



Contents lists available at ScienceDirect

# Vision Research

journal homepage: [www.elsevier.com/locate/visres](http://www.elsevier.com/locate/visres)



## Rapid visual grouping and figure–ground processing using temporally displays

Samuel Cheadle<sup>a,\*</sup>, Marius Usher<sup>a,c</sup>, Hermann Müller<sup>b</sup>

<sup>a</sup>School of Psychology, Birkbeck, University of London, London, UK

<sup>b</sup>Department of Psychology, Ludwig-Maximilian University, Munich, Germany

<sup>c</sup>Department of Psychology, University of Tel Aviv, Israel

### ARTICLE INFO

#### Article history:

Received 29 December 2009

Received in revised form 1 June 2010

Available online xxxx

#### Keywords:

Visual grouping

Figure ground

Temporal structure

### ABSTRACT

We examine the time course of visual grouping and figure–ground processing. Figure (contour) and ground (random-texture) elements were flickered with different phases (i.e., contour and background are alternated), requiring the observer to group information within a pre-specified time window. It was found this grouping has a high temporal resolution: less than 20 ms for smooth contours, and less than 50 ms for line conjunctions with sharp angles. Furthermore, the grouping process takes place without an explicit knowledge of the phase of the elements, and it requires a cumulative build-up of information. The results are discussed in relation to the neural mechanism for visual grouping and figure–ground segregation.

© 2010 Published by Elsevier Ltd.

### 1. Introduction

The human visual system is known to first decompose images into local visual features or elements (Hubel & Wiesel, 1968). In order to then perceive and discriminate objects embedded within a rich environment, the visual system must group the elements that belong to one object and segregate them from those belonging to other objects or the background. Research on figure–ground segregation of contour elements has focused on the relative importance of several factors identified by the early Gestalt psychologists, such as proximity, similarity, and good continuation of the elements that make up the figure (Field, Hayes, & Hess, 1993; Kovács & Julesz, 1993; for reviews, see Hess & Field (1999), Kovacs (1996)). More recently, the mechanism and the time course of figure–ground segregation has been the focus of intensive research (Clifford, Holcombe, & Pearson, 2004; Francis, 2009; Hess, Beaudot, & Mullen, 2001; Holcombe & Cavanagh, 2001; Lachapelle, McKerrall, Jaffret, & Bach, 2008; Neri & Levi, 2007; Polat, 1999; Polat & Sagi, 2006), aimed at revealing the nature of its neural code (Hess et al., 2001; Singer, 1999).

While the time to detect individual elements is short (of the order of 20 ms; Hess et al., 2001), the time needed for contour integration is thought to be much longer and to depend on the curvature of the contour. For example, the detection of a continuous contour composed of Gabor elements within a background of random elements (Path-finder displays) required approximately 100–250 ms for masked stimuli (Hess et al., 2001), suggesting a

slow process that depends on recurrent activity (i.e., feedback from higher areas and lateral connections; Craft, Schütze, Niebur, & von der Heydt, 2007; Li, Piëch, & Gilbert, 2006). Other paradigms, however, have obtained much faster estimates of the time course of visual grouping. For example, the time course of global form perception was investigated by Clifford and colleagues (2004) who designed a novel technique based on alternations between two stimulus displays, each containing a coherent spatial structure which disappears when the two displays (or frames) are summed up. In particular, they used spiral Glass-patterns (Glass, 1969) for which the two frames were mutually exclusive (the simultaneous presentation of both abolishes all global form cues), which were alternated at various frequencies. Consequently, in order to generate a coherent global percept, observers had to extract information relating to the global structure *within each frame presentation*. Discrimination between the patterns contained in the alternating frames was possible at remarkably high frequencies (~25 ms/frame), which demonstrates a much higher sensitivity of the visual system to temporal structure – indicative of a fast neural binding mechanism.

These discrepant estimations may be explained by differing stimuli and task demands, which may have resulted in the measurement of two distinct grouping mechanisms that require different amounts of processing. On the one hand, the detection of the target in the Path-finder displays (Hess et al., 2001) requires contour integration of local elements that belong to the contour, but the detection time could also be affected by additional processes, as the target needs to be selected from a variety of partial contours formed by randomly aligned elements within the background. This may therefore overestimate the time required for perceptual grouping when the background lacks strong competing signals

\* Corresponding author. Address: School of Psychology, Birkbeck, University of London, Malet Street, WC1 7HX London, UK.  
E-mail address: [s.cheadle@bbk.ac.uk](mailto:s.cheadle@bbk.ac.uk) (S. Cheadle).

(based on the same visual property as the target), as is the case in homogeneously structured background arrays. On the other hand, while the Glass-pattern paradigm involves an easier discrimination, it is arguable whether it truly measures global contour *integration* (grouping of the elements), or rather the detection of localized orientation cues, resulting from the orientation summation of multiple neighboring dot pairs, without the need to bind them (Dakin & Bex, 2001). Thus, the dominant orientation within localized regions of the stimulus may be detected via neural mechanisms also involved in texture processing. This process may lead to the general impression or sensation of structure, without the need to localize and bind specific elements involved (as is necessary with Path-finder displays). The additional fact that the modulation of V1 and V2 responses to coherent structure in Glass-patterns is seen to be minimal (Smith, Bair, & Movshon, 2002; Smith, Kohn, & Movshon, 2007) also lends weight to the argument that the local elements are not being strictly bound in Glass-pattern perception; this stands in contrast to modulation (in correlation or firing rate) of V1 and V2 activity in response to collinear stimuli that are likely to form grouped representations (Bauer & Heinze, 2002; Kapadia, Ito, Gilbert, & Westheimer, 1995; Li et al., 2006; Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998).

The time course of visual grouping provides important constraints for the underlying neural mechanism. One interesting suggestion is that grouping (or visual binding) is encoded via neural synchrony (Gray, 1999; Singer, 1999; Singer & Gray, 1995; von der Malsburg, 1999). Accordingly, detectors that respond to grouped elements (belonging to the same object or contour) respond in synchrony (on a timescale faster than that of the psychological moment, estimated to be 50 ms or more; von der Malsburg, 1999), while detectors that respond to background elements respond in a non-synchronized way. This synchronization can be mediated by lateral interactions in the visual cortex, as illustrated in a computational model that accounts for a wide range of psychophysical and physiological data on contour salience (Yen & Finkel, 1998). This scheme also predicts that grouping should be sensitive to the synchrony of visual elements (on a time scale smaller than 50 ms), as shown in a number of psychophysical studies (Leonards, Singer, & Fahle, 1996; Parton, Donnelly, & Usher, 2001; Parton, Donner, Donnelly, & Usher, 2006; Usher & Donnelly, 1998). Other neural schemes for grouping rely on facilitatory interactions between detectors with similar orientations that result in stronger responses to contours (Adini, Sagi, & Tsodyks, 1997; Li, Piech, & Gilbert, 2008; Polat, 1999) or a combination of stronger and more synchronized response onsets (Hancock, Walton, Mitchell, Plenderleith, & Phillips, 2008; Sterkin, Sterkin, & Polat, 2008).

In the present study we investigate the time course of perceptual grouping by extending the approach of Clifford et al. (2004) and Fahle (1993), in which no spatial structure exists once the cyclic display is summed across the frames of 1 cycle. A primary aim is to measure the temporal resolution with which observers can detect contours in such temporal modulated sequences. Of special interest is whether this temporal resolution can be higher than that for detecting the simultaneity of the elements. Such an outcome would support a grouping process that is sensitive to temporal modulations, faster than the 'psychological present' (the time needed to judge simultaneity). Furthermore, if grouping mediates contour detection in this paradigm, we predict that its temporal resolution will depend on the spatial coherence (e.g., smoothness) of the contour. This is in line with the results of previous studies showing that detectability of the target contour changes as a function of the alignment of between elements (Field et al., 1993). This is tested using a range of figure types that vary in spatial coherency.

