Perception & Psychophysics
1997, 59 (1), 142-147

Time course of perceptual grouping

DANIEL D. KURYLO
Bowdoin College, Brunswick, Maine

An investigation was made of the time course of perceptual grouping that is based on two qualitatively different spatial relationships: proximity and alignment. An index of grouping capacity was used to assess the processing time required before a backward pattern mask interfered with grouping. Stimuli consisted of bistable arrays of disjunct dots that were followed by a mask. Grouping cues, either proximity or alignment, were randomly assigned to either the horizontal or vertical orientation, and subjects indicated whether the dots appeared grouped as a series of horizontal or vertical lines. Spatial metrics of the cues were systematically altered until they no longer served as a cue for grouping, thereby determining the grouping threshold. The stimulus onset asynchrony (SOA) of the mask, relative to the test stimulus, ranged from 33.3 to 150 msec. The SOA at which grouping thresholds first became elevated identified the point at which the mask first interfered with the grouping process, thereby identifying the processing time required for grouping by the specified cue. The processing time for grouping by proximity and alignment differed significantly, requiring means of 87.6 and 118.8 msec, respectively, for processing to be completed. These measurements serve to identify the processing time necessary for spatially integrating stimulus elements into unified forms, thereby delineating temporal constraints at this stage of visual processing.

In order to organize complex stimuli in preparation for object recognition, the visual system quickly and automatically associates and segregates elements of the visual scene. To accommodate novel stimuli with speed and accuracy, the process of grouping likely follows established algorithms that mediate grouping strategies. These algorithms operate by extracting relevant information from the stimuli and then applying a set of criteria to that information which specify the grouping arrangement best suited for form identification. Although many of the characteristics associated with perceptual grouping have been described, the principles by which grouping functions, and the time course of its operation, are not well understood.

Grouping strategies are based on several characteristics of stimuli, including feature similarities, such as color or element shape, and spatial relationships, such as element proximity or regularity. The present study examined grouping that is based on two spatial relationships among stimulus elements: element proximity and element alignment. In each case, a single spatial parameter was systematically altered while observers made judgments about the manner in which elements appeared to be grouped. In this way, the association between a specific spatial metric and the process of grouping was isolated.

Grouping by element proximity and element alignment differ qualitatively in the nature of the spatial cue from which grouping is derived. Grouping by proximity (Rush, 1937) is based on the tendency for more proximal images to produce unified configurations, and for elements with greater separation to lose figural integration (Gillam & Grant, 1984). In this case, grouping processes are sensitive to distances between elements, and grouping assignment is based on the ratio of element separation (Gillam, 1981). Grouping by element alignment, which is an example of the Gestalt principle of good continuation (Rush, 1937), reflects the tendency to group elements that form regularities in the pattern. In this case, elements that can been seen as smooth continuations tend to be grouped together (Palmer, 1992), and the strength of grouping is related to a specified angle at which sequences of elements are juxtaposed (Prytulak, 1974). For the case of alignment, elements that cohere to a straight line are perceived as grouped, whereas elements that are randomly dispersed from a line have a reduced associative strength.

Grouping Threshold

A difficulty with studying the time scale of perceptual grouping is the need for an index that directly reflects the operation of grouping processes. Such an index is provided by the perceived grouping of multistable stimuli, in which elements may be grouped in one of several possible arrangements. The present study employs an array of dots that may be perceived as a series of either horizontal or vertical lines. A spatial cue for grouping, either proximity or alignment, was assigned to one of the two orientations, thereby biasing perceptual grouping along the cued orientation. The strength of the grouping cue is described in terms of stimulus metrics. Strong grouping cues correspond to stimuli in which elements along one orientation are highly proximal or highly aligned. Weak grouping cues correspond to stimuli in which the metrics along each orientation are similar. In order to measure the limit at which the cue serves in grouping, the strength of the cue was progressively diminished across trials until the noncued orientation was perceived as grouped. The point at which

I am grateful to Meredith van den Beemt and James Garner for their assistance in this study. Correspondence should be addressed to D. D. Kurylo, Psychology Department, Bowdoin College, Brunswick, ME 04011 (e-mail: dkurylo@polar.bowdoin.edu).

