of which imply that neural signals triggered by visual stimulation outlast the stimulus itself. Because the neural consequence of visual stimulation persists for some time after physical termination of the evoking stimulus, those persisting signals are available for integration with signals arising within about a tenth of a second after termination of the first stimulus.

One can imagine circumstances where temporal integration on this time-scale would be advantageous, but summing signals over time introduces a potentially serious cost: one loses the ability to resolve events occurring very closely in time, the result being a kind of neural blurring of stimulus information (Burr, 1980). Species reliant on rapid unexpected events eschew temporal integration altogether, as evidenced by their very high sensitivity to rapidly occurring events (i.e., they possess very high temporal resolution). One classic example is the fly, renowned for its ability to detect and react to dynamics created by the complex optic flow generated by flying. What is the evidence concerning temporal resolution in human vision? We turn to this question in the next section.

Temporal Differentiation

If all visual information available to perception were only contained in "chunks" integrated over time, it would be impossible to judge the temporal order of events occurring very closely in time (VanRullen & Koch, 2003). Does human vision suffer this limitation? To frame the question in the simplest possible terms, imagine a single stimulus that appears, briefly disappears, and then reappears. Would the brief temporal gap even be noticed? The answer is yes—people can detect temporal offsets as brief as 5 msec (Georgeson & Georgeson, 1985; Smith, Howell, & Stanley, 1982). Or consider another, slightly more complicated stimulus sequence, where a stimulus appears for 100 msec at one location, disappears very briefly, and then reappears for 100 msec at another, nearby location. If the two events occur within 20 msec or so of one another, will they appear as a single event, namely, two stimuli occurring simultaneously at two locations? The answer is no—one readily experiences compelling apparent motion, with a unitary stimulus seen to move from the initial location to the subsequent one (Anstis, 1970). These two examples, as well as others (e.g., Lappin & Bell, 1972), imply that the visual system retains very good information about when in time events occur relative to one another, for otherwise motion or brief disappearance would not be experienced.

Another, related line of evidence pointing to high temporal resolution comes from a task called *temporal order detection* (Exner, 1875; Hirsh & Sherrick, 1961; Sweet, 1953; Wertheimer, 1912; Westheimer & McKee, 1977; Yund & Efron, 1974). Similar to an apparent motion display, two stimuli appear asynchronously at different locations in the visual field, and the observer's task is to indicate which one appeared first. Threshold estimates for this task range from about 20 msec, when the two stimuli are separated by several degrees of visual angle (Hirsh & Sherrick, 1961), down to 2 msec, when the stimuli are immediately next to one another (Sweet, 1953; Westheimer & McKee, 1977).

A third task that taps into temporal resolution involves detecting asynchrony between patterns flickering at the same rate, with asynchrony gauged in terms of temporal phase shift in the flicker of one pattern relative to another. A representative example of this technique is provided by an experiment by Motoyoshi (2004), the methods of which are summarized in Figure 3. He presented four small grating patches arranged in a square configuration, with the distance between patches varied over conditions. All gratings switched between vertical and horizontal, always at the same reversal rate; switches were produced by smoothly changing the contrast of the two orientations in a reciprocal fashion. Reversals in orientation of one of the four gratings was offset in time by varying amounts relative to the other three (which changed orientation in synchrony), and on each trial, observers identified which one of the four gratings was out of synch with the other three. Temporal resolution on this task is thus indexed by the smallest temporal offset between one pattern and three others. Motoyoshi's results showed that under optimal conditions (slowly flickering patterns in close spatial proximity), observers could resolve offsets on the order of 30 msec. Temporal resolution was, however, strongly dependent on both temporal frequency and spatial separation (see also Forte, Hogben, & Ross, 1999), leading Motoyoshi to conclude that perceptual synchrony is governed by mechanisms conjointly sensitive to spatial and temporal factors. Later in our review, we will see that the efficacy of temporal structure in visual grouping is also strongly dependent on spatial interactions. It is worth noting, inciden-

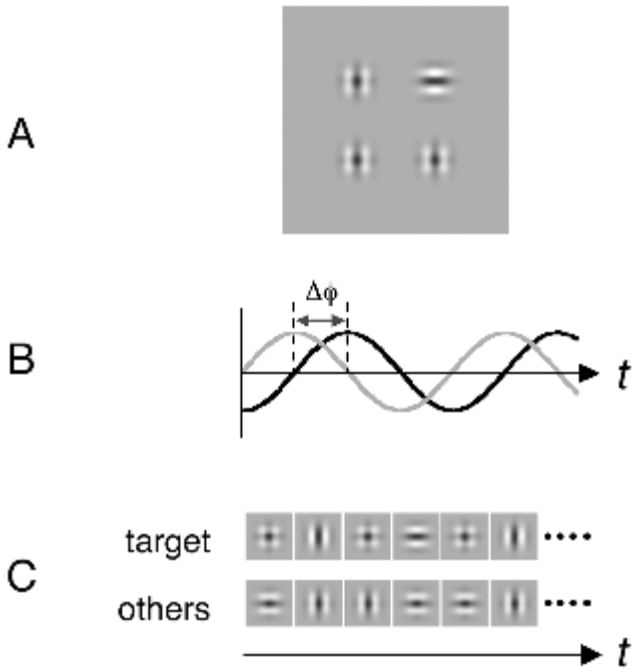

Figure 3: Schematic of Temporal Resolution Task Devised by Motoyoshi (2004).

NOTE: (A) Four grating patches were arrayed in a square configuration (the spatial separation of the patches was varied over conditions). Within each patch, the contours switched back and forth repetitively from horizontal to vertical (with the transition occurring smoothly by sinusoidally modulating the contrast of the two orientations); the switch rate was varied over blocks of trials. (B) and (C) Plots of changes in orientation over time (t). Orientation changes in one of the four grating patches (the "target" shown in light gray in panel B) were slightly out of phase (shown in panel B as $\Delta\phi$) with the orientation changes in the other three patches (all of which switched orientation in perfect synchrony, shown in black in panel B); the location of the target within the array was varied randomly over trials, and following each trial, observers had to indicate which one of the four constituted the target.

tally, that temporal resolution estimated using repetitive flicker tends to be poorer than that indexed by discrete, nonrepetitive events. These two classes of stimuli have substantially different energy distributions within the temporal frequency domain, and repetitive stimulation engages temporal summation in ways that discrete events do not.

Besides the study of visual simultaneity (synchrony), there are also studies aimed at measuring the perception of event synchrony between different aspects of vision. Thus, for example, when the direction of motion of an array of stimulus elements reverses periodically over time and also the color of those elements changes between two values periodically over time, the motion change has to occur slightly in advance of the color change for the two events to appear perceptually simultaneous (Arnold, Clifford, & Wenderoth, 2001; Moutoussis & Zeki, 1997; Nishida & Johnston, 2002). In a similar vein, when visual events and auditory events are paired closely in time, the auditory events typically need to lag the visual events by a brief but reliable duration for the two events to be perceived as simultaneous (e.g., Exner, 1875). VanRullen and Koch (2003) recently reviewed this rich literature on perceptual simultaneity, so we will not cover those studies here. Suffice it to say that this general question of perceptual timing has implications for visual temporal resolution and, therefore, temporal structure and grouping.

In summary, results from studies measuring temporal differentiation indicate that the visual system can resolve sequential visual events with relatively high precision. Although estimates of resolution for temporal differentiation vary depending on the task and stimulus conditions, those values are considerably smaller than the values characterizing temporal integration. How do we account for these seemingly large discrepancies? How is it, in other words, that human vision can realize high temporal resolution while at the same time integrating visual signals over relatively long durations?

One explanation appeals to the involvement of different visual channels, or pathways, distinguished by their temporal properties. As popularized several decades ago, this multichannel hypothesis posited the existence of a sustained channel concerned with form analysis and a transient channel specialized for motion (Breitmeyer, 1984; Burr, 1980; Watson, 1986). Given what we know about these putative channels, it is easy to imagine that the sustained channel could support temporal integration and the transient channel temporal differentiation. Alternatively, one could envision both integration and differentiation being mediated by a single mechanism whose neural response profile varied over time relative to stimulus onset or stimulus offset. To our knowledge, this latter hypothesis has not been formalized, but there are reasonable precedents for this idea in the literature. We know, for example, that the duration of a relatively weak light flash can be adjusted to make that flash just as detectable as a brief, bright light flash—this is the classic example of temporal integration termed Bloch's law. Yet those equally detectable light flashes are nonetheless easily discriminated from one another (Zacks, 1970). Evidently, then, observers have access not only to the magnitude of the neural response triggered by a stimulus but also to the distribution of that response over time, as determined by the temporal structure of the stimulus. We see no reason why the same kind of multicoding could not be involved in the mediation of temporal integration and temporal differentiation.

So, based on studies of temporal differentiation, we conclude that human vision possesses reasonably good temporal resolution. This, in turn, suggests that human vision should be able to register the temporal structure associated with dynamic visual events. But can human vision exploit that temporal structure to promote visual grouping of features possessing common temporal structure? This question brings us to the main theme of this article, spatial grouping based on temporal structure.

SPATIAL STRUCTURE FROM TEMPORAL STRUCTURE

Can common fate in the form of temporal structure promote spatial grouping and figure/ground segregation? The answer is most certainly yes, and in the following paragraphs, we describe the results that substantiate this affirmative answer. As a prelude to reviewing those studies, we should say a word about the different forms of temporal structure that have been used and about the various possible "carriers" of that temporal structure. The general concept of temporal structure is grounded in information theory and signal processing, and readers interested in the quantitative details underlying those conceptualizations are referred to de Coulon (1986) and/or Brook and Wynne (1988). For purposes of our literature review, however, it is sufficient to distinguish between two categories of temporal structure: *deterministic* and *stochastic*.

Deterministic temporal structure refers to predictable time-dependent changes in a stimulus along some visual dimension, meaning that the points in time at which change occurs can be specified a priori by some mathematical expression. To give a few examples, a spot of light flickering on and off repetitively constitutes an event with deterministic temporal structure. So, too, does a grating pattern whose spatial phase is reversed regularly over time or whose contrast is modulated up and down regularly over time. For each of these events,

temporal structure can be summarized by a parameter specifying the flicker rate, the reversal rate, or the modulation rate. Deterministic temporal structure can also take more complex forms, such as that embodied in a frequency modulated counterphase grating (i.e., repetitive phase shifts at a rate that varies predictably over time). Deterministic changes convey very little information, in the sense that these events involve no uncertainty about their time-course once the rate of change has been determined. In its simplest form, repetitive change is monotonous.

