Pergamon

0042-6989(94)00173-1

Vision Res. Vol. 35, No. 6, pp. 853–866, 1995

0042-6989/95 $9.50 + 0.00

Perceptual Grouping by Similarity and Proximity: Experimental Results can be Predicted by Intensity Autocorrelations*

MERCEDES BARCHILON BEN-AV,† DOV SAGI†

Received 8 November 1993; in revised form 22 July 1994

A model for perceptual grouping based on measurements of spatial correlations is proposed and tested. Gestalt-like grouping experiments were performed to study and quantify the effect of element similarity (shape, luminance) and proximity. Observers reported the horizontal or vertical organization of stimuli with proximity and similarity providing conflicting grouping cues. Proximity grouping was found to be perceived much faster than similarity grouping. However, with increasing processing time, similarity was found to dominate grouping. The experimental results can be accounted for by assuming a process that compares horizontal and vertical intensity autocorrelations. The model suggests that correlations are measured across a limited spatial range, and that this range increases with processing time.

Autocorrelation Grouping Proximity Similarity

INTRODUCTION

Perceptual grouping is a process involved in chunking of visual information and in image segmentation, constituting an important aspect of visual processing. However, the rules governing grouping lack a quantitative formulation. Any account of perceptual grouping must still refer to the pioneering work of the Gestalt psychologists (Wertheimer, 1923) and to their "laws of grouping". According to the Gestalt theory of grouping, simple rules such as similarity of elements (shape), proximity, good continuation, common fate and connectedness dominate perceptual grouping by segmenting a visual scene into regions having some internal consistency (Wertheimer, 1923; Koffka, 1935). It is still not clear how to define shape and similarity (Beck, 1966; Olson & Attneave, 1970) and how to deal with multiple cues (e.g. similarity and proximity). Here we present a quantitative model for perceptual grouping, which is based on intensity autocorrelations. The model performance is successfully compared with data from psychophysical experiments, suggesting that at least some of the Gestalt rules of grouping (i.e. similarity and proximity) can be formalized in terms of spatial correlations.

When dealing with perceptual grouping we refer to a class of demonstrations with images consisting of numerous elements distributed in such a way as to generate a perception of some global form (Fig. 1). Although these images are sometimes treated as textures, we make a distinction between psychophysical tasks involving perceptual grouping and tasks involving texture segmentation. This distinction seems to be useful as there are two important differences between the processes underlying the two tasks. The first difference concerns the level of processing involved and the role of visual attention. Texture segmentation seems to be dominated by an early stage of visual processing which is preattentive and is free of resource limitations (Braun & Sagi, 1990), while perceptual grouping seems to be dependent on the availability of attentive resources (Ben-Av, Sagi & Braun, 1992; Mack, Tang, Tuma, Kahn & Rock, 1992).

The second difference between texture segmentation tasks and grouping tasks concerns the range of spatial integration involved. Processes involved in texture segmentation are dominated by short-range interactions (Sagi, 1991), while perceptual grouping seems to require long-range integration (Ben-Av *et al.*, 1992). Thus texture processes are better viewed as border enhancers operating on local differences between image regions, while grouping should be viewed as a process operating

*This paper is a portion of the Ph.D. dissertation of M. B. Ben-Av (1992). Presented in part at the ARVO conference, Sarasota, Fla, May 1992.

†Department of Neurobiology, Brain Research, The Weizmann Institute of Science, Rehovot 76100, Israel [*Email* dubi@nisan.weizmann.ac.il].

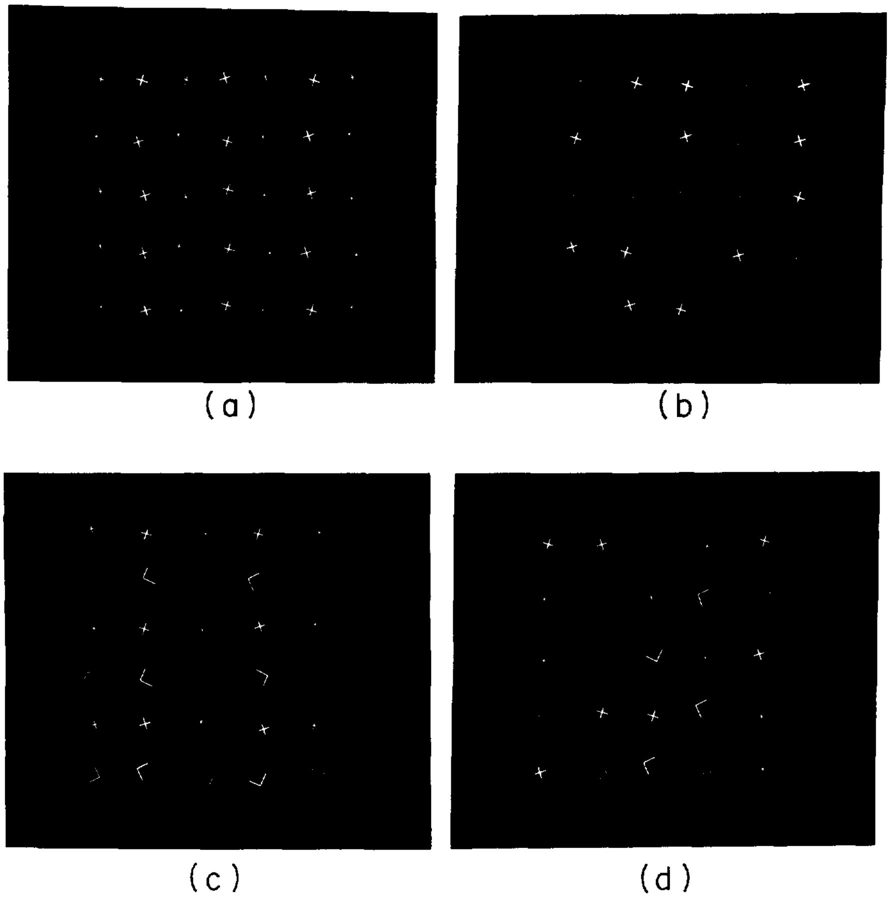

FIGURE 1. Stimuli used in the experiments. (a) Stimulus composed of Xs randomly oriented. The spacing ratio d_h/d_v between rows and columns is equal to 0.75 and the luminance ratio between alternating columns is equal to 2. These particular parameters show one possible stimulus array for the uniform display, the perceptual organization is presumably into horizontal rows. (b) Possible mask for the uniform display. The luminance of the elements of the mask is chosen randomly between the two luminances appeared in the trial to be masked. (c) A possible stimulus for the combined display, Xs or Ls in alternating rows. The spacing ratio d_h/d_v is equal to 1.16 and luminance ratio has the value of 2. (d) Possible mask for the combined display. The mask is composed of randomly positioned and jittered Xs or Ls. The luminance of the elements was chosen in the same manner as for the uniform display mask.

within a region, linking together distant image points, and abstracting some global image properties [e.g. closure (Kovács & Julesz, 1993)]. Models of texture segmentation assume a first stage of spatial filtering (selective for certain image features such as orientation and spatial frequency), followed by local inhibition (sometimes termed as a second stage of filtering) operating on filters having the same properties (Sutter, Beck & Graham, 1989; Fogel & Sagi, 1989; Landy & Bergen, 1991; Malik & Perona, 1990; Rubenstein & Sagi, 1990). These models are able to simulate human behaviour when the spatial filters integrate (linearly) over a relatively small spatial extent (about 1–4 elements in Fig. 1), thus computing local oriented energy. Models assuming global linear integration were shown to be inadequate for texture discrimination, as some textures having equal power spectra were shown to be effortlessly discriminable (Julesz, 1980). On the other hand, large flow-frequency filters were found to be useful in simulating some grouping phenomena (Ginsburg, 1971, 1980), although grouping was also demonstrated with images devoid of low frequency components (Jañez, 1984). Thus grouping seems to require some nonlinear integration over large image portions in order to abstract some long-range correlations (or form) in the image. It is possible that this integration is performed on the output of spatial filters [although the term "grouping" is used sometimes for processes integrating dots into lines by small scale oriented spatial filters (i.e. Zucker, Stevens & Sander, 1983)].