the noncued orientation was selected is referred to as the *grouping threshold.*

In a previous study (Kurylo, 1996), measurements were made of the limit to which arrays of dots could be resolved into perceptual groups. These grouping thresholds were derived from competitive grouping arrays, in which elements can be grouped into one of two possible arrangements. The assignment of grouping to one of the arrangements was based on the spatial relationships among elements. As the spatial metrics among possible arrangements approached equality, the displays became multistable and grouping assignment became inconsistent, reflecting the grouping threshold. For grouping by element proximity, the two possible arrangements differed in the amount of separation among dots. In this case, element proximity served as a cue for grouping only when the element separation for each arrangement exceeded 5.8%. At separation differences below this amount, stimuli became bistable. For grouping by alignment, the two possible arrangements differed in the degree to which they cohered to a straight line. In this case, alignment served as a grouping cue when the alternate arrangement was misaligned by 15.9% of the separation between elements. For misalignments less than this amount, stimuli became bistable. In both cases, grouping assignment was based on ratios of the isolated metric, which remained constant across the size scale of the stimulus array.

In the study reported here, these results were extended with an examination of the time course of grouping processes. To date, analyses of temporal factors associated with grouping have been based on either successive presentation of partial test stimuli, or visual masking. Oyama and Yamada (1978) sequentially presented partial stimuli that together formed a dot array in which elements could be grouped as a series of horizontal or vertical lines. Grouping selection remained consistent when the gap duration between the two stimuli was between 0 and 30 msec. For interstimulus gaps between 30 and 70 msec, grouping ability progressively decreased, as did judgments of simultaneity of stimulus presentation. These measurements correspond to factors of perceptual integration, particularly iconic memory, and therefore do not isolate processing times specific to perceptual grouping.

A more direct measurement of grouping processing time was made by Uttal (1969), who used a visual mask to interfere with the recognition of alphabetic characters. The characters were composed of dots that were disjunct in space, thereby requiring perceptual grouping in order for the letter to be identified. A random dot pattern, presented either before or following the test stimuli, was spatially superimposed on the character pattern. For interstimulus gaps (between the alphabetic characters and the mask) that ranged between 0 and 40 msec, the interference effect of the mask progressively decreased with gap size; that is, the percentage of trials in which characters were correctly identified progressively increased. For interstimulus gaps greater than 40 msec, the mask had little effect on the identification of the characters. It was found, however, that the mask did not completely interfere with subjects' ability to identify the characters, even with simultaneous presentation of the mask and the test stimulus.

Mask Interference Point

In the present study, the time course of grouping processes was measured by means of a backward pattern mask. The test stimulus (containing a grouping cue) was followed by a pattern mask, which disrupted the organization of the test stimulus. The time course of perceptual grouping was analyzed by measuring the point at which the pattern mask elevated the grouping thresholds of the test stimuli. The stimulus onset asynchrony (SOA) between the test stimulus and mask was systematically shortened until a rise in grouping threshold was observed. An elevation in grouping thresholds reflects interference by the mask with the operation of the grouping process, and will be referred to as the mask interference point.

Equivalent mask interference points for proximity and alignment cues would suggest that a common mechanism mediates grouping by each cue. Alternatively, differing mask interference points would indicate that the processing characteristics differ for each cue. Differences may stem from factors such as varied discriminability or attention allocation, or may occur because each cue is mediated by independent processes specifically sensitive to each spatial relationship. These predictions were tested by measuring grouping thresholds in the presence of either the proximity or the alignment cue alone.

METHOD

Subjects

Nine individuals who were experienced with the test procedure served as subjects. The subjects comprised undergraduate students paid for their participation, as well as the author. Each subject had participated in pilot testing of the procedure, during which their performance stabilized. All subjects had a best corrected 14-in. visual acuity of 20/20 (Snellen).

Apparatus

The stimuli appeared on a computer monitor (DEC PCXCV-GA) initialized to 640 × 480 pixel VGA mode. Stimulus presentation, data collection, and contingency algorithms were controlled by computer (DECpc LPv 433dx).