Stochastic temporal structure refers to time-dependent changes in a stimulus that are unpredictable, because the points in time at which change occurs are determined by a random process. For stochastic temporal structures, we can only make statistical predictions about when change will occur, with those predictions based either on learning or on prior probabilities. Stochastic temporal structure "looks" very different from deterministic temporal structure,[2] in the same way that a regularly textured figure looks quite different from a randomly textured one (see Figure 4). Stochastic temporal structure can be defined using what engineers and natural scientists term a *point process*—a time series denoting when state changes occur in a system. Applications utilizing point process modeling range from predicting natural catastrophes such as forest fires and earthquakes to controlling data traffic in electronic communication networks to analyzing the times and locations of criminal events. In the case of vision, point process models can be used to characterize unpredictable changes along any dimension for which stochastic changes occur, including luminance, contrast, and phase. For any given point process that defines stochastic temporal structure, we can derive measures of central tendency and variance, and we can express the degree of unpredictability of change in terms of entropy (a quantity related to the Poisson rate parameter generated in a given time series). We shall return to the notion of entropy shortly, when describing experimental results. For now, it is sufficient to point out that stochastic temporal structure is inherently unpredictable and, therefore, transmits more information than deterministic temporal structure when events *do* occur. Visual changes embodying stochastic temporal structure are not monotonous.

We believe this distinction between deterministic and stochastic temporal structure is potentially important. Unlike most engineered communications systems that use well-defined frequencies to transmit information, biological systems often have to deal with unpredictable environmental events that occur within noisy backgrounds. Elsewhere we have conjectured that grouping from stochastic temporal structure is robust because stochastic change conveys more information and more effectively engages the biological machinery from which vision is crafted (Blake & Lee, 2000).

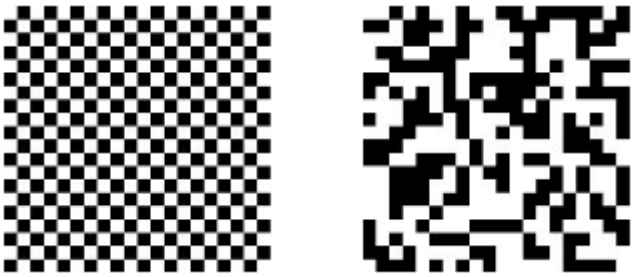

Figure 4: Regular Versus Random Spatial Structure
NOTE: Both figures consist of black and white squares of the same size and the same average density. The two textures look quite different, however, owing to the spatial configuration ("regularity") of the squares. The analogs in the case of temporal structure are predictable, repetitive change in some stimulus feature (e.g., luminance) versus irregular, unpredictable change in that feature.

Before proceeding to a review of experiments, we also should underscore that temporal structure, regardless of whether it is deterministic or stochastic, can be specified independently of the carrier(s) of that temporal structure. Thus, in the examples given above, we saw instances where temporal structure was defined by luminance change (flicker), by phase change, and by contrast change. One could also envision conditions where temporal structure was embodied in changes of the color of a test patch, changes in the orientation of a contour, or changes in the depth plane of a disparity-defined surface. Indeed, a relevant question concerning spatial grouping from temporal structure is the efficacy of different carriers of temporal structure. Given that different aspects of vision exhibit different degrees of temporal resolution (e.g., color is reputed to be temporally "sluggish"), we would expect different carriers to vary in their effectiveness as grouping cues when temporal structure was the defining feature. Some evidence to this effect has been reported (e.g., Guttman, Gilroy, & Blake, in press; Suzuki & Grabowecky, 2002).

With those distinctions in place, we will now move through a survey of experiments bearing on the role of temporal structure in figure/ground segmentation and grouping, organizing them in terms of the nature of the temporal structure embodying it. Within each section, the order of coverage follows a more or less chronological sequence.

DETERMINISTIC TEMPORAL STRUCTURE

One of the simplest ways to manipulate temporal structure is to introduce a temporal phase lag between two sets of repetitively flickering elements. If human vision is sensitive to the phase lag, the two sets of ele-

ments should form separate, distinguishable groups. This is exactly the idea contained in the two-frame animation created by Rogers-Ramachandran and Ramachandran (1998), shown schematically in Figure 5a. Black and white spots are spatially distributed on a gray background creating a texture border in one frame. In Frame 2, the luminance polarity of each spot is reversed. When these two frames are rapidly and repetitively alternated over time, the two groups of dots will flicker in counterphase relative to one another. When the spots flickered at 15 Hz (producing a 33.3 msec phase difference between light and dark elements), observers could still perceive the texture boundary defined by luminance. However, they could not discern whether any two spots were flickering in-phase or out-of-phase. In other words, all the spots appeared to be identical when flickered at 15 Hz, with the border still being distinct. Rogers-Ramachandran and Ramachandran dubbed this a "phantom" contour, because clear texture segregation was perceived even though the texture elements themselves were indistinguishable. Temporal information alone seemed to be creating segregation of the two regions.

In a similar vein, Fahle (1993) created temporal structure among flickering elements by manipulating the stimulus onset asynchrony between two, nonoverlapping groups of flickering stimuli. The stimuli were arrays of regularly or randomly spaced dots (Figure 5b), with the dots within a small "target" region (denoted by a rectangle in Figure 5b) flickered synchronously at a given temporal frequency. The remaining dots outside of the target region were also flickered in synchrony at the same rate, but their temporal phase was delayed relative to that of the target dots. Observers judged the shape or the location of the region defined by "target" dots, and Fahle varied the time delay (phase lag) between the flickering target and surround dots. The threshold time delay varied depending on the flicker frequency, but under optimal conditions observers could perform the task with phase lags as brief as 6-7 msec. Fahle concluded that the visual system can segregate a visual scene into separate regions based on "purely temporal cues" because the dots in the figure and those in the background were homogeneous in static information and differed only in temporal phase. Kojima (1998) also confirmed this finding using spatial-frequency filtered random dot textures. In Kojima's study, observers easily perceived figure from background even when the patterns composing the two subregions were delayed by only 13 msec.

The results just summarized imply that temporal delays as brief as 5-15 msec can effectively promote spatial grouping and texture segregation. However, two other studies have reported seemingly contradictory results. Kiper et al. (1996) examined the role of temporal information in visual segmentation by asking whether onset asynchrony of texture elements can influ-

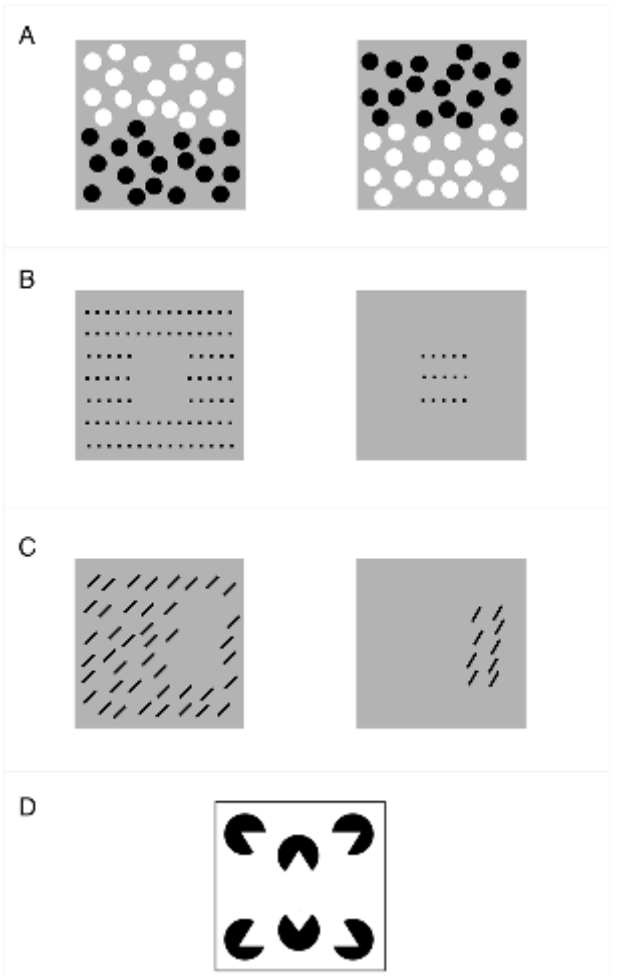

Figure 5: Schematics of Visual Displays Used to Examine the Role of Temporal Cues in Spatial Grouping.

NOTE: (A) The two frames of an animation devised by Rogers-Ramachandran and Ramachandran (1998). When these two frames are alternately presented at 15 Hz (30 frames per second), observers see a distinct, horizontal border even though the spots in the upper and lower halves of the display appear indistinguishable in lightness. (B) Example of two frames from an animation used by Fahle (1993). When these two frames are shown in succession with a slight delay between the offset of one and the onset of the other, observers readily perceive a rectangle. The minimal delay yielding shape perception depends on the rate at which the two frames are repetitively presented. (C) Texture arrays used by Kiper, Gegenfurtner, and Movshon (1996). The "target" region was defined by differences in orientation between "target" and "background" and/or by differences in temporal phase of target and background (i.e., background elements were presented slightly earlier or later than the target elements). (D) Display used by Fahle and Koch (1995), consisting of two sets of partially occluded circles that create the impression of two Kanizsa triangles one on top of the other. The "closer" triangle fluctuates between the two over time. In their experiment, Fahle and Koch flickered the stimulus elements defining one triangle in synchrony while flickering the stimulus elements for the other triangle out of phase.

ence performance in texture segmentation and grouping. In their experiments, observers discriminated the orientation (vertical vs. horizontal) of a rectangular region containing line segments different in orientation from those in a surrounding region (Figure 5c). Spatial similarity and temporal similarity were manipulated as independent variables. The spatial factor was the angular difference in orientation between texture elements in target and surround regions. The temporal factor was the difference in onset time between target and surround elements. Kiper et al. reasoned that if temporal phase can be utilized by human vision for texture segmentation, performance should be enhanced when texture elements in the target region are presented out of phase with those in the surrounding region, because temporal phase provides additional information for the task. However, they found no influence of temporal asynchrony on texture segmentation; performance depended entirely on the magnitude of the difference in orientation between target and surround.

Also arguing against the efficacy of temporal structure as a grouping cue are results from a study by Fahle and Koch (1995). These investigators created an ambiguous stimulus, two overlapping Kanizsa triangles (Figure 5d), that could be seen in either of two configurations, and they examined whether flickering components of a given configuration could bias observers to perceive that configuration. Without flicker, observers experienced perceptual rivalry: one triangle seemed to occlude the other for several seconds, with the "front" triangle switching between the two alternatives every few seconds. As expected, disrupting the spatial configuration of one of the two triangles led to the other, unperturbed triangle being seen predominantly in front. Disruptions in the temporal configuration, however, had no such effect. Over a range of flicker frequencies (10-75 Hz), temporal phase differences among the components of a given configuration did nothing to weaken that configuration's predominance. These results, together with those of Kiper et al. (1996), question whether temporal structure plays a prominent role in spatial grouping.

So at this point in the chronology, we have some results showing robust grouping by temporal synchrony and other results showing no effect of temporal synchrony on grouping. How do we resolve these conflicting results? A clue to this question may come from the interactive relationship between spatial and temporal cues in spatial organization. In the studies showing that flicker does contribute to grouping (Fahle, 1993; Kojima, 1998; Rogers-Ramachandran & Ramachandran, 1998), the displays contained no coherent spatial information for segmentation—all stimulus elements in the displays were identical in terms of form, orientation, and color. Spatial grouping was defined solely by temporal phase lags among elements within distinct regions. On the other hand, prominent spatial cues were always present in the displays used in those studies that failed to find an effect of temporal phase and flicker. (In the study by Kiper et al., 1996, differences in orientation defined figure and ground; in the study by Fahle & Koch, 1995, luminance edges induced contours that conspicuously defined the two competing triangles.) Perhaps, then, the efficacy of temporal structure is constrained by the presence and strength of spatial structure.

