ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 45 (2005) 1021–1030
Vision Research
www.elsevier.com/locate/visres

Mixed messengers, unified message: spatial grouping from temporal structure

Sharon E. Guttman *, Lee A. Gilroy, Randolph Blake

Department of Psychology, Vanderbilt University, 301 Wilson Hall, 111 21st Ave. South, Nashville, TN 37203, United States

Received 15 July 2004; received in revised form 12 October 2004

Abstract

In dynamic visual environments, objects can differ from their backgrounds in terms of their associated temporal structure—the time course of changes in some stimulus property defining object and background. In a series of experiments, we investigated whether different "messengers" of temporal structure group into coherent spatial forms. Observers viewed arrays of Gabor patches in which different temporal structures designated figure and ground regions; extracting the figure required grouping across synchronized orientation, spatial frequency, phase, and/or contrast changes. Observers were able to extract spatial form from temporal structure even when information had to be combined across different messengers. Further, mixing messengers of temporal structure proved cost-free: task performance when grouping across messengers approximated performance when all information resided within a single messenger. Thus, the visual system can abstract temporal structure regardless of the messenger of the dynamic event; a coherent spatial structure emerges from this abstracted temporal structure.

Keywords: Visual grouping; Temporal structure; Spatial structure; Synchrony; Binding problem

1. Introduction

Object perception depends critically on processes of segmentation and grouping. To perceive a meaningful environment, the visual system often must represent spatially distinct regions as belonging to unified objects while representing some spatially contiguous regions as belonging to different objects. This integration across space is one of several instantiations of "the binding problem" that must be solved for coherent visual perception. In recent years, empirical and theoretical work has generated much debate about the mechanisms by which the visual system signals binding (e.g., Crick & Koch, 1990; Farid, 2002; Kiper, Gegenfurtner, & Movshon, 1996; Rogers-Ramachandran & Ramachandran, 1998; Shadlen & Movshon, 1999; Singer, 1999; Singer & Gray, 1995). These debates reveal a superceding question: What visual features trigger the operation of these binding mechanisms? That is, what algorithms does the visual system use to determine which elements belong to a single object and which belong to different objects?

Last century, Gestalt psychologists proposed a handful of principles for visual perceptual organization (Koffka, 1935; Wertheimer, 1923/1938). For most of the organizational principles—including proximity, similarity, and good continuation—grouping depends on the spatial configuration of the image. Nonetheless, temporal structure also affects perceptual organization, as reflected by the principle of *common fate:* objects that move together over time group together over space.

Recent studies have expanded our conceptualization of common fate, as well as the role of temporal factors in grouping more generally (e.g., Alais, Blake, & Lee, 1998; Fahle, 1993; Kandil & Fahle, 2001; Lee & Blake,

* Corresponding author. Tel.: +1 615 322 8768; fax: +1 615 343 8449.

E-mail address: sharon.guttman@vanderbilt.edu (S.E. Guttman).

0042-6989/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2004.10.014

1999b, 2001; Leonards, Singer, & Fahle, 1996; Sekuler & Bennett, 2001; Usher & Donnelly, 1998). For example, Alais et al. (1998) found that temporally correlated contrast changes enhanced the perceptual coherence of moving gratings, whereas temporally uncorrelated contrast changes reduced perceptual coherence. Similarly, segmentation and grouping can occur on the basis of variations in luminance over time; elements that brighten and darken according to the same time course group together, and segment from elements that brighten and darken according to a different time course (Sekuler & Bennett, 2001). Using an array of moving Gabor patches, Lee and Blake (1999b) demonstrated the spatial grouping of moving elements undergoing stochastic but synchronous direction reversals and their segmentation from background elements that underwent similar but unsynchronized direction reversals. This phenomenon differs from common fate as classically defined because the individual elements in the array did not "move together" in the typical sense (i.e., at the same general speed and in the same general direction); rather, the contours of the grouped elements were of random orientation and moved in different directions, sharing only the timing of motion reversals. Together, these experiments suggest that temporal correlation, broadly defined, may be a fundamental principle of visual perceptual organization. The specifics of the mechanisms underlying grouping by temporal structure remain arguable (e.g., Adelson & Farid, 1999; Lee & Blake, 1999a; Shadlen & Movshon, 1999; Singer, 1999; Singer & Gray, 1995).

If correlated change (i.e., correlated temporal structure) plays a central role in solving the binding problem, then the visual system should show a general sensitivity to the temporal structure of a stimulus. Under these circumstances, we predict that the visual system will group together *any* stimulus elements undergoing synchronous, salient changes, regardless of the precise nature of those changes.

In the current paper, we present three experiments investigating the extent to which the visual system abstracts temporal structure from the nature of the dynamic change. Specifically, we asked: Do temporally correlated changes lead to grouping even when the nature of those changes varies across the spatial array? In other words, do different "messengers" of a common temporal structure become perceptually organized into a coherent spatial form?