In this paper we study the Gestalt principles of proximity and similarity, with stimuli composed of discrete elements Xs or Ls (Fig. 1). The elements were arranged in arrays giving rise, in most of the situations, to two possible perceptual organizations, either horizontal rows or vertical columns (some parameter values produced ambiguous perceptual organizations). The effect of inter-elements spacing (proximity), shape similarity (Xs vs Ls), and luminance similarity on perceptual grouping was investigated. We investigated the relationship between these parameters by examining how they interfere with each other, and whether there is any hierarchy of these parameters in facilitating the grouping processes. The parameters were manipulated so as to create cooperative or competitive situations between spacing and luminance. We find high sensitivity for inter-element spacing variations with proximity grouping dominating performance when the stimulus is available for a short time (< 100 msec). For longer processing times, similarity based grouping takes over and dominates performance.

The psychophysical experimental results were simulated by a simple model based on the autocorrelation function (Priestly, 1981) operating on the image intensity values. The model uses a directional autocorrelation function, comparing vertical and horizontal autocorrelations. Since the autocorrelation function operates on the image intensity values (pixels), no assumptions are made on featural differences between the local line elements that give rise to similarity grouping (e.g. X vs L). Departing from classical autocorrelation models of vision (Uttal, 1975), we suggest that the interactions underlying the correlation measurements are time dependent in a way that long-range correlations take more time to be effective, while short range correlations can be detected within a relatively short processing time. This hypothesis is implemented in the model by multiplying the autocorrelation function by an exponential weighting function with a time dependent space constant.

METHODS

Observers

Five practiced observers participated in the experiments. Four of them (MA, IG, AM and SW), were paid high school students and were naive as to the purpose of the study. All enjoyed normal, or corrected to normal, visual acuity.

Apparatus

Stimuli were presented in a dark environment on a Hewlett–Packard 1310B oscilloscope (P31 phosphor). The oscilloscope was driven by custom-designed hardware, which allowed real-time control of the stimulus properties. The experiments and graphic device were controlled by a Sun 3/160 workstation. Screen resolution was $1024 \times 1024 \times 1024$ ($x \times y \times$ luminance). Viewing distance was approx. 150 cm, resulting in a stimulus subtending 6×6 deg of visual angle.

Stimulus

The stimuli consisted of discrete pattern elements in the shape of Xs or Ls, arranged in an array of rows and columns. The Xs and Ls were formed by two perpendicular segments of equal length (30 pixels subtending 0.30 deg of visual angle).

Two basic stimuli were used, the *uniform stimulus* in which the elements were Xs [Fig. 1(a)] and the *combined stimulus* in which the odd rows (or columns) were Xs and the even rows (or columns) were Ls [Fig. 1(c)]. For each of these stimuli, three different experiments were designed depending on the experimental variables.

Experimental variables

Two experimental variables were considered: the *relative distance* (d_h/d_v) between columns and rows and the *relative luminance* between alternating columns (ρ).

Experiment 1. In Expt 1, the relative distance between rows and columns varied, while the relative luminance remained fixed (all the elements of the stimulus had the same luminance). Nine spacing ratios (the horizontal separation between the centres of the elements divided by the vertical separation) were used: $d_h/d_v = 0.50$, 0.75, 0.84, 0.92, 1.00, 1.08, 1.16, 1.25 and 1.50. For each spacing ratio, the maximum separation between centres of elements, either horizontal or vertical, was set to $\max(d_h, d_v) = 1.25$ deg. The separation in the other direction was calculated from the ratio value. The ratio of the pattern size to inter-pattern separation varied between 1:2 and 1:4 (0.62–1.25 deg).

Experiment 2. In Expt 2, the relative distance between elements remained fixed and equal to 1 (all the elements of the stimulus were equidistant) while their relative luminance varied. Seven different luminance ratios between odd and even columns were employed ($\rho = 1.0$, 1.2, 1.5, 2.0, 3.0, 4.0 and 5.0), keeping the total luminance constant. Notice that luminance intensity differences were introduced in alternating columns while shape differences (Xs vs Ls) appeared in alternating rows. Consequently, shape similarity always competed with luminance similarity.

Experiment 3. Finally, in Expt 3, both parameters, relative distance and relative luminance varied simultaneously. Only five relative distances were considered: 0.50, 0.75, 1.00, 1.25 and 1.50. The luminance ratios considered were the same as in Expt 2.

Each element was positioned randomly (jitter) around the centre of its grid position, up to 10 pixels (0.10 deg) in any direction. In addition, all array elements were randomly rotated with one exception: Expt 1 was replicated with upright elements to study the effect of the random orientation in the outcome of the experiment.

The global appearance of the matrix that formed the stimulus was rectangular in most of the cases, due to the different spacing between rows and columns. This appearance may serve as a cue to the observers' judgement. To avoid this bias, the screen was covered with a square window of size 6×6 deg, so that the global form of the stimulus was constant across all experiments. As a consequence of this, the number of elements in the stimulus varied as a function of the relative spacing ratios. In addition, to compensate for any bias toward horizontal or vertical direction, the stimulus was rotated by 90 deg randomly with probability of 0.5. Thus the Ls in the combined stimulus appeared in alternating rows (columns) and the changes in luminance were in alternating columns (rows).

From now on, for simplicity, we will refer to the stimulus as if it was not rotated, that is the luminance difference in alternating columns and the elements' similarity (Xs vs Ls) in alternating rows,

TABLE 1. List of parameters used in the different experiments

Experiment	Spacing ratios	Luminance ratios	SOA (msec)	Orientation
1 (a)	0.5, 0.75, 1, 1.25, 1.5	1	60, 160, no mask	Random
1 (b)	0.5, 0.75, 1, 1.25, 1.5	1	160	Upright
1 (c)	0.84, 0.92, 1, 1.08, 1.16	1	160	Random and upright
2	1	1, 1.2, 1.5, 2, 3, 4, 5	60, 160, no mask	Random
3	0.5, 0.75, 1, 1.25, 1.5	1, 1.2, 1.5, 2, 3, 4, 5	60, 160, no mask	Random

although observers saw them either in alternating rows or columns.

Visual tasks

The displays considered in these experiments were based on the Gestalt demonstrations of proximity and similarity grouping (Koffka, 1935). Depending on the spacing and luminance ratios considered, the perceptual organization of the display was in general horizontal rows or vertical columns except for ambiguous situations. Observers were asked to report the perceptual organization of the display as horizontal rows or vertical columns.

Procedure

Data were collected in blocks of 50 trials in Expt 1, 49 trials in Expt 2 and 35 trials in Expt 3. Each trial was preceded by a fixation mark at the centre of the display until the observer signalled readiness by pressing the space bar on a standard terminal keyboard. Then, after a dark interval of duration 500 msec, the stimulus was briefly presented (for 20 msec), and then masked. The time interval between the onset of stimulus and onset of mask (stimulus onset asynchrony, SOA) was varied. The visual availability of the stimulus was controlled by SOA. The rapidness of this sequence prevented a second eye fixation.

The response of the observer consisted in typing on the keyboard 0 when horizontal organization was perceived or 1 when vertical organization was perceived. Each session lasted approx. 1 hr.

Mask patterns

In order to mask all the relevant aspects of the stimulus, different masks for the different stimuli were generated. The mask was a matrix composed of randomly positioned and jittered elements {Xs in the uniform case [Fig. 1(b)] or of Xs and Ls in the combined case}, with two random luminances (the luminances which appeared in the trial to be masked) [Fig. 1(d)]. The elements of the mask presented a different grouping organization from that of the stimulus. The mask spacing ratios used were $d_h/d_v = 0.66$, 1, 1.33, 1.66 and 2, and the $\max(d_h, d_v) = 1.23$ deg. The same square window as in the stimulus was used.

