

Perceptual grouping based on temporal structure: Impact of subliminal flicker and visual transients

Andrew Parton

*Department of Visual Neuroscience, Division of Neuroscience and
Psychological Medicine, Imperial College London, and Institute of Cognitive
Neuroscience, UCL, London, UK*

Tobias H. Donner

*Department of Psychology, Birkbeck College, London, UK, and Institute of
Neurophysiology and Pathophysiology, Centre for Experimental Medicine,
University Hospital Eppendorf, Hamburg, Germany*

Nick Donnelly

Department of Psychology, University of Southampton, UK

Marius Usher

Department of Psychology, Birkbeck College, London, UK

In two experiments we examined the perceived grouping of grids of equidistant dots, which are rapidly modulated over time so that alternate rows or columns are presented out of phase. In Experiment 1, we report that observers were able to group the grids consistent with the temporal modulation reliably, even at contrasts/frequencies for which flicker was not detectable. Moreover, flicker thresholds decreased with stimulus duration, whilst grouping thresholds did not change. In Experiment 2, we examined the impact of visual transients, by measuring performance when, either a mask or a contrast ramp was presented before and after the stimulus. Performance dropped substantially for both conditions, but remained significantly above chance. The results are discussed in relation to the role of temporal correlations in stimulus modulations and visual transients in grouping.

Please address all correspondence to: Dr. A. Parton, Department of Visual Neuroscience, Division of Neuroscience and Psychological Medicine, Imperial College London, Charing Cross Hospital Campus, London W6 8RF, UK. Email: a.parton@imperial.ac.uk

A. Parton and T.H. Donner have contributed equally and are jointly considered first author.

The grouping and segregation of elements in a visual scene is achieved on the basis of similarities or differences in a variety of stimulus properties, which have been characterized by Gestalt principles of perceptual organization (Wertheimer, 1923). Traditionally, these Gestalt principles describe properties that are readily apparent in the display. For example, factors such as element similarity and good continuity (spatial factors) and common fate in motion (a spatio-temporal factor) involve clearly visible properties that affect the grouping of elements within a scene. In recent years much research has focused on the influence of fast temporal modulations, which are not necessarily visible, on the grouping and segregation of elements in a visual scene (Alais, Blake, & Lee, 1998; Elliott & Müller, 1998; Fahle, 1993; Fahle & Koch, 1995; Forte, Hogben, & Ross, 1999; Kandil & Fahle, 2001; Kiper, Gegenfurtner, & Movshon, 1996; Lee & Blake, 1999a, 1999b; Leonards, Singer, & Fahle, 1996; Morgan & Castet, 2002; Parton, Donnelly, & Usher, 2001; Sekuler & Bennett, 2001; Suzuki & Grabowecky, 2002; Usher & Donnelly, 1998).

Many of these studies were motivated by a theory advanced by von der Malsburg (1981, 1999) according to which the binding of features belonging to one object is represented by synchronous neural discharges. This scheme requires (a) an intrinsic cortical mechanism, which synchronizes neural responses according to the perceptual Gestalt rules, and (b) that this synchronized activity can be “read out” by a processing stage, at which a decision about perceptual groups are formed. There is physiological evidence for both requirements being satisfied (Castelo-Branco, Goebel, Neuenschwander, & Singer, 2000; Singer & Gray, 1995), which is, however, controversial (Lamme & Spekreijse, 1998; Shadlen & Movshon, 1999). Although psychophysical studies can, by definition, not assess the mechanisms of internal response synchronization in the cortex, they can test requirement (b) of the binding-by-synchrony theory. That is, they can test the hypothesis that response synchronization, however generated, can be used by the brain for the grouping features into wholes.

The basic rationale is as follows (Fahle, 1993; Kiper et al., 1996): The features (luminance, contrast, orientation, or motion direction) of some elements of a visual display are modulated in temporal synchrony, and asynchronously to other elements, typically at a high rate. It is well documented that neurons in primary visual cortex of man and monkey follow such external temporal modulations up to rates of around 100 Hz (e.g., Williams, Mechler, Gordon, Shapley, & Hawken, 2004). If a coding based on internal neural synchronization is used for binding elements into objects, then such induced synchronous temporal modulations should facilitate grouping. However, it can be objected that the sensitivity to temporal grouping reported in previous studies can be explained by mechanisms that are neutral to the temporal correlation theory of feature binding, e.g., the results might be attributable to the observers’ ability to identify the spatial configuration contained within a single frame of the display

(Farid, 2002). Therefore a stronger support for the binding-by-synchrony hypothesis would obtain, if it was found that the grouping mechanism is sensitive to temporal modulations, which are too fast to be consciously detectable.

A further motivation for the previous studies of grouping on the basis of stimulus temporal synchrony is to examine the limits of temporal sensitivity that human observers possess in their perceptual organization of visual input and to extend the range of grouping properties beyond the “classical” ones. Studies by Blake and colleagues (Alais et al., 1998; Lee & Blake, 1999a, 1999b), for example, have revealed a number of temporal properties (such as fast, correlated changes of contrast and motion direction) that can influence grouping. It has been argued that grouping on the basis of such properties may be thought of as a “natural extension” of the classical Gestalt rule of “common fate” (see, for example, Lee & Blake, 1999b; Sekuler & Bennett, 2001; Shadlen & Movshon, 1999). However, the grouping characteristics described in these studies, as with those outlined by Gestalt psychologists, are based on temporal properties that are readily apparent when viewing the stimuli. This is not necessarily true for the temporal modulation of stimuli that flicker at high temporal frequencies, as observers may not be able to perceive any change in the stimuli. In fact, a number of studies have claimed that the rapid flicker of their stimuli was not visible. However, we are not aware of any study, which objectively assessed observers’ ability to detect the presence of flicker.¹ Our first aim was to examine if grouping is influenced by temporally correlated changes in stimulus properties even when those changes are not detectable.

To do this, the current experiments examine temporal grouping in the context of a stimulus classically used to examine perceptual grouping: A lattice of circular elements, which may appear to group into a series of rows or columns depending on the ratio of their horizontal/vertical interelement distances (Benav, Sagi, & Braun, 1992; Wertheimer, 1923). Usher and Donnelly (1998) introduced the temporal variant of this grouping stimulus: here, all elements are equidistant but flicker rapidly (rate >30 Hz) in different temporal configurations. In temporal cue conditions, either alternate (a) rows or (b) columns are presented in separate successive stimulus frames each displayed for ~16 ms. Hence, a complete grid was defined over two stimulus frames. In a no cue condition all elements were presented simultaneously and alternated with a blank frame (see General Method section and Figure 1 for details). Usher and Donnelly found that observers grouped the elements in accordance with the temporal manipulation although they reported the stimulus appeared stable, and that their results could not be explained by mechanisms sensitive to apparent motion.