Stimuli

The test stimuli consisted of square arrays, 19.3° on a side, of white pixels (49.3 cd/m^2) on a dark background (0.046 cd/m^2). On each trial, the grouping cue (either greater proximity or greater alignment) was randomly assigned to either the horizontal or the vertical orientation.

Condition 1: Proximity cues. Elements were aligned and spaced regularly, although each orientation differed in the amount of element separation (Figure 1a). Element separation along the orientation containing the cue was set at 3.03°. Separation along the noncued orientation ranged from 150% to 100% of that of the orientation containing the cue, and was altered in increments of 5%. Metrics of the proximity cue are described in terms of a separation ratio, defined as the element separation along the noncued orientation relative to separation along the orientation containing the cue [(separation along

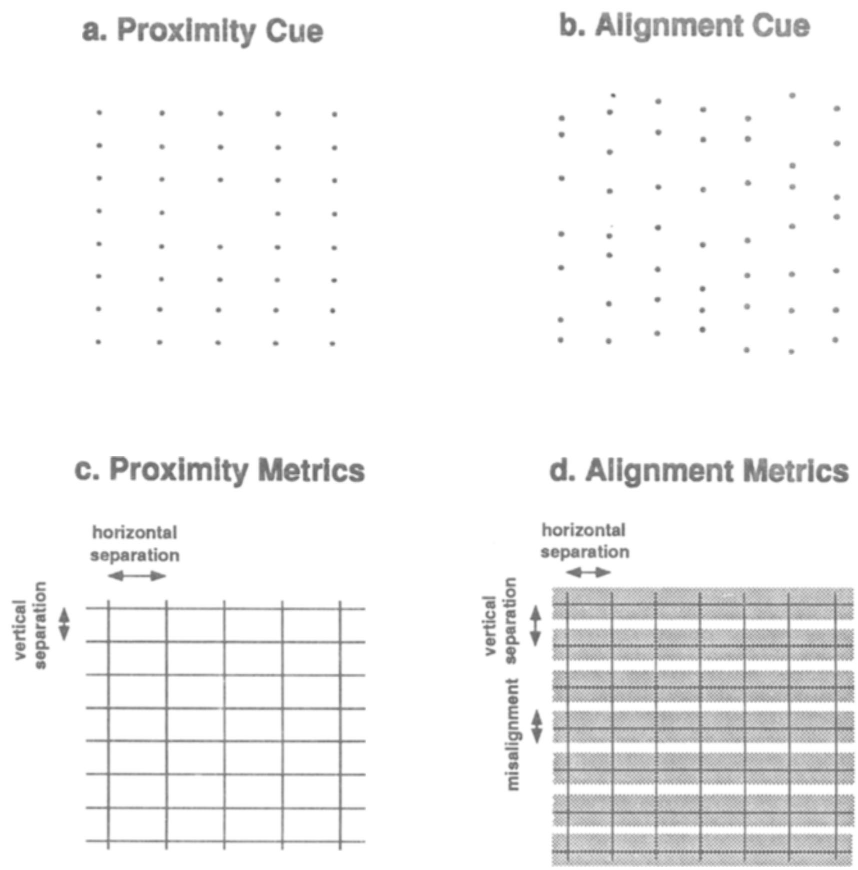

Figure 1. Examples of stimuli (a and b), and their associated spatial metrics (c and d). For each trial, grouping cues were randomly assigned to either the horizontal or the vertical orientation. For examples shown here, grouping cues are assigned to the vertical orientation. (a) Proximity cue: Elements along the horizontal orientation are more separated than those along the vertical. Proximity metrics (c) are defined in terms of the relative element separation of the cued and noncued orientation [(horizontal separation/vertical separation) − 1]. (b) Alignment cue: The mean horizontal and vertical separations are equivalent. Elements along the vertical orientation are perfectly aligned. Elements along the horizontal orientation are misaligned, and laterally displaced to fall within a dispersion field (d). The extent of dispersion field is represented by the shaded area. Alignment metrics (d) are defined in terms of misalignment (extent of dispersion field) relative to separation (misalignment/separation).

noncued orientation/separation along the cued orientation) − 1] (Figure 1c). For example, a grouping threshold of 10% indicates that elements along the noncued orientation were 10% more separated than those along the orientation containing the cue. A grouping threshold of 0% indicates that element proximity was equivalent between the two orientations, and therefore, that a grouping cue was not available.