Although stimulus and mask were presented in a rapid sequence, no percept of global apparent motion was observed. Accordingly, observers could not have identified the stimulus array on the basis of a motion percept.

Although the masks used do have an horizontal or vertical organization (based on proximity grouping), the following remarks have to be considered.

(1) The mask grouping parameters were different from those of the stimulus: different spacing ratios, shape similarity and luminance elements randomly distributed across the entire array.
(2) The random distribution of shape similarity and luminance of the mask elements does not necessarily create a perceptual organization on the basis of proximity grouping.
(3) The horizontal or vertical organization of both the stimulus and the mask were randomly and independently determined.

All the above considerations inhibited the possible motion percept resulting from proximity perceptual organization of the mask.

Table 1 summarizes the different variables used in each experiment.

RESULTS

All the experiments described below were carried out with two different stimuli: the uniform stimulus and the combined stimulus (cf. Methods). With one or two exceptions (see Table 1), experiments were performed for three different SOAs: 60 msec, 160 msec and without mask. All the elements of the stimuli were randomly oriented (except for one case, see Table 1) and positioned (jittered). Observers were asked to report the organization of the display into horizontal rows or vertical columns.

Experiment 1: proximity and similarity

This experiment was conducted in order to determine the effect of the relative position between the stimuli elements in determining grouping perception. Therefore, the variable tested was the spacing ratio between columns and rows d_h/d_v. In addition, the similarity parameter introduced by the combined stimulus allowed us to study the interactions between proximity and similarity. Each observer was tested on 15 blocks of 50 trials each. The curves representing the outcome of the experiment are presented in Fig. 2 (the results for the uniform and combined displays are plotted on the same graph). The graphs show the percentage of vertical grouping as a function of the spacing ratio.

The curves show that the parameter "time" plays a significant role in the perception of grouping since

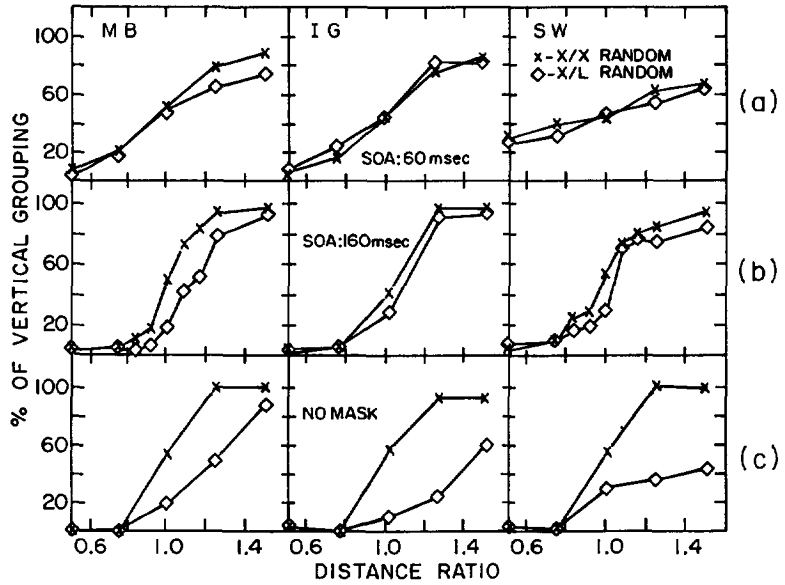

FIGURE 2. Outcome of Expt 1. The results obtained with both the uniform (×) and the combined (◇) displays are plotted on the same axis. The elements of the stimuli were randomly oriented and positioned (jitter). The different graphs depict the results for each of the observers and all of the SOAs considered (a) 60 msec, (b) 160 msec and (c) no mask (practically infinite SOA). The graphs represent the percent of vertical grouping (ordinate) as a function of the distance ratio (abscissa). Note that in (b) and (c) the curves for the combined display are shifted to the right with respect to the curves for the uniform display, showing the effect of similarity. This effect is not present in (a).

different results were obtained for the different SOAs. With the uniform stimulus (Fig. 2, ×s), as expected, elements closely spaced were grouped together and the equilibrium point (50% grouping in each direction) was obtained for spacing ratio of 1 (equally spaced elements). However, the shape similarity present in the combined display and the different spacing ratios created cooperation or competition between both parameters.

The effect of SOA is easy to see if we concentrate on the area between the two curves (uniform and combined displays) in each graph. For short SOA (60 msec), the two curves are close to identical (the area is almost zero) [Fig. 2(a)], consequently, the effect of similarity was very weak or non-existent. As SOA increased, the effect of element similarity became more and more apparent (the area between the curves increases), therefore for SOA of 160 msec [Fig. 2(b)], grouping was affected by element similarity. Thus, for equidistant elements, all observers performed around 75% grouping in the similarity direction. Without mask (practically infinite SOA), the effect of shape similarity was bigger [Fig. 2(c)]. The results as a function of SOA, provide us with information about the possible hierarchy in the parameters which facilitate grouping. For short SOA (60 msec), only grouping by proximity is perceived, there is no effect of similarity. The effect of similarity is built up from 60 to 160 msec SOA. At the no-mask condition, similarity based grouping dominates over proximity grouping.

Experiments 1(a) and 1(b) were performed again [Expt 1(c)] for 160 msec SOA, using upright elements instead of random [Expt 1(c)]. As can be seen in Fig. 3, the outcome of Expt 1(c) is not significantly different from the outcome of Expts 1(a) and (b). This suggests first, that grouping was based on element position rather than on features differences and second,

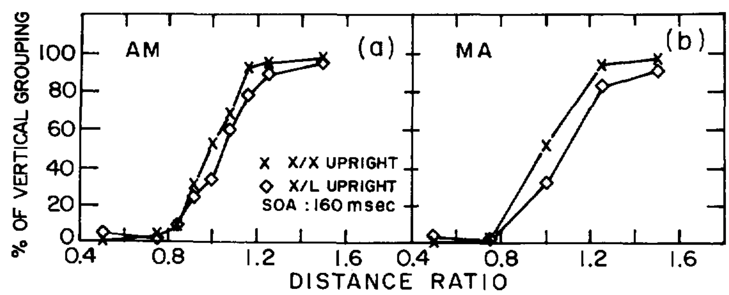

FIGURE 3. Outcome of Expt 1, with upright elements. The figure shows the results of two observers AM and MA. The outcome of both, uniform (×) and combined (◇), displays are plotted on the same axis. The graphs depict the results obtained for 160 msec SOA. Percentage of vertical grouping (ordinate) as a function of the distance ratio (ordinate). There is no significant difference with the results plotted in Fig. 2.