This possibility was explicitly tested by Leonards, Singer, and Fahle (1996), who used a display modeled after the one employed by Kiper et al. (1996) (Figure 5c). Figure and ground regions could be defined by a difference in temporal phase alone, by a difference in orientation alone, or by both temporal phase and orientation differences. Observers were able to identify a "figure" provided that the figure was defined solely by temporal cues or when those temporal cues were consonant with the spatial cue. When, however, the two cues were in conflict, spatial cues dominated.

This interaction between spatial and temporal structure is also revealed in a series of experiments performed by Usher and Donnelly (1998). They created a square lattice display (Figure 6), in which elements could be grouped either into rows or into columns. When the elements in alternating rows (or columns) of the lattice were flickered asynchronously (out of phase), the display was perceived as rows (or columns) correspondingly. Temporal phase, in other words, determined global perceptual organization in this otherwise ambiguous display. Of particular relevance, the efficacy of asynchronous flicker was governed by the shape of lattice elements. The strongest grouping effect from temporal structure was obtained when using circles rather than crosses in the display. In line with the findings of Leonards et al. (1996), this result probably means that the efficacy of temporal structure is constrained by the presence and strength of spatial structure—the lattice of crosses contains abundant collinearity, whereas the lattice of circles does not. In another experiment, Usher and Donnelly (1998) asked observers to detect a target of collinear line segments embedded in an array of otherwise randomly oriented line segments. Performance was better when target and background line segments were flickered asynchronously than when they were synchronized. Note that in this condition temporal and spatial structure were congruent with one another and that temporal structure enhanced the perception of spatial structure. Targets consisting of randomly oriented line segments, and thus defined solely by temporal structure (flicker asynchrony), could also be detected effectively, although the phase lag required for grouping was somewhat longer.

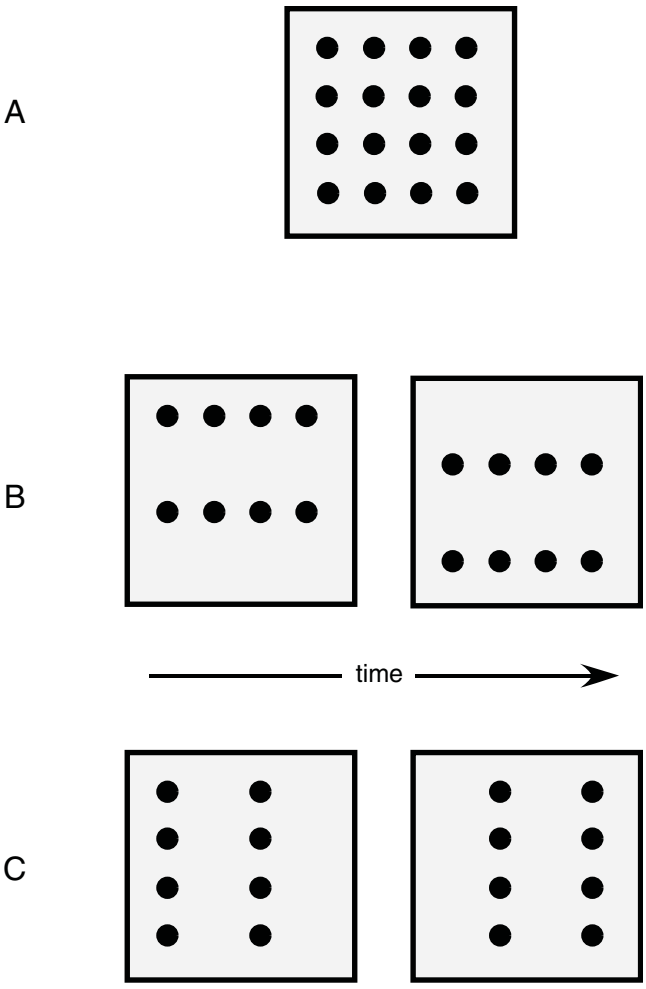

Figure 6: Schematic (Not Exact or to Scale) of One of the Displays Used by Usher and Donnelly (1998) in Their Study of Visual Synchrony and Grouping.

NOTE: (A) An ambiguous "composite" display seen either as rows of black circles or as columns of black circles; over trials, these two alternative perceptual outcomes are equally likely when the array is briefly flashed. (B) and (C) When components of the composite are presented sequentially in time (one frame immediately following the other), perception is biased in favor of rows (panel B) or columns (C), even though the successive exposures were sufficiently close in time to appear simultaneous.

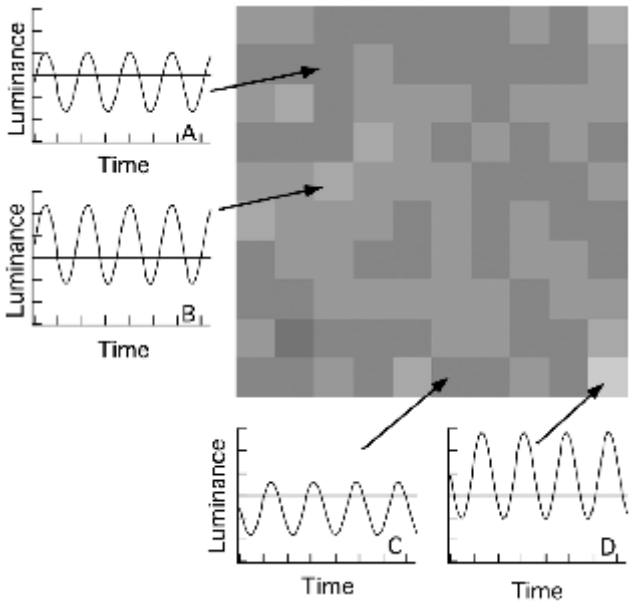

Figure 7: Display Used by Sekuler and Bennett (2001) to Study Grouping by Common Luminance Changes.

NOTE: All squares within a matrix changed in luminance sinusoidally over time within a limited range of luminance values, with the range (and mean) varying among squares. A subset of contiguous squares within the matrix (which defined the "target") became lighter and darker together (two of the target squares are shown as A and B in the figure), whereas the remaining squares also became lighter and darker together but with a slight phase lag relative to the target squares. The average luminance within target and background was the same; only the time difference in luminance modulation distinguished target from background.

SOURCE: Figure reproduced with permission of Sekulet and Bennett (2001) © American Psychological Society.

In yet another demonstration of grouping based on temporal information, Sekuler and Bennett (2001) devised a clever display consisting of a 10×10 array of squares whose individual luminance values varied randomly throughout the array (see Figure 7). The luminance values of all squares varied sinusoidally over time, such that each square went from light to dark in a smooth, repetitive fashion. One small, rectangular-shaped cluster of squares, the "target" region, increased and decreased in-phase (meaning that the peaks and troughs of the sinusoidal modulations were aligned); all the remaining squares outside of this target region also modulated in-phase with one another but not in-phase with the target elements. In other words, the target was defined by a cluster of squares that became progressively lighter and darker together, relative to the direction of the luminance modulations in the background. Observers were able to identify the orientation of the target region even when the depth of modulation was less than 2%, an incredibly small amount of change over time. Sensitivity to phase differences in luminance modulation was best at modulation rates of 5 Hz and greater. Sekuler and Bennett also varied the phase lag between sinusoidal modulations of target and background, a manipulation that varies the time separating the peaks in the two sets of modulating elements. The task remained easy even with phase differences as small as 22 degrees, which translates into a time difference as brief as 6.5 msec at a flicker rate of 9.6 Hz. Sekuler and Bennett interpreted their findings as an indication that the Gestalt notion of common fate extends beyond motion to include synchronized luminance change.

The experiments described so far all utilized dynamic stimuli that, in principle, could activate motion mecha-

nisms differentially within figure and ground regions.[3] Cognizant of this possibility, Kandil and Fahle (2001) set out to design a figure/ground grouping display in which motion was explicitly present but unequivocally insufficient to support grouping. They also wanted to design animations that would prevent grouping based on luminance differences, a lingering concern in some of the studies described earlier in this section. To accomplish their goal, Kandil and Fahle created animations in which each frame contained a large number of regularly spaced dot pairs, or "colons" as they called them (see Figure 8). Each pair of dots flipped their angular orientation by 90 degrees, with these flips occurring repetitively and periodically every other frame of the animation. Dot pairs within a virtual "figure" region flipped on even-numbered frames and dot pairs within the virtual "surround" flipped on odd-numbered frames. Segmentation was thus defined by a phase offset between the flip times of the figure and ground dot pairs. Observers were reliably able to identify the shape of the virtual figural region at flip frequencies just over 20 Hz (which corresponds to a timing difference of 22 msec between figure and ground events). Interestingly, this temporal acuity was approximately halved (i.e., temporal resolution was reduced twofold) in observers older than 50 years of age, although Kandil and Fahle did not speculate on possible reasons for this performance decline. They also modified their animations so that they could vary the phase-lag between figure and ground flips while holding flip frequency constant. With this maneuver, they found that young observers could group dot pairs based on synchronized motion down to differences as small as 11 msec. They interpreted their findings as definitive evidence for the effectiveness of "astonishingly" short temporal delays as a grouping cue, uncontaminated by luminance or motion artifacts. In a follow-up article, Kandil and Fahle (2003) explored boundary conditions for time-based figure/ground segmentation. They found that isoluminant stimuli (dot pairs defined solely by color) had slightly poorer temporal resolution compared to luminance-defined stimuli. They also tested conditions involving dichoptic presentation of stimulus elements (i.e., elements moved from one position in one eye to another position in the other eye). Here the target would be perceived only if the images shown to the two eyes were being matched between the eyes over time. Significantly, dichoptic stimulation abolished figure/ground segmentation, implying that monocular neural mechanisms mediate this form of spatial structure from motion.