Two experimental paradigms were used, testing complementary types of grouping: contours (varying in smoothness) within

a uniform background, and arrow shapes that involve conjunctions of lines, in bi-stable perception. Care has been taken in designing both types of stimuli to rule out possible artifactual cues previously associated with flickering input (Adelson & Farid, 1999; Dakin & Bex, 2002; Fahle, 2001). In Experiment 1, the figure (contour) and background elements were segregated into two alternating frames; integration of these frames resulted in a homogeneous display, for which detection of the contour was not possible. Detection was measured for a range of contours varying in smoothness (spatial coherency). In order to estimate the impact of onset and offset transients in contour grouping, and to test whether the figure/ground processing accumulates across alternation cycles, in Experiment 2 we measured contour detection with both single and multiple cycle displays (1, 3 or 5 cycles). In Experiment 3, we examined the relative importance of periodic modulations, by contrasting a temporally structured, oscillatory modulation with a non-periodic one. In Experiment 4, we extend these results to stimulus displays with bi-stable figures. Here, each frame in the alternation cycle contains visual segments consistent with one of the interpretations, but which becomes invisible once the two frames of the cycle are combined. The visual elements presented in each frame constitute an arrow structure (conjunctions of lines; Fig. 6), posing a more stringent test for visual grouping. To anticipate, we find that grouping can be performed within fast time intervals of 10–50 ms (depending on the complexity of the target structure – smooth contour vs. line junctions), but only for coherent spatial structure.

## 2. Experiment 1 – effect of spatial structure and frame duration

### 2.1. Method

#### 2.1.1. Apparatus

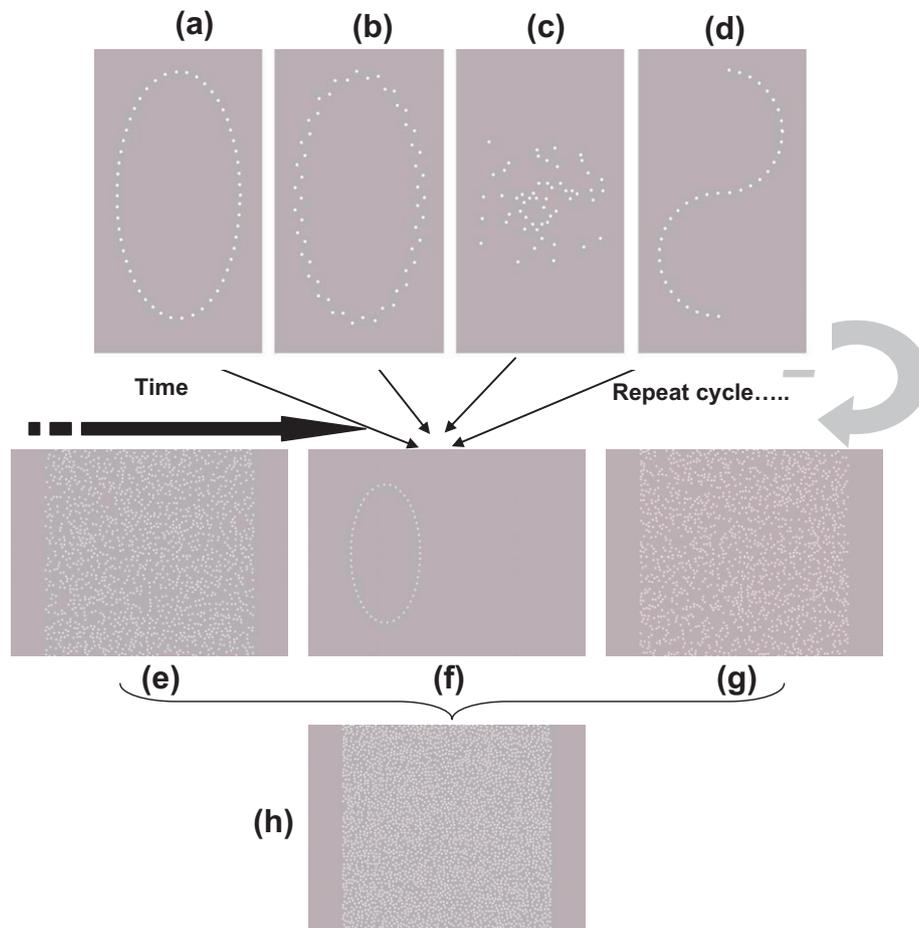
Stimuli were generated using a Microsoft Windows system PC equipped with a VSG 2/5 graphics board (Cambridge Research Systems) and displayed on a gamma-corrected Sony Trinitron Multiscan E450 monitor. For all experiments, the frame rate of the display was set at 140 Hz, the screen resolution was set at 800 × 600 pixels, and the background luminance at 28 cd/m<sup>2</sup>. Observers were tested in a dimly illuminated room, and were required to use a chin rest to minimize head movements and maintain a constant viewing distance of 57 cm. Responses were recorded using the left and right buttons of a CT3 four-button response box (Cambridge Research Systems).

#### 2.1.2. Observers

Five observers (four naive), with normal or corrected-to-normal vision, participated in Experiment 1.

#### 2.1.3. Stimuli

The stimulus display was divided into a three-frame sequence, in which frame B<sub>1</sub> contained a random half (1500) of the background dots (e.g., Fig. 1e), frame T the 52 target figure defining dots (e.g., Fig. 1f), and frame B<sub>2</sub> the remaining 1500 background dots (e.g., Fig. 1g). All dots were white, had identical properties, and were displayed on a gray background with a luminance of 28 cd/m<sup>2</sup>. After setting the coordinates of the target dots, the background dots were positioned sequentially, with the constraint of a minimum inter-element distance of 0.44 deg. This three-frame sequence was presented for a fixed number of cycles (e.g., 5), in immediate succession (no delay between the cycles) and was forward- and backward-masked with a display (M) containing all dots. For example a stimulus sequence of 2 cycles can be denoted in the following way: M, B<sub>1</sub>, T, B<sub>2</sub>, B<sub>1</sub>, T, B<sub>2</sub>, M (where M = mask, B<sub>1</sub> and B<sub>2</sub> = background, and T = targets, and there is no gap between



**Fig. 1.** Stimulus screen shots and display procedure. (a) Smooth-Ellipse target, (b) Fuzzy-Ellipse target, (c) Gaussian-Cloud target, (d) Snake target, (e) first distractor set, (f) ellipse target elements, (g) second distractor set, (h) summed input of the three-frames sequence.

the frames). The mask (M) had a constant duration of 21 ms, across all experimental conditions, while the duration of the other frames was varied. Three target conditions were used: (i) a Smooth-Ellipse, (ii) a Fuzzy-Ellipse, and (iii) a Gaussian-Cloud. In the Smooth-Ellipse condition, the stimulus consisted of an elliptical contour, made up of 52 white dots (e.g., Fig. 1a). The target ellipse had a width of 4.35 deg and a height of 8.68 deg, or vice versa. The orientation of the target ellipse, vertical or horizontal, together with its position, was maintained across all frames of a single trial, but varied randomly between trials. Thus there was no correlation between the type of target and the location when it appears on the screen.

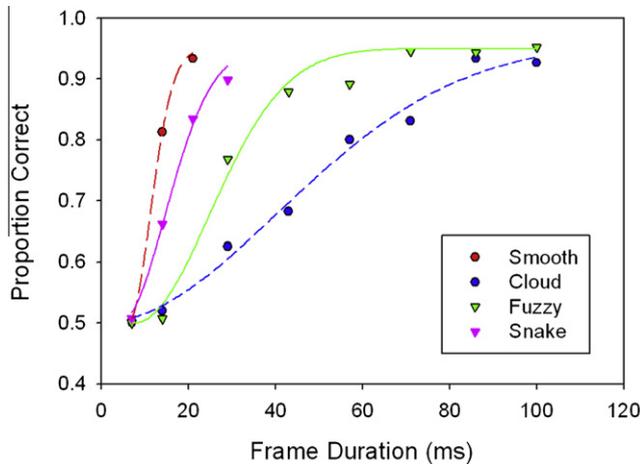
In the Fuzzy-Ellipse condition, the target-forming dots were not placed directly onto the elliptical contour, but offset by a regular distance (0.22 deg) either towards or away from the center of the ellipse, resulting in a jagged, or fuzzy elliptical shape (Fig. 1b). This structure retains the same global properties as the original “smooth” structure, but the orientation signals generated by the local dot pairs are not as easily integrated, resulting in a more difficult binding condition. The third structure was designed with the aim of abolishing, as far as possible, any clear structural information in the target. In this Gaussian-Cloud condition (Fig. 1c), an identical number of dots were randomly positioned without structure according to a 2D Gaussian distribution ( $SD = 2.2$  deg), and positioned in either the left or the right half of the display.