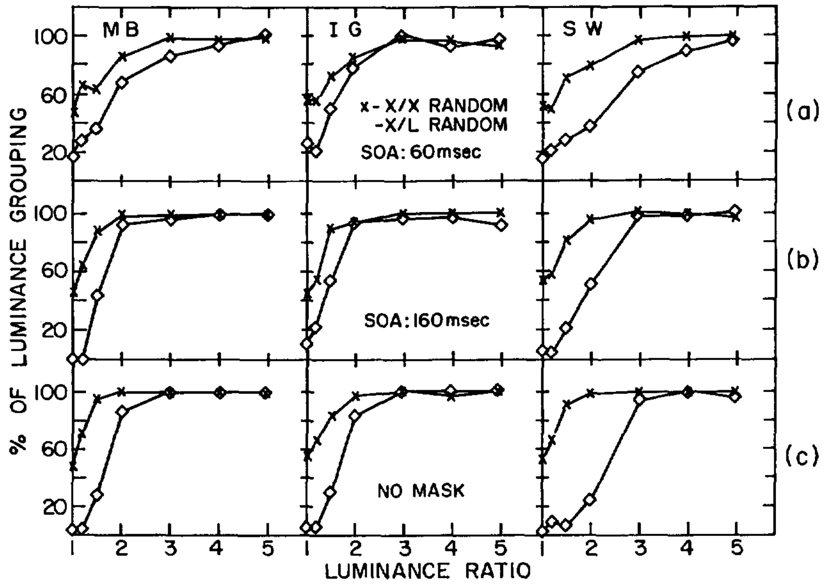

FIGURE 4. Outcome of Expt 2 for three different observers (MB, IG and SW) and three different SOAs, (a) 60 msec, (b) 160 msec and (c) ∞ SOA (no mask). For each SOA and each observer, the corresponding curves representing the data for both displays are plotted in the same graph. ×s represent the data points for the uniform display and ◇s the data points for the combined display. The graphs represent the percentage of luminance grouping (ordinate axis) as a function of the luminance ratio (abscissa axis). Note that the curves depicting the results for the combined display are shifted downwards compared to the curves corresponding to the uniform display, showing the effect of pattern similarity. Notice that as SOA increases, the similarity effect increases for luminance ratios smaller than 3 and decreases for luminance ratios >3.

that grouping processes seem to operate at a stage before shape discrimination. Since the orientation of the elements did not appear to alter significantly the outcome of the experiments, we will consider only randomly oriented elements in the coming experiments.

Experiment 2: similarity and luminance

This experiment was intended to study how different luminance levels affect the grouping perception. Consequently, we considered equidistant elements (spacing ratio of 1) and varied the luminance ratio between alternating columns (rows). Notice that with the combined stimulus, shape similarity and luminance were always brought into competition (cf. Methods). Each observer was tested on 15 blocks of 49 trials each. The experimental results are depicted in Fig. 4 (×s for the uniform stimulus, ◇s for the combined stimulus), Fig. 4(a) for 60 msec SOA, Fig. 4(b) for 160 msec SOA and Fig. 4(c) without mask. The graphs represent the percentage of luminance grouping as a function of the luminance ratio.

As in the previous experiment, different results were achieved for the different SOAs. The equilibrium between luminance grouping and similarity grouping for all SOAs considered was between 1.5 and 2 luminance ratio, except for SW without mask, in which case the equilibrium was approx. 2.5. Although the luminance ratio required to reach equilibrium (50% grouping in either direction) was almost the same for the different SOAs, the slope at equilibrium differed: as the SOA increased, the slope became steeper. The steepness of the slope indicates that the luminance as well as the similarity effects become stronger as SOA increases.

Experiment 3: proximity, similarity and luminance

This experiment was a combination of the previous ones. The variables tested were both the spacing ratio between elements and the luminance ratio between alternative columns (rows). Each observer was tested on 50 blocks of 35 trials. The results of the experiment are depicted in Fig. 5(a, b, c) for three different observers. For each observer, the top graphs represent the results obtained with the uniform stimulus for the different SOAs and the bottom graphs the results obtained with the combined stimulus. For each SOA, six different luminance curves are plotted in the same graph, each one corresponding to a different fixed spacing ratio. The

FIGURE 5 (*opposite*). The graphs show the results of Expt 3 for three different observers [MB (a), IG (b), SW (c)] and for the different SOAs (60 msec, 160 msec and no mask). Two displays were used: the uniform condition in which all the elements were randomly oriented and positioned Xs, and the combined condition in which the elements of the display were all Xs or all Ls in alternating rows. For each observer, the upper graphs correspond to the results obtained with the uniform display, and the lower graphs the results obtained with the combined display. For each SOA, six different curves are plotted, corresponding to the six different spacing ratios. The graphs represent the percentage of vertical grouping (ordinate axis) as a function of the luminance ratio (abscissa axis).

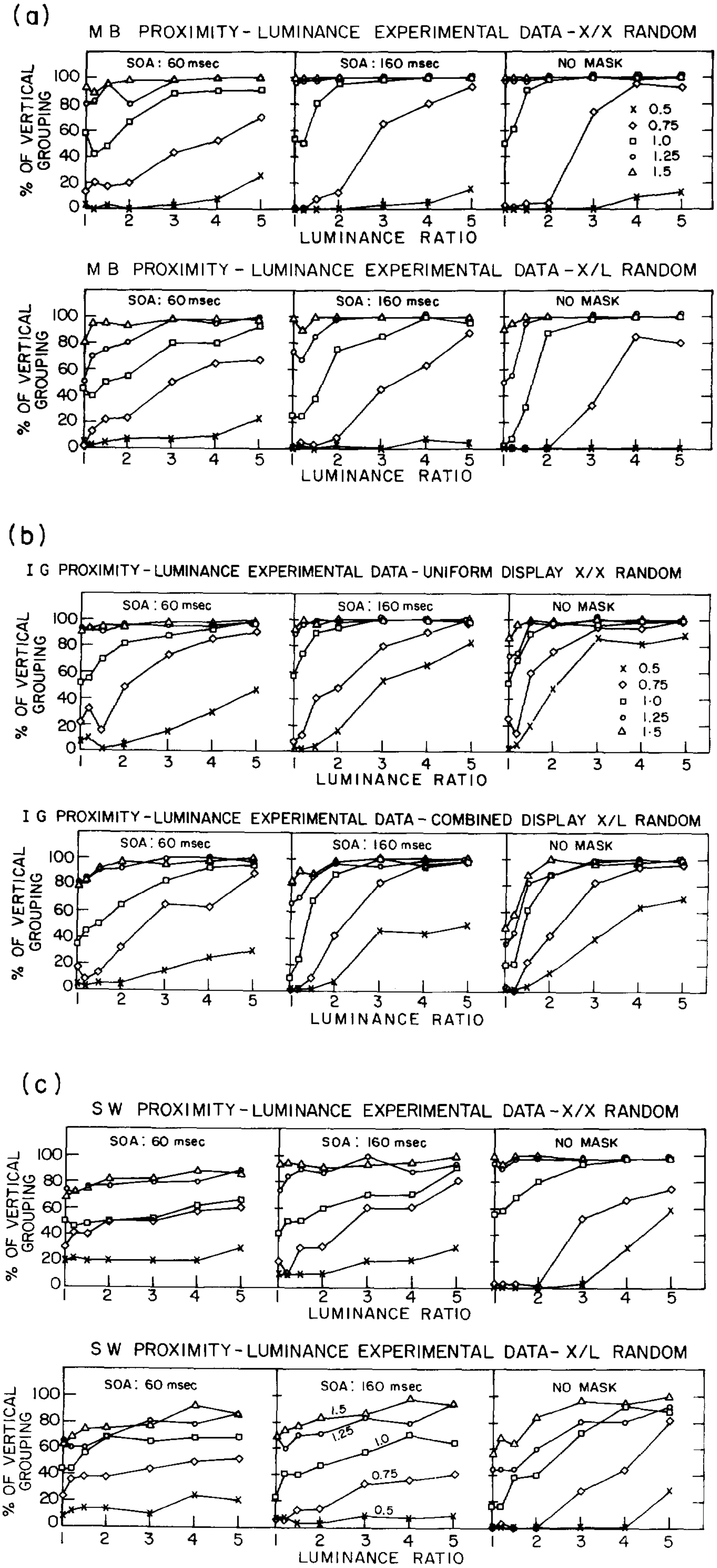

FIGURE 5—*Caption on facing page.*

TABLE 2. Interactions between proximity, shape similarity and luminance for the stimuli considered in the experiments

Spacing ratio	< 1 Prox. & Lum.	< 1 Prox. & Sim.	> 1 Prox. & Lum.	> 1 Prox. & Sim.
Uniform stimuli	Compete		Cooperate	
Combined stimuli	Compete	Cooperate	Cooperate	Compete

luminance curves show the percent of vertical grouping* as a function of the luminance ratio. Notice that for both stimuli considered (uniform and combined), luminance and proximity grouping competed for spacing ratios < 1 and cooperated for spacing ratios > 1. Luminance and shape similarity (combined stimulus) are set such that they always compete.