Considered together, the results summarized in this section indicate that the temporal structure created by repetitive luminance flicker can promote grouping and segmentation, with the efficacy of temporal structure

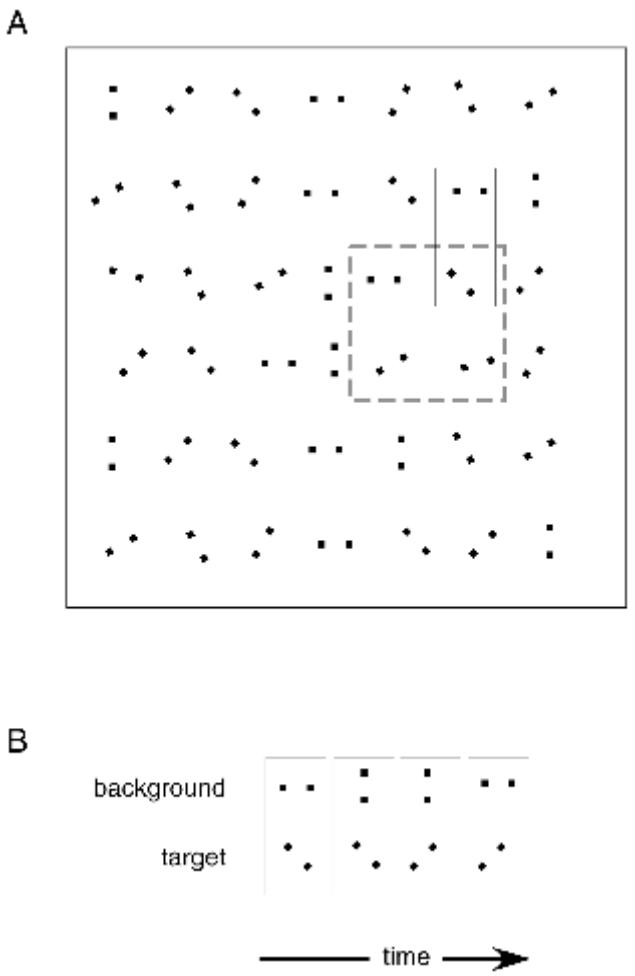

Figure 8: Schematic of the Display (Not Exact or to Scale) Used by Kandil and Fahle (2001) to Study Figure-Ground Segregation Based on Temporal Phase.

NOTE: On each frame of an animation, the observer saw an array of "colon"-shaped elements (pairs of dots whose virtual orientations varied irregularly throughout the array). Every other frame of the animation, each dot pair flipped 90 degrees in virtual orientation, with the point of rotation being the imaginary midpoint of the dot pair. Dot pairs within a "target" region flipped in synchrony, and dot pairs within the background flipped in synchrony out-of-phase with the target dot pairs (see Panel B for an example of one target dot pair and one background dot pair shown on four successive frames). The target dots and background dots flip at different points in time, providing a temporal cue for figure/ground segmentation.

modulated by the existence and strength of spatial structure within the display. This conclusion is not limited to figure/ground organization defined by luminance flicker. There are also a couple of studies that have examined grouping of spatially distributed features based on repetitive contrast modulation as the source of temporal structure. Those studies are summarized in the following paragraphs.

In one study, Alais, Blake, and Lee (1998) created a display comprising four spatially distributed apertures

each of which contained a sinusoidal grating that when viewed alone, always appeared to drift in the direction orthogonal to its orientation (Figure 9a). When all four gratings were viewed together, however, they intermittently grouped to form a unique global motion whose direction corresponded to the vector sum of the component motions (e.g., Lorenceau & Shiffrar, 1992). With extended viewing, then, the display appeared bistable: Observers experienced perceptual fluctuations between global motion and local motion. To evaluate the role of temporal structure in grouping, Alais and colleagues independently modulated in time the contrast levels of the four gratings and assessed the influence of this time-varying event on the incidence of global motion. When contrast modulations were correlated (meaning that the timing and direction of contrast changes were the same among all four gratings, even though their absolute contrast values differed), the four gratings were much more likely to group into a single, coherent global object moving in the vector sum direction; when contrast modulations were uncorrelated, local component motion was much more likely. Alais and colleagues obtained similar results in another experiment using two superimposed, drifting gratings (Figure 9b). Under appropriate conditions, this display, too, is perceptually bistable, appearing either as two transparent gratings drifting in different directions or as a coherent "plaid" pattern moving in the direction defined by the vector sum of the two component velocities (Adelson & Movshon, 1982). Again, the incidence of coherent motion was enhanced by correlated contrast modulations and was reduced by uncorrelated contrast modulations.

Temporal patterning of contrast modulation and grouping has also been examined using another form of bistable perception, binocular rivalry. When dissimilar patterns are imaged on corresponding areas of the two eyes, they compete, or rival, for perceptual dominance. When one views multiple pairs of rival targets spatially distributed within the visual field, dominance among those multiple targets can become entrained if those targets are similar in color, orientation, or motion (Alais & Blake, 1998; Kovacs, Papathomas, Yang, & Feher, 1997; Whittle, Bloor, & Pocock, 1968). Alais and Blake examined whether correlated contrast modulations of local rivalry patterns can also promote simultaneous dominance during piecemeal rivalry. Figure 10 shows schematics of the displays used in their study. During 60-sec viewing periods, observers viewed these dichoptic rival displays and pressed buttons to track conjoint dominance of the two gratings. Correlated contrast modulation between the two gratings promoted joint predominance of those gratings more than did uncorrelated modulations. Moreover, the effectiveness of correlated contrast modulation was dependent on the spatial configuration of the two gratings, being maximum when the two were collinear. This latter outcome is in line with the results described above, showing that the efficacy of temporal structure is affected by existing spatial structure.

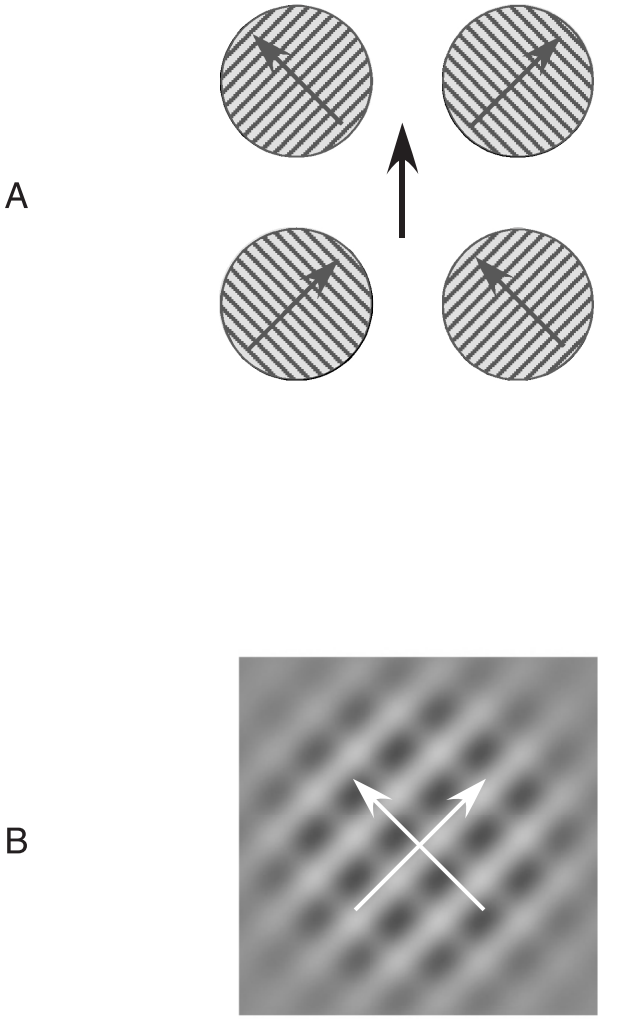

Figure 9: Examples of Visual Stimuli Used by Alais et al. (1998) to Study the Role of Common Temporal Structure in Visual Grouping.

NOTE: (A) Four circular grating patches arrayed in a square configuration. Within each patch, contours of the grating drifted steadily in one direction (indicated by arrows). Over time, observers sometimes see the four separate directions of motion ("local" motion) and other times see the four motion vectors group and appear to form a single, partially occluded grating that drifts upward ("global" motion). Correlated contrast modulation of all four gratings enhances perception of global motion, implying that common temporal structure promotes grouping. (B) A plaid produced by superimposition of two sinusoidal gratings. The two gratings drift upward in a direction orthogonal to their contour orientations (indicated by the white arrows). Perceptually, the two gratings sometimes cohere and appear to move as a single "object" upward. Correlated contrast modulation of the two gratings enhances perception of coherent, global motion.

In a recent article, Suzuki and Grabowecky (2002) studied the role of temporal synchrony on grouping

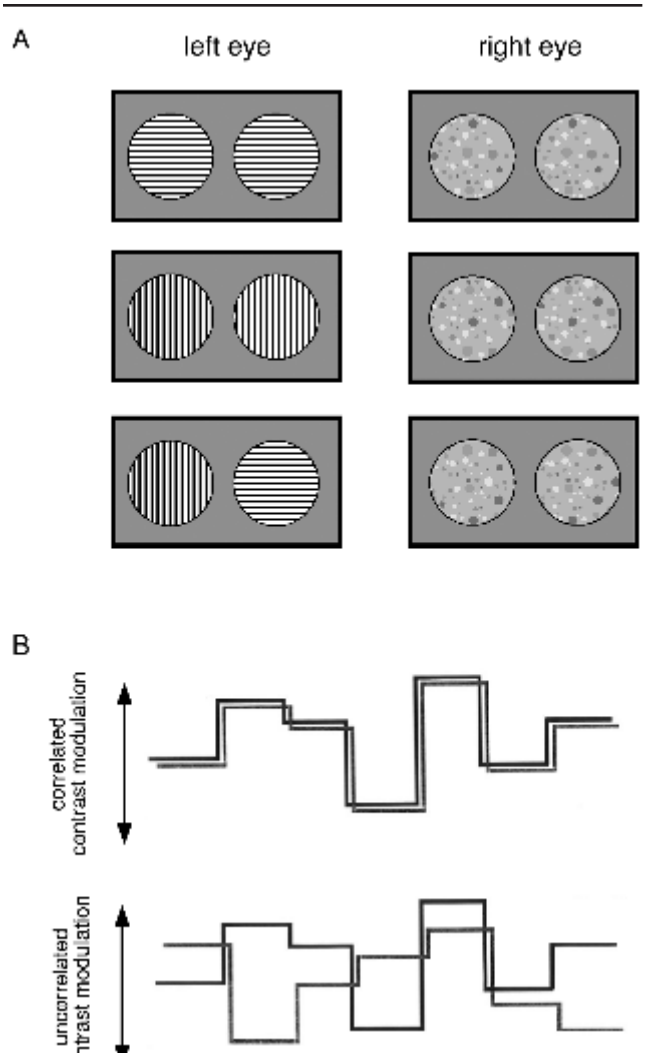

Figure 10
NOTE: (A) Three Examples of dichoptic stimuli (left-eye and right-eye images viewed through a mirror stereoscope) used by Alais and Blake (1999) to study spatial and temporal factors influencing predominance during binocular rivalry. Observers viewed a pair of half-images and pressed a button to indicate when both circular patches of grating were dominant simultaneously ("joint predominance"). The incidence of joint predominance was highest for the rival pair with collinear orientations (top panel) and lowest for the pair with orthogonal orientations (bottom panel). (B) When the contrast levels of both gratings were modulated over time in a correlated fashion (direction and magnitude identical for both), the incidence of joint predominance was elevated for all three grating configurations, relative to conditions where contrast levels were modulated in an uncorrelated fashion.

using overlapping visual features that flickered periodically, one being two pairs of orthogonally oriented bars and the other a set of small circles. The bars switched repetitively from diagonal left to diagonal right, and the set of circles flashed on periodically in synchrony with one of the two bar orientations. The rate of change in bar orientation was varied beyond a value at which the switches were perceptible—beyond this exchange rate, observers saw two, superimposed sets of oriented bars. However, the two orientations waxed and waned in visibility, with diagonal left being dominant for several seconds only to be replaced by diagonal right for a few seconds. This outcome is not surprising (see Atkinson, Campbell, Florentini, & Maffei, 1973). What is remarkable is that the set of dots appeared to be affixed to the oriented bars with which those dots were synchronously flashing; when the orthogonally oriented bars were dominant, the dots appeared to form a separate cluster of features seen transparently in relation to the bars. This "binding" of the dots to the bars flashing in synchrony was not observed when the dots were equiluminant with the background, and thus defined by color alone. The failure of temporal structure to group color may well be attributable to the temporal sluggishness of the chromatic system.

