

Visual display recognition and the duplication of inspection sequences

JOHN A. WHITESIDE

Clark University, Worcester, Massachusetts 01610

In a test of Noton and Stark's (1971) contention that visual display recognition depends upon the duplication of the fixation sequence initially used in viewing the display, recognition performance and reaction times were recorded for six subjects in a visual recognition task. The task involved initial sequential presentation of elements of a display followed by presentation of a recognition probe that either did or did not duplicate the initial presentation sequence. Either performance or reaction times, or both, suffered when the initial presentation sequence was not duplicated, lending some support to Noton and Stark's position. However, these differences were less pronounced when the recognition probe was displayed for a longer time, indicating that non-sequence-dependent processes may become increasingly important at longer viewing times.

Noton and Stark have proposed a model of visual display recognition that is inspired by the observations that when viewers inspect various visual materials, their saccadic eye movements tend to describe cyclic, repeating patterns and further that these patterns tend to be duplicated during attempts at recognition of previously viewed figures (Noton, 1970; Noton & Stark, 1971a, 1971b, 1971c). They have attributed these regularities to what they argue is the sequence-dependent nature of the processes underlying visual recognition. There follows below a description of Noton and Stark's model, a brief summary of empirical work relating to it, and the rationale for the current experiment as a test of one of the model's predictions.

According to Noton and Stark's account, when a viewer is presented with an unfamiliar display, the fixation or attention shifts used to inspect the display will form a stable, repeating pattern called a scanpath. It is hypothesized that during the execution of the scanpath the viewer is laying down a memory trace of the display. The organization of this memory trace reflects the organization of the scanpath. The scanpath is a sequence or chain in which fixations on portions of the display alternate with fixation shifts; in a like fashion, the memory trace of the scanpath (termed a feature ring) consists of a sequence in which the representations of the features examined during fixations alternate with traces of the eye movements or attention shifts necessary to move from feature to feature.

The author would like to thank Marc B. Mandler, Leland S. Goodman, and Clifford M. Karmiohl for help with this work. Reprint requests should be sent to: John A. Whiteside, Department of Psychology, Clark University, Worcester, Massachusetts 01610.

Scanpaths underlie display recognition, according to the theory. Once the viewer has generated a stable scanpath (and so a memory trace) for a particular display, the same scanpath must be used in subsequent attempts to recognize the display or to distinguish it from other displays. That is, access to the sequentially organized memory trace depends upon the viewer duplicating the original sequence of fixations. As the viewer reexecutes the scanpath during an attempt at recognition, he/she sequentially checks the features as they are viewed against the memory traces of the features established by previous execution of the scanpath. The critical thing is that the order of inspection must be the same for both familiarization and recognition; otherwise, there is no basis for the recognition judgment.

Noton and Stark make no formal distinction between overt eye movements and covert attention shifts; both share the same sequential properties in their model. Whether one or the other occurs depends mainly on viewing conditions: the size of the field, peripheral acuity of the viewer, degree of illumination, and so forth.

That overt scanpaths exist seems certain. Jennerod, Gerin, and Pernier (1968), Locher and Nodine (1974), Noton and Stark (1971a, 1971b, 1971c), Whiteside (1974), and Yarbus (1967) have all reported that, during free inspection of various materials, subjects' eye movements tended to form stable, repeating patterns. Luria and Strauss (1975) reported that subjects did not exhibit reliable scanpaths during execution of a visual search task; however, their use of a search task is an inappropriate test of a model of visual familiarization and recognition.

Empirically, the role of scanpaths in recognition is unclear. Noton and Stark's data indicate that successful recognition was accompanied by duplication of

the scanpath originally used during familiarization. However, it is not clear that there were any instances of unsuccessful recognition in their experiment. Locher and Nodine (1974), in a recognition task, failed to find any correlation between successful performance and duplication of scanpaths during attempts at recognition.