Performance differences across observers were observed for spacing ratios of 0.5 and 0.75, a range in which luminance and proximity competed. For the spacing ratio of 0.5, observers MB and SW did not show any luminance effect, even for luminance ratio of 5, irrespectively of the SOA and of the stimulus [Fig. 5(a, c)]. Observer IG [Fig. 5(b)], for the same 0.5 spacing ratio demonstrated grouping by luminance for some of the SOAs and from some luminance ratio values.

Luminance and proximity equilibrium

For each of the spacing ratios that brings proximity and luminance into competition (spacing ratio $\leqslant 1$), we evaluated the amount of luminance required to achieve equilibrium between luminance and proximity, i.e. to get 50% grouping in either direction at 160 msec SOA. The results are shown in Fig. 6(a) for the uniform stimulus and in Fig. 6(b) for the combined stimulus. The horizontal axis represents the spacing ratio while the vertical axis shows the luminance ratio. The different symbols depict the corresponding equilibrium values for the different observers and the continuous line corresponds to their average. In both cases (uniform and combined stimuli), as the spacing ratio decreases more luminance is required to reach equilibrium. If we superimpose both graphs, we see that the graph corresponding to the uniform stimulus [Fig. 6(a)] is always below the graph representing the results for the combined stimulus [Fig. 6(b)]. This behaviour is due to the similarity effect. Since on the one hand luminance and shape similarity always compete with each other (see Methods), and on the other we are considering the cases in which proximity competes with luminance, it follows that in this situation, shape similarity and proximity cooperate. Therefore, for each spacing ratio, more luminance is required to reach equilibrium with the combined stimulus as compared to the uniform stimulus. Consequently, the graph depicting the results with the uniform stimulus is always below that with the combined stimulus.

Similarity as a function of the SOA

We considered it of interest to evaluate the similarity effect as a function of SOA. Consequently, we compared the luminance curves of Fig. 5 for the different SOAs corresponding to spacing ratio of 1 (equidistant elements) in the combined stimulus, and in those curves, the performance corresponding to the luminance ratio of 1 (equiluminant elements). The corresponding performances for different SOAs were plotted in Fig. 7, where the abscissa represents the SOA and the ordinate shows the percent of similarity grouping. The different symbols depict the data from the different observers and the dashed line shows their average. Since we considered the "neutral" spacing ratio and the "neutral" luminance ratio, the only factor influencing the results is the shape

*For simplicity, we refer to the stimuli as if luminance varied across vertical columns and shape similarity varied across horizontal rows (see Methods). Accordingly, we will refer to vertical grouping as "luminance grouping" and to horizontal grouping as "similarity grouping".

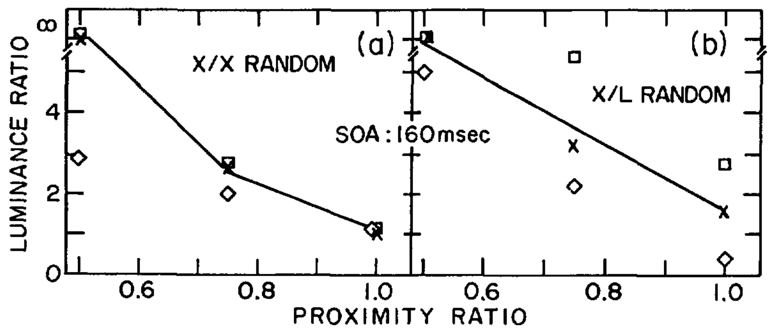

FIGURE 6. The graphs represent the luminance ratio required for each spacing ratio in order to reach equilibrium between luminance and proximity, that is 50% grouping in either direction. Accordingly, the horizontal axis represents the spacing ratio and the vertical axis the luminance ratio. (a) The uniform display, (b) the combined display. The different symbols depict the data for the different observers, the continuous line is the average across observers.

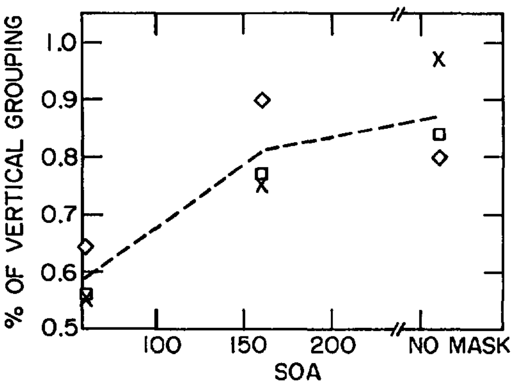

FIGURE 7. The graph represents the effect of similarity as a function of the SOA. The data was taken from the results obtained for the combined display in Expt 3. For each SOA, the percentage of vertical grouping corresponding to spacing ratio of 1 (equidistant) and luminance ratio of 1 (equiluminant) is plotted. The different symbols depict the data from the different observers and the dashed line the average across observers. Note that the dashed line is an increasing function of the SOA, indicating an enhancement of the grouping organization in the shape similarity direction.

similarity. Figure 7 shows that as SOA increases, so does the similarity effect. For small SOA, the average performance is very close to 50% (equilibrium between proximity and luminance), meaning that there is no effect of similarity. The similarity effect appears at the larger 100 msec SOA and continues to increase as SOA increases. This result suggests that grouping based on shape similarity is not an automatic process. Since approx. 100 msec are required to build the similarity effect, this might be an indication that some top-down processes are involved (see also Ben-Av *et al.*, 1992). Does luminance similarity behave the same as shape similarity?

Luminance as a function of the SOA

As previously with shape similarity, we evaluated the luminance effect as a function of SOA. For both stimuli, we considered the luminance curve of Fig. 5 corresponding to the spacing ratio of 0.75 and found the luminance ratio leading to equilibrium, i.e. 50% performance. We then considered the slope at the threshold luminance. Figure 8 shows the threshold luminance slopes as a function of the SOA [Fig. 8(a) depicts the results for the uniform display, Fig. 8(b) for the combined stimulus]. The different symbols are the results obtained for the different observers and the dashed line is the average across observers. For both stimuli, the slope becomes steeper as SOA increases, which means that the effect of luminance increases with SOA. The effect of shape similarity is also demonstrated in these graphs. If Fig. 8(a) and (b) are superimposed, it is easy to see that the graph depicting the results for the uniform stimulus [Fig. 8(a)] is above the graph depicting the results for the combined stimulus [Fig. 8(b)]. This means that the effect of luminance is weaker in the case of the combined stimulus presumably due to the competing similarity. However, notice that for the smallest SOA considered (60 msec), the slope for both stimuli is practically the same, showing again that for short SOA there is no similarity effect.

Summary

All the experiments show that perceptual organization is time dependent. Proximity grouping can be perceived much faster than similarity (shape or luminance) based grouping. Our experimental paradigm shows that shape similarity and luminance similarity are built up between 60 and 160 msec SOA while proximity is built up in less than 60 msec. While proximity grouping is a fast process, it can be taken over by similarity cues when they are later perceived. The system appears to be very sensitive to small variations in spacing ratios.