To address this question, we used a paradigm in which observers had to segment and group visual stimuli on the basis of temporal structure alone (e.g., Lee & Blake, 1999b); that is, stochastically changing figure and ground elements could be distinguished solely on the basis of the timing of those changes. In the current experiments, different "messengers" (i.e., different types of changes) carried the temporal structure of different elements within the array: some elements underwent orientation changes, whereas other elements underwent spatial frequency changes, phase changes, or contrast changes. Extracting the relevant spatial structure required grouping across different messengers of temporal structure. Therefore, task performance should be proficient only to the extent that the visual system abstracts the notion of "change," and utilizes the temporal structure of this abstracted messenger in solving the binding problem.

2. General methods

2.1. Stimuli

Fig. 1A schematizes a single frame of a stimulus used in this experiment. The stimuli consisted of arrays of Gabor patches on a mid-gray background (16.5 cd/m^2). Each Gabor patch had a randomly assigned orientation, phase, spatial frequency, and contrast.

Within the array, a rectangular subset of elements was designated as the figure region and all remaining elements comprised the ground. The orientation of the figure (horizontal or vertical) varied randomly across trials. The figure and ground regions differed only in the temporal structure by which their component elements changed. Gabor patches within the figure region changed at times designated by one stochastic point process, whereas Gabor patches within the ground region changed at times designated by a second point process (Fig. 1B). Both point processes operated at a rate of 30 Hz (i.e., changes occurred at 33.3 ms intervals or some multiple thereof). The correlation between the two point processes, which specified the level of temporal structure available for segregating figure and ground, varied across trials. Constraints ensured that changes occurred on precisely 50% of the frames and that no more than four successive frames contained either all changes or no changes (a constraint that minimizes contrast artifacts owing to temporal summation over frames).

In these experiments, a "change" involved assigning each Gabor patch in the designated region a new orientation, phase, spatial frequency, or contrast. To ensure detectability of the changes, the newly assigned parameters differed from the previous values by at least a designated minimum, but were otherwise random. For orientation, the minimum change was 15°; phase changes had to exceed $\pi/4$; spatial frequency and contrast were required to increase or decrease by at least 33% of the previous value.

2.2. Procedure

Each trial consisted of a fixation point until key press, followed by a dynamically changing array of Gabor patches. The displays appeared for 2033 ms (initial frame

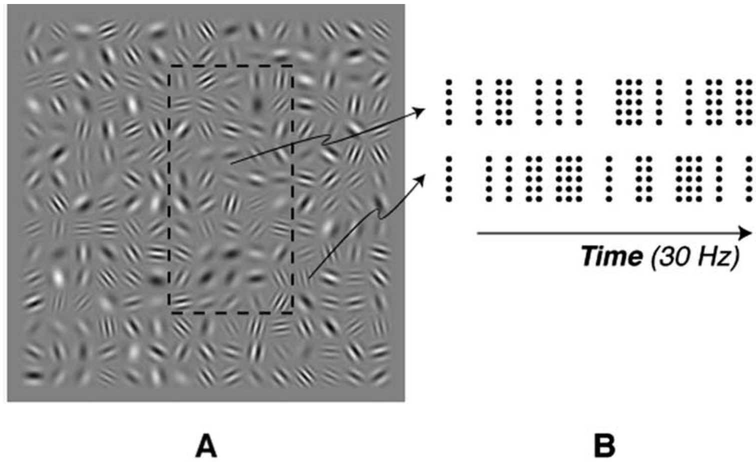

Fig. 1. (A) Schematic illustration of a single stimulus frame. The dotted rectangle (not present in an actual stimulus display) denotes the figure region; all elements outside the dotted rectangle comprised the ground. Note that each Gabor patch had randomly assigned orientation, phase, spatial frequency, and contrast, such that the figure and ground region could not be segregated on the basis of static cues. (B) Dots indicate times at which the Gabor patches changed. Elements within the figure region changed at times designated by one point process, whereas Gabor patches within the ground region changed at times designated by a second point process. The two point processes illustrated here have a temporal correlation of 0.

plus 60 subsequent frames); both the figure and ground elements underwent 30 changes during each trial. Observers judged whether the rectangular figure region, as defined by the temporal structure of the changes, was oriented horizontally or vertically. Judgements were indicated by pressing one of two keys. An auditory "ping" provided feedback for incorrect responses.

2.3. Apparatus

The stimuli and experimental trials were generated on an Apple Macintosh G4 computer and presented on a Mitsubishi Diamond Pro 2020u monitor (20 in.; 1024×768 pixels; 120 Hz). The monitor provided the only source of illumination in an otherwise darkened testing room. Observers sat 80 cm from the screen with their heads stabilized in a chinrest.

2.4. Observers

Four observers with normal or corrected-to-normal vision participated in all experiments reported herein. Two observers (SEG and LAG) were authors on this paper; the other two observers (DAB and CYK) had considerable experience as psychophysical observers but were the naïve to the experimental hypotheses.

