*Perception & Psychophysics*
*1979, Vol. 25 (3), 225-231*

# The order of visual processing: "Top-down," "bottom-up," or "middle-out"

R. A. KINCHLA and J. M. WOLFE
*Princeton University, Princeton, New Jersey 08540*

This paper deals with the order in which different levels of form are recognized in a visual image. An experiment is reported in which the size of a tachistoscopically viewed image was varied. The results suggest neither an invariant "top-down" (gross shapes first followed by lower-order details) or "bottom-up" (the opposite) sequence. Rather, they seem to suggest a sort of "middle-out" sequence: forms at some intermediate level of structure having an optimal size or spatial-frequency spectrum are processed first, with subsequent processing of both higher and lower levels of form.

Contemporary theories of visual information processing have been strongly influenced by work on computer pattern recognition, particularly the concept of image structure employed in syntactic scene analysis (see Fu, 1974). The idea is that an image (a "scene") can be parsed into hierarchical levels of form, much as a paragraph can be parsed into sentences, phrases, words, etc. The value of such a representation stems from the redundancy (correlation) between different levels of structure in most scenes. For example, the fact that a scene contains a "head" is highly correlated with the presence of such lower-order components as "eyes," "ears," and "nose," etc. Knowledge of such structural redundancy can facilitate computer processing of images, just as syntactical redundancy facilitates the processing of language.

This redundancy can be utilized in a "top-down" sequence of processing, where the identification of a higher-order form (e.g., head) facilitates its subsequent analysis into lower-order components (e.g., eyes, ears, etc.) or in a "bottom-up" sequence, whereby identification of the lower-order components facilitates their subsequent synthesis into the higher-order form. Redundancy has also been employed in more complicated sequences involving both top-down and bottom-up components, for example, "analysis by synthesis," in which a tentative synthesis of components into a higher-order form facilitates subsequent analysis of previously ambiguous components (e.g., Halle & Stevens, 1962).

Theories of visual perception were influenced by this work. For example, the electrophysiological analysis of "receptive fields" seemed to suggest a bottom-up mode of processing: Cells associated with progressively more complex fields (visual stimuli) were found as one went from the retina to the visual cortex, as if systems for detecting low-order "features" ("points," "edges," etc.) fed into systems for detecting progressively more complex patterns. A whole family of psychophysical models was developed depicting visual information processing as a sequence of successive stages, beginning with low-order feature extraction (e.g., Rumelhart, 1970). While Neisser (1967) proposed an essentially bottom-up ("constructive") mode of visual processing, he also suggested that operations analogous to analysis by synthesis might play an important role.

More recently, it has been suggested that the order of visual processing is best described as a top-down process, with higher-order forms processed first, followed by subsequent analysis of progressively lower-order forms (e.g., see Broadbent, 1977; Kahneman, 1973). Of particular relevance here is evidence presented by Navon (1977) for what he termed a "global-to-local" sequence of processing.[1] He employed stimuli of a type originally suggested by this author (Kinchla, 1974) for studying the perception of different levels of structure. They are large letters made up of smaller ones, as illustrated in Figure 1. The advantage of using such stimuli rather than more naturalistic scenes, is that the relation between the "higher-" (large-letter) and "lower-order" (small-letter) forms may be manipulated freely, whereas the relation between levels of form in most images is constrained by natural laws; e.g., there is a single biologically correct relation between the components of a face and its parts. Furthermore, the familiarity and complexity of the forms at each level should be about the same.

One of Navon's most impresive experimental findings was obtained when he asked subjects to respond as rapidly as possible to stimuli like those

---

This paper was originally presented at the annual meeting of the Psychonomic Society in Washington, D.C., in November 1977. The work was supported by NSF Grant BNS-76 1082, on which the senior author is principal investigator. Requests for reprints should be sent to R.A. Kinchla, Department of Psychology, Princeton University, Princeton, New Jersey 08540.