Condition 2: Alignment cues. Both orientations were equivalent in separation ($M = 3.03°$), but they differed in the amount of misalignment from a straight line. Elements along the orientation containing the cue were perfectly aligned but irregularly staggered. Along the aligned orientation, the amount to which some pairs of elements became crowded was equal to the amount to which other pairs became widely separated, thereby producing a mean element separation that was equivalent to the misaligned orientation. Elements along the misaligned orientation were misaligned laterally and distributed randomly within a dispersion field (Figure 1b). The distance between the borders of the dispersion field ranged from 100% to 0% of element separation and was altered in increments of 10%. The metrics of alignment cues are described in terms of the extent of the dispersion field relative to the separation of elements (lateral dispersion field borders/element separation) (Figure 1d). For example, a grouping threshold of 20% indicates that the noncued orientation was dispersed laterally within a field that extended 20% of the separation between elements.

Disproportionately high acuity in the central visual field allows the possibility that grouping could be based on local cues, thereby presented a potentially confounding factor. Therefore, a 2.5° area centered in the array was devoid of illuminated pixels in order to produce greater uniformity in resolution across the stimulus array. Eliminating dots from the foveal viewing area precluded the possibility that responses were based on local cues derived from adjacent elements in the central area; that is, the distance between a pair of adjacent elements along the vertical orientation was compared with the distance between a pair oriented horizontally. This procedure eliminated no more than a few dots from the center of the stimulus arrays.

Masking stimuli consisted of an array of crosses (plus signs) arranged in eight rows by eight columns, evenly spaced. Each cross was made from orthogonally bisecting line segments that were composed of a string of five dots.

Procedure

Subjects fixated a central target on the computer monitor at a viewing distance of 46 cm. A test stimulus appeared briefly and was followed immediately by a masking stimulus. For each trial, subjects indicated whether the dots in the test stimulus appeared to be grouping as a series of horizontal or vertical lines, by entering a response on the computer keyboard. The subjects were instructed to make their judgments regardless of whether the stimuli appeared more or less proximal, or whether they appeared aligned or misaligned. At the beginning of each test series, the difference in the spatial metrics between the cued and noncued orientations were most extreme, which elicited perceptual grouping along the orientation containing the cue. As the trial series progressed, the metrics of the noncued orientation was made increasingly similar to those of the orientation containing the cue.

Psychophysical thresholds were determined by means of a two-alternative forced choice staircase procedure. The cued orientation was randomly assigned to either the horizontal or the vertical orientation on each trial. Trials began with a maximum amount of cue strength. Selection of the cued orientation on 2 consecutive trials resulted in decreasing the cue strength, whereas each selection of the noncued orientation resulted in increasing the cue strength. A reversal was marked when the noncued orientation was chosen after at least 2 consecutive trials in which the cued orientation was selected. Following 4 reversals, the strength of the cue was increased by 10 increments, and the descending series was repeated. Three descending series were performed for each condition. The mean of 12 reversals (three descending series with 4 reversals each) was used as the grouping threshold, which required approximately 90 trials in order to be achieved. The computer algorithm that controlled the direction of change of the cue strength, which was contingent upon subjects' responses, converged on a level at which subjects would select the orientation containing the cue with a long-run probability of 71% (Levitt, 1971). Five subjects received Condition 1 first, followed by Condition 2, and the other 4 subjects had the reverse order.

Mask Interference Point

Immediately following the offset of the test stimulus, a pattern mask appeared for 200 msec. The SOA (stimulus onset asynchrony between the onset of the test stimulus and the onset of the masking stimulus) ranged from 150 to 33.3 msec, in increments of 16.7 msec. In order to allow the visual system full opportunity to process stimuli before the interference effect of the mask, the duration of the test stimulus was equivalent to the SOA. A fixed-duration stimulus would have produced a variable length gap between the test stimulus and the mask, thereby introducing poststimulus effects (e.g., iconic storage) that would vary with gap duration. Filling the interval before the appearance of the mask maintained experimental control over the test stimulus. For each test condition, grouping thresholds were initially measured for an SOA of 150 msec. The SOA was then shortened by 16.7 msec, and threshold measurements were repeated. This procedure was repeated until grouping thresholds increased by approximately 100%. The mask interference point was defined as the last SOA before the grouping threshold elevated beyond two standard deviations (*SD*) above the mean of thresholds measured at longer SOAs. Calculations were based on data from individual subjects.