A number of important properties of the stimulus should be noted: (1) all elements of the stimuli are flickering and performance cannot be based on detection of the flicker. Neither can per-

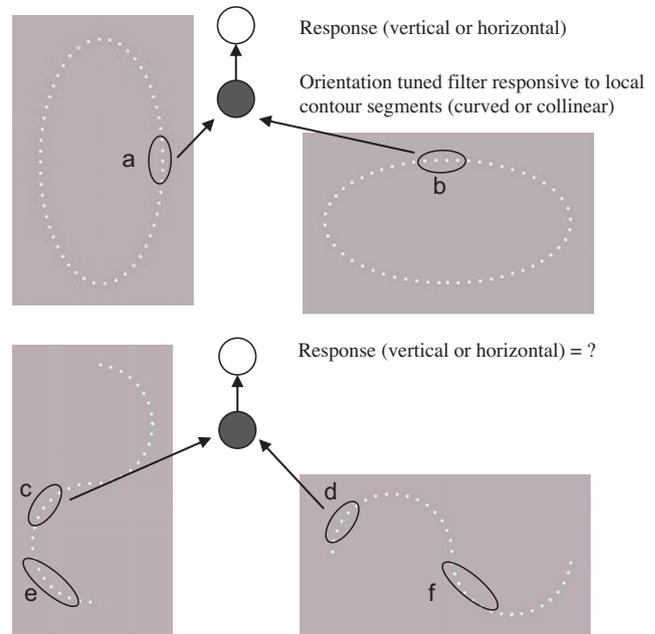
formance be based on the perception of motion signals at the border of figure-ground regions (Kandil & Fahle, 2001), because any such signals are effectively masked by additional motion signals between background elements. (2) As both target and background regions are presented periodically and have identical properties on each presentation, detection cannot be based on perceived contrast differences between figure and ground regions (Adelson & Farid, 1999). (3) Most importantly, no spatial structure exists in the time averaged stimulus sequence. Thus, detection of the figure requires the visual system to utilize temporal structure for grouping.

#### 2.1.4. Procedure

Observers were required to make binary orientation discriminations (vertical or horizontal) in the Smooth-Ellipse and Fuzzy-Ellipse conditions and binary localization judgments (left or right half of screen) in the Gaussian-Cloud condition. They were informed that a target (contour or cloud of dots) was present in the display on every trial, and that they should try their hardest to detect it. They were instructed to fixate centrally and avoid eye movements. Auditory feedback was given for errors. The three spatial-grouping conditions were run in separate blocks, in each of which a range of different frame durations were randomly intermixed. Each observer completed 60 trials per frame duration. For the Fuzzy-Ellipse and Gaussian-Cloud targets, seven different frame durations were used: 14, 29, 43, 57, 71, 86, and 100 ms. For the Smooth-Ellipse condition, three different frame durations were used: 7, 14, and 21 ms. For all grouping conditions, the stim-



**Fig. 2.** (A) Displays mean proportion-correct results (five subjects) for the Smooth-Ellipse, Fuzzy-Ellipse, Gaussian-Cloud, and Snake (see Section 2.3) conditions. The 0.5 proportion-correct level indicates chance performance. The data were fitted with a four-parameter Weibull function (Wichmann & Hill, 2001).<sup>12</sup>



**Fig. 3.** Schematic illustration of the detection of local-contour regions, and their potential influence on task performance. For the ellipse figure, short sections of the contour may activate orientation-tuned filters (e.g., the ovals a and b), which will be sufficient for orientation discrimination. In the Snake condition, this problem is eliminated as almost all local-contour sections (e.g., c and d) are replicated in both horizontal and vertical conditions. Even detectors with larger receptive (e.g., e and f) fields will not respond selectively to one orientation of the snake stimulus.

2.3. Discussion

Using the temporal-grouping paradigm with alternating frames of figure/background elements, we found, that the visual system is able to form groupings rapidly. At short frame durations (7 ms for the Smooth-Ellipse and 14 ms for the other two conditions) the ability of the observers to detect the target (i.e., to group the target elements and discriminate them from the background) was at chance. This is despite the fact that the stimulus was repeated for 5 cycles (i.e., 35 ms for Smooth-Ellipse and 70 ms for the other conditions, excluding the presentation time of the background). Furthermore, we observed that further increasing the number of cycles at such fast presentation rates does not help to boost target detection (not reported). In all conditions, the important parameter that affects target detection rate is the frame duration for the presentation of target and background elements. We thus follow, Clifford et al. (2004) in considering this as the critical variable (or limiting factor) for the temporal resolution of visual grouping (but see Section 3 for a discussion on the role of the number of cycle repetitions).

The time course of visual grouping obtained (Fig. 2) depends strongly on the spatial coherency of the to-be-grouped elements. In particular, we found that the highest temporal resolution (12 ms) was obtained for the Smooth-Ellipse condition in which the elements strictly followed the Gestalt law of good continuation. In the other two conditions, the strength of grouping between the constituent elements was weakened, so that that structure was either less clearly defined (Fuzzy-Ellipse) or not present at all (Gaussian-Cloud). In consequence, these conditions resulted in a reduced temporal resolution of 32 and 53 ms, respectively.

We thus found a high temporal resolution of visual grouping for smooth contours, in a task in which the background is unstructured and does not contain partial targets that could act as distractors. An important question, however, is whether the mechanism

ulus sequence was cycled five times for each presentation. The stimulus sequence was both forward- and backward-masked, to prevent the influence of onset and offset transients, by a ~20-ms display containing all (target and background) dots. The order of the blocks was counter-balanced across subjects. Subjects 1–3 performed the conditions in the following order: Fuzzy, Smooth, Cloud. The reversed order was used for subjects 4 and 5: Cloud, Smooth, Fuzzy. Prior to starting each block, 50 practice trials were given.

2.2. Results

Fig. 2 shows the target discrimination accuracy in the three conditions (Smooth-Ellipse, Fuzzy-Ellipse, and Gaussian-Cloud) as a function of frame duration. For the Smooth-Ellipse condition accuracy remains at chance level at 7 ms, and until 14 ms for Fuzzy-Ellipse and Gaussian-Cloud conditions. Performance then improves with frame duration. Moreover, performance is seen to be clearly superior for the Smooth-Ellipse condition (red-circles),<sup>1</sup> relative to the other two, “weaker” spatial-grouping conditions (green-triangles and blue-circles).

For the Smooth-Ellipse condition, ceiling performance was reached with 21 ms per frame, which is why longer frame durations were not tested. Orientation discrimination for a Fuzzy-Ellipse, whose constituent elements were not coherently linked, required longer frame durations and improved more gradually, but performance was consistently higher compared to Gaussian-Cloud localization.

Statistical comparisons between conditions were made by computing 75% thresholds for individual subjects (Fig. 3). For the Smooth-Ellipse condition, this threshold was reached with a frame duration of only 12 ms demonstrating the high sensitivity of the visual system to spatially coherent stimuli. Relative to the Smooth-Ellipse, temporal thresholds were longer for the Fuzzy-Ellipse condition: 32 ms ( $t(4) = 7.0, p = .002$ ), and for the Gaussian-Cloud condition: 53 ms ( $t(4) = 12.3, p < .0001$ ). The latter condition had the longest temporal threshold, and it also differed significantly from that of the Fuzzy-Ellipse condition ( $t(4) = 4.0, p = .017$ ).