0031-5117/79/030225-07$00.95/0

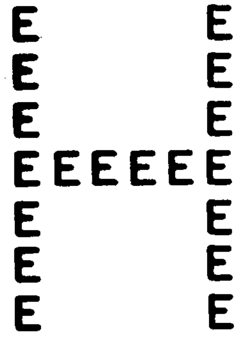

**Figure 1. An example of the type of stimulus patterns employed in the experiment: a large letter made up of smaller ones.**

in Figure 1 under two different conditions. In a "global-directed condition," they were told to base their responses solely on the large letter, pressing one key if it was an H or another key if it was an E. In the other, "local-directed," condition, they were told to make a similar decision based solely on the smaller letters. Navon found that subjects could successfully ignore the small letters in the global-directed condition, responding as rapidly to the large letter whether it was consistent with the smaller letters or not. However, subjects in the local-directed condition couldn't ignore the large letter, responding faster to the small letters when they were the same as the large one. Navon argued that this demonstrated the "inevitability of global processing": subjects had to process the large letter first in both conditions. This explained why they could initiate their responses before the interfering effects of the small letters in the global-directed condition, but couldn't avoid the interfering effect of the large letter in the local-directed condition.

However, it seems clear that there must be some limit to the precedence of the higher-order form. Shapes which are very large must project identifying contours on low-resolution regions of the retina, while the constituent lower-order forms fall on high-acuity regions near the fovea. A study by Vurpillot, Ruel, and Castrec (1977) utilized large forms made up of smaller ones presented to infants at various viewing distances. They found that prolonged free viewing produced habituation (a reduction in the child's tendency to fixate on a pattern) primarily to the large form at longer viewing distances and to the small form at closer distances. This suggests that the infants' tendency to recognize either the larger or constituent forms depended on viewing distance.

The experiment reported in this paper seems to imply that a global-to-local sequence of processing isn't inevitable and that the consistent pattern of results in Navon's study is simply due to the limited range of visual angles subtended by his stimuli. The present results suggest a different sequence of visual processing, one which could be described as neither "top-down" nor "bottom-up," but as "middle-out."

It should be emphasized that the phrase "middle-out" is meant to imply that the subject initially accesses some intermediate level of structural knowledge with subsequent associative activation of both higher and lower levels of hierarchically organized structural information. It is *not* meant to refer to the middle stage in a sequence of processing stages of the sort found in many current theories of information processing (e.g., Rumelhart, 1970). This argument will be developed further in reference to the following experiment.

## AN EXPERIMENT

This experiment was designed to assess how the angular size of an image determines the order in which its components are perceived. The basic experimental task consisted of a series of trials. At the beginning of each trial, a subject heard the name of a target letter for that trial: E, H, or S. Next, a stimulus pattern of the sort shown in Figure 1 (a large letter made up of smaller ones) appeared for 100 msec on a screen directly in front of the subject. His task was to decide as quickly as possible whether the target letter appeared in the display, pressing one button if *either* the large or the small letters corresponded to the target (a "yes" response) and another button otherwise (a "no" response). Of principle interest was the relative speed of "yes" responses to large and small target letters as a function of the angular size of the display. During each testing session, the height of the large letter subtended, with equal probability on each trial, 4.8°, 6.7°, 8.0°, 10.3°, or 22.1° visual angle.

### Method

**Stimuli.** Each stimulus pattern was a large letter, E, H, or S, composed of an appropriate six-column by seven-row array of smaller letters (as in Figure 1). The small letters were all the same, E, H, or S, and different from the large letter they defined. Thus, there were a total of six possible patterns. Each was typed in conventional uppercase font on white paper and photographed from various distances to produce the desired range of angular sizes when rear-projected onto a screen 64 cm in front of the subject. The projection slides were printed as negatives, so the stimuli appeared as white letters on a dark background. The white region had a luminance of 822.3 $cd/m^2$, and the dark 30.8 $cd/m^2$. The large letter was twice as tall as it was wide and 4.8 times as tall as each small letter. The small letters were 1.25 times as tall as they were wide. Each of the six patterns was photographed from five different distances to produce a set of 30 slides.

Each subject sat in a dimly illuminated room, 64 cm in front of a frosted glass screen on which the stimuli could be rear-projected. A dim, but clearly visible fixation point, consisting of a .1°-diam circle of light, was present in the center of the screen throughout each session. The projected test stimuli were centered on this point.

**Procedure.** On each trial, the visual stimulus was presented 3 sec after the subject heard the target letter defined.Each subject was told to respond as "quickly as possible after the stimulus appears, pressing the 'yes' button if either the large or small letter corresponds to the target letter, and the 'no' button otherwise." The subjects were also told to "avoid making errors, which under no circumstances should occur more than once or twice a session."