3. Experiment 1: Temporal structure defined by random changes

3.1. Method

In the first experiment, we investigated whether elements undergoing different sequences of changes but with the same temporal structure would be grouped together by the visual system. Thus, the type of change that each element underwent (i.e., orientation change, phase change, spatial frequency change, or contrast change) varied randomly both across elements and over the course of a trial. That is, one Gabor patch might change successively by orientation, contrast, phase, and then orientation again, whereas an immediately adjacent element might undergo a different sequence of changes but at precisely the same times. [1]

The arrays for this experiment contained 25×25 Gabor patches (SD = 0.125°; visible area $\approx$ 0.4°); the center-to-center distance between elements measured 0.5°, for a total stimulus size of $12.5° \times 12.5°$. The figure consisted of a horizontal or vertical 9×15 rectangular subset of elements that varied randomly in location. Spatial frequency was constrained between 2 and 8 cpd; contrast ranged from 0.1 to 1.0. The correlation between the point processes for the figure and ground regions (0, 0.2, 0.4, 0.6, or 0.8) was the key independent variable. All observers participated in two experimental sessions, each consisting of 10 randomly ordered trials with each level of correlation.

3.2. Results and discussion

Fig. 2 plots observers' ability to correctly judge the orientation of the figure region as a function of figure–ground correlation. Small correlation values signify conspicuous differences in temporal structure between figure and ground regions; as correlation increases, differences

[1] Illustrative stimulus sequences for each experiment are available online as supplementary files.

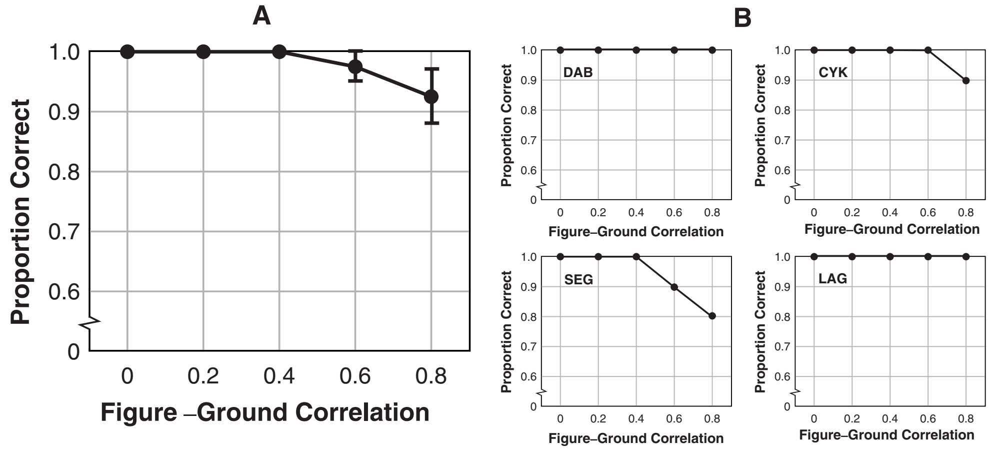

Fig. 2. Results for Experiment 1: group average (A) and individual data (B). The proportion of responses on which observers correctly reported the orientation of the figure appears as a function of figure–ground correlation, where high correlation values denote weak differences in temporal structure between figure and ground. Error bars represent ±1 SE across observers.

in temporal structure diminish and figure–ground segmentation becomes more difficult. Clearly, performance approached ceiling levels even when the temporal structure available for segmentation and grouping was minimized (i.e., high correlation between the figure and ground point processes). As the temporal structures of the figure and ground regions would be identical at a correlation of 1.0, performance would no doubt have declined if the range of tested correlations were extended further. Regardless, this experiment establishes that observers can effectively segment a visual array based on temporal information carried by multiple messengers.

These results are consistent with the idea that observers can group across different messengers of temporal structure to perceive a coherent spatial form. Alternatively, however, the spatial form may have emerged through dynamically changing groupings within individual messengers. That is, all elements momentarily undergoing orientation changes, for example, may be grouped together; the subset of these elements subsequently changing in phase may then group with other elements undergoing a phase change at the same time. Though brief and varying over time, these groupings could theoretically provide sufficient spatial information to distinguish horizontal from vertical figures. Therefore, we investigated the idea of grouping across different messengers of change more systematically in two further experiments.

4. Experiment 2: Mixing messengers systematically

In Experiment 2, we sought conclusive evidence that the visual system abstracts temporal structure from the messenger of the dynamic change. To this end, we utilized displays in which the task-relevant spatial structure could emerge *only* if the visual system grouped across different messengers of temporal structure. If grouping occurs within individual messengers but not across messengers, then task performance will be at chance level.