THE AUTOCORRELATION MODEL

The model description

The stimulus is considered as a grid of L_x by L_y pixels. Let $f(x, y)$ be the input intensity of a pixel at a (x, y) position on the grid, and let ξ be a shift unit ($\xi \geqslant 0$). The autocovariance in the direction of x as a function of y is defined as:

$$g_x(y, \xi) = \frac{1}{L_x - \xi}\left[\sum_{x=1}^{L_x - \xi} f(x, y) f(x + \xi, y)\right]. \quad (1)$$

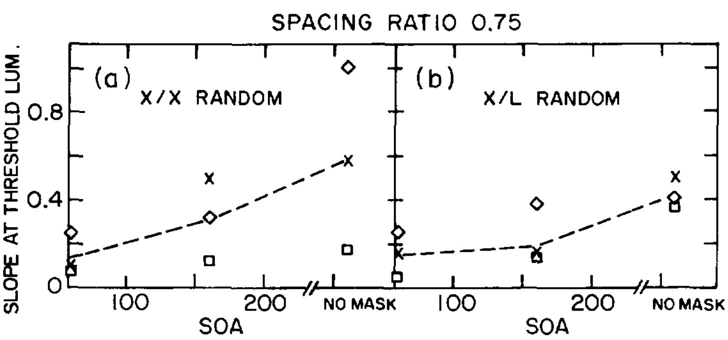

FIGURE 8. The graphs represent the slope at threshold luminance. (a) Outcome of the uniform display, (b) outcome of the combined display. The data is taken from Expt 3. The graphs depict the slope at 50% grouping performance for the luminance curves corresponding to 0.75 spacing ratio as a function of the SOA. As the SOA increases, so does the slope. The curve (a) is above (b) showing the effect of similarity. Notice however, that for the smallest SOA, both curves coincide, meaning that for that SOA value there is no similarity effect.

Analogously, the autocovariance in the direction of y as a function of x is defined as:

$$g_y(x, \xi) = \frac{1}{L_y - \xi}\left[\sum_{y=1}^{L_y - \xi} f(x, y) f(x, y + \xi)\right]. \quad (2)$$

In this manner, two families of autocovariance functions were built: $\{g_x(y, \xi)\}_{y=1,L_y}$ in the direction of x and $\{g_y(x, \xi)\}_{x=1,L_x}$ in the direction of y.

Since the global organization of the elements of the grid is to be modeled, the next step will be to average $\{g_x(y, \xi)\}$ over L_y and $\{g_y(x, \xi)\}$ over L_x respectively obtaining one autocovariance function $w_x(\xi)$ for the x direction and another one $w_y(\xi)$ for the y direction. Therefore, $w_x(\xi)$ and $w_y(\xi)$ are defined as follows:

$$w_x(\xi) = \frac{1}{L_y}\sum_{y=1}^{L_y} g_x(y, \xi) \quad (3)$$

$$w_y(\xi) = \frac{1}{L_x}\sum_{x=1}^{L_x} g_y(x, \xi). \quad (4)$$

The corresponding autocorrelation functions were obtained by normalizing the last two quantities $w_x(\xi)$ and $w_y(\xi)$. Consequently, the autocorrelation functions, $v_x(\xi)$ and $v_y(\xi)$, of $\{f(x, y)\}$ in the direction of x and y respectively, are:

$$v_x(\xi) = \frac{w_x(\xi)}{w_x(0)} \quad (5)$$

$$v_y(\xi) = \frac{w_y(\xi)}{w_y(0)}. \quad (6)$$

Next, we introduce a correlation distance weighting function. Since closer elements are more likely to be "similar" while distant elements are more likely to be "different" in a natural environment, a larger weight has to be given to short distances than to large distances in order to reduce effects of noise. Therefore, a weighted sum of the autocorrelations was considered. Several weights were analysed, the best results being produced by an exponential weighting function ($e^{-\xi/\xi_0}$), that emphasizes short-range correlations and strongly attenuates those that are of longer range. Consequently, the *total weighted correlation* in each direction is defined as the following weighted sums:

$$\lambda_x = \Delta\xi_x \cdot \sum_{\xi=0}^{L_y - 1} e^{-\xi/\xi_0} \cdot v_x(\xi) \quad (7)$$

$$\lambda_y = \Delta\xi_y \cdot \sum_{\xi=0}^{L_x - 1} e^{-\xi/\xi_0} \cdot v_y(\xi). \quad (8)$$

$\Delta\xi_x$ and $\Delta\xi_y$ are the distance between two consecutive pixels in the directions of x and y respectively. In the display used in the psychophysical experiments, $\Delta\xi_x$ was equal to $\Delta\xi_y$. ξ_0 is the only free parameter of the model. As will be seen later, ξ_0 is time dependent.

Decision stage

At this point, where the total autocorrelations of both the horizontal and the vertical directions are provided by λ_x and λ_y respectively, a decision has to be made. A quantitative grouping organization direction is given by the ratio λ_x/λ_y as follows:

if $\lambda_x/\lambda_y > 1 \Rightarrow$ horizontal grouping;
if $\lambda_x/\lambda_y < 1 \Rightarrow$ vertical grouping;
if $\lambda_x/\lambda_y = 1 \Rightarrow$ either horizontal or vertical with equal probability.

Computer simulations

The psychophysical display (see Methods) was simulated in the model in the following manner: at each position (x, y) of the grid, the input intensity function $f(x, y)$ was defined as follows:

$$f(x, y) = \begin{cases} 0 & \text{blank pixel in the psychophysical display} \\ i & \text{otherwise} \end{cases}$$

where $i \geqslant 1$ simulates the luminance ratio of the corresponding pixel. Accordingly, i varied in the set of values $\{1, 1.2, 1.5, 2, 3, 4, 5\}$, which corresponds to the luminance ratios ($\rho = I_o/I_e$) used in the psychophysical experiments (see Methods). The "neutral" luminance ratio, that is the luminance ratio when all the elements of the display were equiluminant (see Results), was simulated by setting $i = 1$.

The size of the display used in the psychophysical experiments was 600 × 600 pixels corresponding to 6 × 6 deg (see Methods). Since this grid size would require considerable amount of computation time, we chose to scale down the grid (for the model implementation), by one-third, resulting in a grid of 200 × 200 pixels. Element size, jitter and max (d_h, d_v) (see Methods) were scaled accordingly. Table 3 sets out the parameters used in the psychophysical experiments as well in the model implementation.

We simulated the three experiments described above (see Results). As in the psychophysical experiments, we used both displays (uniform and combined). The elements of the display were randomly rotated and their positions were jittered. Thus, the model predictions are not exact, but rather represent the outcome of a stochastic process, repeated a few hundred times in each case. We present next, the outcome of the model for 160 msec SOA, with $\xi_0 = 50$. Figures 9 and 10 show the outcome of the psychophysical Expts 1 and 2 as well as the model performance for 160 msec SOA. The ordinate of the graphs shows the percent of vertical grouping, the abscissa represents the spacing ratio in the

TABLE 3. Psychophysical and simulation parameters

Parameters	Grid size	Equidistant elements	Element size	Jitter	max (d_h, d_v)
Psychophysics	600 × 600 (6 × 6 deg)	5 × 5	30 pixels	20 pixels	120 pixels
Model	200 × 200 (6 × 6 deg)	5 × 5	11 pixels	6 pixels	40 pixels

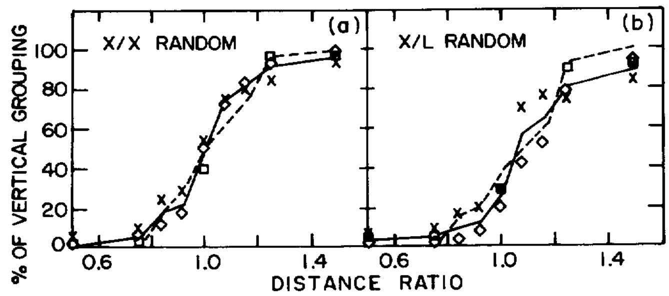

FIGURE 9. The graphs show the percentage of vertical grouping as a function of the distance ratio for the uniform display X/X (a) and for the combined display X/L (b). The different symbols represent the data for the different observers, the solid line is their average and the dashed line represents the prediction of the model.

case of Expt 1 (Fig. 9) and the luminance ratio in the case of Expt 2 (Fig. 10). Figures 9(a) and 10(a) depict the data for the uniform display, Figs 9(b) and 10(b) for the combined display. The different symbols stand for the outcome of the psychophysical experiment for the different observers, the continuous line is the average across observers and the dashed line represents the model prediction. For each simulation performance, the model ran through five blocks of 50 trials for Expt 1 and three blocks of 50 trials for Expt 2.