Three, paid, male college-student subjects performed the task in sessions of 90 randomly ordered trials, with each of the 30 stimulus slides presented three times within each session. Prior to each session, there were 20 unrecorded practice trials which were a random subset of the subsequent 90 test trials. The practice trials plus the 90 test trials in each session took about 25 min to complete. Each subject completed five sessions of trials for a total of 450 recorded trials. Thus, at each angular size, there were 30 trials when the large letter was the target, 30 when the small letters were, and 30 when neither was the target; i.e., a "yes" response was appropriate on two-thirds of the trials.

## Results

Figure 2 summarizes each subject's reaction time data, as well as the average of these data for the three subjects. Each graph presents the mean reaction time for "no" responses (open circles), for "yes" responses to large target letters (closed squares), and for "yes" responses to small target letters (closed triangles). The average and standard deviation of the latencies of each response at each visual angle for each observer are presented in Table 1. Subjects were virtually always correct; the proportions of error trials for subjects 1,2, and 3 were .006, .018, and .024, respectively. Thus, each data point on the individual subject graphs in Figure 2a is based on approximately 30 correct responses.

The general pattern of results was the same for all three subjects (and is readily apparent in the average data graph in Figure 2b). "No" responses generally took longer than "yes" responses, and there is a crossover interaction between the speed of a "yes" response to large and small targets and the visual angle of the display. At the smaller visual angles, a large target letter evoked the fastest "yes" response, while at the larger angles, the small target letters did.

A two-factor (target size and visual angle) within-subjects analysis of variance was performed on the mean latencies of "yes" responses for the three subjects. There were three means in each of the 10 cells defined by large or small target and the five visual angles. While neither of the main effects (target size and visual angle) was statistically significant, the interaction between these two factors was significant [$F(4,8) = 21.28$, $p < .001$]. The "no" responses were not evaluated in the analysis of variance. However, 14 of the 15 latencies for "no" responses shown in Figure 2a were greater than "yes" responses to either large or small targets, which indicates a statistically significant difference between latencies for "no" and "yes" responses by a simple sign test ($p \leq .01$).

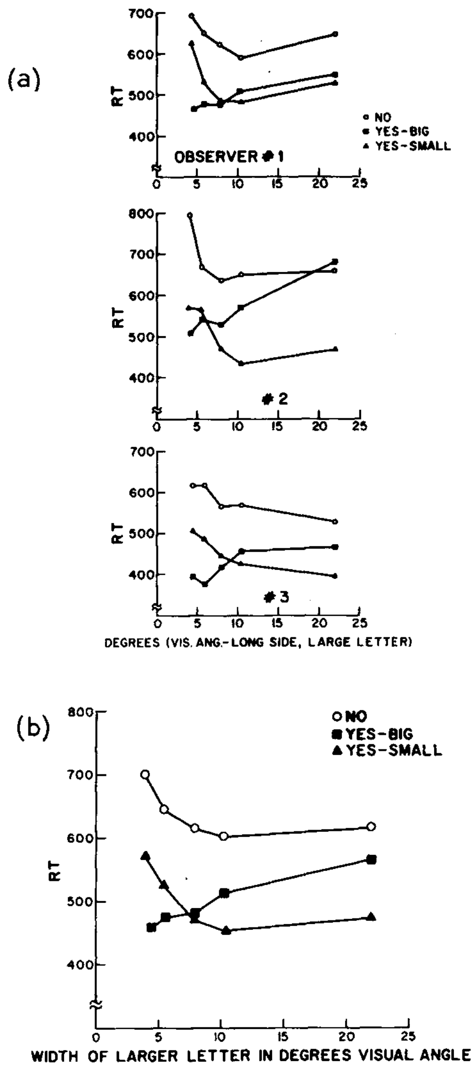

**Figure 2 (a) The average time to respond "yes" at each angular display size when the large letter was the target (solid squares), when the small letters were the target (solid triangles), and to respond "no" when neither was the target (open circles). (b) The average of the three performances shown in Figure 2a.**

## Discussion.

While the experiments conducted by Navon (1977) differed in a number of ways from the present experiment, one fact seems particularly relevant. In none of his experiments did the visual stimuli subsume more than 5.5° visual angle, and in most cases they were considerably smaller. The present results suggest that the large letters were processed faster than the small ones only when the display was smaller than about 6° to 9°. When the display was larger than this, the order appears reversed; small target letters seem to have been recognized faster than the large one. Thus, Navon's conclusion regarding the "inevitability of global processing" seems wrong.