4.1. Method

Fig. 3 depicts a single frame of a stimulus for this experiment. The stimuli consisted of 6×6 arrays of Gabor patches (SD = 0.2°; center-to-center distance = 1.0°; total stimulus size = 6.0° × 6.0°). The central 2×4 rectangle, oriented either horizontally or vertically,

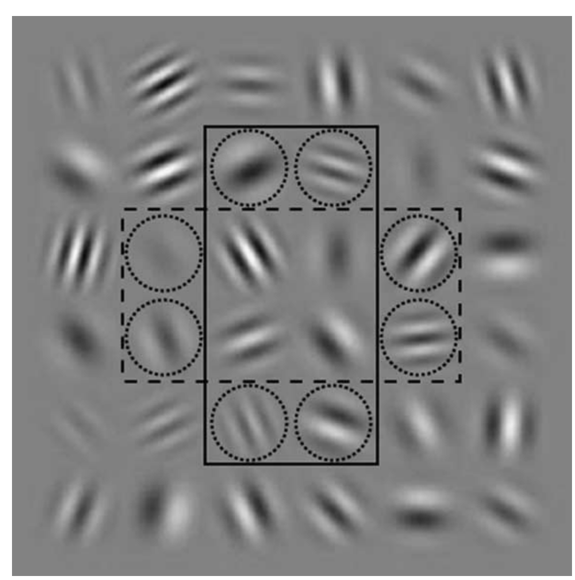

Fig. 3. Single frame of a stimulus for Experiment 2. A central 2×4 rectangle either vertically (solid lines) or horizontally (dashed lines) constituted the figure. Regardless of the figure orientation, the eight elements flanking the central square, indicated by the dotted circles, had their temporal structure carried by one messenger; all remaining elements had their temporal structure carried by a second messenger.

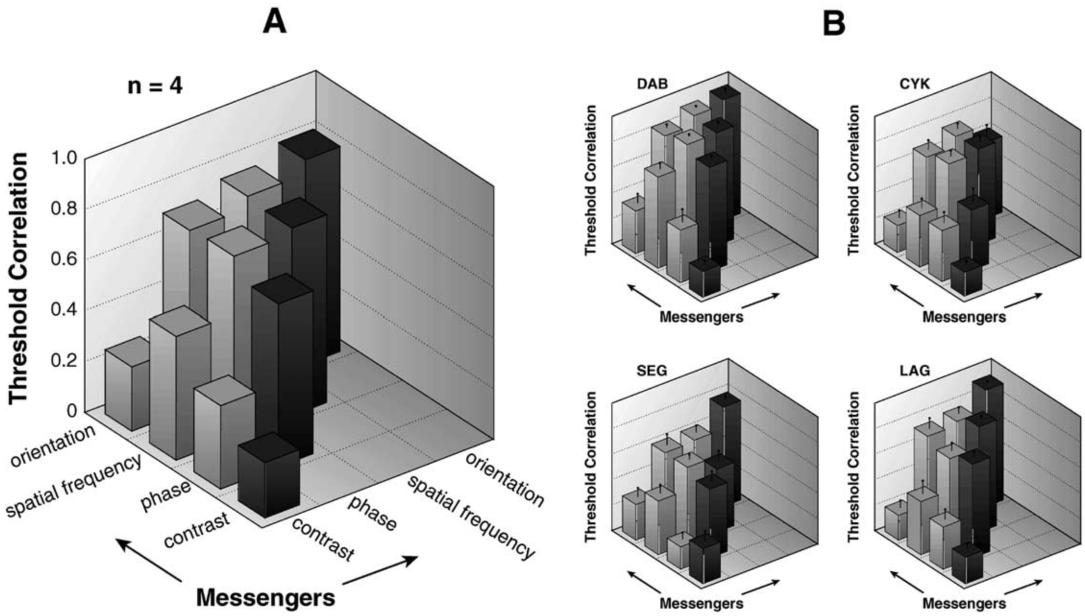

Fig. 4. Results for Experiment 2: group average (A) and individual data (B). For all graphs, threshold correlation (i.e., figure–ground correlation leading to .71 correct responses) is plotted as a function of the combination of messengers in the display. Dark bars indicate conditions in which a single messenger carried all temporal structure; light bars indicate conditions in which task performance required grouping across two different messengers of temporal structure. Error bars on the individual plots represent 1 SE across the eight staircases measured for each messenger combination.

constituted the figure region. Initially, the elements varied randomly in orientation, phase, spatial frequency (1.0–4.0 cpd), and contrast (0.1–1.0).

As before, two different point processes determined the times at which figure elements and ground elements underwent change. However, the messenger of change varied systematically, rather than randomly, across the stimulus array and remained constant for each element during the course of a trial. The eight elements flanking the central square had their temporal structure defined by one messenger (i.e., underwent one type of change), whereas the remaining 28 elements had their temporal structure defined by a second messenger (i.e., a second type of change; see Fig. 3). Note that this arrangement of messengers is independent of the conveyed temporal structure that defines the figure and ground regions. In this manner, determining whether the figure was oriented horizontally or vertically required grouping across messengers; the relevant spatial structure would emerge only if two pairs of flankers became grouped with the central square on the basis of common temporal structure carried by different messengers, while the other two pairs of flankers became grouped with the surrounding elements.

In separate experimental sessions, we tested stimuli containing all possible pairwise combinations of messengers as well as stimuli in which all elements had their temporal structure carried by a single messenger. Figure–ground correlation varied according to two interleaved 2-up/1-down staircases. Correlation started at 0.4, changed at 0.2 increments for four reversals, and then changed at 0.067 increments for 10 reversals. For one of the staircases, one of the two messengers in the combination being tested defined the temporal structure of the flankers and the second messenger defined the temporal structure of the central square and surrounding elements; for the second staircase, the opposite configuration was used. We averaged the correlation values leading to the final four reversals in each staircase to yield a measure of threshold correlation for the combination of carriers in question. [2] All observers participated in four experimental sessions for each individual carrier, followed by four sessions for each combination of two different carriers. Session ordering was counterbalanced across subjects.