So to sum up so far, studies using luminance-defined temporal structure and contrast-defined temporal structure all support the idea that human vision *can* use temporal structure for spatial grouping. The efficacy of temporal structure is greatest when existing spatial cues define ambiguous spatial structure or when preexisting spatial structure does not exist in the absence of temporal structure. When spatial and temporal cues are in conflict, the relative salience of the two cues determines the grouping outcome.

Grouping by Stochastic Temporal Structure

In many of the studies reviewed in the last section, local elements in one region of the display (the "figure") flickered on and off repetitively in phase while, at the same time, elements in the rest of the display (the "ground") flickered together but out of phase with the flickering figure elements. This means, in other words, that only figure elements were visible in some frames of the animated sequence and only ground elements were visible in other frames. This point is dramatized by slowing the flicker rate to the point where individual frames can be inspected—in this case, one can easily discern the figure from ground because the two regions are explicitly defined by spatial discontinuities in luminance, a potent cue for figure/ground segregation. As flicker rate increases, this luminance cue becomes less conspicuous because of temporal summation across frames; consequently, figure/ground organization becomes less salient although discernible. Thought of in this way, one realizes that the studies reviewed above do not unequivocally test whether human vision can group visual features based strictly on temporal structure. In fact, what those studies were measuring is the upper temporal limit for differentiating sequential images each of which contains spatial structure defined by luminance discontinuities.

In this respect, those studies resemble the form/part integration task described earlier in this article.

Besides not unequivocally isolating temporal structure, all of those studies utilized periodic flicker, a highly predictable, deterministic signal. For several reasons, we question whether deterministic temporal sequences are the preferred choice for studying grouping. First, deterministic temporal structure is quite rare in the natural environment where the visual system more often confronts irregular and unpredictable events. Second, from an information theoretic viewpoint, periodic flicker conveys very little information, which suggests that flicker may underestimate the ability of human vision to utilize temporal cues for grouping. Finally, flicker leaves little room for manipulating temporal relationships among local elements. When all elements undergo periodic fluctuation in luminance or in contrast, the only way to manipulate the temporal relationship among those elements is to vary their relative phase. But human vision may be sensitive not only to phase differences but also to coherence/incoherence in global temporal structure among local visual elements.

Is there an alternative means for producing temporal structure that bypasses these limitations inherent in deterministic visual signals? This requires creating animations in which individual frames contain no static cues whatsoever about spatial structure and in which temporal structure is unpredictable. How can this be done?

A key for overcoming these challenges can be found from an understanding of "temporal structure" conveyed by time-varying signals. To illustrate, consider a single spot of light that changes in luminance irregularly over time. We may plot the actual luminance values as a function of time, producing a graph like the one shown in Figure 11a. Alternatively, we may produce a time series like the one in Figure 11b that conveys just the points in time at which changes in luminance happen, irrespective of the direction and magnitude of those changes. This latter plot is termed a "point process," and it portrays the temporal structure conveyed by this series of events. Now, if we can create an animation display in which groups of local elements differ only in their point processes, and not any other stimulus properties when considered frame by frame, that display can be said to be devoid of static cues for grouping. Moreover, because the point process defines a stochastic event, the information content of the event would be greater, and perhaps more effective, than the information content of deterministic events like flicker.

With those considerations in mind, Lee and Blake (1999a) devised a novel class of visual displays in which global spatial form was defined exclusively from differences in point process among elements undergoing

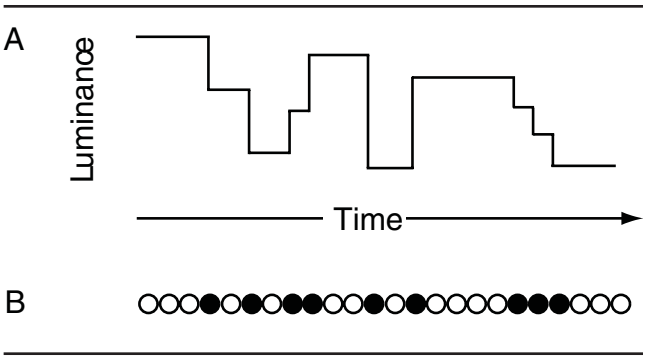

Figure 11: Alternative Ways to Portray Irregular Changes in Luminance of a Stimulus Spot Over Time.
NOTE: (A) The actual level of luminance is plotted as the function of time; this kind of graph is comparable to the plots of deterministic change shown in Figure 7. (B) Time series showing the points in time at which luminance level remains constant (open circles) and points in time at which it changes (filled circles); the resulting time series is called a point process, and it specifies temporal structure irrespective of absolute luminance value.

rapid irregular changes. In our original version of these animations, each individual frame contained an array of many small sinusoidal gratings each windowed by a stationary, circular Gaussian envelope (Gabor patch; see Figure 12). All of the Gabor patches had the same contrast, and their orientations were randomly established. As the animation was played, grating contours within each small, stationary Gabor patch moved in one of two directions orthogonal to their orientation. Each grating reversed direction of motion irregularly over time according to a random (Poisson) process. Temporal structure of each Gabor element was described by a point process consisting of points in time at which motion reversed direction (see Figure 12). All Gabor elements within a virtual figure region could be assigned the same point process, in which case each of those figural Gabors would reverse their directions of motion simultaneously; elements in the virtual surround would all be assigned a single point process that differed from the one carried by the figural Gabors, meaning that all the surround Gabors would reverse direction of motion simultaneously but often at different times than the figural Gabors. When displays like these are viewed, even for just a few hundred milliseconds, the figure is easily distinguished from the background. Here, the figure region and the ground region differed only in their point processes and not in terms of the contrast or orientation of individual elements. Furthermore, there was no information about the figure in any two successive frames of the animation, because all contours moved from frame to frame. Therefore, only the differences in the two sets of point processes (point processes specifying *when* motion direction changes) distinguished figure from background.

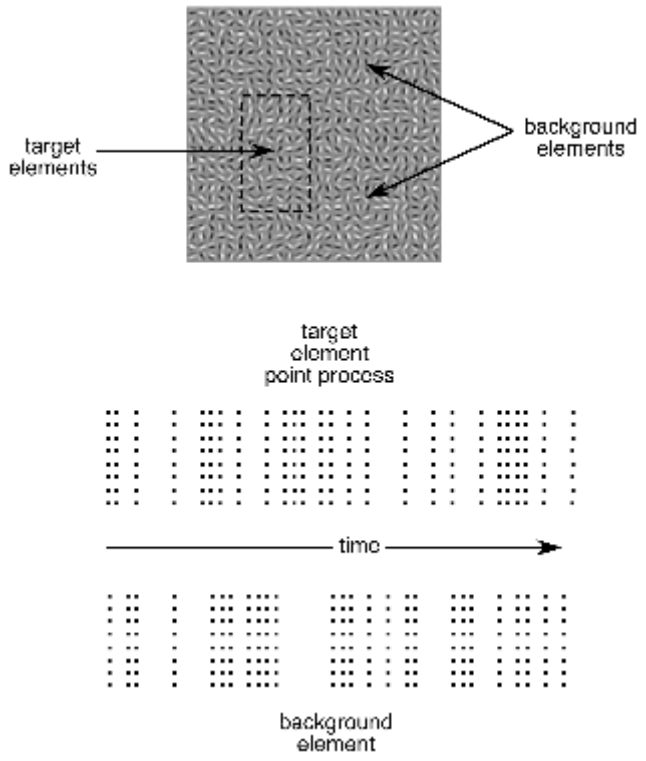

Figure 12: Schematic of Stochastic Temporal Structure Display Devised by Lee and Blake (1999a) To Study Spatial Grouping from Temporal Cues.

NOTE: (A) Animations are made using an array of Gabor patches (Gaussian-windowed sinusoidal gratings), with the orientation of the grating in each patch being randomly determined (thereby precluding any texture cue based on orientation). Over time, each small grating drifts in a direction orthogonal to the orientation of its contours, and at randomly determined times, the direction of motion reverses (e.g., upward drifting contours begin drifting downward). It is important to realize that the patch itself remains stationary—only the contours within the patch move, and there is no coherent motion within the array because motion directions are tied to the random orientations of the contours. All gratings within a virtual "target" region change their directions of motion at the same time, independently of changes in motion direction among the remaining "background" gratings falling outside this region. (B) Temporal structure in these stochastic animations is defined in terms of a "point-process" shown here as a series of black "dots" denoting points in time at which motion direction changes. Each row corresponds to a given Gabor patch, and it can be seen that motion direction changes within the target region and within the background occur in synchrony (i.e., their change times are perfectly correlated even though their directions of motion are random). The correlation in change times between target and background regions is uncorrelated in this example, and when viewed on a video monitor, the target region is easily distinguished from the background. There are several variants of these stochastic displays, including ones where the background elements have uncorrelated point processes and the target elements have correlated point processes. The correlation between target and background point processes can also be manipulated to vary the salience of the target. Finally, it is possible to assign all elements throughout the array the same point process but introduce a temporal phase lag between target and background elements.

Because all local elements throughout the display change directions of motion irregularly over time, we were able to manipulate two potentially important properties governing temporal structure. First, the "predictability" (or "randomness") of temporal structure conveyed by individual elements was manipulated by changing the probability distribution designating the two alternative directions of motion. Borrowing a concept from information theory, this predictability was quantified by computing the *entropy* of temporal patterns among the elements. Time-varying signals with high entropy convey more dynamic or finer temporal structure, which means that systematic manipulation of entropy of all the elements enabled us to measure how accurately human vision can register fine temporal structure. Second, the temporal relationship among elements in the figure region could be manipulated by varying the extent to which all possible pairs of those elements are correlated. Because the time points at which the elements change direction of motion could be represented by point processes, the index of temporal relationship among those elements could be quantified by computing the correlations among their point processes. By varying this index systematically, we were able to measure the efficiency with which people can utilize temporal structure. In our initial work (Lee & Blake, 1999a), we found that increases in entropy and increases in correlation among figure elements systematically enhanced the perceptual quality of spatial form created by temporal structure (as assessed by performance on a forced-choice form identification task). We took this to mean that human vision can register fine temporal structure with high fidelity and can efficiently construct spatial structure solely based on the temporal relations among local elements distributed over space. Readers are encouraged to visit the following Web site to see demonstrations of spatial structure from stochastic temporal structure: http://www.psy.vanderbilt.edu/faculty/blake/TS/TS.html.