<sup>1</sup> For interpretation of color in Figs. 1–6 and 9, the reader is referred to the web version of this article.

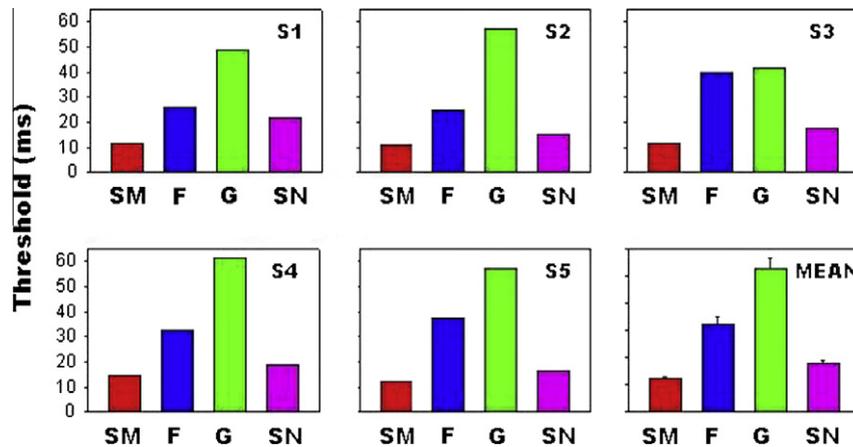


Fig. 4. Seventy-five percent correct thresholds for individual subject, plus mean results with error bars denoting 1 SE (between subjects). SM = Smooth-Ellipse, F = Fuzzy-Ellipse, C = Gaussian-Cloud, and S = Snake.

underlying this fast process is full-fledged grouping or whether it involves mediation by local detectors that respond to elongated elements and that are sensitive to synchrony of their inputs, as illustrated in Fig. 4.

In order to examine this, we tested a group of five new observers with the same paradigm, except that the contours were not elongated ellipses, but rather S (or snake) shapes (see Fig. 1d). The results are shown as a pink curve in Fig. 2, indicating a fast temporal resolution (75%-threshold of 18 ms). We consider this figure to provide a more realistic estimate of the time needed to group visual elements that fall along a smooth contour.

There is one important aspect of the experiment that was not addressed: the number of cycles for target-background presentation. This has important implications for the process by which grouping is achieved. For example, if grouping is triggered by visual transients, one may expect that it will not improve with repetitions. If, on the other hand, it involves an accumulating process triggered by the detection of synchronous (at the frame rate scale) contour elements, then detection should improve with the number of cycles. We examine this issue in Experiment 2.

### 3. Experiment 2 – performance for single vs. multiple cycle presentations

In the second experiment we measured detection accuracy for the same class of stimuli used in Experiment 1, but for a range of stimulus presentation durations. This was done by varying the number of times the three-frame cycle was repeatedly presented (either 1, 3 or 5 cycles). This manipulation allows us to investigate the importance of different factors in the detection process; detection could either be mediated by the onset/offset transients (as is the case in judgements of simultaneity; Guttman, Gilroy, & Blake, 2007), or by a grouping process that accumulates across repetition cycles.

#### 3.1. Method

##### 3.1.1. Observers

Five observers (four naive), with normal or corrected-to-normal vision, participated in Experiment 2.

##### 3.1.2. Stimuli and procedure

The stimuli and procedure were identical to that used in Experiment 1, apart from the following details. Only a single target (the Smooth-Ellipse) and a single frame duration were tested (14 ms).

This frame duration was used because it resulted in sufficiently high performance for 5 cycles, in the previous test. The number of stimulus cycles was varied between 1, 3 and 5, and the order of presentation randomised within sessions.

#### 3.2. Results

Fig. 5 displays the target discrimination accuracy for the five observers in the Smooth-Ellipse condition, as a function of the number of alternation cycles.

One can see that, for all observers, the accuracy of grouping improves with the number of cycles, from chance level with a single cycle towards relatively high accuracies at 3 or 5 cycles. Comparisons showed that the accuracy increased significantly from 1 to 3 cycles (mean difference = 30%;  $t(4) = 5.1, p = .007$ ), and increased further between 3 and 5 cycles, although the latter increase (mean difference = 5%) was only marginally significant ( $t(4) = 2.7, p = .054$ ).

#### 3.3. Discussion

The results of Experiment 2 have shown that the ability to utilize temporal structure at short frame durations can accumulate over the cycles. This is not an obvious result, since sensitivity to

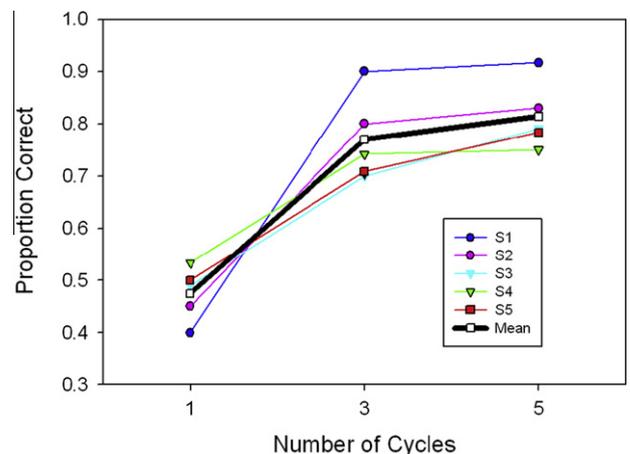
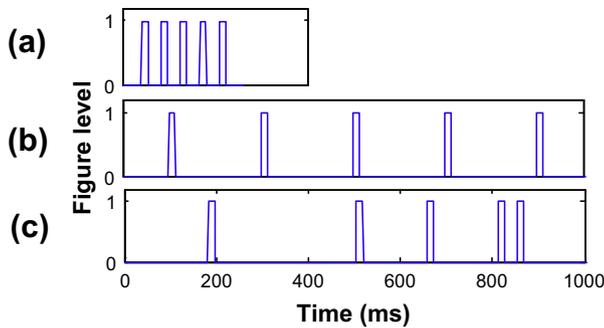


Fig. 5. Proportion correct, for each of the five observers, in detecting a smooth ellipse (from background) and reporting its orientation (vertical/horizontal) at 14 ms frame duration for 1, 3, and 5 alternation cycles.



**Fig. 6.** Example presentation sequences used for the comparison of short periodic (a: original sequence), long periodic (b), and long non-periodic (c) temporal structures. The blue bars represent the time points at which the target figure is displayed, among background frames.

onset/offset transients), while at the same time preventing detection of the target. Long intervals between target frames were filled with this new two-frame cycle. To create a periodic and non-periodic sequences with the same number of figure frames, it was necessary to lengthen the presentation time of the non-periodic displays to 1 s. Long stimulus sequences were created by either equally spacing (every 200 ms; periodic condition) or randomly positioning (non-periodic condition) the five target frames. The remaining elements of the display array were filled with the two-frame sequence (described above), in which the target (T) was undetectable. Example sub-sequence for the periodic condition: B<sub>1</sub>, T, B<sub>2</sub>, B<sub>1</sub>, T, B<sub>2</sub>, etc. Example sub-sequence for the long periodic and non-periodic conditions: B<sub>1</sub>, T, B<sub>2</sub>, B<sub>3</sub>, B<sub>4</sub>, B<sub>3</sub>, B<sub>4</sub>, B<sub>1</sub>, T, B<sub>2</sub>, etc.

4.2. Results

The results indicated that the presence of a periodic structure or its frequency do not influence the efficiency of visual grouping, which was equivalent across conditions (performance levels of 86%, 86%, and 89% for conditions (i), (ii), and (iii), respectively, with no significant differences). The lack of difference between short and long display conditions indicates that the ISI between target frames is not critical, and that any partial, or incomplete grouping representation can be maintained for >100 ms, and integrated with future grouping signals.