Figure 11 shows the results of Expt 3 for only one observer. For each distance ratio (different symbols), the corresponding luminance curve is plotted, representing the percentage of vertical grouping as a function of the luminance ratio I_o/I_e. From bottom to top, the corresponding spacing ratios are 0.5, 0.75, 1, 1.25 and 1.5 respectively. The continuous line (for each symbol) represents the outcome of the psychophysical experiment, the dashed line shows the model performance. The other observers presented similar behaviour.

In order to summarize all the luminance curves obtained in Expt 3, for each spacing ratio, the amount of luminance required to get 50% grouping in either direction was calculated (see Results). This luminance value is named the "equilibrium luminance" (EL). For each observers, the corresponding EL was calculated. Figure 12 depicts EL as a function of the spacing ratio for each observer (different symbols), the mean across observers (solid line) and the prediction of the model (dashed line). EL was computed from the data obtained with both stimuli (uniform and combined).

All data above were fitted by adjusting ξ_0. This parameter, which is the only free parameter of the model, was found to be SOA dependent, and determined the effective range of correlation measurement. For small SOA (60 msec), the best fit was with $\xi_0 = 0.5$ deg, i.e. 1.6 times the element size. For 160 msec SOA, the optimal value of ξ_0 was found to be 1.5 deg, which corresponds approximately to 5 times the element size (50 pixels). Finally, for the case in which no mask was used, ξ_0 was found to be equal to 6 deg, that is, 20 times the element size.

DISCUSSION

We studied the role of proximity and similarity cues (shape and luminance) and their interactions in visual tasks involving perceptual organizations. Psychophysical experiments demonstrated that all different cues affect grouping performance, but on different time scales. Thus, proximity grouping seems to evolve faster than similarity grouping. The experimental data were successfully modeled by an autocorrelation function assuming an increased weight for short-range correlations.

Psychophysical results

Grouping processes appeared to be very sensitive to changes in spacing ratio. As for shape similarity,

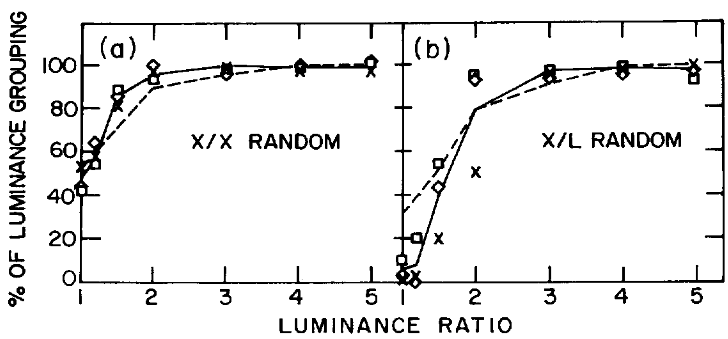

FIGURE 10. The graphs represent the percentage of luminance grouping as a function of the luminance ratio. Uniform display X/X (a), combined display X/L (b). The different symbols represent the data from the different observers, the solid line being their average and the dashed line representing the prediction of the model.

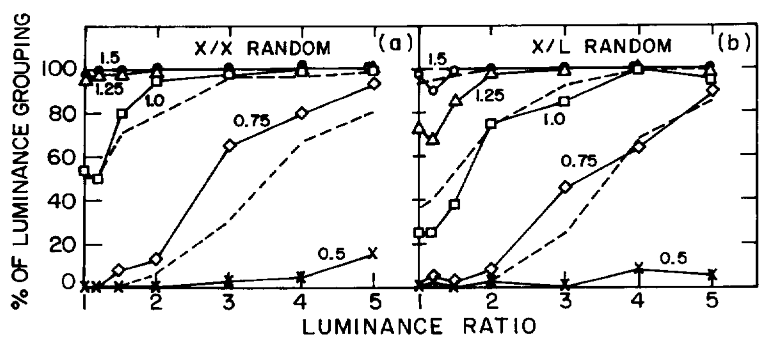

FIGURE 11. The graphs represent the percentage of luminance grouping as a function of the luminance ratio. Each curve corresponds to a different spacing ratio (different symbols, see the text). (a) The data for uniform display, and (b) for the combined display. The solid line describes the psychophysical data obtained for one of the subjects and the dashed line describes the prediction of the model.

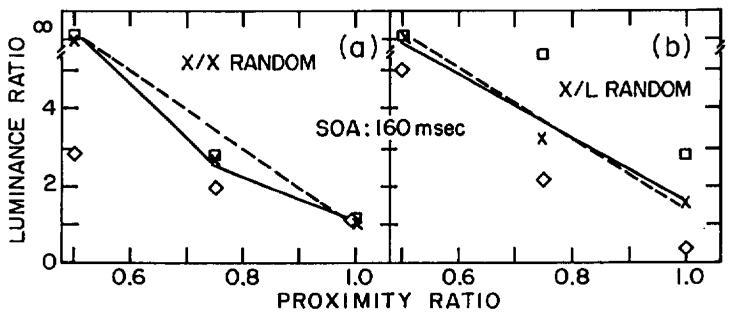

FIGURE 12. The graphs represent the luminance ratio required for each spacing ratio in order to get equilibrium (50% grouping in either direction). Consequently, the horizontal axis represents the spacing ratio and the vertical axis the luminance ratio. (a) The X/X display, (b) the X/L display. The different symbols depict the data for the different observers, the solid line is the average across observers and the dashed line represents the model prediction.

interestingly, the two micropatterns used, which produce effortless texture segmentation and error free pop-out detection at SOA of 100 msec (Bergen & Julesz, 1983) did not produce error free performance in our case. The dissimilarity of these two micropatterns was detected at SOA of 160 msec and at the no mask condition. Although shape similarity is a slowly developing process (relative to proximity), it becomes dominant once it is fully developed. It is possible that the fast proximity based phase of grouping is carried out by a parallel preattentive system with a high spatial resolution (Sagi & Julesz, 1985), while the slower, shape based grouping process is taken care by a resource limited attentive system. Indeed, recently we have shown that grouping processes involve attentive processing (Ben-Av *et al.*, 1992).

As for luminance grouping, the results indicate good grouping performance for contrasts above 20% (luminance ratio of 1.5) when SOA is more than 160 msec. Shape similarity and luminance similarity were set so that they were always competing with one another. When using equidistant elements (spacing ratio of 1), the experiments show that the equilibrium between shape similarity and luminance is reached at a luminance ratio of approx. 2 (average across observers). With spacing ratios different from 1, competitive and cooperative situations are created between proximity grouping and one of the similarity parameters (shape or luminance). For example with a spacing ratio of 0.5 even a luminance ratio of 5 does not allow for luminance grouping. Different performances were obtained for different SOAs.

The autocorrelation model

Autocorrelation may not be the most obvious model. An obvious approach would have been to model perceptual grouping organization on the basis of a similarity metric, that is, by grouping together elements that share common features (Beck, 1966). However, we had more success when using a model based on a covariance metric: the autocorrelation function. In preliminary studies, we investigated a "similarity metric" that was basically constructed by replacing the multiplication in the definition of the autocorrelation function by a subtraction. The absolute value of the difference was considered, this similarity metric did not account for the full experimental data. An analogous result was reported by Werkhoven, Snippe and Koenderink (1990), in an attempt to model low-level motion perception. This emphasizes again the importance of the autocorrelation function in modeling human performances (Uttal, 1975). The model proposed in this paper compares the weighted sums of the horizontal and vertical autocorrelations of the pixel intensities. Although shape similarity is not defined explicitly, the model still accounts for the shape similarity effect. A critical ingredient in our model is the spatial correlation weighting function. The exponential weighting function used ($e^{-\xi/\xi_0}$) enhances the connections between close elements and strongly decreases the effective correlation between distant elements. This model simulates successfully the psychophysical data obtained, i.e. the relationship between the different spacing ratios, the similarity (X/X vs X/L) and the different luminance ratios when brought into competition or into cooperation.