¹But see, Elliott and Müller (1998), who assessed the conscious detectability of their “synchronous premask” in psychophysical control experiments, that is, although there was visible stochastic flicker, observers were unable to consciously discern synchronous from random premask displays.

In the current paper we examined if grouping on the basis of temporal structure can be performed with such stimuli even when the modulation of individual elements is not visible. We compared directly the grouping performance with temporally modulated grids of dots and the ability to detect the flicker of the elements within the matched grids, using a criterion-free measure of psychophysical performance derived from signal detection theory (Macmillan & Creelman, 1991). In our first experiments, we tested whether observers could accurately classify the rows–columns stimuli even when they were unable to detect the flicker. To further determine the relationship between the mechanisms underlying flicker detection and grouping of elements to rows or columns, we characterized and compared these mechanisms along the dimensions of contrast, frequency, and duration. The effect of stimulus duration on performance within these two tasks was of particular interest as Usher and Donnelly (1998) found that performance in the rows–columns task decreased with presentation time; by contrast, sensitivity in flicker detection has been shown to increase with presentation time (van der Wildt & Rijdsdijk, 1979) as typically observed in low level vision tasks. This increase is likely due to probability summation over time (Watson, 1979).

A secondary aim of the current studies was to examine whether transient signals generated by the initial onset and final offset of the stimulus are sufficient to account for observers' performance in this task. Dakin and Bex (2002) have shown that the ability of observers to use the temporal structure in order to detect a target made of elements arranged on a path among a texture of random elements is much diminished when the visual transients are eliminated by masking or contrast ramping. Dakin and Bex concluded that, since the effect of synchrony modulations on performance is triggered so rapidly (i.e., at transients), this “would seem to be problematic for an account of contour binding based on synchrony” (p. 684). They proposed, instead, that the influence of temporal modulations on observers' performance in that task may be largely based on the output of a neural mechanism that possesses a large receptive fields (encompassing two or more elements) with narrow orientation bandwidth and is particularly sensitive to the initial transient onset or final offset of the stimulus.² Although, such a mechanism needs to be sensitive to the synchrony of the elements during the first two stimulus frames, predicting that masking the stimulus onset/offset (or minimizing the size of the contrast change generated by the initial/final stimulus frames) will abolish the grouping effect. Here we follow Dakin and Bex in testing the impact of pre/post masking and of contrast ramping on temporal grouping, to examine if such a mechanism could explain the results in the task described above.

²A related explanation is based on priming by the first stimulus frame (Beaudot, 2002).

In summary, the current experiments assess whether observers can use correlated temporal modulations of stimulus elements to judge the spatial structure of that stimulus even when the temporal modulation in the stimulus is not detectable. They then assess the degree to which such judgements of spatial structure are based upon the transients generated by the first and final stimulus frames. The results of the studies will be discussed in relation to their general implications for grouping and with reference to the debate surrounding von der Malsburg's (1999) proposal regarding the role of internal neural synchrony codes in visual grouping.

GENERAL METHOD

Observers

Two observers participated in Experiment 1a and five observers took part in Experiment 1b. Three observers in Experiment 1b were naïve to the hypothesis and had not previously participated in a psychophysical experiment. All observers had normal, or corrected to normal, visual acuity. Five observers took part in Experiment 2a and three subsequently participated within Experiment 2b. All had normal, or corrected to normal, visual acuity. Two (ED and SP) observers were naïve to the experimental procedure.

Stimuli and apparatus

The experimental stimuli were generated and presented on a Silicon Graphics O2 workstation with a 21-inch monitor with a screen refresh rate set to either 60 or 75 Hz (the temporal reliability of the system for 60 and 75 Hz modulations was verified with an oscilloscope). The spatial components of each of the stimuli were fully specified by two successive screen frames, defined using the SGI screen buffering facility, and hence one complete stimulus cycle was displayed at an effective rate of 30 (Experiment 1b) and 37.5 Hz (Experiment 1a). The stimuli were presented against a darker grey background (8.7 cd/m²) and their contrast varied between experimental conditions and observers (from 4.4% to 30.1%).

The stimuli used were 8 × 8 symmetrical grids of circular elements viewed using a chinrest positioned 70 cm from the screen at which distance 1 pixel subtended 1.2 arc min of visual angle. Each element had a radius of 9.6 arc min with its midpoint separated from the neighbouring elements by 1 degree. The stimulus contrast was calculated using the formula:

$$(L_{\max} - L_{\min}) / (L_{\max} + L_{\min}) \quad (1)$$

Two spatiotemporal configurations were used for each task: (a) in the rows–columns task either alternate (i) columns or (ii) rows were oscillated in counter

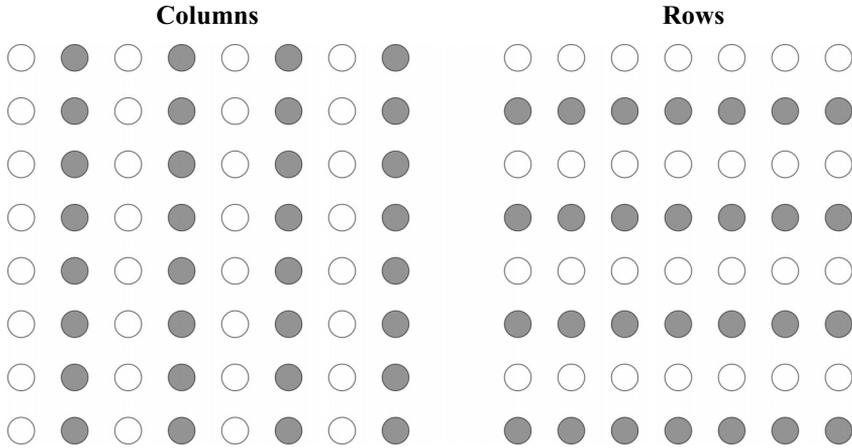


Figure 1. A diagram of the stimuli used in Experiment 1 for the rows-columns classification task. For purposes of illustration the temporal structure is indicated by dot colour, i.e., white dots represent the elements in frame 1 and grey dots those in frame 2. However, in the actual experiment elements in both frames were the same colour, i.e., white against a grey background. The stimuli are for illustrative purposes and are not to scale, i.e., the gaps between dots were larger than their diameters.

phase (see Figure 1); and (b) in the flicker detection task either (iii) the whole grid oscillating in phase or (iv) a stable nonflickering grid was presented.

A brief calibration experiment was run prior to the main experiments. Each observer was asked to adjust the contrast level of a stable (nonflickering) configuration of dots until they matched the perceived contrast to that of an adjacent flickering stimulus (using the same flicker rates employed in the experiments). They made 10 settings for 10 contrast levels. A quadratic expression was fitted to the resulting data ($r_2 = .998$) and it was used to construct a lookup table to match the contrast of the stable grid to the flickering grid during the course of the Experiment.