4.3. Discussion

The results of this experiment do not support a special role for stimulus evoked oscillatory modulations (at least at the frequency tested) in the mediation of visual grouping. The results of the first three experiments, however, indicate that grouping is sensitive to fast (non-periodic) modulations of visual elements. In all these experiments we used alternating displays, in which the figure and the ground elements are presented in successive frames, such that when summed together they result in a spatially homogenous display lacking spatial structure. To detect the figure, observers had to group the elements during a frame of the display or at least to do so *partially* and then accumulate this information across temporal cycles of presentation (see further discussion in Section 6). It could be possible to argue, however, that target detection is partially mediated by motion cues due to the phase difference between the target and the background elements (Kandil & Fahle, 2001). We believe this is unlikely for two reasons: first, the segregation of randomly positioned background elements into two frames, introduced motion throughout the display, masking any special motion cue at the target. Second, the fact that the temporal resolution depends on the smoothness of the elements indicates that indeed the task measures grouping (motion cues would have been equally effective for detection of a Smooth-Ellipse or Gaussian-Cloud targets. Nevertheless, we aimed in Experiment 4 to use a design that removes any contribution of motion cues leading to detection of the target.

5. Experiment 4: grouping in ambiguous figure-ground displays

To rule out any contribution of motion cues we switched from a display that contains a figure within a homogenous background, to a bi-stable display that contains an image, whose elements (line segments) can be grouped according to two alternative figure interpretations: leftward/rightwards arrows (Fig. 7e). In this case one type of binding of the lines results in left-arrows, while a different binding results in right arrows. Moreover, each of these perceptual organizations includes all the line elements; the difference

temporal information is known to be high for stimulus onsets, so that, conceivably, the ability to perform visual grouping at such fast alternation rates is primarily due to stimulus onset (or possible offset) transients (Guttman et al., 2007). This has been ruled out by the finding of chance level (50%) discrimination performance for a single cycle. Performance is seen to increase dramatically from 1 to 3 cycles, but also, although to a smaller degree, from 3 to 5 cycles. This may indicate that either the visual system is sensitive to rhythmic, oscillatory structure in the input signal (consistent with the reported oscillatory activity in visual grouping; Brosch, Bauer, & Eckhorn, 1997; Busch, Herrmann, Muller, Lenz, & Gruber, 2006; Castelo-Branco, Goebel, Neuenschwander, & Singer, 2000; Vidal, Chaumon, O'Regan, & Tallon-Baudry, 2006), or that there is another type of cumulative build-up of the grouping computation over a number of cycles. To distinguish between these two alternative interpretations, we carried out a further test, contrasting a temporally structured, oscillatory signal with a non-periodic signal.

4. Experiment 3 – importance of periodicity in the input signal

Experiment 3 investigates whether the reported grouping effect is dependent on a sustained periodic signal. This is achieved by contrasting three types of alternation protocols that vary in the temporal structure of the stimuli: (i) a periodic sequence identical to the one used before (Fig. 6a); (ii) a non-periodic sequence (Fig. 6c) in which target frames appeared irregularly, with five target frames randomly positioned throughout a 1-s display sequence; and (iii) a lengthened periodic sequence, with five target frames evenly positioned (every 200 ms) throughout a 1-s display sequence (Fig. 6b).

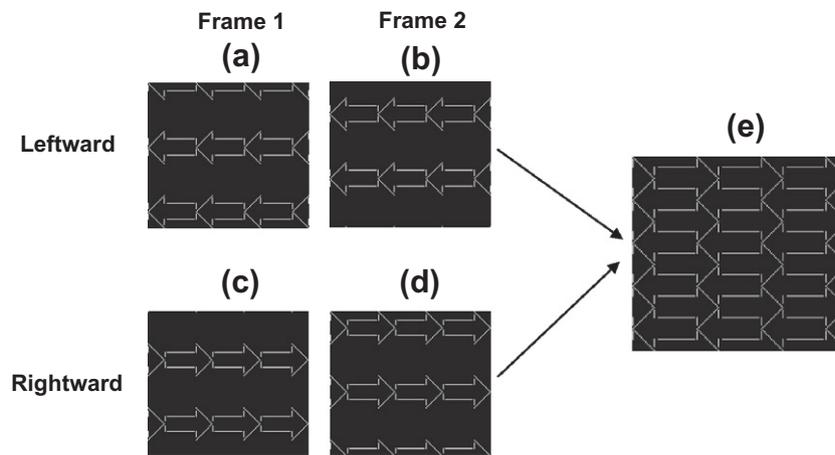
4.1. Method

4.1.1. Observers

Five observers (four naive), with normal or corrected-to-normal vision, participated in Experiment 3.

4.1.2. Stimuli and procedure

The stimuli and procedure were identical to that used in Experiment 1, apart from the following details. As in Experiment 2 a single target (ellipse) and frame duration (14 ms; ~75% threshold level in the original experiment) was used, with the conditions randomized within blocks. All sequences contain the same number of figure frames (namely, 5), but for the long periodic and non-periodic conditions, an additional display cycle was created which consisted of a two-frame background (B) sequence (B<sub>3</sub> and B<sub>4</sub>) each containing a random selection of distractor and target dots. This was used to maintain perceptible flicker (important for preventing



**Fig. 7.** Screen shots of stimuli used for the bi-stable arrow experiment. (a and b) Show two consecutive frames, in which alternate lines of leftward pointing arrows are displayed. (c and d) Show the equivalent for rightward pointing arrows. (Panel e) displays the resulting (bi-stable) composite image, and is equivalent to the observers' percept at high presentation frequencies of the two-frame sequence. In this image, groups of either left- or rightward pointing arrows can be perceived.

only involves the perception of the interior of the arrows as figure or background. As before, we alternate frames, which present, half of the elements each, so that the synchrony of line elements within a frame provides a bias for one of the interpretations, but this bias vanishes when the frames are summed up. In order for such alternation to affect the figure/ground assignment, the presentation time of each frame needs to be such that one can bind its line elements. Note that now, while up–down motion may be perceived in the display, this does not distinguish between the two alternative interpretations; all elements belong to both perceptual organizations, and thus motion cues do not indicate the location of the figure.

Previous studies that examined the impact of temporal modulations (phase of flicker) on bi-stable perception of symmetric arrays, found an effect of temporal phase on perception (rows vs. columns) at fast resolutions of 13–16 ms (Parton et al., 2001; Usher & Donnelly, 1998). That affect, however, could be explained as a result of the contribution of detectors with elongated receptive fields (vertical or horizontal), which are sensitive to the synchrony of their input. Note that such detectors could not account for the left/right arrows in the present display, both interpretations rely on activation of the same orientation detectors. Thus this task is likely to be more difficult, measuring the temporal resolution for grouping of elements that involves conjunctions of intersecting lines.

Additionally, we investigate the relationship between the temporal resolution for grouping and the temporal resolution for simultaneity judgments, using a variant of the bi-stable arrows stimulus. The ability to perform simultaneity judgments implies that the observer has knowledge of which elements appear within the same frame, and which appear in different frames. To rule out the possibility that performance in the grouping task relies upon this type of explicit temporal knowledge (reflecting a relatively trivial process in which elements can be “tagged” as belonging to a particular phase; this would surely be the case for very slow presentations) it is necessary to establish a higher temporal resolution for grouping than for judgement of simultaneity.

## 5.1. Method

### 5.1.1. Observers

Nine observers (eight naive) participated in the arrows discrimination task, and nine different observers (eight naive) participated in the line discrimination control task. All had normal or corrected-to-normal vision.