Note that the model assumes only one free parameter, the effective correlation range ξ_0. As grouping behaviour was found to change with the processing time (SOA) given to the observers, we found it necessary to use increasing values of ξ_0 for increasing SOA values. The optimal ξ_0 values found were 0.5, 1.5 and 6 deg for SOA values of 60 msec, 160 msec and infinity (no mask) respectively. This implies that the range of neuronal interactions required to establish correlations between two retinal positions is limited by activity duration, probably due to a finite speed of lateral activity propagation in the visual system. A crude estimate of the speed of activity propagation yields a velocity of about 10–20 deg/sec. The estimate seems somewhat low when considering information propagation in neuronal axons, however, it is possible that long range transmission is established through multiple links [e.g. synapses (see Polat & Sagi, 1994b)]. Interestingly, a recent study of the cortical point-spread function using real-time optical imaging of Macaque monkey primary visual cortex, (Grinvald, Lieke, Frostig & Hildesheim, 1994) indicates a velocity of cortical activity spread of between 10 and 20 cm/sec, supporting a relatively slow propagation of activity.

It is quite surprising that this simple weighted autocorrelation model accounts so well for the detailed psychophysical data presented here. It is known that the visual system filters the incoming luminance data by using local spatial filters, but here we needed to make no assumptions about this early filtering stage. However it is clear that there exists a large set of reasonable filters whose application would not affect our model predictions. Also, it would be possible to use the global frequency power spectrum of the input image, which is the Fourier transform of the autocorrelation function (Bracewell, 1986). In this case the correlation weighting function would be replaced by a frequency weighting function (the Fourier transform of the spatial weighting function), operating on the power spectrum of the image, thus requiring the existence of global Fourier analysers operating on regions of visual field as large as (6×6 deg). This type of global analysis is not consistent with the known structure of early vision. Rather, it is more plausible to assume that a full account of grouping processes involves the operation of local filters with long-range interactions between them. On this account, one has to consider multiplicative operations (formalized in terms of the autocorrelation function) applied to images resulting from a convolution between the input image and the spatial filters. The autocorrelation of the resulting image is a convolution between the autocorrelation functions of the input image and the corresponding filter (Papoulis, 1962). The total (non-weighted) directional correlations, obtained by integrating across all correlation distances in a given direction, differ from those of the original image (without filtering) by a scaling factor given by the total directional autocorrelation of the filter applied (i.e. its power at $\omega = 0$). [When considering both filtering and correlation weighting, one should look at the "equivalent" weighting function obtained by convolving the original weighting function with the filter directional autocorrelation function.] Thus oriented filters, like the ones used to model early vision (e.g. Gabor filters), can be used for estimations of directional autocorrelations, with the necessary addition of long range multiplicative interactions. It is possible that the same mechanisms underlying the recently observed collinear facilitation between local oriented filters (Polat & Sagi, 1994a b) can be used to obtain the directional correlations used in the present model. Also, it still remains to be seen whether the proposed grouping model can be unified with models of other early visual processes, such as of texture segmentation. Texture segmentation, being a fast short range process, may share some of the short-range interactions with the grouping process and thus may be involved in the initial phase of grouping.

REFERENCES

Beck, J. (1966). Perceptual grouping produced by changes in orientation and shape. *Science, 154*, 538–540.

Ben-Av, M. B. (1992). The role of similarity, proximity and attention in visual grouping. Doctoral dissertation, The Weizmann Institute of Science, Rehovot, Israel.

Ben-Av, M. B., Sagi, D. & Braun, J. (1992). Visual attention and perceptual grouping. *Perception Psychophysics, 52*, 277–294.

Bergen, J. R. & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature (London), 303*, 696–698.

Bracewell, R. N. (1986). *The Fourier transform and its applications.* New York: McGraw-Hill.

Braun, J. & Sagi, D. (1990). Vision outside the focus of attention. *Perception & Psychophysics, 48*, 45–58.

Fogel, I. & Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics, 61*, 103–113.

Ginsburg, A. P. (1971). Psychological correlates of a model of the human visual system. Proceedings NAECON, Dayton, Ohio. *IEEE Transactions on Aerospace and Electronics Systems, 71-C-24-AES*, 283–290.

Ginsburg, A. P. (1980). Specifying relevant spatial information for image evaluation and display design: An explanation of how we see certain objects. *Proceedings of the SID, 21*, 219–227.

Grinvald, A., Lieke, E. E., Frostig, R. D. & Hildesheim, R. (1994). Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *Journal of Neuroscience, 14*, 2545–2568.

Jáñez, L. (1984). Visual grouping without low spatial frequencies. *Vision Research, 24*, 271–274.

Julesz, B. (1980). Spatial nonlinearities in the instantaneous perception of textures with identical power spectra. *Philosophical Transactions of the Royal Society of London B, 290*, 83–94.

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

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 U.S.A.*, *90*, 7495–7497.

Landy, S. L. & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research*, *31*, 679–691.

Mack, A., Tang, B., Tuma, R., Kahn S. & Rock, I. (1992). Perceptual organization and attention. *Cognitive Psychology*, *24*, 475–501.

Malik, J. & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, *7*, 923–932.

Olson, R. & Attneave, F. (1970). What variables produce similarity-grouping. *American Journal of Psychology*, *83*, 1–21.

Papoulis, A. (1962). *The Fourier integral and its applications.* New York: McGraw-Hill.

Polat, U. & Sagi, D. (1994a). The architecture of perceptual spatial interactions. *Vision Research*, *34*, 73–78.

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

Priestley, M. B. (1981). *Spectral analysis and time series.* London: Academic Press.

Rubenstein, B. S. & Sagi, D. (1990). Spatial variability as a limiting factor in texture discrimination tasks: Implications for performance asymmetries. *Journal of the Optical Society of America A*, *7*, 1632–1643.

Sagi, D. (1991). Spatial filters in texture segmentation tasks. In Blum, B. (Ed.), *Channels in the visual nervous system: Neurophysiology, psychophysics and models.* London: Freund.

Sagi, D. & Julesz, B. (1985). "Where" and "What" in vision. *Science*, *228*, 1217–1219.

Sutter, A., Beck, J. & Graham, N. (1989). Contrast and spatial variables in texture segregation: Testing a simple spatial-frequency channels model. *Perception & Psychophysics*, *46*, 312–332.

Uttal, W. R. (1975). *An autocorrelation theory of form detection.* Hillsdale, N.J.: Erlbaum.

Werkhoven, P., Snippe, H. P. & Koenderink, J. J. (1990). Metrics for the strength of low-level motion perception. *Journal of Visual Communication and Image Representation*, *1*, 176–188.

Wertheimer, W. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung 4*, 301–350. Abridged translation: Principles of perceptual organization. In Beardsler, D. C. & Wertheimer, M. (Eds), *Reading in perception.* New York: Van Nostrand.

Zucker, S. W., Stevens, A. K. & Sander, P. (1983). The relation between proximity and brightness similarity in dot patterns. *Perception & Psychophysics*, *34*, 513–522.

Acknowledgements—We thank Drs Maggie Shiffrar and Daniel Glaser for helpful comments on the manuscript, Yehuda Barbut for figure preparation, Oded Smikt, Leah Mory and Nava Shaya for technical assistance.