Psychophysical performance

All tasks involved the classification of a single stimulus display into one of two possible alternatives. The grouping task required discrimination between rows and columns stimuli, while the flicker task required discrimination between flickering and static stimuli.³ Therefore both tasks required a judgement on different stimulus dimensions. We employed a one-interval, rather than a two-

³ However, it should be noted the nature of CRT displays means the static stimulus was temporally modulated at the screen refresh rate, which was twice the frequency of the flickering stimulus.

interval discrimination paradigm for both tasks, because percepts in the rows–columns task are likely to be correlated between sequential stimulus intervals, due to hysteresis effects. Such correlations may then lead to an underestimation of the true sensitivity. Despite the formal symmetry in the method used for the two tasks, differences in decision criteria may be expected. In order to generate a criterion-free measure of sensitivity for all tasks, correct and incorrect responses were counted separately for each stimulus alternative and two fractions of correct responses, f_A and f_B , were obtained for each block of trials. The fractions f_A (“hits”) and $1 - f_B$ (“false alarms”) were combined to yield a criterion-free measure of sensitivity p_{\max} , expressed in units of proportion correct (Macmillan & Creelman, 1991):⁴

$$p_{\max} = 100 * F[0.5 (F^{-1}(f_A) - F^{-1}(1 - f_B))] \quad (2)$$

where F denotes the normal distribution and F^{-1} its inverse. Psychophysical performance is expressed as p_{\max} in all experiments, except for Experiment 1b.

EXPERIMENT 1: COMPARING FLICKER DETECTION AND GROUPING

In Experiment 1a we compared the relationship between observers’ ability to perform the rows–columns classification task and a flicker detection task at a range of stimulus contrasts levels. Specifically, we were interested in whether observers could reliably perform the row–columns task when performing at chance level in the flicker detection task. The aim of Experiment 1b was to examine the relationship between observers’ ability to perform the rows–columns classification task and a flicker detection task across a range of different presentation times. This replicated and extended the results of Experiment 1a by examining if there was evidence for different mechanisms underlying performance in the two tasks. Previous studies have reported that flicker detection thresholds improve as the stimulus presentation time increases (van der Wildt & Rijdsdijk, 1979). This is a typical finding across a range of visual properties and presumably reflects probability summation over time in the detection process (Watson, 1979). In contrast, Usher and Donnelly (1998) reported that observers’ thresholds in the rows–columns classification task were inversely related to the presentation time. If the two tasks behave differently with respect to time this would add to the case for different underlying mechanisms.

⁴ p_{\max} is a transform of the well-known discriminability index d' of signal detection theory, which equals the term in the inner brackets of equation 1.

Method

Design and procedure. In Experiment 1a, observers' performed two 2AFC tasks classifying the stimuli on the basis of their temporal characteristics using the left and right mouse buttons to indicate their decision. The tasks required the observers to discriminate between either (a) a rows condition (alternate rows of elements were presented 180° out of phase) and a columns condition (alternate columns of elements were presented 180° out of phase) or (b) a flicker condition (all elements oscillated in phase) and a static condition (all elements were displayed for the entire presentation time). The order of presentation of the tasks was counterbalanced across observers. Before commencing the experiment both observers performed a series of practice blocks for both tasks to familiarize them with the experimental procedure and to determine the four contrast levels at which they performed the task during the experiment. These contrast levels were selected such that observers demonstrated different levels of accuracy but their performance was neither perfect nor entirely at chance within the rows–columns task for stimuli. We performed both tasks at a higher temporal frequency (37.5 Hz) than used within earlier studies (Parton et al., 2001; Usher & Donnelly, 1998) to maximize the dynamic range at which observers were performing at chance in the flicker detection task. The same contrast values were used within both tasks and presentation rates so that performance could be directly compared. Observer AP performed 4 blocks and TD performed 10 blocks of trials. Each block contained 80 trials comprised of 10 trials of all stimulus/contrast combinations. The total presentation time of each trial was 106 ms during which time four stimulus cycles (or eight screen frames) were presented.

In Experiment 1b, thresholds were measured using three interleaved staircases that converged to different points on the psychometric function (50%, 70%, and 84%) as this was considered the most suitable method for use with psychophysically naïve observers. Each staircase terminated after 12 reversals and the thresholds were calculated from the last eight reversal points. Two sets of staircases were completed and averaged together at each of the three stimulus durations (100, 133, and 200 ms). Both tasks were performed at 30 Hz. Comparisons were made using data from the 70% point although results were qualitatively similar across thresholds.

Results

Figure 2 illustrates the results of Experiment 1a for both of the observers within the rows–columns and flicker detection tasks performed at the four tested contrast levels. The error bars in the graphs represent the 95% confidence limits and are used as the basis for comparing the observers' performance across the experimental conditions. At all of the tested stimulus contrast levels, TD failed to reliably detect the presence of flicker but he was, nonetheless, still able to

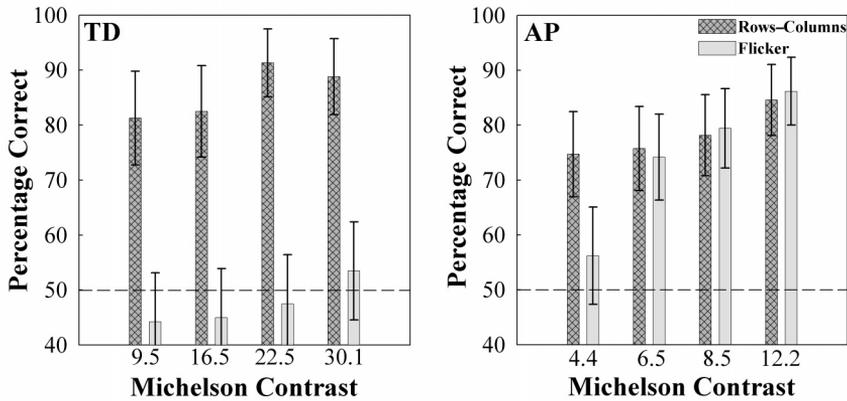


Figure 2. Illustrates the performance of the two observers within the rows-columns grouping task and the flicker detection task. The error bars represent the 95% confidence limits and the dashed line indicates chance (50%) performance.

perform the row-columns grouping task extremely well (from 79% to 91% correct). Similarly, when observer AP performed at chance levels in flicker detection (contrast level 4.4) he was still able to discriminate between the rows-columns stimuli (74% correct). Therefore, the results indicate that performance within the rows-columns task is not dependent on the explicit detection of flicker.

Additionally, the results also indicate that performance within each task demonstrates a monotonic increase with contrast.⁵ To explore the dependence of task sensitivity on contrast more completely, psychometric functions relating contrast level and performance in the flicker detection and rows-columns task were derived for observer TD. Data for this purpose were collected using the same procedure except a different range of contrast values were employed within each of the two tasks (flicker range, 10.4–22.6; rows-columns, range 2.3–16) allowing observations to be made across their full dynamic range. Figure 3 depicts TD’s correct responses at all contrast levels within each task collapsed across stimulus type (generating 160 observations at each data point). The data clearly show for both tasks that performance increases monotonically with contrast and that varying contrast can be used to generate psychometric functions (and thus thresholds). This validates the use of contrast variations in a staircase procedure in order to determine efficiently thresholds in Experiment 1b (see below). Additionally, the results again demonstrate the consistent superiority of TD’s performance in the rows-columns grouping task compared to

⁵ This is to be expected because in both tasks the relevant signals are variations of contrast over time.