4.2. Results and discussion

Fig. 4 depicts threshold correlation—the correlation between the figure and ground point processes leading to .71 correct horizontal/vertical judgment—as a function of the combination of carriers in the display. A higher threshold correlation corresponds to better task performance; as correlation increases, less temporal difference exists to segregate figure and ground. However,

[2] No systematic differences arose between the two different configurations for each messenger combination. Therefore, all presented data average across configuration.

one should note that even a threshold correlation of zero represents above-chance task performance for the easiest judgments.

The dark bars indicate conditions in which all Gabor patches had their temporal structure carried by the same messenger. These data indicate that observers can effectively discriminate figure from ground whenever these regions are defined by differentially synchronized changes within a single stimulus attribute. For all observers, performance levels with each individual messenger significantly exceeded chance level (i.e., the staircases converged readily). However, orientation changes, spatial frequency changes, phase changes, and contrast changes supported spatial grouping to different relative extents. The relative effectiveness of the four messengers, which varied somewhat across observers, may be attributable to differences in the salience of neural signals accompanying change (i.e., signals that we think of as transient responses in neural elements possessing high temporal fidelity).

An examination of the data for the various mixed-messenger conditions (light bars in Fig. 4) reveals an interesting pattern. First, all messenger combinations led to effective extraction of spatial form, even though no relevant spatial structure existed unless observers combined temporal structure information across different messengers. Second, the level of performance for each combination of messengers approximated the level of performance achieved when all information resided within the weaker of the two individual messengers. In some cases (e.g., the spatial frequency and contrast combination), we even observed a performance advantage in the mixed-messenger conditions, relative to performance with the component messengers. This advantage may be attributable to a boosting of a weakly signaled change (i.e., change in contrast) by a more strongly signaled change.

In sum, this experiment clearly demonstrates that observers extract spatial structure from temporal structure, even when temporal information must be combined across different messengers of change. Further, mixing messengers of temporal structure is cost-free: to the extent that the visual system can extract the temporal structure from changes within a given messenger, this information can be combined with the temporal structure from changes within a different messenger.

5. Experiment 3: Distinguishing different messengers of change

In a final experiment, we investigated whether the combination of information across different messengers precedes or follows the extraction of temporal structure. Based on results discussed thus far, the temporal structure used for spatial segregation could be extracted independently from the different messengers, then subsequently compared by a higher-level mechanism. Alternatively, the notion of "change" may be abstracted from the nature of the messenger, such that temporal structure is extracted directly from a universal messenger.

5.1. Method

To distinguish these possibilities, temporal structure defined a 9×15 rectangular figure within a 25×25 array of Gabor patches, as in Experiment 1. However, the stochastic point processes for figure and ground determined not *whether* a change would occur on any given frame, but the *type* of change that would occur. On every frame, each element in the array underwent either a phase change or a spatial frequency change. All elements within a given region changed by the same messenger on any given frame; across the two regions, temporal structure had a correlation of 0, meaning that different types of change occurred in the figure versus the ground on 50% of frames. Thus, the temporal structure of relevance for segregation and grouping depended on the different patterns of change in the figure and ground regions over time.

In previous experiments, "changes" involved randomly sampling new values for the messenger in question. In Experiment 3, the step sizes of the phase and spatial frequency changes were held constant within a trial, with only the direction of the change randomly determined. Following random determination of the initial values for each Gabor patch (within the limits previously reported), spatial frequency changed by either 33.3% or 66.7% on each frame; phase changed in steps of 0 (i.e., no change), $\pi/6$, $\pi/3$, $\pi/2$, $2\pi/3$, or $5\pi/6$. All observers completed eight experimental sessions, each consisting of five randomly ordered trials with each crossed combination of spatial frequency and phase step size. Additionally, each observer participated in four sessions in which spatial frequency was held constant, such that the figure was defined only by the temporal structure of phase changes ($\pi/6$, $\pi/3$, $\pi/2$, $2\pi/3$, or $5\pi/6$; 10 trials of each level per session).

5.2. Results and discussion

If the extraction of temporal structure precedes the combination of information across messengers, then two strong sources of temporal structure define the figure region in the stimuli for this experiment. That is, both the timing of spatial frequency changes and the timing of phase changes would cause the same figure region to segregate from the ground region. With these two strong sources of information being extracted independently then subsequently combined, task performance should be excellent regardless of the sizes of the changes (Fig. 5A). By contrast, if the visual system extracts temporal structure from an abstracted messenger, then a different pattern should emerge. When one messenger indicates a large change relative to the other messenger, grouping and segregation may be possible. However, when the sizes of the two changes match perceptually, the displays carry no differentiated temporal structure for spatial segregation: both the figure and ground elements simply ''change'' on every frame. Thus, for each spatial frequency step size, there should be a perceptually matched phase step size at which task performance is at chance level (Fig. 5B).