In sum, Noton and Stark have proposed that both the creation of and access to internal representations of visual displays are governed by sequentially organized processes. The inspection sequence used in initial viewing of display limits the means of access to the memory of relations among features in the display; recognition can occur only if the initial sequence is duplicated. The purpose of the present experiment was to determine whether the processes underlying visual recognition are organized in the sequence-dependent way that Noton and Stark suggest. Their model holds that visual recognition proceeds sequentially, using the same order of fixation or attention shifts that was used during familiarization. From this it follows that recognition should suffer when a different order is used during recognition. The present experiment was designed to test this prediction, using both recognition performance and reaction time as measures of recognition ability.

To provide this test of Noton and Stark's model of visual recognition, it was necessary to insure that viewers inspect sets of features according to predetermined sequences; thus, the experiment required techniques for creating, rather than simply observing, inspection sequences and for controlling the availability of peripheral visual information. This was done by flashing various portions of a display in sequence, at a rate close to that of naturally occurring eye movements. Whiteside (1976) has shown that adults will make eye movements reliably in response to peripheral flashes of light and will continue to do so for many trials. It was hoped, by sequential presentation of features, to approximate the manner in which features are presented to the fovea as a result of naturally occurring fixation shifts. In addition, pilot testing suggested that the time available for inspection of the recognition stimulus played a critical role; therefore, this was included as a variable.

METHOD

Subjects

Six volunteers, five male and one female, between the ages of 21 and 30 served as subjects.

Apparatus

Stimuli were presented on a VR-12 cathode ray display device controlled by a PDP-12 computer, both manufactured by Digital Equipment Corporation. Two response buttons, also interfaced to the computer, were provided.

Stimuli

Four shapes (cross, plus, square, and triangle) projected on the display screen were used as stimuli. These are referred to below

as features. Each feature subtended 30' of visual angle, with a luminance of approximately 5 mL.

On any given trial, three of the features, randomly selected from the full set of four features, were flashed one by one in succession on the otherwise blank screen. Only three of the four were used in order to insure that the subject remain uncertain about the identity of the last feature until it was actually viewed. These three features are termed the familiarization stimulus below. Each feature was presented for 256 msec with no delay between successive presentations. The locations in which the features were presented formed the vertices of an equilateral triangle with each side subtending 2° of visual angle. The particular triangle flashed on any given trial was randomly selected from a set of eight possible equilateral triangles defined by 14 possible feature locations distributed throughout a 4° vertical × 8° horizontal viewing area. Figure 1 shows a representation of the locations in which features could be presented. The upper portion of the figure shows the locations used for the familiarization stimulus. The locations are labeled X, Y, and Z, and this order corresponds to the order in which features appeared in the respective locations. An arrow pointing from one location to another indicates that features could be presented in those locations in the order implied by the arrow. The first of the three features was always presented in one of two locations labeled X. The second feature then appeared in one of two locations labeled Y and also connected to the previously displayed X location. Similarly, the third feature, Z, was one of the two connected to whichever Y location had been used. In this way, the subject was always uncertain about which location would be presented next. Notice that the orientations of the familiarization triangles was such that no leg was either vertical or horizontal with respect to gravity. Each triangle occurred an equal number of times throughout the experiment according to a random schedule.

Following presentation of the familiarization stimulus, there was a 256-msec delay, after which two of the three familiarization features were presented, again one by one, in succession. The presentation of these two features is termed the probe below. The probe always duplicated two of the three familiarization features. The initial feature of the probe was presented 5° of