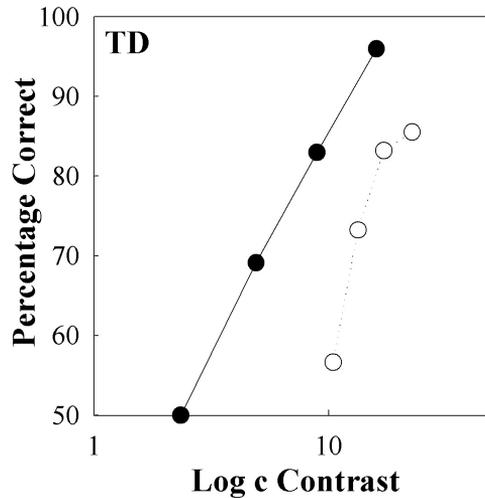


Figure 3. Illustrates the performance of observer TD in the flicker detection (white circles) and rows-columns (black circles) grouping across a range of contrast levels used to derive psychometric functions.

flicker detection. The 75% thresholds were 0.81 (0.75–0.87 95% CI) and 1.4 (1.37–1.44 95% CI) for the rows–columns and flicker detection tasks respectively, i.e., thresholds in both tasks differ significantly. Furthermore, as the gradient of the flicker function (3.4) is steeper than that for the rows–columns stimuli (1.9) the difference between the two tasks would increase if the threshold were defined by a lower criterion.

In Experiment 1b, we examined the relationship between presentation duration and performance within the two tasks for a group of observers. The results for all observers are presented in Figure 4 both individually and as a group mean. The data were analysed using a 2 (task type: Flicker vs. rows–columns) \times 3 (stimulus duration: 100 ms vs. 133 ms and 200 ms) ANOVA repeated over all factors. Observers demonstrated a greater sensitivity within the rows–columns classification task than the flicker detection task, $F(1, 4) = 28.64$, $p < .01$. There was no main effect of presentation time but there was a significant interaction with task type, $F(2, 6) = 15.6$, $p < .01$. Figure 4 indicates that sensitivity in the flicker task tends to increase with time (indicated by a negative slope in the sensitivity function), whilst in the rows–columns task sensitivity tends remain constant or decrease slightly (indicated by the flat or positively sloped sensitivity functions). Therefore, we attribute the interaction between presentation time and task type to an increasing convergence between performance in each task as the presentation time increases, Pages $L = 56$, $p < .01$. There was a mean difference in contrast of 14.8, 7.4, and 2 at the 100, 133, and 200 ms presentation times, respectively.

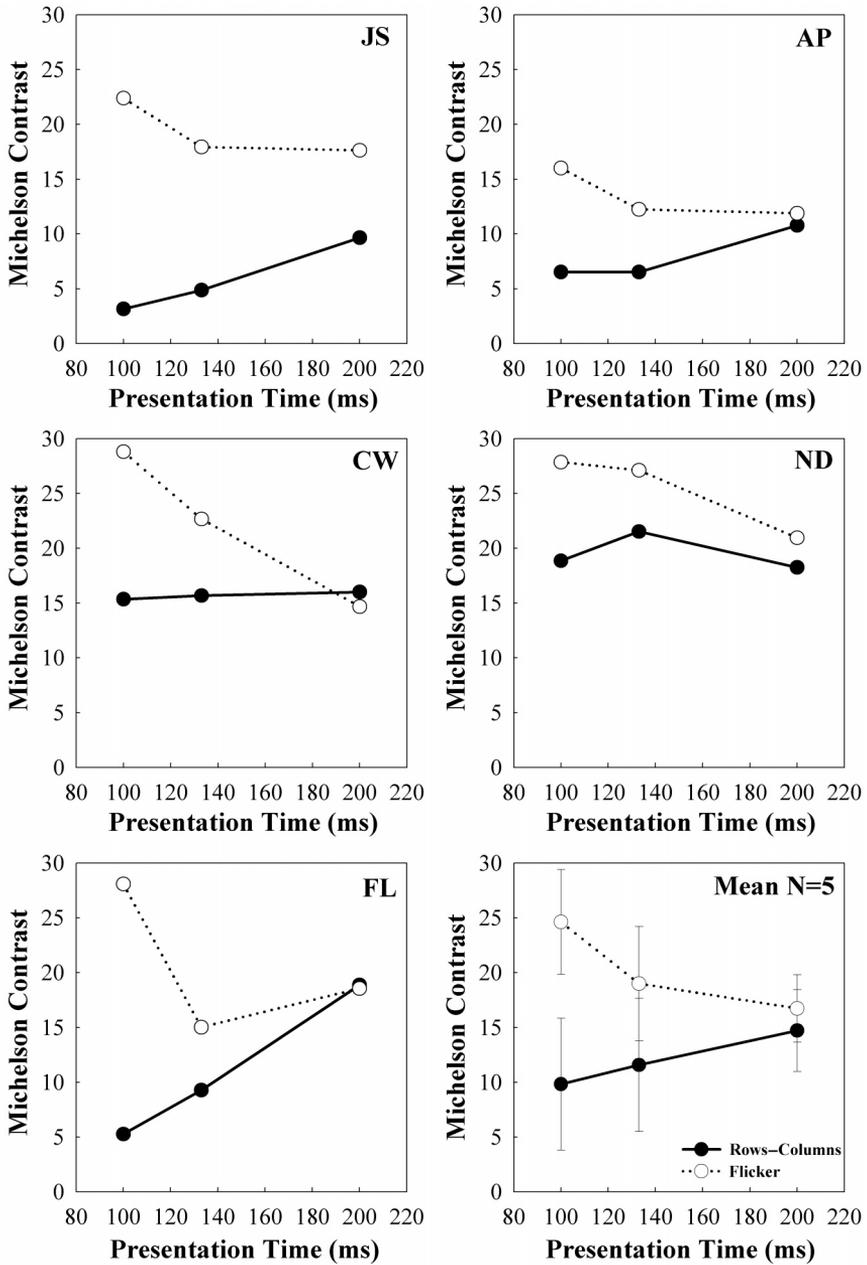


Figure 4. Illustrates the contrast levels at which each of the five observers achieved 70% correct performance within the rows-columns grouping task and the flicker detection task for three different presentation times. The final graph shows the mean for all observers with the error bars indicating 95% confidence limits.