It should be emphasized that the largest stimuli employed in this experiment had an angular extent appreciably smaller than a conventional sheet of typing paper viewed at arms' length. Furthermore, the subjects had no difficulty recognizing the largest letter, making virtually no errors and responding

**Table 1**

**Average ($\bar{X}$) and Standard Deviation ($\sigma$) of Response Latencies for Each Type of Response, at Each Visual Angle, for Each Observer, in Milliseconds**

| Response | Angle* | Observer 1 |  | Observer 2 |  | Observer 3 |  |
|---|---|---|---|---|---|---|---|
|  |  | $\bar{X}$ | $\sigma$ | $\bar{X}$ | $\sigma$ | $\bar{X}$ | $\sigma$ |
| Yes-Big | 25.0 | 553 | 151 | 682 | 85 | 466 | 108 |
|  | 11.5 | 509 | 105 | 571 | 82 | 459 | 104 |
|  | 9.0 | 478 | 105 | 531 | 75 | 418 | 94 |
|  | 6.4 | 480 | 94 | 542 | 114 | 378 | 72 |
|  | 5.4 | 467 | 103 | 507 | 89 | 393 | 95 |
| Yes-Small | 25.0 | 532 | 109 | 472 | 75 | 399 | 131 |
|  | 11.5 | 484 | 72 | 434 | 60 | 427 | 82 |
|  | 9.0 | 487 | 66 | 473 | 87 | 448 | 98 |
|  | 6.4 | 533 | 63 | 563 | 140 | 489 | 133 |
|  | 5.4 | 626 | 230 | 562 | 166 | 508 | 207 |
| No | 25.0 | 651 | 100 | 658 | 140 | 532 | 81 |
|  | 11.5 | 592 | 64 | 651 | 113 | 568 | 122 |
|  | 9.0 | 632 | 78 | 636 | 92 | 565 | 124 |
|  | 6.4 | 648 | 70 | 669 | 129 | 618 | 168 |
|  | 5.4 | 691 | 150 | 794 | 177 | 617 | 185 |

**In degrees.*

only about 100 msec more slowly than to the small ones, even though the display was presented for only 100 msec. Thus, the large stimuli were nowhere near the limits of the viewer's visual field and were in the range of objects which can be easily recognized with a single brief fixation. The small letters were also perfectly recognizable, even at the smallest visual angle, the only difference being that it took about 100 msec longer to recognize a small target letter in the smallest displays than it did in the largest.

## MIDDLE-OUT-PROCESSING

The preceding results imply that processing is neither consistently top-down nor bottom-up, but, rather, that forms having an optimal size in the visual field are processed first, with subsequent processing of *both* larger and smaller forms. Even though both the large and small letters were always perfectly perceptible, the large letters were processed more slowly when they were too large and the small letters more slowly when they were too small. Thus, it would seem that the initially processed components of a natural image would correspond to some intermediate level of structure with subsequent processing of forms at both "higher" and "lower" levels. This is what the phrase "middle-out processing" is intended to convey.

Of course, there is no question that factors such as the familiarity, complexity, and a priori probability of a form influence recognition latency. The point here is simply that if such factors are held constant there should still be an optimal angular size at which the form will be recognized most rapidly. If the form is made larger, its primary identifying characteristics may fall in regions of lower acuity. If it is made smaller, some of these characteristics may approach the limits of even foveal acuity.

It has been suggested that the "features" of our visual system are hierarchical spatial frequency bands (e.g. Ginsburg, 1976). Furthermore there is considerable evidence that we respond most rapidly to low spatial frequencies and progressively more slowly to higher frequencies (e.g., see Breitmeyer, 1975; Lupp, Hauske, & Wolf, 1976). Thus, one might argue that we process information in low spatial frequency bands first and in progressively higher bands later. This would be generally consistent with an invariant top-down sequence of processing, since progressively lower levels of form (as defined by syntactic scene analysis) would become recognizable as progressively higher spatial frequency bands were successively processed. In fact, this is a basic part of the argument developed by Navon (1977) to support his view of an invariant global-to-local sequence of visual processing.