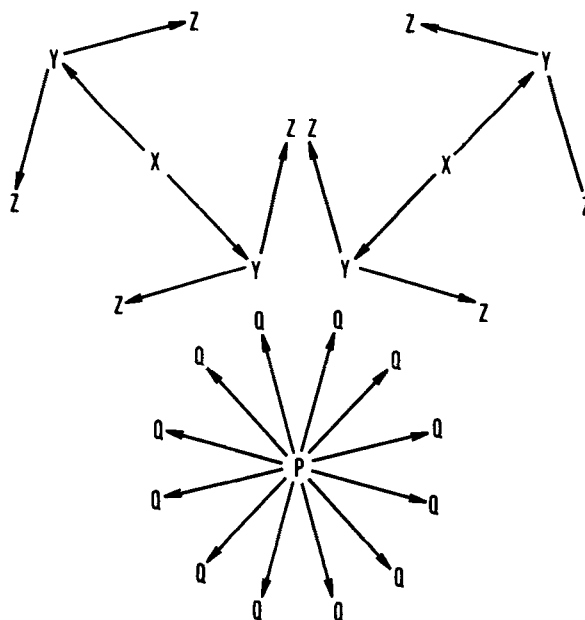


Figure 1. Representation of viewing area showing locations in which features were presented.

visual angle below the center of the $4^\circ \times 8^\circ$ area within which the familiarization stimulus was presented. The absolute location of the initial probe feature was constant throughout the experiment. The initial probe feature was presented for either 256 or 1,024 msec. After termination of the initial probe feature, the second and final probe feature was presented for the same duration as the initial probe feature. There were 12 possible locations for the second probe feature, arranged as though on the circumference of a circle around the fixed location of the first probe feature. This arrangement is shown in the lower portion of Figure 1. The location labeled P shows the fixed location of the first probe feature. The second probe feature occurred in one of the locations labeled Q, depending on which familiarization triangle had been used and on the nature of the trial. Notice that the lines drawn from the 12 points labeled Q to the point P have the property that every leg of each of the eight familiarization triangles has a line segment congruent (identical in length and orientation) to it in the probe display.

The identities, locations, and sequence of the probe features were varied in such a way as to create six probe conditions: the early duplicating, late duplicating, early reversing, late reversing, indirect, and implied conditions. These terms are merely intended to be descriptive, and are not intended to imply a priori logical relations among the conditions. In what follows, X, Y, and Z represent the three locations in which familiarization features were presented. X', Y' and Z' represent probe locations congruent, respectively, to X, Y, and Z, but displaced from them. A, B, and C represent three features. The notation X(A) \rightarrow Y(B) indicates that first feature A was presented in location X and then B was presented in location Y. Assume that the familiarization presentation involved the sequence: X(A) \rightarrow Y(B) \rightarrow Z(C). Given this, the probe sequence X'(A) \rightarrow Y'(B or C) constituted the *early duplicating* condition, while the sequence Y'(B) \rightarrow Z'(C or A) made up the *late duplicating* condition. These sequences are termed "duplicating" because both involve exact duplications of sequences of locations that were presented in direct succession during the familiarization stimulus. The early duplicating probe corresponded to the first two familiarization locations to be presented and the late duplicating probe corresponded to the last two familiarization locations. The probe sequences Y'(B) \rightarrow X'(A or C) and Z'(C) \rightarrow Y'(B or A) were designated the *early reversing* and *late reversing* conditions, respectively, because these probes involved sequences of locations that were the reverse of the sequences provided during the familiarization presentation. The probe sequence X'(A) \rightarrow Z'(C or B) made up the *indirect* condition, because this sequence, while not directly traversed in the familiarization triangle, was indirectly reached (via location Y). Finally, the probe sequence Z'(C) \rightarrow X'(A or B) constituted the *implied* condition, so-called because this sequence of locations, while neither directly nor indirectly scanned during the familiarization presentation, would close the triangle if the scan was continued. Notice that, in each probe condition, two alternatives are given for the second probe feature. The first alternative corresponds to probe trials in which the features presented actually existed in that spatial configuration in the familiarization stimulus. As detailed below, subjects were to answer "yes" following such trials. The second alternative involves a spatial configuration of probe features that was not present in the familiarization stimulus and a correct answer on such a trial was "no."