In summary, the greater contrast sensitivity of rows–columns grouping than of flicker detection suggests that even the detection of the mere presence of a temporal modulation of the stimuli is not necessary in order to use this modulation efficiently for grouping. Moreover, the results also show that the performance in the two tasks exhibit a different relationship with presentation duration, which is indicative of differences in the underlying mechanisms used to derive the observers' decision for each type of task. This is a remarkable finding as the sensitivity for detecting and discriminating a wide range of simple visual properties increases over time as a consequence of probability summation (Watson, 1979).

EXPERIMENT 2: THE CONTRIBUTION OF VISUAL TRANSIENTS

Two recent studies have claimed that the enhanced performance for asynchronous flickering stimuli in figure–ground segregation tasks are attributable to the effects of the onset and/or offset stimulus frames (Beaudot, 2002; Dakin & Bex, 2002). Both of these studies assessed performance using a contour integration task that required observers to locate a path (a group of collinearly oriented elements) amongst a field of randomly ordered distractor elements (Field, Hayes, & Hess, 1993). In the temporal variant of this task reported by Usher and Donnelly (1998) performance was compared in synchronous (paths and distractors presented together) and asynchronous (paths and distractors presented in separate frames) conditions and they attributed the enhanced performance in the asynchronous conditions to the effects of external synchrony. Beaudot (2002) found the effect of asynchrony was greatly reduced in conditions where the distractors and not the path were presented in the first frame, and so attributed the results to a priming effect by the first stimulus frame.⁶ In another study, Dakin and Bex (2002) reported that the effect of stimulus asynchrony was removed (except at long presentation durations) when the transients during the stimulus onset and offsets were eliminated by pre- and postmasking with a phase randomized version of a typical contour stimulus or by contrast ramping. They argued that this indicates that the asynchrony effects are due to either visual persistence or the identification of transient structure from the initial onset or final offset of the stimulus. Dakin and Bex argue that this latter process could be achieved from the output of a single filter with a large orientation bandwidth, which is not modulated across stimulus cycles. If such mechanisms were the sole explanation of the current phenomenon then the effects should be abolished by

⁶Note that Beaudot's (2002) conclusions were based on the interpretation of a *nonsignificant* ANOVA with three observers and a small number of observations. Furthermore, Dakin and Bex (2002) did not find any effect of path/distractor order in a paradigm more closely matched to that of Usher and Donnelly (1998).

the elimination of large contrast changes generated by the onset and offset of the stimulus.

The aim of the following experiments was to test whether such explanations could account for the effects found in the rows–columns grouping tasks. This task was originally presented alongside the temporal path detection and the temporal grouping effects could in principle be accounted for by the same explanation. Therefore, in two experiments we assess the effects of minimizing the transients at the onset and offset of the stimulus using masking and contrast ramping.

Method

Design and procedure. The stimuli were the same as those used in the rows–columns task in Experiments 1a and b except for the following changes: In the mask condition, a two frame mask preceded the first rows–columns frame and followed the final frame. Masks were generated by randomly distributing all of the elements in the grid across two screens. Different onset and offset masks were generated for each trial.

Additionally, a small (10.8 arc min) fixation cross was displayed at the centre of the screen throughout the experiment. Observers performed two rows–columns tasks (masked and unmasked) at three contrast levels (4.4, 12.2, and 15.6%). The Experiment was run in 50 trial blocks for each task and contrast level with the order of presentation randomized. Three observers (AP, MU, and ED) performed three blocks in each condition, SP performed two blocks in each condition, and TD performed four blocks in each condition.

In the contrast ramping experiment instead of having an abrupt onset/offset the stimulus appears and disappears gradually as the contrast of alternate stimulus frames was slowly increased (onset) or decreased (offset). The onset was ramped by increasing the contrast of successive stimulus frames by one-tenth of the maximum contrast. Thus, the maximum stimulus contrast would be reached after the presentation of 10 frames (166 ms). The offset was ramped by gradually decreasing the contrast of successive frames in a similar manner. The offset was ramped by gradually decreasing the contrast of successive frames in a similar manner. Observers performed 300–400 trials at three different durations of maximum stimulus contrast (32, 64, and 128 ms). Two observers (TD and ED) performed the task at 15.6% contrast and AP performed it at 12.2% contrast.

Results

The results for masking are presented for each observer in Figure 5 both individually and as a group mean. The data were analysed using a 2 (task type: Masked vs. unmasked) \times 3 (contrast level: 5.5% vs. 13.1% and 16.5%) ANOVA repeated over all factors. Observers demonstrated a significantly greater sensitivity within the unmasked rows–columns classification task than the masked task, $F(1, 4) = 23.67$, $p < .01$. The increase in performance with