From our results, we concluded that perception of spatial structure in these displays required multiple computational steps: (a) registration of changes in direction within local regions of the visual field, (b) registration of the points in time at which those direction changes occur, and (c) identification of boundaries defined by discontinuities in temporal structure (a process that must operate globally over space). We noted that Steps a and b could plausibly be accomplished by transient neural signals generated by motion-selective neurons, but we offered no suggestions about how Step c might be accomplished. In some later work (Lee & Blake, 2001), we found evidence suggesting that grouping could be mediated, in part, by lateral connections among spatially neighboring neurons, with the strength of those connections dependent on the similarity in preferred orientation among those neurons.

Shortly after we published our initial description of this novel technique and the results obtained using it, Adelson and Farid (1999) published a critique of this technique. In that critique, they questioned whether temporal structure per se was responsible for the visibility of a figure within these dynamic displays. They speculated that observers might instead rely on a luminance-based cue that could occasionally become available in these stochastic displays, owing to contrast summation (a possibility we wrote about but rejected in our original article). According to their argument, elements in the target region, but not the background region, may reverse directions several times in succession and, through temporal blurring, momentarily create a target region of heightened contrast. (Alternatively, of course, it could be elements in the surround that summate in relation to the target.) At other times during the sequence, target but not background elements might continue moving in the *same* direction for a number of successive frames, thereby creating through temporal blurring a washed-out region relative to the background. Adelson and Farid confirmed their intuitions by creating new animations from one of our original ones, this time making the contrast of each element in each frame a weighted average of a number of the immediately preceding frames, thereby mimicking the output of a lowpass temporal filter. Inspection of those hybrid animations did indeed reveal infrequent instances where overall contrast within the figure differed from contrast in the surround.

Adelson and Farid's (1999) simulation utilized the simplest possible means for computing the outputs from their filter: averaging luminance over time on a pixel-by-pixel basis, hardly what the visual system would do if lowpass temporal filtering were involved. Nonetheless, their analysis and simulation motivated us to perform several tests of their hypothesis (Lee & Blake, 1999b). First, we indexed animations from our experiment in terms of the magnitude of the theoretical luminance cue created by temporal blurring, using the Adelson and Farid lowpass filtering model. We then computed the correlation between this index and the actual psychophysical performance associated with given displays. We found no correlation between the two, implying that observers were not using this luminance cue even if it were indeed available. Second, we created new stochastic displays from which we removed multiple reversals ("jitter") and extended periods without reversals ("runs"), the putative culprits implicated by Adelson and Farid's analysis. We also introduced differences in average luminance among elements throughout the display, randomly and independently of temporal structure. Finally, we randomized the contrast of each moving element within the array on a frame-by-frame basis. These maneuvers thoroughly eliminate any potential luminance-based cues, as verified by simulations using the Adelson and Farid lowpass filter. Despite these modifications, form from temporal structure was clearly visible. This latter observation is notable, because randomizing contrast and luminance actually creates visual noise that would conflict with form created by temporal structure. Nonetheless, temporal structure was an effective cue for grouping.

Farid and Adelson (2001) next challenged the efficacy of temporal structure in grouping based on results using a modified temporal structure display. In their new animation, drifting dots within a "target" region simultaneously changed directions of motion over time, whereas dots within the remaining "background" region simultaneously changed their directions at times uncorrelated with target change times. Only when angular changes in direction of motion were large did the target dots perceptually group to form a figure whose shape could be accurately identified; angular changes 120 degrees or less yield near-chance performance despite the presence of temporal asynchrony between target and background regions.

Farid and Adelson (2001) reported that perceptual performance covaried with the strength of the output from a temporal filter whose response $h(t)$ is given by:

$$h(t) = (kt/\tau)^n e^{-kt/\tau} (1/n! - (kt/\tau)^2/(n+2)!).$$

This equation defines a *biphasic* temporal filter, which is well suited for registering abrupt, transient visual changes. Indeed, in our original article, we hypothesized that just such "transient detectors" may be involved in signaling temporal structure in stochastic displays involving spatially distributed, time-varying events. Farid and Adelson characterized the output of these filters as providing a "temporal contrast cue" that obviates the need for positing a role for temporal synchrony in visual grouping. In our view, temporal synchrony *in the stimulus* is precisely what produces the structured output in the array of transient filters. Again, as in their previous publication (Adelson & Farid, 1999), Farid and Adelson used static pictures to portray the outputs of biphasic filters during a given instant in time. Such a portrayal gives the impression that these stochastic displays generate *luminance* contrast visible in the spatial domain. But the "contrast" cue arises in the temporal domain, which is exactly the point we made in our descriptions of these stimuli. Moreover, the energy associated with transients (i.e., the outputs of biphasic filters) fluctuates very rapidly over time, in a pattern strongly correlated ($r > .8$) with the temporal structure of the animation displays. This implies that dynamic, not static, temporal patterns of transient signals underlie figure-ground segregation in these

displays. So, we believe Farid and Adelson have identified new conditions under which the strengths of the outputs of transient detectors are closely correlated with human observers' performance on a visual grouping task. In our view, this finding reinforces the hypothesis that human vision possesses the ability to group local features based on temporal structure.

Shortly after the appearance of Adelson and Farid's articles, Morgan and Castet (2002) published an article describing two experiments using stochastic displays much like those schematized in Figure 12. In one experiment, small Gabor elements shifted phase randomly over time, with the elements in the target region changing direction at the same time and elements in the background region changing directions at random times relative to one another. Morgan and Castet found that performance was comparable regardless of whether all Gabor elements were equal in contrast or were randomized across the entire array (a maneuver designed to "camouflage" a figure based on contrast grouping). This aspect of their article replicates the findings of Lee and Blake (1999b), further undermining the hypothesis that form from temporal structure results from contrast artifacts in these displays. In a second experiment, Morgan and Castet sought to test the hypothesis that differences in motion appearance between target and background regions were responsible for the visibility of the target region. By "motion appearance," they were referring to the runs and jitters discussed above—Morgan and Castet observed that targets comprising sequences with long stretches of unchanging motion direction produced "strong motion," whereas targets comprising sequences with a series of reversals appeared to "shiver." To minimize motion appearance as a cue for distinguishing target and background elements, Morgan and Castet created an array of Gabor elements all with the same point process (and, hence, all with the same temporal structure). Gabor elements within the target region all reversed in direction at the same time (i.e., the point processes for all elements were in phase), and Gabor elements within the background had starting frames that were randomized among the elements (i.e., the phases of the point processes for all background elements were uncorrelated). Based on inspection of these animations, Morgan and Castet concluded that coherent motion cues were not visible except in sequences that, by chance, had a string of runs followed by a string of jitter, or vice versa. They also observed that the target area was not visible, except in those unusual sequences with strings of runs and jitters. From these observations, they concluded that temporal structure is at best a weak cue operating only at relatively low temporal frequencies. Morgan and Castet acknowledged that their procedure, because of the limited duration of their sequences,

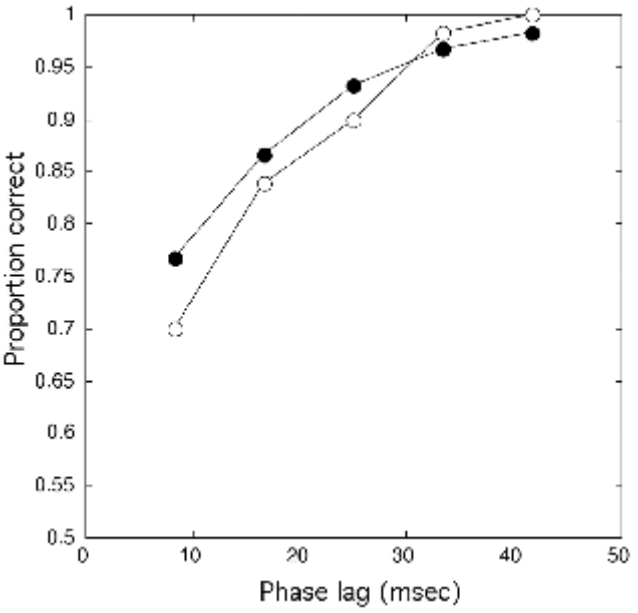

Figure 13: Results From A Two-Alternative, Forced-Choice Experiment in Which Observers Were Required to Report the Orientation ("Vertical" vs. "Horizontal") of a "Target" Region Defined by Common Temporal Structure.

NOTE: The display was much like the one shown in Figure 11, except that all Gabor patches had identical point processes, with the only distinguishing feature being that the point process of the Gabor patches defining the target were phase-shifted in time relative to the point process of the Gabor patches defining the background. The graph plots percent-correct (chance performance = 50%) as the function of the duration of the phase shift between target and background.

increased the opportunity for synchronized changes within target and background regions, which is bound to compromise discriminability of the target region. They concluded their article by writing, "We agree (as is obvious on logical grounds) that synchrony of relatively low temporal frequency modulations of motion and contrast can be the basis for segregation . . . but we argue that the temporal grain is not at the 1000 Hz rate that would be required for synchrony of individual neural spikes" (Morgan & Castet, 2002, p. 516).

We have no quarrel with this conclusion, although we are puzzled by Morgan and Castet's failure to perceive spatial grouping when target and background were distinguished only by phase. That observation is inconsistent with work summarized earlier (e.g., Sekuler & Bennett, 2001) and with results obtained in our lab using forced-choice testing and the same kind of display as that used by Morgan and Castet. In his dissertation, Lee (2002) performed an experiment in which grating elements in the figure and background regions all had exactly the same point process (and, hence, the same first-order temporal statistics). The only cue defining the figure was a temporal phase shift of the "target" grating

elements relative to the background elements. He found that observers performed well above chance levels on a 2AFC task with phase-lags as brief as 16.7 msec (see Figure 13). Those results, which have been published in abstract form (Blake & Lee, 2002), indicate that synchrony per se can support figure/ground segmentation. This phase shift value also squares very nicely with the value of 15 msec found in other studies (e.g., Fahle, 1993; Leonards et al., 1996) including a very recently published article by Kandil and Fahle (2004) using phase-lagged, deterministic temporal structure.

As an aside, we are grateful for the interest in grouping and temporal structure expressed by Adelson and Farid (1999) and Farid and Adelson (2001) and by Morgan and Castet (2002)—their critiques of our technique and ideas have sharpened current thinking about the problem. There is no disagreement that temporal structure can support spatial grouping—the debate now centers around the nature of the mechanisms responsible for that grouping and the temporal "grain" of those mechanisms. It is correct to say that the articles by these two groups of investigators have drawn wider attention to the problem of temporal structure. In so doing, their views on our work have stimulated additional studies on the problem, studies that we turn to next.