Procedure

Subjects were seated 45 cm from the viewing screen and told that on each trial they would view an initial sequence of three flashed features forming an equilateral triangle. Immediately following, two of the three features would be presented, in sequence, in such a way as to represent one of the legs of the initial triangle. The subjects were asked to indicate, as quickly and accurately as possible, whether or not the two probe features occupied locations congruent (that is identical except for displacement) to those

occupied by the same features in the original triangle. To indicate their responses, the subjects pressed one of two response buttons. Which button they pressed and the time between the onset of the second probe feature and the button push were automatically recorded.

Each subject was exposed to a total of 2,304 trials, but the first 1,152 trials were considered practice trials to allow the subjects to become thoroughly familiar with the task and the response. The subjects were tested over a period of 4 days, receiving 576 trials on each day. Half of the trials involved a long probe duration (1,024 msec per feature) and the other half involved a short probe duration (256 msec per feature). In addition, for half of the probe trials, the spatial arrangement of features was not congruent to the arrangement of the corresponding features in the familiarization stimulus (configuration-absent condition); for the other half of the probe trials, the arrangement was congruent (configuration-present condition). There were equal numbers of probe trials relating to each of the three legs of the familiarization triangle. The identities of the features and their assignment within the familiarization triangle was also randomly determined, as was the order in which the conditions were presented.

The subjects were given no prior information about the nature of each next trial. From the subjects' point of view, at any instant during a trial there was only a single feature on an otherwise blank screen. The nature of the presentation conditions outlined above insured that the subjects could not predict with certainty the absolute location, orientation, or direction of each familiarization triangle. Furthermore, no inference (other than a guess) concerning the exact identity of any probe trial could be made prior to the presentation of the second probe feature. The subjects were instructed to perform as quickly and accurately as possible on all conditions.

RESULTS

To analyze both correct performance and reaction time as a function of probe condition, an orthogonal set of planned contrasts was devised. These involved: (1) contrasting the duplicating and reversing (averaged for early and late conditions) conditions with the indirect and implied conditions; (2) contrasting the indirect with the implied condition; and (3) contrasting the duplicating with the reversing condition. These contrasts were chosen to reflect the experimental hypotheses and also to try to eliminate any confounding with time of presentation; in each case, the contrasts involve comparing conditions for which the average time between the presentation of the familiarization locations to be probed and the onset of the initial probe feature was constant. The two remaining degrees of freedom were used in a test of secondary interest from the viewpoint of the experimental hypotheses—a contrast between early and late presentations.

Figure 2 shows mean correct responses as a function of probe feature duration for the duplicating, reversing, indirect, and implied conditions. The duplicating and reversing curves represent the average of early and late presentations. Performance clearly improved with an increase in probe duration, $F(1,5) = 68.13, p < .001$. With respect to the overall effect

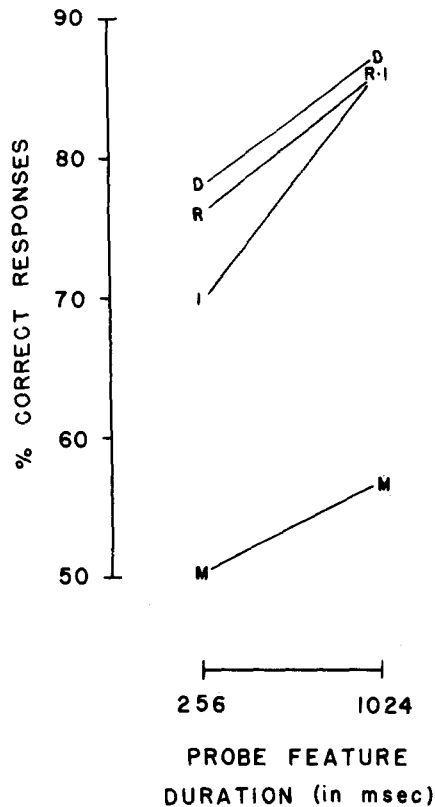


Figure 2. Mean percentage of correct responses as a function of probe feature duration and probe condition. In this figure and in Figure 3, probe conditions are identified by letters as follows: D = duplicating; R = reversing; I = indirect; and M = implied condition.