### 5.1.2. Stimuli

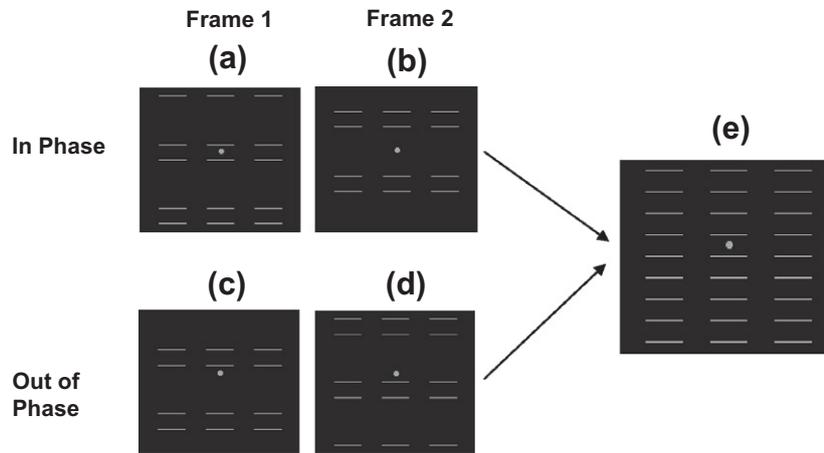
The stimulus used in the arrows detection task was generated by iteratively placing simple “arrow” shapes (Fig. 7) at regularly spaced intervals on the screen. The structure of these arrows is such that, when multiple arrows are drawn adjacently with the appropriate alignment, the stimulus becomes bi-stable, that is: multiple instances of either leftward or rightward pointing arrows can be perceived (Fig. 7c). Each arrow is formed from the conjunction of eight different line segments, occupying an area of  $3.9 \times 2.6$  deg measured from the flanking tips of the arrow. Arrows were spaced at regular distances along the horizontal axis (every 3.9 deg), and extended to the edge of the display area ( $35.8 \times 26.0$  deg), to avoid any biasing of the bi-stable pattern.

For the purposes of this experiment and in order to investigate the role of temporal structure, this bi-stable display was split into two parts and assigned to different frames. Each frame contained alternating rows (Fig. 7), so that adjacent rows were never drawn in the same frame. The resulting two frames are no longer bi-stable – they contain groups of arrows that will either point leftward or rightward. When alternated at slow speeds the stimulus is not ambiguous, and points only in a single direction. However, when this two-frame sequence is presented in alternation at high frequencies, the distinction is no longer apparent. We aimed to investigate the time course of grouping the separate line sections into a coherent object (an arrow), by varying the frame durations for this two-frame sequence.

For the temporal (line) judgement control task, a corresponding stimulus was constructed in which only the horizontal line segments of the original display were present (Fig. 8). All vertical and diagonal line segments were removed. This abolished spatial structure from the display and prevented the grouping of line segments into coherent objects. This modification to the stimulus was accompanied by a modification to the task. As discrimination based on spatial structure was no longer possible, observers were required to make a temporal judgment regarding the relative phase of two target line segments.

### 5.1.3. Procedure

Participants were required simply to view the bi-stable stimulus and report the orientation (either leftward or rightward) of the perceived arrows. They were instructed to maintain fixation at a central position, during a 5-cycle presentation. The display sequence consisted of, first, a 20-ms “mask” composed of all arrows (Fig. 7e) and therefore not biasing the observer to perceive a particular orientation, followed by 5 cycles of the two-frame stimulus



**Fig. 8.** Example stimulus screen shots for the temporal (line) judgement task. For each of the two conditions (in-phase/out of phase), the constituent two-frame sequence is presented. (a and b) Show a sequence in which the first frame (a) contains both target lines (above and below the fixation dot) and the second frame (b) contains neither, resulting in a synchronous (in-phase) presentation. For the out of phase presentation condition, each frame (c and d) contains one of the target lines, resulting in an asynchronous display. The combination of all lines (at high frequencies when observers can no longer segregate the frames) results in e.

sequence (described above), followed by an additional 20-ms mask. A range of frame durations were tested and randomly intermixed within blocks. Each observer completed 50 trials per frame duration. Observers were informed that, occasionally, they may perceive a heterogeneous group of arrows (pointing both left and right), and in this case they should respond according to the strongest percept. Feedback was not given.

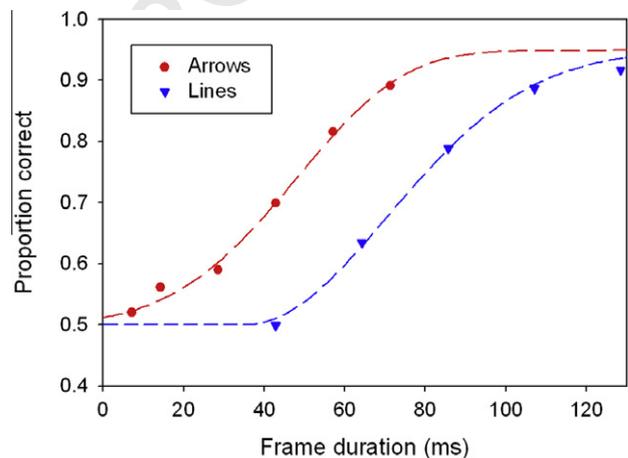
For the temporal (line) judgment task, observers were required to make a judgment about the phase properties of two target line sections. These targets were defined as the two closest segments to a centrally located fixation dot, and could either be presented in-phase (same frame) or 180 deg out of phase (different frames).

5.2. Results and discussion

Fig. 9 shows the proportion of responses for which the observers' percept agreed with the arrows' direction contained in individual frames (and which corresponds to the accuracy in the detection of the arrow direction of the individual frames), as a function of the frame duration (red symbols and line). Although this task has a slower 75%-correct temporal threshold (48 ms) than that established for the Smooth-Ellipse condition (Experiment 1), this interval is nevertheless much faster than the intervals associated with slower attentional processes, which usually exceed 100 ms, as reported in previous studies (Hess et al., 2001).

One objection to the interpretation of the arrow-perception as reflecting grouping, could be that performance is driven by the ability to tag the line elements as belonging to various frames, and responding on the basis of elements that are perceived as simultaneous. In order this alternative interpretation we tested another group of observers on the same task, except that only the horizontal line elements were presented (Fig. 8). With such displays, the to-be-grouped elements are not spatially contiguous (parallel lines), minimizing the spatial coherency of the percepts.

As can be seen in Fig. 9 (blue symbols), this task has a slower temporal threshold (78 ms) and differed significantly from the arrows threshold ( $t(16) = 7.8, p < .001$ ), indicating that the ability to group and segment information on the basis of temporal phase in such displays is enhanced when the figure has continuity. In particular, one can observe that for a frame duration of around 40 ms, observers report the arrow consistent with the temporal manipulation 70% of the time, although their ability to say if two adjacent lines flicker in/out of phase is at chance level (see Parton et al. (2001), for a similar result, in the domain of flicker detection).



**Fig. 9.** Proportion-correct scores for the arrows task in which observers were required to report perceived orientation of an array of bi-stable arrows (fitted with a Weibull curve; see Experiment 1 for details), and the lines task in which observers were required to make a temporal judgment on the phase (synchronous or asynchronous) of two discrete line sections. Nine observers for the arrows condition and nine (different) observers for the lines condition.

6. General discussion

Using a stimulus alternation paradigm, in which figure and background elements are alternated for a number of cycles, we investigated the temporal resolution of visual grouping. This temporal resolution reflects the time it takes for the visual system to compute some aspects of the grouping process that can be maintained and integrated across time (despite interruption by noise). Thus, the results reflect the time it takes to build-up a basic unit of grouping that can then be extended. Two stimulus types were used that varied in the complexity of the figural grouping and in the nature of the background. In the first three experiments, the elements were dots, the background corresponded to a homogeneous dot field, and the figure consisted of a set of dots that were arranged either as a contour object (Smooth/Fuzzy-Ellipse or S-Shape) or as a non-contour object (Gaussian-Cloud). First, we found that the critical duration that permits grouping depends on the spatial coherency of the figure. This critical duration was lowest for the smooth contour condition (12–18 ms), and it increased considerably for non-smooth contour (32 ms) and for non-contour figures (53 ms). Second, the results show that the

accuracy of grouping increases with the number of figure/background alternation cycles. Third, we found that this process is not dependent on the periodicity (or frequency) of the target frames among background frames. The second type of stimuli (Experiment 4) involved line elements that created an ambiguous figure/ground assignment, with the figure consisting of more complex conjunctions of lines (arrows; Fig. 7). Using these stimuli, we found a slower temporal resolution for grouping (48 ms), but this was nevertheless much faster than previous estimates.