However, in the present experiment, the larger the visual angle subtended by the large letter, the lower the spatial frequency band in which its identifying contours should fall. Yet the response latency to large target letters *increased* with their size. Of course, the overall relationship between response latency and size is certainly not monotonic. There is no question that latency to respond to large or small target letters (and error rates) would eventually increase if the letters were made sufficiently large or small. What is shown in Figure 2 is simply the point at which the latency function for large letters intersects that for the small letters. Extending the range of visual angles would eventually produce an increase in the latency to both types of targets; i.e., both functions must be essentially U-shaped, with the minimum for the large letters at a smaller visual angle than that of the small letters. Roughly speaking, one would expect a minimum latency to large-letter targets at visual angles one-fifth as large as the minimum for small-letter targets, since the small letters were about one-fifth the size of the larger ones. The average results shown in Figure 2b are not inconsistent with this prediction. The latency function for large targets (solid squares) seems to reach a minimum around 10°. This would suggest that the minimum latency for large letters should be around 2°, which is not inconsistent with the latency function for large targets in Figure 2b.

Of course, one would *not* expect the latency function for large targets to be a simple transformation of that for small targets. The small letters are surrounded by other small letters, while the large letters are not (thus the effect of lateral interactions should not be identical). The small letters are also made up of continuous contours, while the large letters are not (they are made up of small letters). Nevertheless, the present results suggest that observers may be fastest at identifying letters when they subtend about 2° visual angle in height, although this conclusion is highly tentative and undoubtably depends on factors such as the particular letter font as well as luminosity, contrast, etc.

It also seems worth commenting on the relationship between the optimal size of a letter in degrees visual angle and the idea of optimal spatial frequency bands. The spatial frequency power spectrum of a particular letter specifies the energy present at each orientation and frequency of its simple sinusoidal components (as well as phase information). Varying the size of a letter produces a systematic shift in its power spectrum. For example, if the letter were made five times larger, each energy component would be associated with a spatial frequency one-fifth as large. If a letter subtending a certain visual angle is identified most rapidly, one could argue its power spectrum at that viewing size was optimal for fast identification. However, this would not indicate the specific contribution of each orientation and frequency to the identification process. Not only would different components contribute more or less to the identification of certain letters, but the information in these components would be highly redundant; e.g., two nonoverlapping spatial frequency bands might each contain information sufficient for identification of the letter.

Thus, while the present results suggest there is an optimal letter size for *fast* identification, and this is associated with a particular spatial frequency power spectrum, it is not possible to specify exactly which components of this spectrum were involved in the identification process. While the modulator transfer function (MTF) of the human eye indicates that we are most sensitive to (need the least contrast to detect) sinusoidal gratings around 1 or 2 Hz, we don't know which frequency components were used in identifying the letters in the present study. Furthermore, our task involved *speeded identification* of letters rather than simply *detecting* the presence of a sinusoidal grating.[2]

## IS STRUCTURAL KNOWLEDGE HIERARCHICAL?

It seems useful to distinguish between the *structure of an image* and our *knowledge of structure.* The techniques of syntactic scene analysis clearly allow one to represent the structure of an image in a hierarchical form, one which has a "top" and a "bottom." For example, consider a newspaper photograph of a face. The highest-order form in the scene, the "face," defines the "top" of the hierarchy of forms, while the lowest-order forms, the individual dots which make up a printed photograph, define the "bottom." However, one's *knowledge* of such a structure hasn't any "top" or "bottom." One knows the "face" is undoubtedly part of a higher-order structure termed a "person" even though only the face is visible. One also knows about very low-order facial structures such as "skin pores" and "facial hairs," which may be much finer details than are actually represented in the printed photograph.

The same argument can be made concerning the "images" our visual system transmits to the brain. They have a "top" and "bottom" just like the photograph. There is an upper limit to the visual angle a form can subsume in our visual field, and a limit to the fine detail our eyes can actually resolve. Yet our knowledge of structure allows us to infer the nature of forms both above and below these limits, just as one may infer the presence of a "person" and "skin-pores" from the photograph of a "face," even if neither level of form is represented directly.

While one's knowledge of structure has no "top" or "bottom," it does seem useful to consider it as hierarchically organized much like the syntactically defined structure of a scene. It is proposed here that a major reason for such an organization is the life-long sequential pattern of our visual experience whereby recognition of a form at one level of structure is an almost invariant precursor of the recognition of forms at levels slightly higher or lower. This sequence may even occur during the processing of a single "look" or fixation, since it seems likely that those forms whose identifying features occupy an optimal size in our visual field may be recognized prior to those which are either larger or smaller.