of probe condition, performance was poorer in the indirect and implied conditions than in the reversing and duplicating conditions, $F(1,25) = 55.91, p < .001$; however, these differences did not vary significantly as a function of probe duration. Performance in the implied condition was clearly worse than performance in the indirect condition, $F(1,25) = 53.76, p < .001$; the extent of this difference varied with probe duration, $F(1,25) = 6.82, p < .05$, in that performance in the indirect condition benefited more from the longer probe than did performance in the implied condition. Performance in the duplicating condition did not differ significantly from that in the reversing condition for either probe duration, and there was no significant difference between the early and late presentations.

Figure 3 shows mean reaction time in milliseconds as a function of probe speed and probe condition. Only reaction times from trials on which the subjects answered correctly are included. In computing these means, preliminary means were first computed for each subject and condition. All reaction times more than two standard deviations above or below the preliminary means were discarded and new means

were computed. The reaction times in Figure 3 are based on these corrected means. Less than 5% of the reaction times were discarded under this rule. The left-hand set of curves are derived from trials on which the probe configuration was actually present in the familiarization stimulus; the right-hand set refers to trials in which the probe configuration was not present in the familiarization stimulus. Reaction times were, on average, longer on the configuration-absent trials than on the configuration-present trials, $F(1,5) = 201.91, p < .001$. Increasing the probe duration had the general effect of lowering reaction times, $F(1,5) = 10.64, p < .05$. Overall, reaction times on the duplicating and reversing trials were faster than times on the indirect and implied trials, $F(1,25) = 67.10, p < .001$. This advantage was more pronounced on the configuration-present than on the configuration-absent trials, $F(1,25) = 6.92, p < .05$, and also more pronounced, on average, on the long than on the short probe trials, $F(1,25) = 15.24, p < .001$. Overall, reaction times were faster in the duplicating than in the reversing condition, $F(1,25) = 10.11, p < .01$, especially so for the fast as opposed to the slow probe, $F(1,25) = 5.05, p < .05$, and also for the configuration-present more than for the configuration-absent condition, $F(1,25) = 6.76, p < .05$. Reaction times on the implied trials were significantly longer than times on the indirect trials, $F(1,25) = 8.72, p < .01$, but the extent of this difference did not vary significantly with either probe

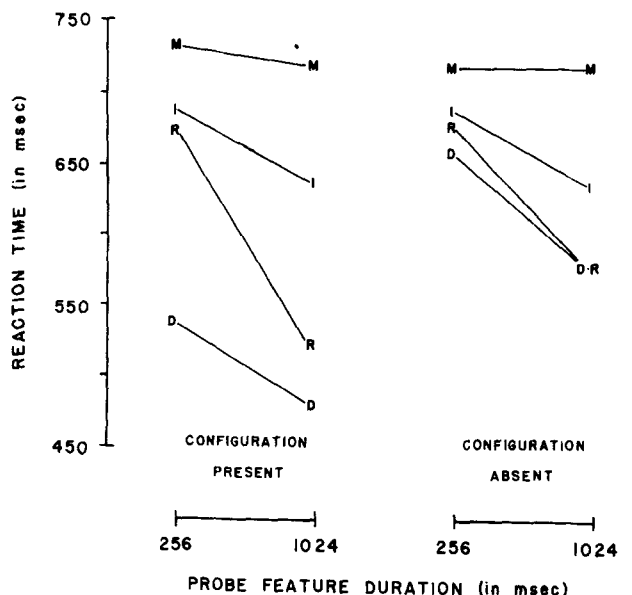


Figure 3. Mean reaction time in milliseconds as a function of probe feature duration and probe condition shown separately for configuration-present and configuration-absent conditions. Probe conditions are identified by letters according to the key given for Figure 2.

speed or configuration. Reaction times for the early reversing and duplicating trials did not differ significantly from the times on the late reversing and duplicating trials (these differences are not indicated in Figure 3).