These results are consistent with those of Clifford et al. (2004), who reported Glass-pattern grouping at a temporal resolution of ~20 ms, and extends them to stimuli which necessitate binding of the elements before target detection can occur, rather than stimuli which may be categorized using a mechanism that pools orientation signals over a local area, similar to a texture processing mechanism. Thus, our stimuli are likely to probe grouping more directly. The critical time resolution for this process was in the range of 10–50 ms, which is much faster than previous estimates using contour displays (Hess et al., 2001). One likely cause of the higher temporal resolution in our experiments compared with those of Hess et al. (2001) may have to do with the nature of the background. While we used a uniform background, this was not the case in the Path-finder paradigm, where partial contours are randomly present in the background and therefore the contour detection becomes a task of selecting the stronger among a set of potential contours. Additionally, the figure elements in the Path-finder paradigm were strongly masked after each frame presentation (by randomly oriented elements in identical positions), thus minimizing the possibility of a cumulative build-up of grouping information. It is, thus, possible that such a process requires a longer duration for its resolution, which is limited by feedback connections and attentional processes. Our results, however, suggest that the temporal resolution of visual grouping can be much higher, when no extra time is needed for selecting among multiple figures. Such a fast time scale is in contrast with the reduced time needed for temporal judgments of simultaneity of flickering elements (Cheadle et al., 2008; Forte, Hogben, & Ross, 1999; Motoyoshi, 2004) – indicative of the temporal resolution of a slow attentional mechanism required to temporally isolate the flickering elements. A fast time scale for grouping relative to that for attentional deployment is also consistent with theoretical claims that grouping needs to precede visual attention (in order to guide it; Craft et al., 2007; Qiu, Sugihara, & von der Heydt, 2007), and with studies that have demonstrated grouping effects in the absence of visual attention (e.g., Lamy, Segal, & Ruderman, 2006).

The results also have potential implications for the nature of the neural mechanism that mediates grouping and figure-ground discrimination. It is important to note that although the limiting factor (frame duration) was relatively fast (<50 ms) for the grouping of elements along continuous contours, detection was facilitated by the repetition of the stimulus sequence for at least 3 cycles. One way to interpret these results is by assuming that the critical frame rate reflects a minimum processing time, such that grouping information can only be extracted if processing time exceeds this critical duration. For Experiments 1–3, one possibility is that, at the neural level, first-stage orientation detectors are activated more strongly by synchronous pairs of dots (that are co-present within a critical duration (10–20 ms), especially if they are supported by lateral connections along a contour (Adini et al., 1997; Polat, 1999). This activation, however, is likely to be interrupted by the background frames, especially in Experiment 3, where the frame targets are rare, making it unlikely that target detection is mediated only by sustained (across cycles) activity of target orientation detectors. Thus, we believe that a secondary process (characterized by a minimal time) of *binding* comes into play, by which relations between these orientation selective units can be

computed. Note this *binding* process is even more essential in Experiment 4, where all line elements are activated equally (in terms of response strength) and it is only the co-activation during a critical interval that carries relational information. This *binding* process can then accumulate/consolidate with additional target presentations, which enhance target detection.

Further work is required in order to understand the nature of the accumulation process that takes us from the first-stage orientation responses to full grouping. One possibility (subject to some debate; for objections see Shadlen & Movshon (1999) is based on the idea that visual binding involves fast learning of synaptic connections between co-active (temporally correlated) representations, possibly via top-down feedback connections (von der Malsburg, 1981, 1999; see also Li et al. (2008), Polat & Sagi (1994), for data supporting synaptic learning in contour integration). The temporal resolution of visual grouping obtained in our study is consistent with this idea: it was important to have the figure elements presented simultaneously within a brief frame duration (18–48 ms) and several cycles were enough to complete the grouping. It is important to note, however, that we did not find an advantage for periodic (relative to stochastic) contour sequences, suggesting that the mechanism that binds contour elements does not depend on externally induced neural oscillations. Although temporal binding is often formulated as implying such oscillatory activity, this is not a necessary condition for temporal binding (Niebur & Koch, 1994). As discussed above, it is possible that, with each co-activation of the figure elements, an incremental binding process is set in place and accumulates across multiple cycles (Fig. 7). This interpretation is consistent with other data showing that aperiodic synchrony can drive grouping (Lee & Blake, 1999), and more recent studies pointing to the importance of response onset similarities for perceptual grouping, be it in the visual input (Hancock et al., 2008) or the neural responses to visual input (Sterkin et al., 2008).

The slower resolution of grouping reported in Experiment 2 is likely to be related to a number of factors. First, as discussed above, the stimulus used in Experiment 2 does not give the figure elements any advantage (in term of synchrony-dependent saliency) relative to background elements, and thus it poses a more stringent requirement on visual binding. Second, the necessity of grouping non-smooth junctions of lines may involve additional neural circuitry that is not needed for continuous contours. For example, it is possible that while the continuous contour integration is mediated by lateral connections within a visual area (e.g., V1), the grouping of non-smooth line junctions may additionally require the involvement of extrastriate areas (e.g., V2; Zhaoping, 2005). In the latter case, signals must travel a greater distance, resulting in an increased critical duration for grouping.

In previous studies with bi-stable perception, the effect of temporal modulations on perceptual interpretation has produced variable conclusions. While Kiper, Gegenfurtner, and Movshon (1996) found very little impact of temporal modulation at a range of frequencies (15–60 Hz, but a constant frame duration of 8 ms) on the visual interpretation of the display, Usher and Donnelly (1998) observed that fast modulations of 13–15 ms/frame were sufficient to bias the perceptual interpretation of symmetric dot lattices (rows vs. columns organization; see also Parton et al., 2001). In the latter study, however, the result could be interpreted as being due to detectors with elongated receptive fields, which are sensitive to synchrony of their input. The display used in our Experiment 4 was constructed so as to avoid such an interpretation, thus posing a more stringent temporal resolution for visual binding.

Future studies may extend the range of stimuli used here and reveal more complex dependencies of temporal sensitivity on stimulus structure. As we have demonstrated, large sensitivity

differences exist for simple (smooth-contour) figures vs. complex (line conjunction) figures. Slower grouping for conjunctions is likely to be due to the required involvement of multiple visual areas in resolving grouping, whereas in the case of contour binding, physiological and anatomical evidence suggests that lateral connections within a single visual area (e.g., V1) are sufficient. It will also be important to fully determine the influence of higher-level band-pass filters (described in the Introduction) because, although we have sought to minimize the influence of any such filters (detection cannot be based on the activation of single detectors!), we cannot be certain such filters are not involved in the grouping processes in our tasks.

7. Uncited reference

Polat and Sagi (1993).

Acknowledgments

Thanks to Frank Bauer for help planning and running the experiments, and to two anonymous reviewers for helpful criticism. This work was supported by a grant of the German Research Council (DFG) to H. Müller and M. Usher.

References

Adelson, E. H., & Farid, H. (1999). Filtering reveals form in temporally structured displays. *Science*, 286(5448), 2231+.

Adini, Y., Sagi, D., & Tsodyks, M. (1997). Excitatory-inhibitory network in the visual cortex: Psychophysical evidence. *Proceedings of the National Academy of Sciences of the United States of America*, 94(19), 10426–10431.

Bauer, R., & Heinze, S. (2002). Contour integration in striate cortex. Classic cell responses or cooperative selection? *Experimental Brain Research*, 147, 145–152.

Brosch, M., Bauer, R., & Eckhorn, R. (1997). Stimulus-dependent modulation of correlated high-frequency oscillations in cat visual cortex. *Cerebral Cortex*, 7, 70–76.

Busch, N. A., Herrmann, C. S., Müller, M. M., Lenz, D., & Gruber, T. (2006). A cross-laboratory study of event-related gamma activity in a standard object recognition paradigm. *NeuroImage*, 33(4), 1169–1177.