Perception of shape from stochastic temporal structure is reminiscent of perception of depth and form from random dot stereograms (where form is defined solely by retinal disparity): For both information sources (temporal structure and disparity), the emergent form can take some time to materialize perceptually when one first views examples of these unusual displays. Practice helps. In the case of temporal structure, Aslin, Blake, and Chun (2002) showed that the people get better at identifying form from temporal structure when given daily training sessions with feedback. Performance reached asymptotic levels within a week or less. Significantly, this improvement in the ability to identify the orientation of a target region defined by temporal structure did not transfer to a task requiring the identification of luminance-defined form. This failure of transfer was found even though both luminance- and temporal structure–defined animations comprised identical stimulus elements, and the task itself was the same. Evidently, the cue used during the temporal structure phase of the experiment had nothing to do with luminance contrast. What did observers actually learn while practicing this task? Aslin et al. entertained two possibilities. First, training might produce experience-dependent increases in the temporal resolution of neural elements registering changes in direction of motion. Recall that both Lee and Blake (1999a) and Farid and Adelson (2001) speculated that synchronized changes in motion direction of the sort used in these temporal structure displays might stimulate neurons selectively responsive to stimulus transients. Experience-dependent changes in the time constants of such neural elements could alter the fidelity with which temporal structure is registered. However, there is another component to the task, namely, the extraction of the spatial distribution of those stimulus elements embodying common temporal structure. Thus, training could also increase the efficiency with which synchronous activity is grouped across distributed neuronal populations representing different regions of visual space.

In some very recent work, Guttman, Gilroy, and Blake (in press) investigated the extent to which shape from stochastic temporal structure depends on the similarity among the "messengers" signaling that temporal structure. In their experiments, observers viewed arrays of Gabor patches in which figure and ground were designated by different temporal structures; the signals defining temporal structure could be changes in orientation, in spatial frequency, in phase, and/or in contrast. Results from several experiments showed that observers could perceive shape from temporal structure even when the defining events had to be combined across different messengers (i.e., changes within different dimensions). Moreover, mixing messengers of temporal structure proved to be cost-free: Grouping across messengers produced performance approximately the same as did grouping within a single messenger. These findings show that vision can abstract temporal structure regardless of the messenger of the dynamic event; a coherent spatial structure emerges from this abstracted temporal structure.

Before concluding our survey of work on temporal structure and spatial grouping, we want briefly to consider possible neurophysiological implications from the work summarized here. This we do in the following section.

IMPLICATIONS FOR NEUROPHYSIOLOGY

The psychophysical studies reviewed in the previous sections indicate that the human visual system possesses two important abilities: (a) the ability to resolve temporal asynchrony down to stimulus onsets differing by less than 5 msec and (b) the ability to utilize information about temporal structure among spatially distributed visual features for the extraction of spatial structure. This, in turn, naturally leads to questions about neural substrates underlying those psychophysical abilities. How do visual neurons encode fine temporal structure of dynamic visual stimuli? How does the brain compute the temporal relationship among neural populations registering visual features distributed over space?

Detailed consideration of these questions lies beyond the scope of this article, but we do want to offer a few tentative conclusions about possible neural mechanisms of temporal resolution and grouping from temporal structure. Our conclusions are tempered by the realization that there exists no consensus among neuroscientists concerning how stimulus information is coded within neural spike trains (Bullock, 1968). Within neurophysiology, two candidate neural coding schemes are typically contrasted, namely, rate coding and temporal correlation. According to the rate coding hypothesis, information is conveyed by the average firing rates of neurons (Barlow, 1972; Shadlen & Newsome, 1994); according to the temporal correlation hypothesis, it is the timing of individual action potentials that embodies stimulus information (Abeles, 1982; Mainen & Sejnowski, 1995; Softky & Koch, 1993). We shall consider the issue of temporal resolution and grouping within the context of these two possible codes.

Neural Coding of Temporal Visual Structure

The excellent temporal acuity/resolution evidenced by human vision (e.g., Westheimer & McKee, 1977) implies that neurons can modulate their responses in a manner time-locked to external visual events. How is this time-locked modulation of neural response achieved? How, in other words, do neurons carry information about the fine temporal structure portrayed by dynamic visual stimuli?

Temporal correlation is certainly a feasible possibility for registering temporal structure. Indeed, it has been shown that individual neurons can reliably reproduce essentially the same spike trains when the same time-varying stimulus is repeatedly presented (Bair & Koch, 1996; Berry, Warland, & Meister, 1997; Mainen & Sejnowski, 1995). It has also been shown that the fine temporal structure contained in external stimulation can be reconstructed solely on the basis of information given in the spike timings of single neurons (Bialek, Rieke, de Ruyter van Steveninck, & Warland, 1991; Buracas, Zador, DeWeese, & Albright, 1998).

Can the rate code hypothesis also account for the fine temporal resolution of human vision? The discharge rate of a single neuron is unlikely to modulate reliably enough to encode the fine temporal structure of time-varying stimuli, because information about the timings of individual spikes is not preserved in firing rate coding (rate, by definition, involves integration over time). If we assume, however, that a given stimulus feature is redundantly encoded by an ensemble of neurons with similar receptive field properties, the observed temporal fidelity of human vision can be explained by the average firing rate among such an ensemble of neurons: The average firing rate of the ensemble *can* fluctuate in a time-locked fashion to time-varying stimuli with temporal precision of 5-10 msec, which corresponds to an effective sampling rate suggested by the rate coding hypothesis (Shadlen & Newsome, 1994). For a more complete discussion of rate coding versus neural synchrony, see Shadlen and Movshon (1999).

Therefore, psychophysical performance itself does not favor one coding hypothesis over the other—both coding schemes can explain the temporal acuity of human vision.

Neural Representation of Spatial Structure From Temporal Structure

Several research groups have argued forcefully that synchrony in spiking activity among groups of visual neurons is responsible for grouping, or "binding," of the local features activating those neurons. Engel and Singer (2001) provided a detailed description of this hypothesis, and they summarized much of the neurophysiological evidence favoring it. Roelfsema, Lamme, and Spekreijse (2004) presented evidence against the neural synchrony hypothesis, including multiunit recordings from the primary visual cortex of monkeys performing a grouping task. It is fair to say that neural synchrony's role in feature binding remains controversial and unresolved. Can we conclude that the psychophysical evidence showing spatial grouping based on temporal structure provides support for the neural synchronization hypothesis? For the reasons explained in the following paragraphs, we believe such a conclusion—though it may ultimately prove correct—would be premature.

According to the temporal correlation hypothesis, local visual elements flickering in synchrony are grouped together into a coherent percept because the action potentials of neurons responsive to those elements are synchronized in a stimulus-locked fashion. These synchronized spike trains constitute the glue forming a neural assembly representing a coherent percept. This synchronization hypothesis is based on two critical assumptions linking temporal modulation of external stimuli to changes in neural responses over time: (a) temporal modulation of external stimuli evokes neural responses whose spike timing is entrained, that is, synchronized, with the temporal structure associated with stimulus modulation; (b) those synchronized neural responses are similar to the synchronous activity of cortical neurons mediating perceptual grouping. This linking hypothesis thus requires endorsement of these two assumptions. Does evidence warrant that endorsement? For the following reasons, we believe neither of those assumptions currently rests on sufficiently firm ground.

The first assumption is tantamount to endorsing the temporal correlation hypothesis. According to this hypothesis, which claims that variations of spike trains of neurons are locked to variations of external stimuli with high temporal precision, two external stimuli with identical temporal structure will produce action potentials in two neural populations that are highly correlated over time. According to the rate coding hypothesis, however, responses of two neural populations responsive to external stimuli modulating in synchrony can be correlated not in spike timing but in averaged firing rate. Although spike counts of a single neuron cannot reliably follow variation of an external stimulus, average firing rates within an ensemble of neurons representing the same stimulus can represent instantaneous changes of the stimulus in a stimulus-locked manner. Thus, synchronized external stimulation could result either in correlated spike timings or in correlated firing rates among neural populations. Consequently, the ability of common temporal structure to promote grouping of local visual features could be mediated by synchronized spike timings or synchronized firing rates. The psychophysical evidence is not definitive with respect to either neurophysiological hypothesis.

Even if Assumption 1 were to prove valid, there remains a serious problem with the second assumption. According to the neural synchronization hypothesis, the entrainment of neural discharges responsible for grouping of visual features is achieved by intrinsic circuitry within the brain, not necessarily by temporal modulation imposed by external stimulation (Singer & Gray, 1995). In the psychophysical experiments, of course, the synchronization of neural discharges (if such exists) arises from phase locking of individual neural responses to the temporal structure of external stimuli. It is premature to assume that externally induced synchronization of neural activity serves the same function as internally generated synchronization. Therefore, the conclusions derived from psychophysical demonstrations of grouping by temporal structure do not necessarily bear on the nature of synchronization embodied in the temporal correlation model. This is a point that we should have stressed more forcefully in our earlier work (Lee & Blake, 1999a), for it is the aspect of our work that has drawn the greatest criticism (Morgan & Castet, 2002).

CONCLUSION

So where do matters stand with respect to the capacity of temporal structure to promote spatial grouping? There is general agreement that synchronized visual events tend to be perceptually grouped, whether those events constitute flicker, changes in motion direction, or changes in contrast. In this respect, the work summarized in this article has expanded the domain of "common fate" beyond that envisioned by the Gestalt psychologists. Moreover, the efficacy of temporal structure depends on the spatial properties of the stimulus features that carry information about spatial structure. Common fate, in other words, interacts synergistically with other Gestalt grouping principles such as good continuation (e.g., Lee & Blake, 2001).

The major disagreements in the literature concern the mechanism(s) responsible for grouping from temporal structure and the temporal resolution underlying grouping from temporal structure. We have no doubt that future work can and will resolve these debates.

Regardless how these issues are resolved, we continue to believe that stochastic temporal structure of the sort introduced by us (Lee & Blake, 1999a) provides a robust means for probing perceptual grouping based on common fate; it may constitute "uncontrolled randomness" (to borrow the phrase coined by Morgan & Castet, 2002), but that is precisely its virtue. We say this because the optical input to vision is replete with complex, unpredictable temporal structure associated with the movement of objects within the environment. Our eyes and brains have evolved in a dynamic visual world, so it stands to reason that vision would evolve mechanisms to exploit this rich source of information.

In closing, we would like to reiterate a point made in an earlier essay on temporal structure and spatial grouping, a point having to do with the role of temporal structure in solving the problem of grouping, or "binding" as it is sometimes called, visual features into coherent object descriptions:

> We are led to speculate whether the rich temporal structure characteristic of normal vision may, in fact, imprint its signature from the outset of neural processing. If this were truly the case, then concern about the binding problem would fade, for there would be no need for a mechanism to reassemble the bits and pieces comprising visual objects. Perhaps temporal structure insures that neural representations of object "components" remain conjoined from the very outset of visual processing. Construed in this way, the brain's job is rather different from that facing the King's horses and men who tried to put Humpty Dumpty back together. Instead of piecing together the parts of a visual puzzle, the brain may resonate to spatio-temporal structure contained in the optical input to vision. (Blake & Lee, 2000, p. 647).

NOTES

1. Not all models of masking invoke temporal integration. For example, a recent model by Enns and Di Lollo (2000) posits that masking results from a mismatch between stimulus representations in different modules within the visual stream.

2. To see demonstrations of stochastic temporal structure, navigate to the first author's Web site and follow the research links to "temporal structure."

3. Motion detectors activated by luminance can respond to changes in luminance within neighboring spatial regions. Thus, flickering or luminance-modulated elements can activate motion detectors, thereby creating motion-defined boundaries (see Kandil & Fahle, 2003).