DISCUSSION

The results indicate that spatial relations between features are best recognized when the inspection sequence employed during recognition duplicates that used during familiarization. Either performance or reaction times, or both, are better in this case than when the recognition sequence does not duplicate the familiarization sequence. Noton and Stark's theory of visual display recognition emphasizes that recognition is accomplished by duplicating the same sequence of fixations as during familiarization; therefore, the present results are in accordance with the predictions of their theory.

Noton and Stark's model predicts that recognition should suffer when the recognition sequence does not duplicate the familiarization sequence, but the model does not distinguish between various forms of non-duplication. The present results, however, suggest an ordering of relationships between differing familiarization and recognition sequences with respect to their ease of processing. Reversal of the sequence does not affect correct performance, but it does increase reaction time, especially when limited time to view the recognition sequence is available. The fact of reaction time differences implies that the reversal trials are handled differently than the duplicating trials by the visual recognition system. In the case of reversing trials, the investment of additional processing time appears to pay dividends in improved performance. However, when the recognition inspection sequence involves two features between which no direct shift has occurred in familiarization, both correct performance and reaction times suffer substantially, especially when the recognition shift is an implied continuation of the familiarization shifts.

These results strongly suggest that, at least for the relatively short viewing times of the present experiment, the representation of spatial relations between features is organized according to the sequence in which the features were viewed. If this is the case, a possible explanation for the occurrence of overt scanpaths is provided. Suppose a viewer is inspecting a display composed of features each of which requires an overt fixation shift to inspect. In such a case, portions of the display become available for inspection in a manner quite similar to the conditions of the present experiment. At the point in scanning where a previously viewed feature is refixated for the first time, the viewer has a choice on the next fixation shift between making one that he/she has pre-

viously made or one that he/she has not. The present results argue that if a duplicating shift is made, verification of the feature relationships involved will be faster and more accurate than if a nonduplicating shift is made. Thus, a scanpath, or repeating sequence, would be the most efficient means of viewing the display. Another implication of the present results is that, when rapid comparison of two similar displays is necessary, the best strategy involves direct duplication of the inspection sequence from one display to the other.

That there are reaction time differences between the duplicating and reversing conditions for the short probe but not for the long probe suggests that the properties of the representation of spatial relations between features change rapidly with time, becoming less sequence dependent. This begins to suggest that scanpaths may be more important for recognition for short viewing times than for long ones. Locher and Nodine (1974) failed to find any relation between the occurrence of scanpaths and recognition; however, their subjects were allowed to inspect the displays for as long as they wanted. The present results suggest that employing short viewing times might have altered their findings.

Stated in extreme form, Noton and Stark's model states that duplication of inspection sequences is critical for recognition: "That the subject follows a fixed path from feature to feature suggests that the eye-movement motor components involved in perception are not merely involved in moving the pattern over the retina, but are an integral part of the memories on which recognition is based. It suggests that the subject's internal representation or memory of the pattern is an alternating sequence of sensory and motor memory traces, recording alternately a feature of the pattern and the eye movement required to reach the next feature. During occurrences of the scanpath in the pattern-learning phase, these memory traces are being laid down. And, when the scanpath appears in the initial eye movements during recognition, the subject is matching the internal representation with the pattern, by reproducing the successive eye-movement memories and verifying the successive feature memories, a successful completion of the scanpath indicating recognition of the pattern" (Noton & Stark, 1971c, p. 310). The present results argue that duplication is important but not essential. Referring to Figure 2, the .05 confidence limit for performance significantly above chance is 56.0% (one-tailed). This means that performance was above chance levels for all conditions, except for the implied condition with the fast probe. With the slow probe, performance was above chance for all probe sequences. These results imply that, while recognition is not impossible without the duplication of sequences, it is more difficult and takes longer under

these conditions. Thus, while the execution and duplication of scanpaths may not be the only way in which visual recognition can operate, the present results argue that it may be the preferred way in some situations, especially when time is limited.