Fig. 6 plots the proportion of correct responses as a function of the size of the phase change in the display. When temporal structure emerged only through spatial frequency changes (i.e., phase step size was 0) or only through phase changes (i.e., spatial frequency step size was 0), task performance approached ceiling levels; this finding indicates that all amounts of change were sufficiently detectable to support spatial grouping. However, the spatial frequency changes and the phase changes clearly did not give additive sources of temporal structure information for spatial grouping. Rather, spatial frequency changes and phase changes opposed one another, as predicted by the ''abstracted messenger'' model: for both levels of spatial frequency change, there existed a level of phase change for which near-chance performance was observed. These levels presumably represent the points at which the salience of the two change types matched perceptually.

Therefore, it appears that the visual system abstracts ''change'' signals from individual messengers before

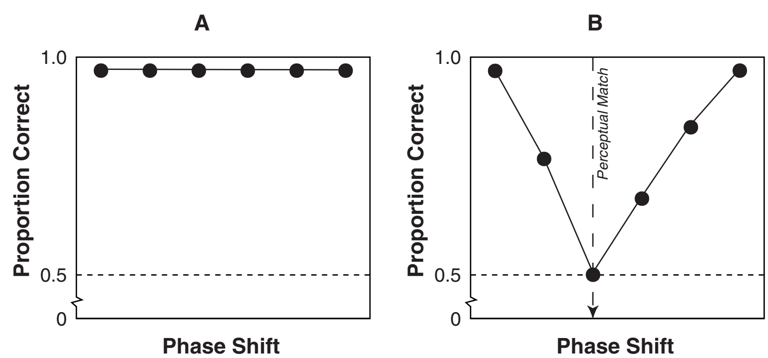

Fig. 5. Predictions for Experiment 3: (A) extraction of temporal structure *precedes* the combination of information across different messengers; (B) extraction of temporal structure *follows* the combination of information across different messengers (i.e., occurs through an abstracted messenger). See text for details.

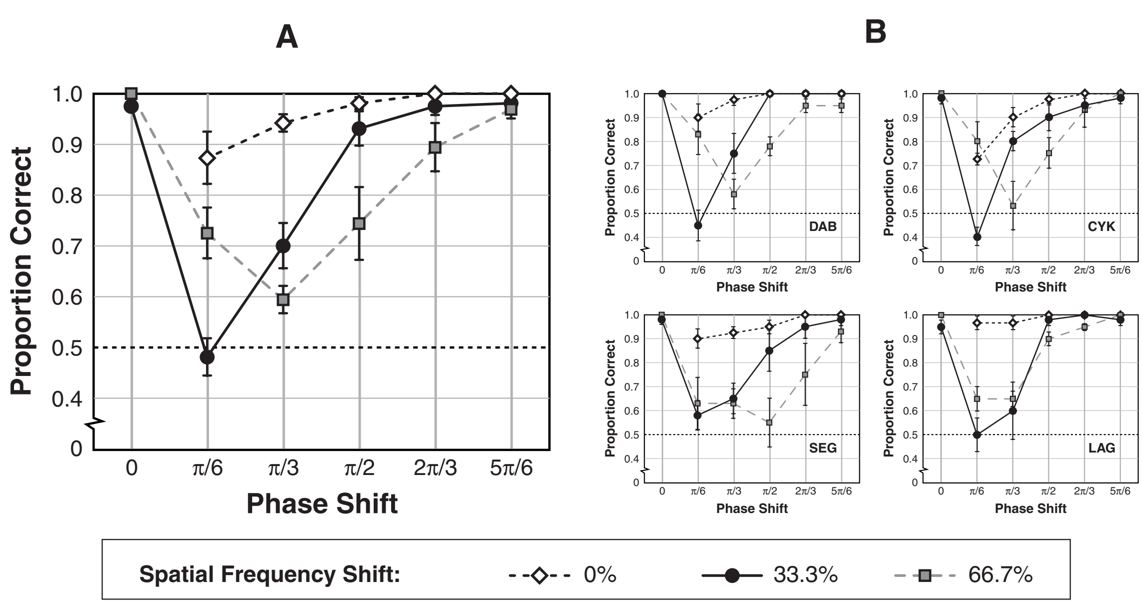

Fig. 6. Results for Experiment 3: group average (A) and individual data (B). For all graphs, proportion correct appears as a function of the size of the phase shifts in the display. Error bars on the individual plots represent ±1 SE across sessions.

recovering the temporal structure of the stimulus. That is, the temporal structure information used for spatial grouping and segregation depends on a universal messenger that simply indicates when change has occurred, disregarding the type of change.

6. General discussion

In the natural environment, temporally coincident events usually can be attributed to a single underlying cause. The visual system appears to be able to capitalize on this ecological knowledge by grouping on the basis of common fate. The results of the current experiments fit with recent tradition in expanding the Gestalt notion of common fate, and establish temporal synchrony, broadly defined, as a central algorithm for visual binding. Any elements undergoing detectable visual changes according to the same temporal pattern appear to organize perceptually into a coherent spatial form. That is, the visual system abstracts the notion of change from the nature of the change, ultimately using the temporal structure of this abstracted messenger as a cue for grouping.