Castelo-Branco, M., Goebel, R., Neuenschwander, S., & Singer, W. (2000). Neural synchrony correlates with surface segregation rules. *Nature*, 405, 685–689.

Cheadle, S., Bauer, F., Parton, A., Müller, H. J., Bonneh, Y. S., & Usher, M. (2008). Spatial structure affects temporal judgments: Evidence for a synchrony binding code. *Journal of Vision*, 8(7), 1–12. <<http://journalofvision.org/8/7/12/>>. doi:10.1167/8.7.12.

Clifford, C. W., Holcombe, A. O., & Pearson, J. (2004). Rapid global form binding with loss of associated colors. *Journal of Vision*, 4(12), 1090–1101. <<http://journalofvision.org/4/12/8/>>. doi:10.1167/4.12.8.

Craft, E., Schütze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. *Journal of Neurophysiology*, 97(6), 4310–4326.

Dakin, S. C., & Bex, P. J. (2001). Local and global visual grouping: Tuning for spatial frequency and contrast. *Journal of Vision*, 1(2), 99–111.

Dakin, S. C., & Bex, P. J. (2002). Role of synchrony in contour binding: Some transient doubts sustained. *Journal of the Optical Society of America A*, 19, 678–686.

Fahle, M. (1993). Figure-ground discrimination from temporal information. *Proceedings Biological Sciences/The Royal Society*, 254(1341), 199–203.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local association field. *Vision Research*, 33(2), 173–193.

Forté, J., Hogben, J. H., & Ross, J. (1999). Spatial limitations of temporal segmentation. *Vision Research*, 39(24), 4052–4061.

Francis, G. (2009). Cortical dynamics of figure-ground segmentation: Shine-through. *Vision Research*, 49(1), 140–163.

Glass, L. (1969). Moire effect from random dots. *Nature*, 223, 578–580.

Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration: Still alive and well. *Neuron*, 24(1).

Guttman, S. E., Gilroy, L. A., & Blake, R. (2007). Spatial grouping in human vision: Temporal structure trumps temporal synchrony. *Vision Research*, 47(2), 219–230.

Hancock, P. J. B., Walton, L., Mitchell, G., Plenderleith, Y., & Phillips, W. A. (2008). Segregation by onset asynchrony. *Journal of Vision*, 8(7), 1–21. <<http://journalofvision.org/8/7/21/>>. doi:10.1167/8.7.21.

Hess, R. F., Beaudot, W. H., & Mullen, K. T. (2001). Dynamics of contour integration. *Vision Research*, 41(8), 1023–1037.

Hess, R. F., & Field, D. J. (1999). Integration of contours: New insights. *Trends in Cognitive Sciences*, 3(12), 480–486.

Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience*, 4(2), 127–128.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195(1), 215–243.

Kandil, F. I., & Fahle, M. (2001). Purely temporal figure-ground segregation. *The European Journal of Neuroscience*, 13(10), 2004–2008.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron*, 15, 843–856.

Kiper, D. C., Gegenfurtner, K. R., & Movshon, J. A. (1996). Cortical oscillatory responses do not affect visual segmentation. *Vision Research*, 36(4), 539–544.

Kovacs, I. (1996). Gestalten of today: Early processing of visual contours and surfaces. *Behavioural Brain Research*, 82(1), 1–11.

Kovács, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences of the United States of America*, 90(16), 7495–7497.

Lachapelle, J., McKerral, M., Jauffret, C., & Bach, M. (2008). Temporal resolution of orientation-defined texture segregation: A VEP study. *Documenta Ophthalmologica*, 117(2), 155–162.

Lamy, D., Segal, H., & Ruderman, L. (2006). Grouping does not require attention. *Perception and Psychophysics*, 68, 17–31.

Lee, S.-H., & Blake, R. (1999). Visual form created solely from temporal structure. *Science*, 284(5417), 1165–1168.

Leonards, U., Singer, W., & Fahle, M. (1996). The influence of temporal phase differences on texture segmentation. *Vision Research*, 36(17), 2689–2697.

Li, W., Piéch, V., & Gilbert, C. D. (2006). Contour saliency in primary visual cortex. *Neuron*, 50(6), 951–962.

Li, W., Piéch, V., & Gilbert, C. (2008). Learning to link visual contours. *Neuron*, 57(3), 442–451.

Motoyoshi, I. (2004). The role of spatial interactions in perceptual synchrony. *Journal of Vision*, 4(5), 352–361. <<http://journalofvision.org/4/5/1/>>. doi:10.1167/4.5.1.

Neri, P., & Levi, D. M. (2007). Temporal dynamics of figure-ground segregation in human vision. *Journal of Neurophysiology*, 97(1), 951–957.

Niebur, E., & Koch, C. (1994). A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *Journal of Computational Neuroscience*, 1(1–2), 141–158.

Parton, A., Donnelly, N., & Usher, M. (2001). The effects of temporal synchrony on perceptual grouping in the row-column ambiguous figure. *Visual Cognition*, 8, 637–654.

Parton, A., Donner, T. H., Donnelly, N., & Usher, M. (2006). Perceptual grouping based on temporal structure: Impact of subliminal flicker and visual transients. *Visual Cognition*, 13, 481–502.

Polat, U. (1999). Functional architecture of long-range perceptual interactions. *Spatial Vision*, 12(2), 143–162.

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., & Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, 391, 580–584.

Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research*, 33(7), 993–999.

Polat, U., & Sagi, D. (1994). Spatial interactions in human vision: From near to far via experience-dependent cascades of connections. *Proceedings of the National Academy of Sciences of the United States of America*, 91(4), 1206–1209.

Polat, U., & Sagi, D. (2006). Temporal asymmetry of collinear lateral interactions. *Vision Research*, 46(6–7), 953–960.

Qiu, F. T., Sugihara, T., & von der Heydt, R. (2007). Figure-ground mechanisms provide structure for selective attention. *Nature Neuroscience*, 10(11), 1492–1499.

Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron*, 24(1).

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24(1), 49–65.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, 18(1), 555–586.

Smith, M. A., Bair, W., & Movshon, J. A. (2002). Signals in macaque striate cortical neurons that support the perception of Glass patterns. *Journal of Neuroscience*, 22, 8334–8345.

Smith, M. A., Kohn, A., & Movshon, J. A. (2007). Glass pattern responses in macaque V2 neurons. *Journal of Vision*, 7(3), 1–15. <<http://journalofvision.org/7/3/5/>>. doi:10.1167/7.3.5.

Sterkin, A., Sterkin, A., & Polat, U. (2008). Response similarity as a basis for perceptual binding. *Journal of Vision*, 8(7), 1–12. <<http://journalofvision.org/8/7/17/>>. doi:10.1167/8.7.17.

Usher, M., & Donnelly, N. (1998). Visual synchrony affects binding and segmentation in perception. *Nature*, 394(6689), 179–182.

Vidal, J. R., Chaumon, M., O'Regan, K. J., & Tallon-Baudry, C. (2006). Visual grouping and the focusing of attention induce gamma-band oscillations at different

936 frequencies in human magnetoencephalogram signals. *Journal of Cognitive*  
937 *Neuroscience*, 18(11), 1850–1862.  
938 von der Malsburg, C. (1981). *The correlation theory of brain function (Internal report*  
939 *no. 81–2)*. Göttingen, Germany: Max-Planck-Institute for Biophysical Chemistry,  
940 Department of Neurology.  
941 von der Malsburg, C. (1999). The what and why of binding: The modeler's  
942 perspective. *Neuron*, 24(1).

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting,  
sampling, and goodness of fit. *Perception and Psychophysics*, 63(8),  
1293–1313.  
Yen, S. C., & Finkel, L. H. (1998). Extraction of perceptually salient contours by striate  
cortical networks. *Vision Research*, 38(5), 719–741.  
Zhaoping, L. (2005). Border ownership from intracortical interactions in visual area  
V2. *Neuron*, 47(1), 143–153.

943  
944  
945  
946  
947  
948  
949  
950

UNCORRECTED PROOF