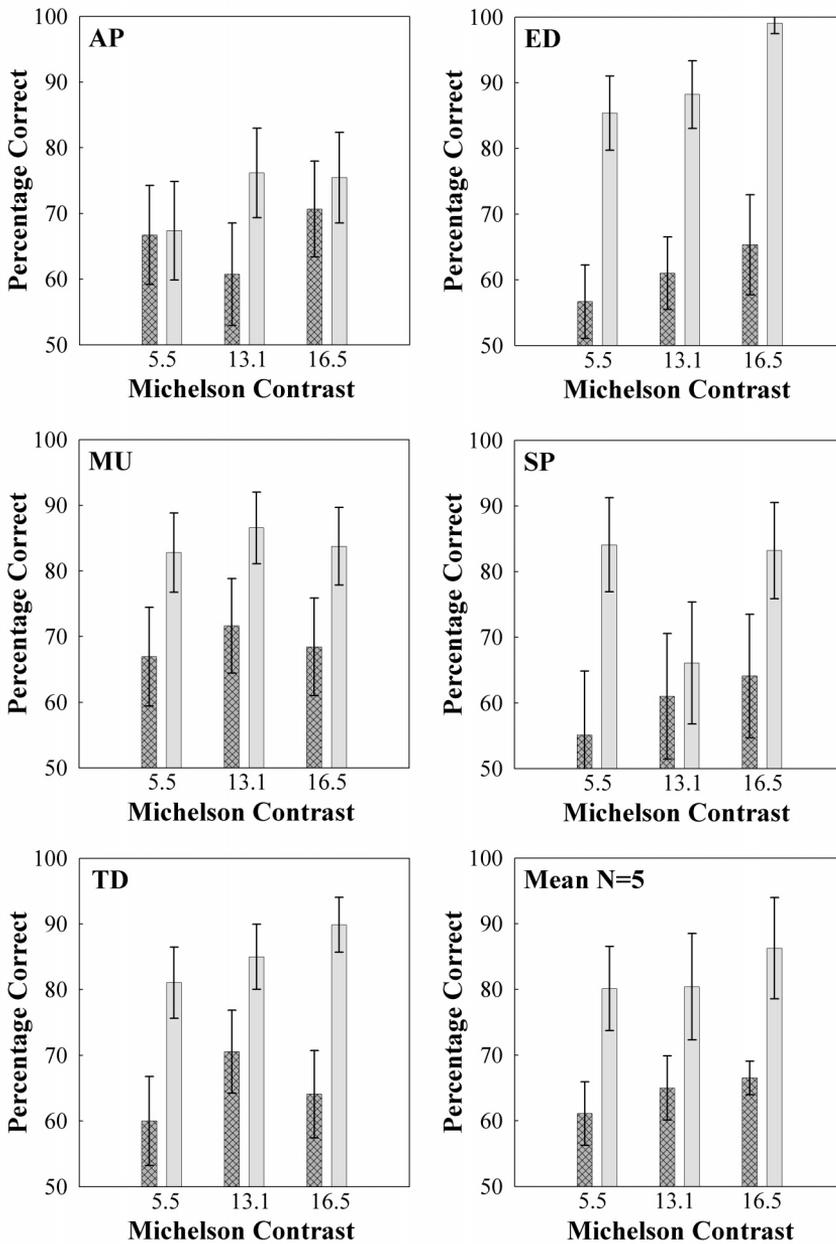


Figure 5. Illustrates the performance levels for the five observers within the masked (dark grey) and unmasked (pale grey) rows-columns grouping task at each contrast level. Error bars represent 95% confidence limits.

contrast level approached significance, $F(2, 8) = 3.74$, $p = .07$. The results clearly show that transients created by the onset and/or the offset of the stimuli, contribute to the observers' performance in the rows–columns classification task. However, even in masked stimulus condition all observers were still consistently able to achieve above chance performance within the task at all contrasts (denoted by their lower limits of the 95% confidence intervals lying above 50% point in the graph), the only exception being observer SP at the lowest contrast level. Consequently, the results of the masking experiment suggest the stimulus transients have an impact on, but are not the only factor determining performance within the rows–columns grouping task.

However, the strength of any conclusions is obviously dependent on the effectiveness of the mask and it is possible that the mask used in the current experiment itself generated a grouping cue in the first and last frame of the stimulus cycle. The second mask frame and the first rows–columns frame will contain a number of elements in common. As a consequence, these elements will be displayed for twice the duration of the other elements in the frame and their apparent contrast may be enhanced relative to other elements due to temporal integration. If two adjacent elements receive such a contrast enhancement then the observer could receive a cue as to the nature of the stimulus. Therefore, in the next experiment we further minimize the transients at the onset and offset by contrast ramping the stimulus on and off over 166 ms. The results for three observers are presented in Figure 6. They clearly indicate that contrast ramping does diminish performance but, nonetheless, observers can still perform the task at significantly above chance levels (denoted by the 95% confidence limits on the error bars). Only one observer (ED) performed at chance within one condition (64 ms). Observers' performance was at very similar levels for all three of the stimulus durations. This probably reflects an asymptotic level of performance that is attributable to the overall increase in stimulus duration caused by the contrast ramping.

In summary, both masking and ramping results indicate that, although the performance in the rows–columns task cannot be *entirely* explained by the effects of stimulus transients, such transients make a large contribution to the effects reported by Usher and Donnelly (1998). The origin of the small residual ability to perform this task in the absence of transients and its relation with evoked oscillatory responses during the plateau requires further investigations, as we discuss below.

GENERAL DISCUSSION

We have examined the impact of two temporal processes, the perception of flicker and the existence of visual transients, on the ability to group the elements of a symmetric grid (the rows–column classification task), according to the phase of high frequency temporal modulations. The results of the first experi-

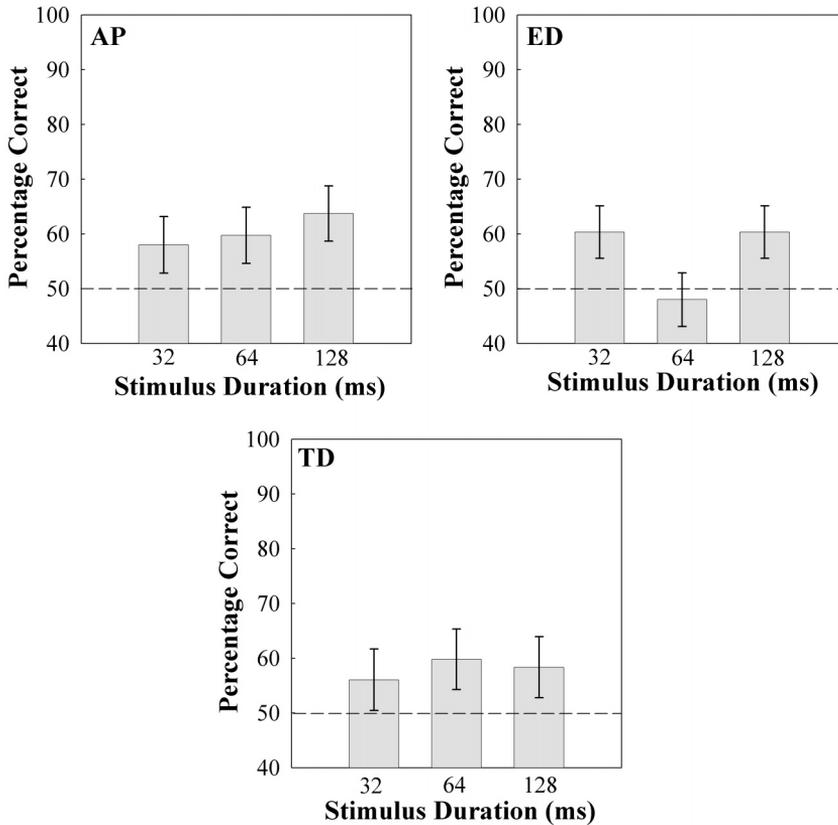


Figure 6. Illustrates the performance levels for the three observers within the ramped rows-columns grouping task at each stimulus duration. The error bars represent the 95% confidence limits and the dashed line indicates chance (50%) performance.

ment indicate that this ability prevails even under conditions where the observer cannot indicate the presence of the element flicker. Additionally, we found that the process of flicker detection dissociates from the rows-columns classification task, with regards to the impact of processing time. While flicker detection shows a clear improvement with processing time, as predicted by the probability summation over time, this is not the case for the rows-columns classification task, where a moderate decrease in performance is found over the range of 100–200 ms.