In response to earlier work on grouping by temporal structure, Adelson and Farid (1999; see also Farid & Adelson, 2001) suggested that low-pass temporal filtering of stimulus sequences akin to those used here may reveal static contrast cues that, in principle, could be used to perform the spatial discrimination task (cf., Lee & Blake, 1999a). By this argument, the binding of Gabor patches into a coherent form occurs not because of temporal synchrony, but because there are moments in the stimulus sequences during which the figure elements integrate to a noticeably higher or lower contrast than the ground elements. Note that Adelson and Farid are not discounting the potency of correlated temporal structure within a dynamic visual scene, but they are proposing that this temporal structure generates luminance contrast boundaries owing to temporal integration. Their arguments are based on the outputs of hypothetical neural filters operating on selected stimulus sequences.

To evaluate whether our mixed-messenger stimuli contain embedded contrast artifacts, we performed a comprehensive analysis of the stimulus sequences used in the current study (see Appendix A). Fig. 7 depicts the results of our analyses for Experiment 2. In these graphs, positive values indicate contrast cues supporting the "correct" orientation, whereas negative values (denoted by dark bars in Fig. 7B) indicate contrast cues supporting the alternative, "incorrect" orientation. In individual trials (Fig. 7A), the potential contrast cue fluctuated rapidly over time between positive values, which may support segregation of the figure, and negative values, which oppose its segregation. Further, when averaged over time and across trials (Fig. 7B), there emerged no consistent contrast cue to support figure–

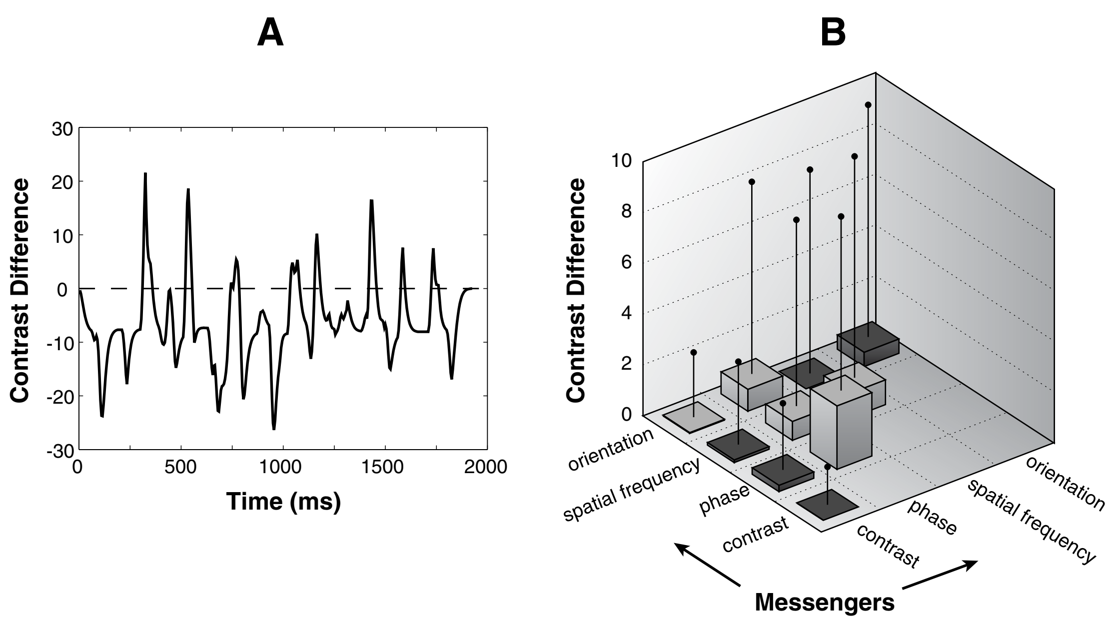

Fig. 7. Potential contrast artifacts for the stimulus sequences used in Experiment 2. (A) Contrast cue over time for a typical trial (phase/spatial frequency combination). Positive values indicate contrast cues supporting the "correct" orientation; negative values indicate contrast cues suggesting the orthogonal orientation. Note that the suggested orientation of the figure, based on the contrast cue, fluctuated rapidly between the two alternatives. (B) Average contrast cue for each individual messenger and combination of messengers at threshold correlation values. Light bars denote positive values; dark bars denote negative values. Error bars indicate 1 SD across stimuli. The substantial variation, readily encompassing zero, indicates that there existed no consistent contrast cue to support performance.

ground segregation for any combination of messengers. Thus, we can attribute performance in the grouping task to the only available information: the differing temporal structures of the figure and ground regions.

If temporal structure provides a strong cue for perceptual organization, even when extracted across different messengers, then one may question why some experiments investigating the influence of synchrony on grouping produced negative results. In some cases, effective grouping based on spatial factors may have precluded any further influence of temporal factors (e.g., Fahle & Koch, 1995; Kiper et al., 1996). However, this explanation cannot account for a lack of perceptual organization in displays containing only temporal cues for grouping. Farid and Adelson (2001), for example, demonstrated that the visual system does not group together small drifting dots that simultaneously change directions.