RESULTS

The nature of the effect of backward pattern masking on grouping thresholds was consistent across subjects. At relatively long SOAs, thresholds remained consistent and appeared to be unaffected by the mask. As the SOA was shortened, there was a point at which thresholds elevated above the stable value, reflecting the mask interference point. At SOAs that were shorter than the mask interference point, subjects reported that the test stimulus could still be detected, and that it appeared to be either an evenly spaced grid of dots, or dots that were randomly distributed on the screen. Under these conditions, although elements of the test stimulus could be detected, organization of the stimulus array was not apparent.

Condition 1: Proximity Cues

The mean grouping threshold (in terms of separation ratios) derived from the stable portion of the masking function was 12.3% ($SD = 6.8$). Grouping thresholds found here were higher than those found previously (Kurylo, 1996), likely because measurements in the previous study were acquired from a small number of highly experienced subjects, and therefore the values do not directly correspond. The mean mask interference point occurred at an SOA of 87.6 msec ($SD = 13.8$). A trend existed for a correlation between grouping thresholds and mask interference points [$r(7) = 0.650, p = .058$]. Mask interference points across subjects ranged from 67 to 117 msec. In Figure 2, mean grouping thresholds are plotted as a function of SOA. In order to better demonstrate the interference effect of the mask on grouping SOA, Figure 2 is plotted relative to the mask interference point, with individual masking functions shifted along the abscissa to align individual mask interference points. It can be seen that grouping thresholds were relatively stable above the mask interference point SOA, and that they rose sharply below this value.

Condition 2: Alignment Cues

The mean grouping threshold derived from the stable portion of the masking function was 21.7% ($SD = 6.7$), and the mean mask interference point occurred at an SOA of 118.8 msec ($SD = 26.9$). Grouping thresholds corre-

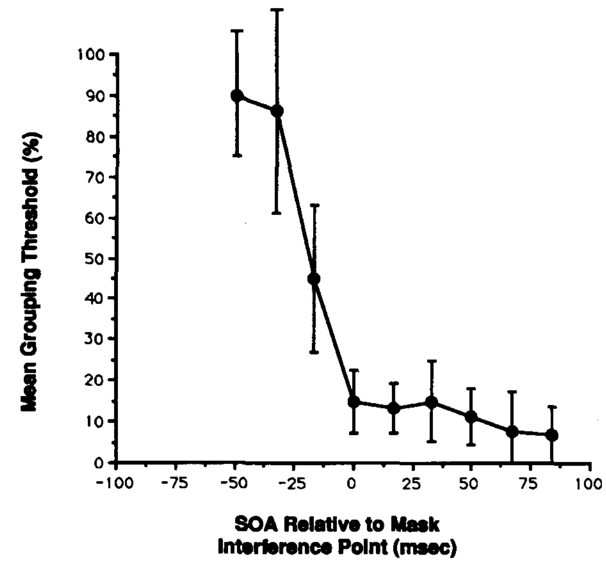

Figure 2. Mean grouping thresholds (in percent) for proximity cues as a function of masking SOA, relative to individual mask interference points. Error bars represent one standard deviation.

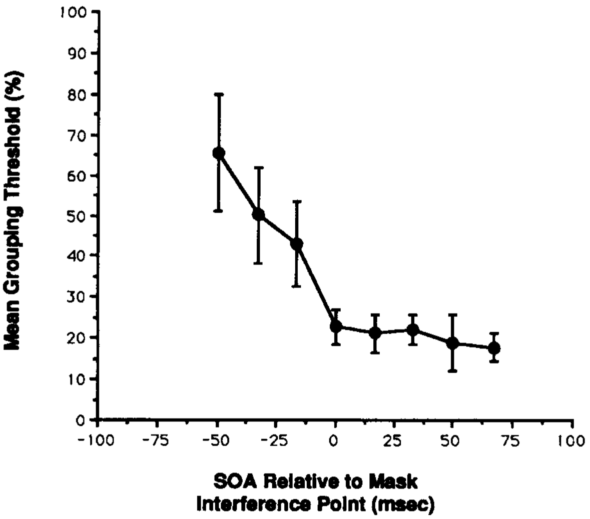

Figure 3. Mean grouping thresholds (in percent) for alignment cues as a function of masking SOA, relative to individual mask interference points. Error bars represent one standard deviation.