This dissociation indicates that the detection of flicker is not a prerequisite for performing the rows-columns classification task. Moreover, it implies that the observers do not perceive the alternation of the complementary stimulus frames. This, in turn, suggests that they cannot access single stimulus frames (which contain a salient static grouping cue; see Figure 1). This conclusion is also supported by previous experiments (Parton et al., 2001). Thus, it is unlikely that

the temporal grouping reported here simply reflects the ability to base the judgement about the element grouping on single stimulus frames (Farid, 2002). Moreover, the fact that the grouping of elements in these displays emerges even when the temporal property of flicker of individual elements is not detectable is interesting as it may indicate that unlike in other classical Gestalt cases of grouping, where a visual property of the individual elements (say, line orientations or motion direction) can be categorized by separate local detectors (say, V1 cells tuned to vertical/horizontal or MT cells tuned to a specific motion profile), in our stimuli the local detectors are less likely to categorize (with their spike rate) a distinguishing property. (In fact for our stimuli, the only distinguishing property is temporal phase, which requires a comparison between a minimum of two elements). This suggests that perhaps, in our stimuli, the temporal modulation is encoded only in the temporal profile of activation of the local detectors, which does not provide access to conscious report, but can effect the response and categorization of nonlocal detectors in higher areas (V4, etc.), that mediate the holistic perception of grouping into rows or columns.

The mechanism mediating this process was further examined in the second experiment on the impact of visual transients generated by the onset and offset of the stimuli on the perceived grouping of elements into rows or columns. Here, consistent with Dakin and Bex (2002), who examined this effect in another grouping task (contour integration), we found that the initial onset and/or final offset transients have a large impact on the rows–columns classification performance. This effect was most apparent when minimizing the transients using ramps. Nevertheless, even in this case we found a residual ability to perform the task, which suggests that mechanisms sensitive to these (*initial* and *final*) transients do not, *exclusively*, account for the performance in the task. This residual performance initially seems at odds with the results of Dakin and Bex's similar manipulation in the context of contour integration. However, there are important differences between the paradigms used in the current paper and those of Dakin and Bex. Specifically, one important difference is that to eliminate the possibility of a contribution from eye movements Dakin and Bex used a dual task procedure, which enabled them to identify trials where observers made eye movements.⁷ This possibility is supported by the results of their fifth experiment

⁷Dakin and Bex (2002) aimed at excluding trial contaminated with eye movements by positioning a Bessel patch rapidly flickering in counter phase at the centre of fixation. The Bessel patch is an area in which luminance is modulated according to a damped two-dimensional circular sinusoid. In simple terms it appears to be a series of alternately light and dark concentric circles. When two spatially counter phase patches are presented in consecutive alternation over time at the same location they are not visible. An eye movement shifts the retinal location of one patch relative to the other and it consequently becomes visible. Consequently, Dakin and Bex rejected trials in which observers detected the Bessel patch and concluded that the remainder of trials reflected temporal grouping performance in the absence of eye movements. Even if successful in eliminating the eye movement confound, this dual task procedure may have led to an underestimation of performance in the peripheral grouping task by consuming attentional resources for the central detection task.

in which they controlled for eye movements by stabilizing the retinal image of the stimuli and consequently only used the primary task. Both subjects now exhibited a small difference between temporally synchronous and asynchronous conditions even in the presence of contrast ramping (although it was only significant for subject PJB). Nonetheless, it is possible that the residual performance in our study may have been mediated by small eye movements, such as microsaccades. Such low amplitude ($<0.5^\circ$) gaze shifts are not detectable with common eye tracking devices, but they may generate strong transients in the retina, which may underlie the residual temporal grouping performance after masking/ramping. Clearly, further studies with high resolution eye tracking systems are required to definitively the effects of microsaccades in our temporal grouping paradigm directly. However, the results of Experiment 1b render this explanation unlikely. If grouping performance was mediated by eye movements then it should improve with a rising cumulative probability of the occurrence of an eye movement, and thus with rising presentation time. We found the opposite to be the case.

We now consider two potential mechanisms that might explain the observers' ability to make use of the phase of high frequency temporal modulations to perform the rows–columns classification task and the impact of visual transients. The first possibility is that performance is mediated by a mechanism with low orientation bandwidth and high transient selectivity (Dakin & Bex, 2002), i.e., orientation tuned neurons in higher areas (e.g., V4) with receptive fields large enough to encompass two (or more) of the elements in stimulus column or row. It is important to note that in order to account for the pattern of results found by Dakin and Bex and within our study such a mechanism must show predominant sensitivity to the transients caused by the initial stimulus onset (i.e., the first stimulus frame) and only negligible sensitivity to the repeated transients caused by the periodic stimulus modulation.⁸ Furthermore, note that in order to distinguish between a row and a column stimulus, the response of such a mechanism needs not only to be sensitive to transients at the onset of the stimulus (i.e., the initial frame) but it also needs to be sensitive to the temporal correlation between these transients. Assume, for example, that a V4 neuron of the type outlined above receives only input from cells lower in the visual hierarchy (V1/2), each producing spikes only (or mainly) at the initial onset of the corresponding element. If the V4 neuron responds by summing the input on a time scale of two temporal frames (26 ms) there could be no differential response of neurons tuned to vertical or horizontal orientations. What is needed to in order to distinguish among the two stimuli conditions is that detectors are

⁸ If grouping performance was mediated by a mechanism responding to the onset of each frame during the ongoing stimulus modulation, performance should (a) be hardly affected by stimulus masking and ramping and (b) increase with increasing stimulus duration, as was pointed out by an anonymous reviewer.

sensitivity (in their firing rate response) to the synchrony of the input on this time scale. Despite some counterarguments (Shadlen & Movshon, 1999), recent neurophysiological data and biophysically plausible models are consistent with the notion that cortical neurons are highly sensitive to the temporal correlation of their inputs (see Salinas & Sejnowski, 2001, for a review). This scenario shares with the binding-by-synchrony theory the essential property that cortical synchrony (on a scale of 10–20 ms) can affect further processing, and ultimately, behaviour (Fries, Reynolds, Rorie, & Desimone, 2001). However, such a scenario relies on single neurons ultimately representing the object, and hence, a sort of “grandmother cell”, the circumvention of which is in fact the major achievement of the binding-by-synchrony theory (von der Malsburg, 1981). The existence of such object-selective neurons may be plausible for simple “objects” as rows or columns and thus it may not represent a general solution to the combinatorial problem confronted by the visual system in natural scenes.

A second possibility is that the external synchrony of the visual elements is directly inducing the synchronization of neurons with small receptive fields responding to the individual elements (say, in V1), which in turn engage (online) some higher level units to bind with the subsets of synchronous elements; such units could then be reallocated to other bindings when the need arises (von der Malsburg, 1981, 1999). This provides a solution to the binding problem for more realistic, i.e., static, stimulus conditions with less regular object shapes, where internal synchrony is thought to be generated by lateral connections and to encode grouping. As outlined in the introduction, synchronized temporal modulations in stimuli are expected to facilitate grouping, according to this scheme. The precise nature of this facilitation and its time course may, however, depend on the interaction between internal synchronization mediated by lateral connections in response to Gestalt properties and stimulus induced synchronization. Detailed computational models are needed to examine, whether such a scheme also predicts a strong contribution of the initial transient in experiments of the types we discussed. This is likely to happen because only at transients may the external synchrony signal be robust enough to time lock with internally generated synchronous activation patterns.⁹

⁹ In order for the synchrony of the visual elements to facilitate grouping, according to the logic of the grouping-synchrony hypothesis, the internal synchronization generated by lateral connections needs to be boosted by the external modulation. Clearly, however, it is the transient that produces the strongest signal that can align the internal and the external temporal modulations, as cells responding to an initial change will exhibit a synchronized increase from a resting state which will be of a greater magnitude than the fluctuations in firing rates introduced by synchronous stimulus changes later in the stimulus presentation cycle. Input synchrony in later stimulus cycles, on the other hand, is less likely to boost the response synchrony due to its lower signal to noise ratio and to the fact that it may come in a random phase relative to the internally generated neural synchrony.