REFERENCES

Abeles, M. (1982). Role of the cortical neuron: Integrator or coincidence detector? *Israeli Journal of Medical Science, 18*, 83-92.

Adelson, E. H., & Farid, H. (1999). Filtering reveals form in temporally structured displays. *Science, 286*, 2231a.

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

Alais, D., & Blake, R. (1998). Interactions between global motion and local binocular rivalry. *Vision Research, 38*, 637-644.

Alais, D., & Blake, R. (1999). Grouping visual features during binocular rivalry. *Vision Research, 39*, 4341-4353.

Alais, D., Blake, R., & Lee, S.-H. (1998). Visual features that covary together over time group together over space. *Nature Neuroscience, 1*, 163-168.

Anstis, S. M. (1970). Phi movement as subtraction process. *Vision Research, 10*, 1411-1430.

Arnold, D. H., Clifford, C. W. G., & Wenderoth, P. (2001). Asynchronous processing in vision: color leads motion. *Current Biology, 11*, 596-600.

Aslin, C., Blake, R., & Chun, M. M. (2002). Perceptual learning of temporal structure. *Vision Research, 42*, 3019-3030.

Atkinson, J., Campbell, F. W., Fiorentini, A., & Maffei, L. (1973). The dependence of monocular rivalry on spatial frequency. *Perception, 2*, 127-133.

Bair, W., & Koch, C. (1996). Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Computation, 8*, 1185-1202.

Barlow, H. B. (1972). Single units and perception: A neuron doctrine for perceptual psychology. *Perception, 1*, 371-394.

Berry, M. J., Warland, D. K., & Meister, M. (1997). The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences USA, 94*, 5411-5416.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., & Warland, D. (1991). Reading a neural code. *Science, 252*, 1854-1857.

Blake, R., & Lee, S. H. (2000). Temporal structure in the input to vision can promote spatial grouping. In S.-W. Lee, H. H. Bülthoff, & T. Poggio (Eds.), *Biologically motivated computer vision* (pp. 635-653). Berlin: Springer-Verlag.

Blake, R., & Lee, S. H. (2002). Temporal precision of visual grouping from temporal structure. *Journal of Vision, 2*, 233a. Retrieved from http://journalofvision.org/2/7/233/

Breitmeyer, B. G. (1978). Disinhibition in metacontrast masking of vernier acuity targets: Sustained channels inhibit transient channels, *Vision Research, 18*, 1401-1405.

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. New York: Oxford University Press.

Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological Review, 83*, 1-36.

Brook, D., & Wynne, R. J. (1988). *Signal processing: Principles and applications*. London: Edward Arnold.

Bullock, T. H. (1968). Representation of information in neurons and sites for molecular participation. *Proceedings of National Academy of Sciences USA, 60*, 1058-1068.

Buracas, G. T., Zador, A. M., DeWeese, M. R., & Albright, T. D. (1998). Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron, 20*, 959-969.

Burr, D. (1980). Motion smear. *Nature, 284*, 164-165.

Corfield, R., Frosdick, J. P., & Campbell, F. W. (1978). Greyout elimination: The roles of spatial waveform, frequency and phase. *Vision Research, 18*, 1305-1311.

de Coulon, F. (1986). *Signal theory and processing*. Dedham, MA: Artech House.

Dixon, P., & Di Lollo, V. (1994). Beyond visual persistence: An alternative account of temporal integration and segmentation in visual processing. *Cognitive Psychology, 26*, 33-63.

Efron, R. (1957). Stereoscopic vision I: Effect of binocular summation. *British Journal of Ophthalmology, 41*, 709-730.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences, 5*, 16-25.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4*, 345-352.

Eriksen, C. W., & Collins, J. F. (1967). Some temporal characteristics of visual pattern perception. *Journal of Experimental Psychology, 74*, 476-484.

Exner, S. (1875). Experimentelle Untersuchungen der einfachsten psychischen Processe. *Pflugers Arch Ges Physiol, 11*, 403-432.

Fahle, M. (1993). Figure-ground discrimination from temporal information. *Proceedings of the Royal Society of London, B, 254*, 199-203.

Fahle, M., & Koch, C. (1995). Spatial displacement, but not temporal synchrony, destroys figural binding. *Vision Research, 35*, 491-494.

Farid, H., & Adelson, E. H. (2001). Synchrony does not promote grouping in temporally structured displays. *Nature Neuroscience, 4*, 875-876.

Forte, J., Hogben, J. H., & Ross, J. (1999). Spatial limitations of temporal segmentation. *Vision Research, 39*, 4052-4061.

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

Georgeson, M. A., & Georgeson, J. M. (1985). On seeing temporal gaps between gratings: A criterion problem for measurement of visible persistence. *Vision Research, 25*, 1729-1733.

Guttman, S., Gilroy, L. A., & Blake, R. (in press). Mixed messengers, unified message: Spatial grouping from temporal structure. *Vision Research*.

Hirsh, I. J., & Sherrick, C. E., Jr. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology, 62*, 423-432.

Hoffman, D. D. (1998). *Visual intelligence: How we create what we see*. New York: W. W. Norton.

Hogben, J. H., & Di Lollo, V. (1974). Perceptual integration and perceptual segregation of brief visual stimuli. *Vision Research, 14*, 1059-1069.

Hogben, J. H., & Di Lollo, V. (1985). Suppression of visible persistence in apparent motion. *Perception & Psychophysics, 38*, 450-460.

Julesz, B. (1971). *Fundamentals of cyclopean perception*. Chicago: University of Chicago Press.

Julesz, B., & White, B. W. (1969). Short term visual memory and the Pulfrich Phenomenon. *Nature, 22*, 639-641.

Kahneman, D. (1968). Method, findings, and theory in studies of visual masking. *Psychological Bulletin, 70*, 404-425.

Kandil, F. I., & Fahle, M. (2001). Purely temporal figure-ground segmentation. *European Journal of Neuroscience, 13*, 2004-2008.

Kandil, F. I., & Fahle, M. (2003). Mechanisms of time-based figure-ground segmentation. *European Journal of Neuroscience, 18*, 2874-2882.

Kandil, F. I., & Fahle, M. (2004). Figure-ground segregation can rely on differences in motion direction. *Vision Research, 44*, 3177-3182.

Kiper, D. C., Gegenfurtner, K. R., & Movshon, J. A. (1996). Cortical oscillatory responses do not affect visual segmentation. *Vision Research, 36*, 539-544.

Kojima, H. (1998). Figure/ground segregation from temporal delay is best at high spatial frequencies. *Vision Research, 38*, 3729-3734.

Kovacs, I., Papathomas, T. V., Yang, M., & Feher, A. (1997). When the brain changes its mind: Interocular grouping during binocular rivalry. *Proceedings of the National Academy of Sciences USA, 93*, 15508-15511.

Lappin, J. S., & Bell, H. H. (1972). Perceptual differentiation of sequential visual patterns. *Perception & Psychophysics, 12*(2A), 129-134.

Lee, S.-H. (2002). *Temporally correlated visual input and perceptual grouping*. Unpublished dissertation, Vanderbilt University, Nashville, TN.

Lee, S.-H., & Blake, R. (1999a). Visual form created solely from temporal structure. *Science, 284*, 1165-1168.

Lee, S.-H., & Blake, R. (1999b). Reply to Adelson and Farid. *Science, 286*, 2231.

Lee, S.-H., & Blake, R. (2001). Neural synergy in visual grouping: When good continuation meets common fate. *Vision Research, 41*, 2057-2064.

Leonards, U., Singer, W., & Fahle, M. (1996). The influence of temporal phase differences on texture segmentation. *Vision Research, 36*, 2689-2697.

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

Mainen, Z. F., & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science, 268*, 1503-1506.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

Meyer, G. E., & Maguire, W. M. (1977). Spatial frequency and the mediation of short-term visual storage. *Science, 198*, 524-525.

Morgan M., & Castet, E. (2002). High temporal frequency synchrony is insufficient for perceptual grouping. *Proceedings of the Royal Society of London, B, 269*, 513-516.

Motoyoshi, I. (2004). The role of spatial interactions in perceptual synchrony. *Journal of Vision, 4*, 352-361. Retrieved from http://journalofvision.org/4/5/1/

Moutoussis, K., & Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London, B, 264*, 393-399.

Nishida, S., & Johnston, A. (2002). Marker correspondence, not processing latency, determines temporal binding of visual attributes. *Current Biology, 12*, 359-368.

Ogle, K. N. (1963). Stereoscopic depth perception and exposure delay between images to the two eyes. *Journal of the Optical Society of America, 53*, 1296-1304.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature Neuroscience, 7*, 982-991.

Rogers-Ramachandran, D. C., & Ramachandran, V. S. (1998). Psychophysical evidence for boundary and surface systems in human vision. *Vision Research, 38*, 71-77.

Ross, J., & Hogben, J. H. (1974). Short-term memory in stereopsis. *Vision Research, 14*, 1195-1201.

Sekuler, A. B., & Bennett, P. J. (2001). Generalized common fate: Grouping by common luminance changes. *Psychological Science, 12*, 437-444.

Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron, 24*, 67-77.

Shadlen, M. N., & Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology, 4*, 569-579.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience, 18*, 555-586.

Smith, G., Howell, E. R., & Stanley, G. (1982). Spatial frequency and the detection of temporal discontinuity in superimposed and adjacent gratings. *Perception & Psychophysics, 31*, 293-297.

Softky, W., & Koch, C. (1993). The irregular firing of cortical cells is inconsistent with temporal integration of random EPSP's. *Journal of Neuroscience, 13*, 334-350.

Suzuki, S., & Grabowecky, M. (2002). Overlapping features can be parsed on the basis of rapid temporal cues that produce stable emergent percepts. *Vision Research, 42*, 2669-2692.

Sweet, A. L. (1953). Temporal discrimination by the human eye. *American Journal of Psychology, 66*, 185-198.

Treisman, A. (1999). Solutions to the binding problem: Progress through controversy and convergence. *Neuron, 24*, 105-110.

Usher, M., & Donnelly, N. (1998). Visual synchrony affects binding and segmentation in perception. *Nature, 394*, 179-182.

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences, 7*, 207-213.

Watson, A. B. (1986). Temporal sensitivity. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1: Sensory processes and perception, pp. 6.1-6.43). New York: John Wiley.

Wertheimer, M. (1912). Experimentelle Studien Ïber das Sehen von Bewegungen'. *Zeitschrift fur Psychologie und Physiologie der Sinnesorgane, 61*, 161-265.

Westheimer, G., & McKee, S. P. (1977). Perception of temporal order in adjacent visual stimuli. *Vision Research, 17*, 887-892.

Whittle, P., Bloor, D., & Pocock, S. (1968). Some experiments on figural effects in binocular rivalry. *Perception & Psychophysics, 4*, 183-188.

Yund, E. W., & Efron, R. (1974). Dichoptic and dichotic micropattern discrimination. *Perception & Psychophysics, 15*, 383-390.

Zacks, J. L. (1970). Temporal summation phenomena at threshold: Their relation to visual mechanisms. *Science, 170*, 197-199.