The results of the current experiments show that not all messengers of temporal structure provide equal support for grouping mechanisms. Contrast changes, for example, result in less effective segmentation and grouping than do orientation, phase, or spatial frequency changes (Experiment 2). Therefore, it appears that different messengers contribute to a generalized temporal structure only to the extent that they individually provide salient, temporally localized information that a change has occurred. The direction-changing dots in Farid and Adelson's (2001) displays may simply fail to provide an effective "change" cue for temporal grouping mechanisms; the inability of these stimuli to trigger a response in band-pass filters supports this conjecture. Simultaneous responses in band-pass filters—which signal the occurrence of change—may be a neural cue that triggers the operation of binding mechanisms; further research is needed both to test this hypothesis and to determine the nature of the mechanism by which the visual system binds the responses of spatially distributed band-pass filters into a coherent representation of form.

Do the current results reveal anything about binding mechanisms? In general, grouping based on extrinsic temporal synchrony—induced by the temporal structure of a stimulus—could be mediated by correlated changes in firing rates (Shadlen & Movshon, 1999; Shadlen & Newsome, 1998) or by synchronous neural responses (Singer, 1999; Singer & Gray, 1995). However, the current results would seem to pose difficulties for the firing rate hypothesis. Recall that elements undergoing temporally correlated contrast changes promote grouping both in isolation and in combination with elements undergoing simultaneous orientation, spatial frequency, or phase changes. Moreover, those contrast changes were random with respect to both magnitude and direction across the field of elements. It is well established that, within limits, increases in contrast produce increases in firing rate, whereas decreases in contrast produce decreases in firing rate. Therefore, figure–ground segmentation in any of the displays with contrast changes required the grouping of elements that were having very different effects on firing rate. Thus, it is difficult to see how the successful grouping of elements changing in contrast could be based on correlated firing rate. However, we see no obvious reason why synchronized contrast changes could not produce synchronized neural events that could promote grouping.

Regardless of the underlying mechanism, the current results highlight the central role of temporal structure in perceptual organization. In determining which spatially distinct elements belong to a unified spatial form, the visual system capitalizes on the dynamic nature of the environment. Our findings indicate that different messengers of temporal structure interact to produce a unified, abstracted message for the binding of local features into global objects. Ultimately, this abstracted message results in the grouping of synchronously changing elements without regard to the nature of the change.

Acknowledgments

Portions of this work previously were presented at the 2003 Annual Meeting of the Psychonomic Society and the 2004 Meeting of the Vision Sciences Society. The authors thank David Bloom and Chai-Youn Kim for acting as experimental observers, and two anonymous reviewers for helpful comments on an earlier version of this paper. Supported by NEI EY07760 to RB, by NIH Core Grant EY08126 awarded to the Vanderbilt Vision Research Center, and by NIH Training Grant EY07135 that supports LAG.

Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.visres.2004.10.014.

Appendix A

To test for contrast artifacts in our stimuli, each stimulus sequence was filtered with a temporal low-pass kernel of the form given by Adelson and Farid (1999):

$$h(t) = (t/\tau)^2 e^{-t/\tau}, \quad \tau = 0.01.$$

For each frame of the filtered sequence, the average root-mean-square contrast was calculated for the figure and ground regions. The absolute value of the contrast difference for the figure versus ground region indicates the extent to which contrast may promote segregation of the figure within each static image; large values indicate conspicuous segregation by contrast whereas small values indicate little or no segregation.

For Experiment 2, the absolute value of the figure–ground contrast difference additionally was calculated for the alternative (orthogonal) figure and ground regions. Taking the difference between these estimates for the correct figure orientation (as designated by temporal structure) versus the orthogonal figure orientation quantifies the contrast available at each point in time potentially to support performance on the horizontal/vertical task. Positive values indicate contrast cues favoring the "correct" orientation, whereas negative values indicate contrast cues favoring the alternative orientation. Finally, averaging over the entire stimulus sequence provides an estimate, for each trial, of the orientation and strength of figure–ground segregation based on contrast cues.

The results of these analyses indicated that contrast artifacts cannot account for performance in the current experiments. See main text for details of the filtering results for Experiment 2.

References

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

Alais, D., Blake, R., & Lee, S.-H. (1998). Visual features that vary together over time group together over space. *Nature Neuroscience, 1*, 160–164.

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences, 2*, 263–275.

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 asynchrony, destroys figural binding. *Vision Research, 35*, 491–494.

Farid, H. (2002). Temporal synchrony in perceptual grouping: A critique. *Trends in Cognitive Sciences, 8*, 284–288.

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

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

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

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

Lee, S.-H., & Blake, R. (1999a). Filtering reveals form in temporally structured displays: Response. *Science, 286*, 2231a.

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

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.

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

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. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *The Journal of Neuroscience, 18*, 3870–3896.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron, 24*, 49–65.

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

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

Wertheimer, M. (1923/1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). Highland, NY: Gestalt Journal Press.