Subjectively, viewers in this experiment unanimously reported a sense of certainty and immediacy in giving their judgments on the duplicating trials. Nonduplicating trials, especially the indirect and implied trials, were said to produce a sense of confusion and hesitation; the correct answer did not seem to be directly given in immediate perceptual experience. Such reports, while not conclusive in themselves, substantiate the quantitative data and imply that the procedure may be tapping a basic property of the visual recognition system.

The magnitude of the effects reported here is large for a procedure involving a small number of familiar forms and a simple judgment. For example, in the short probe condition, subjects attained 78% correct on the duplicating trials as opposed to 50% correct (chance level) on the implied trials, a difference of 28%. This was accompanied by a 205-msec difference in reaction times: 531 msec for the duplicating as opposed to 736 msec for the implied conditions. This implies that a viewer, by choosing a poor fixation strategy in a free-viewing situation with similar stimulus materials, could be requiring him or her to spend as much as 200 msec extra processing time on each fixation.

Several limitations to the generalizability of the present results must be noted. It is not clear that attention is necessarily a sequential process in situations where viewers have ample peripheral visual information available. In such situations, it is not even clear that the point of fixation corresponds to the focus of attention. Thus, the present results can only be generalized to situations where viewers are compelled to inspect features sequentially. Another issue not addressed in the current experiment is whether, assuming the existence of sequentially organized traces, something about the eye movement itself, as opposed to its effect of rendering features visible, is important in establishing the trace. This question cannot be posed within the framework of

Noton and Stark's model, since they postulate formal equivalence between overt movements and covert attention shifts.

The results presented here imply several potentially fruitful directions for future research. The role of inspection sequences could be more precisely defined by using a wider variety of stimulus configurations and exposure times. It seems that all the sequence effects reported here must disappear if the viewing time is long enough. Manipulation of viewing time might provide clues to alternate viewing strategies that may become possible with longer viewing. As an adjunct to these experiments, the relation between recognition and subject-initiated eye movements should be further investigated, but using shorter viewer times and a more detailed analysis of fixation sequences than investigators have typically used.

REFERENCES

- JENNEROD, M., GERIN, P., & PERNIER, J. Déplacements et fixations du regard dans l'exploration libre d'une scène visuelle. *Vision Research*, 1968, 8, 81-97.
- LOCHER, P. J., & NODINE, C. F. The role of scanpaths in the recognition of random shapes. *Perception & Psychophysics*, 1974, 15, 308-314.
- LURIA, S. M., & STRAUSS, M. S. Eye movements during search for coded and uncoded targets. *Perception & Psychophysics*, 1975, 17, 303-308.
- NOTON, D. A theory of visual pattern perception. *IEEE Transactions on Systems, Science and Cybernetics*, 1970, 6, 349-357.
- NOTON, D., & STARK, L. Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 1971, 2, 929-942. (a)
- NOTON, D., & STARK, L. Eye movements and visual perception. *Scientific American*, 1971, 224(6), 34-43. (b)
- NOTON, D., & STARK, L. Scanpaths in eye movements during pattern perception. *Science*, 1971, 171, 308-311. (c)
- WHITESIDE, J. A. Eye movements of children, adults, and elderly persons during inspection of dot patterns. *Journal of Experimental Child Psychology*, 1974, 18, 313-332.
- WHITESIDE, J. A. Peripheral vision in children and adults. *Child Development*, 1976, 47, 290-293.
- YARBUS, A. L. *Eye movements and vision*. New York: Plenum Press, 1967.

(Received for publication April 11, 1977;
revision accepted February 9, 1978.)