The results we report are consistent with both alternative mechanisms. On the one hand, as rows and columns can be discriminated by monitoring the output of a fixed unit (sensitive to the onset transient and the synchrony of their input) our data do not require the computation of novel/online binding, but could use preassigned detectors, which, however, should be sensitive to the synchrony of their inputs. On the other hand, a von der Malsburg-like binding-by-synchrony mechanism may also be especially sensitive to transients. Other data from paradigms, in which the grouped features cannot be picked by detectors with elongated receptive fields (Elliott & Müller, 1998, 2001; Suzuki & Grabowecky, 2002), provide additional support for a scheme of this type.

In conclusion, the data we presented shows that classifying grid stimuli according to the phase of fast temporal modulations is possible even when observers cannot detect the presence of the modulations and that this ability is to a large extent (but not completely) mediated by transients. This may substantiate the notion of either a novel Gestalt rule or the *extension* of “common fate”, and it indicates that the visual system in some way exploits (transient) synchronous responses (on a time scale of 10–20 ms) for perceptual organization. Although consistent with elements of the binding-by-synchrony theory, the results could be explained by several mechanisms. It is possible that a combination of psychophysics with neurophysiological techniques such as multiunit recordings in animals (Castelo-Branco et al., 2000) or EEG/MEG recordings in humans as well as formal neurocomputational modelling is required to distinguish among various mechanisms underlying these phenomena and answer the elusive question of how the visual cortex solves the binding problem.

REFERENCES

- Alais, D., Blake, R., & Lee, S. H. (1998). Visual features that vary together over time group together over space. *Nature Neuroscience*, *1*(2), 160–164.
- Beaudot, W. H. (2002). Role of onset asynchrony in contour integration. *Vision Research*, *42*(1), 1–9.
- Benav, M. B., Sagi, D., & Braun, J. (1992). Visual-attention and perceptual grouping. *Perception and Psychophysics*, *52*(3), 277–294.
- Castelo-Branco, M., Goebel, R., Neuenschwander, S., & Singer, W. (2000). Neural synchrony correlates with surface segregation rules. *Nature*, *405*(6787), 685–689.
- 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. Optics, Image Science, and Vision*, *19*(4), 678–686.
- Elliott, M. A., & Müller, H. J. (1998). Synchronous information presented in 40-Hz flicker enhances visual feature binding. *Psychological Science*, *9*(4), 277–283.
- Elliott, M. A., & Müller, H. J. (2001). Effects of stimulus synchrony on mechanisms of perceptual organization. *Visual Cognition*, *8*(3–5), 655–677.
- Fahle, M. (1993). Figure-ground discrimination from temporal information. *Proceedings of the Royal Society of London: Series B. Biological Sciences*, *254*(1341), 199–203.
- Fahle, M., & Koch, C. (1995). Spatial displacement, but not temporal asynchrony, destroys figural binding. *Vision Research*, *35*(4), 491–494.

- Farid, H. (2002). Temporal synchrony in perceptual grouping: A critique. *Trends in Cognitive Sciences*, 6(7), 284–288.
- 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.
- Forte, J., Hogben, J. H., & Ross, J. (1999). Spatial limitations of temporal segmentation. *Vision Research*, 39(24), 4052–4061.
- Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291(5508), 1560–1563.
- Kandil, F. I., & Fahle, M. (2001). Purely temporal figure–ground segregation. *European Journal of Neuroscience*, 13(10), 2004–2008.
- Kiper, D. C., Gegenfurtner, K. R., & Movshon, J. A. (1996). Cortical oscillatory responses do not affect visual segmentation. *Vision Research*, 36(4), 539–544.
- Lamme, V. A., & Spekreijse, H. (1998). Neuronal synchrony does not represent texture segregation. *Nature*, 396(6709), 362–366.
- Lee, S. H., & Blake, R. (1999a). Detection of temporal structure depends on spatial structure. *Vision Research*, 39(18), 3033–3048.
- Lee, S. H., & Blake, R. (1999b). 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.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user’s guide*. Cambridge, UK: Cambridge University Press.
- Morgan, M., & Castet, E. (2002). High temporal frequency synchrony is insufficient for perceptual grouping. *Proceedings of the Royal Society of London: Series B. Biological Sciences*, 269(1490), 513–516.
- Parton, A., Donnelly, N., & Usher, M. (2001). The effects of temporal synchrony on the perceived organization of elements in spatially symmetric and asymmetric grids. *Visual Cognition*, 8(3–5), 637–654.
- Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews: Neuroscience*, 2(8), 539–550.
- Sekuler, A. B., & Bennett, P. J. (2001). Generalized common fate: Grouping by common luminance changes. *Psychological Science*, 12(6), 437–444.
- Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron*, 24(1), 111–125.
- Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, 18, 555–586.
- Suzuki, S., & Grabowecky, M. (2002). Overlapping features can be parsed on the basis of rapid temporal cues that produce stable emergent percepts. *Vision Research*, 42(24), 2669–2692.
- Usher, M., & Donnelly, N. (1998). Visual synchrony affects binding and segmentation in perception. *Nature*, 394(6689), 179–182.
- Van der Wildt, G. J., & Rijdsdijk, J. P. (1979). Flicker sensitivity measured with intermittent stimuli: I. Influence of the stimulus duration on the flicker threshold. *Journal of the Optical Society of America*, 69(5), 660–665.
- Von der Malsburg, C. (1981). *The correlation theory of brain function* (Internal Report 81–2). Göttingen, Germany: Dept. of Neurobiology, Max-Planck-Institute for Biophysical Chemistry. Available online at <http://www.neuroinformatik.ruhr-uni-bochum.de/VDM/1981/index.html>
- Von der Malsburg, C. (1999). The what and why of binding: The modeler’s perspective. *Neuron*, 24(1), 95–104.
- Watson, A. B. (1979). Probability summation over time. *Vision Research*, 19(5), 515–522.
- Wertheimer, M. (1923). Untersuchungen zur Lehre der Gestalt. *Psychologische Forschung*, 4, 301–350.

Williams, P. E., Mechler, F., Gordon, J., Shapley, R., & Hawken, M. J. (2004). Entrainment to video displays in primary visual cortex of macaque and humans. *Journal of Neuroscience*, *24*(38), 8278–8288.

Manuscript received July 2004
Revised manuscript accepted February 2005