More important, perhaps, is the manner in which forms pass in and out of our range of vision as we move toward or away from particular objects. As we move toward a person, we may see only his general outline. As we move closer, lower-order forms such as the general outlines of his face become apparent. If we move very close to the person, even lower-order forms such as eyes, mouth, nose, etc., may become most easily recognizable. A similar, but opposite, sequence occurs as we move away from things: progressively higher order forms successively replace those previously most easily recognized.

This succession of "most recognizable" forms may be thought of in terms of size in the visual field or in terms of an optimal band of spatial frequencies. Specific contours of an object will be represented by progressively higher spatial frequency components as one moves away from it and by progressively lower ones as one moves toward it.

As this author (Kinchla, 1977) and others (e.g., Ginsburg, 1976; Palmer, 1976) have pointed out, there is a close relationship between the hierarchy of forms one defines through syntactic scene analysis and the hierarchy one could define by progressively removing lower and lower spatial frequency components of an image (much as you do by progressively "blurring" a projected image when you slowly defocus a projector's lens). The order in which components of the image become unrecognizable as you progressively remove lower frequency components is very similar to the levels of structure in syntactic scene analysis. The lowest-order forms become unrecognizable first when only the highest frequencies have been removed (the image is slightly blurred). The highest-order forms remain recognizable even when only the lowest-frequency components are present (at high levels of "blurring"). Thus, as you move toward an object, progressively lower-order components of its structure are successively recognized as the spatial frequency ranges of their contours become high enough to be seen. Exactly the opposite sequence is generated as you move away from something. Thus, recognition of one level of form becomes a reliable cue for the subsequent recognition of both higher and lower levels.

Use of such cues is such an automatic, highly practiced, and usually accurate part of normal perception that it is often difficult to decide whether a particular form was actually "seen," in the sense that the presence of its identifying contours on the retina determined our perception, or whether its presence was simply inferred from the recognition of a highly correlated (redundant) form. For example, suppose you photographed a drawing of a face after erasing some detail such as an eye, you could then briefly flash the photograph on a screen so that it was clearly recognizable as a face even though the details of the face were barely perceptible. It would be easy to show that some subjects would quite confidently report "seeing" the eye which had actually been erased. It would be clear that they had simply inferred the presence of this lower-order form from their recognition of the higher-order form, the "face," i.e., the "face" served as an associative cue for the perception of the eye without the subject's awareness.

A mathematical model of the manner in which structural redundancy influences the processing of visual information has been developed by Kinchla (1977). Subjects evaluated visual arrays consisting of two large letters, "T" and "L," composed of many different small letters. The subjects' task was to decide whether a small letter "F" was among the small letters present in each display. Structural redundancy was introduced by making the target letter (F) twice as probable in one large letter as the other during each testing session. The model characterized the subjects as evaluating imperfect ("noisy") impressions of the small letters in each large letter through a *weighted integration process,* assigning more weight to the impressions from the large letter most likely to contain the target. This interpretation was shown to account for the data without the introduction of any "attentional" process whereby the subject "allocated more attention" to that large letter which was (a priori) most likely to contain the target. The results of the present study suggest that the large letters may have been recognized as much as 100 msec earlier than the small ones. This would have provided sufficient time for "reallocation of attention" according to some earlier estimates of such mechanisms (e.g., see Eriksen & Hoffman, 1974). Nevertheless, it may be that processing of the larger letters required a commitment of processing capacity to an extent which makes reallocation impractical. This view suggests that the attentional mechanisms postulated by other authors, such as Sperling and Melchner (in press), Shaw (1978), and Posner, Nissen, and Ogden (1978), may require more time to reallocate attention than that suggested by the Eriksen and Hoffman (1974) results cited earlier.

## REFERENCES

Breitmeyer, B. G. Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision Research*, 1975, 15, 1411-1412.

Broadbent, D. E. The hidden preattentive processes. *American Psychologist*, 1977, **32**, 109-118.

Eriksen, C. W., & Hoffman, J. E. Selective attention: Noise suppression or signal enhancement? *Bulletin of the Psychonomic Society*, 1974, **4**, 587-589.

Fu, K. S. *Syntactic methods in pattern recognition*. New York: Academic Press, 1974.