lated significantly with mask interference points [$r(7) = 0.680, p < .05$]. The mean grouping threshold as a function of masking SOA, aligned to individual mask interference points, is depicted in Figure 3. The nature of the function is similar to that found in Condition 1.

Comparison of Proximity and Alignment Cues

The metrics for the proximity and the alignment cues describe two different aspects of the stimulus, and therefore the values for grouping thresholds from Conditions 1 and 2 cannot be directly compared. However, the mask interference point reflects the SOA at which stable grouping thresholds begin to elevate because of the presence of the mask. A comparison of mask interference points between Conditions 1 and 2 therefore reflects the SOA at which a change occurs in the grouping threshold, and it does not represent a comparison of the metrics between these conditions.

In all cases, the mask interference point for the proximity cue occurred at a shorter SOA than did the mask interference point for the alignment cue. Across individual subjects, the difference in mask interference point between the two conditions ranged from 16.7 msec to as much as 66.7 msec. Analysis of mask interference points was performed with within-subjects paired tests, and data were not converted to accommodate individual differences, as was done for Figures 2 and 3. A paired t test indicated that mask interference points for the proximity and the alignment conditions differed significantly [$t(8) = 5.33, p < .001$]. Across subjects, mask interference points for the proximity and the alignment conditions correlated significantly [$r(7) = 0.817, p < .01$].

DISCUSSION

A new method for quantifying perceptual grouping was used to measure the time course of grouping processes. The index used here (grouping threshold) reflects the maximum capacity of the grouping process. At a specific duration of stimulus exposure, a backward pattern mask diminished the effectiveness of grouping, indicating interference with the operation of the grouping process by the mask. The main finding of this study is that the process of grouping disjunct elements into unified forms required a distinct and measurable duration, below which elements lacked a coherent organization. The mask interference point represents the SOA below which the mask first had an effect on the grouping threshold, which enabled the identification of processing time required before functional limitations were observed. At SOAs longer than the mask interference point, grouping thresholds were stable, and a minimal cue strength was required to elicit grouping. At SOAs below the mask interference point, the strength of the cue needed to be increased in order to perceive grouping. In this regard, the mask interference point represents the time required to derive the maximum effectiveness from the cue.

The conditions employed here required the organization of relatively complex and ambiguous stimuli. Under these conditions, grouping entailed the selection of a particular arrangement from among several possibilities that possessed highly similar metrics. Mask interference points thereby reflect grouping processes that were operating near the limitations of their capacities. With a more robust delineation between grouping arrangements, less processing time was required to resolve the stimuli. This effect is revealed at SOAs near to and shorter than the mask interference point, in which consistent grouping assignment occurred, although only when a more salient distinction of one grouping arrangement existed. This effect indicates that without significant competition from a noncued grouping arrangement, perceptual grouping is completed more quickly.

A second finding is that the two grouping strategies studied here required significantly different amounts of time in order to be completed. For each condition, information is derived from different aspects of the spatial relationships among elements, which suggests that distinct algorithms are applied for each condition. Grouping by alignment required a significantly longer amount of time than did grouping by proximity, and the mask appeared to reduce information about proximity less than it reduced information about alignment. These results may be interpreted as indicating that grouping by alignment, which involves the identification of pattern regularities, is computationally more intensive than grouping by proximity, which involves ratio comparisons of element distances. In this regard, ratio comparisons of element separations may be viewed as a low-level mechanism, based more on data-driven processing, whereas grouping by alignment may rely more on top-down information. Although the neural mechanisms that mediate perceptual grouping are not known, these results indicate that a distinction exists in the processing of each of these perceptual functions.