Ginsburg, A. The perception of visual form: A two-dimensional filter analysis. In V. D. Gleger (Ed.), *Proceedings of the Fourth Symposium on Sensory System Physiology* (Information Processing in the Visual System). Leningrad: Russian Academy of Sciences, 1976.

Halle, M., & Stevens, K. Speech recognition: A model and a program of research. *IRD Transactions on Information Theory*, 1962, IT-8, 155-159.

Kahneman, D. *Attention and effort*. Englewood Cliffs, N.J: Prentice-Hall, 1973.

Kinchla, R. A. Detecting target elements in multi-element arrays: A confusability model. *Perception & Psychophysics*, 1974, **15**, 149-158.

Kinchla, R. A. The role of structural redundancy in the perception of visual targets. *Perception & Psychophysics*, 1977, **22**, 19-30.

Lupp, U., Hauske, G., & Wolf, W. Perceptual latencies to sinusoidal gratings. *Vision Research*, 1976, **16**, 969-972.

Navon, D. Forest before the trees: The precedence of global features in visual perception. *Cognitive Psychology*, 1977, **9**, 353-383.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Palmer, S. E. Visual perception and world knowledge. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco: Freeman, 1976.

Pomerantz, J. R., & Sager, L. C. Asymmetric integrality with dimensions of visual pattern. *Perception & Psychophysics*, 1975, **18**, 460-466.

Posner, M. I., Nissen, M. J., & Ogden, W. C. Attended and unattended processing modes: The role of set for spatial location. In H. L. Pick & I. J. Saltzman (Eds.), *Modes of perceiving and processing information*. Hillsdale, N.J: Erlbaum, 1978.

Rumelhart, D. E. A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, 1970, **7**, 191-218.

Shaw, M. L. A capacity allocation model for reaction time. *Journal of Experimental Psychology: Perception and Human Performance*, 1978, **4**, 586-598.

Sperling, G., & Melchner, M. J. Visual search, visual attention, and the attention operating characteristic. In J. Requin (Ed.), *Attention and performance VII*. New York: Academic Press, in press.

Sterling, N., & Coltheart, M. Stroop interference in a letter naming task. *Bulletin of the Psychonomic Society*, 1977, **10**, 31-34.

Vurpillot, E., Ruel, J., & Castrec, A. L'organisation perceptive chez le nourrisson: Réponse au tout ou à ses éléments. *Bulletin de Psychologie*, 1977, **30**, 396-405.

## NOTES

1. Sterling and Coltheart (1977) independently conducted an experiment similar to Navon's, in which subjects were told to rapidly say the name of a large letter composed of smaller ones. They found a Stroop-like interference effect of the small letters on speed of vocalization: subjects were slower to name the large letter when the small letters were a different letter. Since the largest array was about 3° of visual angle in height, their results seem inconsistent with Navon's. However, their large letters were composed of 5 by 5 or 5 by 7 arrays of small letters rather than the 6 by 7 arrays used by Navon. Furthermore, they employed a voice-key to measure vocalization latency rather than a two-alternative keypressing response. Thus, the studies aren't directly comparable.

Another study, one by Pomerantz and Sager (1975), required subjects to sort cards on which were printed large forms (letters and other forms) made up of small letters. At normal viewing distances, the large forms subtended about 2.5° of visual angle. The authors found interference from the irrelevant level of form whether the subject was asked to sort on the basis of the large or small forms. Again, however, it is difficult to compare their results to Navon's because of the numerous procedural differences.

2. The apparently positive monotone relationship between reaction time and spatial frequency should be interpreted with caution. First of all, the cited studies (Breitmeyer, 1975; Lupp et al., 1976) employed a detection (simple) reaction time procedure in which subjects responded as soon as they noticed any change in the display. It is not clear that a recognition (choice) reaction time task would imply a similar relationship between speed and spatial frequency, e.g., a task in which subjects made one response to a vertical sinusoidal grating and another to a horizontal grating. This is an important consideration, since the type of task considered in this paper clearly involved recognition rather than simply detecting the onset of a stimulus presentation. Furthermore, it should be noted that Lupp et al. (1976) report a preliminary study in which a "blocked" presentation of different spatial-frequency stimuli (rather than randomizing frequency from trial to trial as in the other studies) resulted in *faster* responses to the higher-frequency stimuli.

(Received for publication January 31, 1978;
revision accepted January 2, 1979.)