The degree of individual variability in mask interference points was considerable. All subjects were practiced at the task, and their performance had stabilized before measurements were made. Therefore, a learning effect cannot account for this individual variability. Perhaps mask interference points reflect the effectiveness of grouping capacities for individual subjects. Individuals with shorter mask interference points may process stimuli with greater efficiency than do individuals with longer thresholds. Consistent with this idea were the correlations between mask interference points and grouping thresholds for proximity and alignment cues; that is, subjects with lower mask interference points tended to have lower grouping thresholds.

It should be noted that the mask did not interfere with the detection of the test arrays, but instead interfered with the perceived grouping of the elements. In this regard, at SOAs that were shorter than the mask interference point, and with closely matched spatial metrics between orientations, subjects were able to detect the stimulus array, but were unable to perceive grouping among the dots. The reported detection of the dots eliminates the possibility that impaired perceptual organization at brief SOAs resulted from interference of the reception and encoding of the stimulus elements.

These measurements reflect not only grouping processes, but other visual functions, such as signal transmission through the retina and processes associated with form recognition. However, transmission and processing durations apply similarly to the masking stimulus, and therefore mask interference points measured here closely reflect processing time specific to grouping.

Because the grouping processes examined here were derived entirely from spatial relationships, accurate information about element position is critical for grouping to occur. Interference of grouping may result from a lack of adequate information of element position. This possibility appears unlikely, however, for two reasons. First, the extraction of spatial information must precede the algorithms that mediate grouping assignment, since grouping assignment is based on spatial information. When the mask follows the test stimulus at a long SOA, the mask has no effect on processing the test stimulus. As the SOA is progressively reduced, the mask encroaches on the end stages of processing the test stimulus. As the SOA is progressively reduced, the mask will first interfere with the late stages of processing the test stimulus before interfering with early stages. The sequence of processing the test stimulus is as follows: first, location information is extracted, and then, grouping is assigned. Therefore, the progressive reduction of SOA will likely interfere with grouping assignment before interfering with extracting location information. Thus, backward masking should first interfere with the grouping assignment algorithms before interfering with the acquisition of spatial information. The second argument against a lack of spatial information causing grouping impairment is that at SOAs below mask interference points, grouping still occurs, although at a higher grouping threshold. This effect indicates that spatial information exists below mask interference points, but that criteria for grouping assignment has changed, requiring a more salient distinction between grouping possibilities.

However, to view the extraction of spatial information as an independent stage in the process of grouping may be an oversimplification. Grouping assignment is based on spatial relationship information, not absolute position. Ratio comparison or the identification of pattern regularities is extracted from relationships among elements. Grouping assignment may act recursively with relationship comparisons, in that the algorithm for assignment continuously samples spatial relationships among elements. Multiple assignment possibilities with closely matched spatial metrics require a more precise evaluation of spatial relationships, thereby requiring longer processing time to complete. In this regard, masking would have interfered with both information extraction as well as the application of criteria for grouping assignment.

REFERENCES

GILLAM, B. (1981). Separation relative to length determines the organization of two lines into a unit. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 884-889.

GILLAM, B., & GRANT, T., JR. (1984). Aggregation and unit formation in the perception of moving collinear lines. *Perception*, **13**, 659-664.

KURYLO, D. D. (1996). *Effects of element separation and alignment on perceptual grouping*. Manuscript submitted for publication.

LEVITT, H. (1971). Transformed up–down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **49**, 467-477.

OYAMA, T., & YAMADA, W. (1978). Perceptual grouping between successively presented stimuli and its relation to visual simultaneity and masking. *Psychological Research*, **40**, 101-112.

PALMER, S. E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology*, **24**, 436-447.

PRYTULAK, L. S. (1974). Good continuation revisited. *Journal of Experimental Psychology*, **102**, 773-777.

RUSH, G. P. (1937). Visual grouping in relation to age. *Archives of Psychology*, **31** (Whole No. 217).

UTTAL, W. R. (1969). Masking of alphabetic character recognition by dynamic visual noise (DVN). *Perception & Psychophysics*, **6**, 121-128.

(Manuscript received August 14, 1995;
revision accepted for publication March 5, 1996.)