PERGAMON

Vision Research 40 (2000) 1349–1364

Vision Research

www.elsevier.com/locate/visres

The spread of attention and learning in feature search: effects of target distribution and task difficulty

Merav Ahissar [a,*], Shaul Hochstein [b]

[a] *Department of Psychology, Hebrew University, Mt. Scopus Campus, 91905 Jerusalem, Israel*
[b] *Department of Neurobiology, Institute of Life Sciences, Hebrew University, Givat Ram Campus, 91904 Jerusalem, Israel*

Received 17 May 1999; received in revised form 14 December 1999

Abstract

We examined the roles of two determinants of spatial attention in governing the spread of perceptual learning, namely, stimulus location distribution and task difficulty. Subjects were trained on detection of a target element with an odd orientation imbedded in an array of light bars with otherwise uniform orientation. To assess the effects of target distribution on attention and learning, target positions were distributed so that attention was allocated not only to the target positions themselves, but also to intermediate positions where the target was not presented. Target detection performance substantially improved and improvement spread to match the induced window of spatial attention rather than only the actual target locations. To assess the effect of task difficulty on the spread of attention and learning, the target-distractor orientation difference and the time interval available for processing were manipulated. In addition, we compared performance of subjects with more versus with less detection difficulty. A consistent pattern emerged: When the task becomes more difficult, the window of attention shrinks, and learning becomes more localized. We conclude that task-specific spatial attention is both necessary and sufficient to induce learning. The spread of spatial attention, and thus of learning, is determined by the integrated effects of target distribution and task difficulty. We propose a theoretical framework whereby these factors combine to determine the cortical level of the focus of attention, which in turn enables learning modifications. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Attention; Learning; Perception; Feature search; Pop-out

1. Introduction

Recent studies found that practice dramatically improves perceptual task performance of adult observers. Learning characteristics appear to result from an interaction between top-down attention-driven processes and bottom-up stimulus-dictated effects. On the one hand, learning of even the simplest skills requires task-selective attention, so that when subjects attend the stimuli but not their appropriate aspect, task performance does not improve (Shiu & Pashler, 1992; Ahissar & Hochstein, 1993; Harris & Fahle 1998). On the other hand, improvement is often substantially specific to the trained spatial conditions. When tested with a novel retinal position, size, spatial frequency or orientation, performance may be severely hampered with respect to the attained asymptote (Ramachandran & Braddick, 1973; Fiorentini & Berardi, 1980; Ball & Sekuler, 1987; Karni & Sagi, 1991; Poggio, Fahle & Edelman, 1992a; Poggio, Edelman & Fahle, 1992b; Shiu & Pashler, 1992; Ahissar & Hochstein, 1993, 1996a,b,c, 1997a, 1998; Polat & Sagi, 1994; Beard, Levi & Reich, 1995; Schoups & Orban, 1995; Schoups, Vogels & Orban, 1995; Rubin, Nakayama & Shapley, 1997; Ahissar, Laiwand & Hochstein, 1998a; Ahissar, Laiwand, Kozminski & Hochstein, 1998b). These specificities indicate that learning in these cases involves levels of processing which retain separation along basic spatial dimensions. The interpretation of these combined effects in terms of the underlying neuronal site(s) is a puzzle. The cortical sites at which neuronal receptive fields are spatially selective are relatively low in the visual-system hierarchy (Desimone & Ungerleider, 1989; but, see Mollon & Danilova, 1996). Yet the cortical sites at which greater top-down effects were found are higher along

* Corresponding author. Fax: +972-2-5881159.
E-mail address: msmerava@pluto.huji.ac.il (M. Ahissar)

PII: S0042-6989(00)00002-X

these pathways (Desimone & Ungerleider, 1989, but see Motter, 1993; Rosenthal, 1995).

Another perplexing, though repeatedly found aspect of perceptual learning is that the degree of specificity to training spatial parameters, though sometimes complete, is rather variable, and changes substantially between tasks, subjects, or training conditions (Ahissar & Hochstein, 1996a,b,c, 1997a and references cited there). We previously studied this variability systematically and found a consistent pattern: Training under easy conditions leads to generalized learning, while difficult condition training leads to more specific learning (Ahissar & Hochstein, 1999). An additional observation concerning the characteristics of learning under intermixed easy and difficult training conditions was that improvement follows an easy-to-difficult condition cascade, reminiscent of the well known 'learning along a continuum' phenomenon (Pavlov, 1927; Lawrence, 1952; Sutherland, Mackintosh & Mackintosh, 1963).

We proposed the Reverse Hierarchy Theory to account for these perceptual learning phenomena (Ahissar & Hochstein, 1997a,b, 1999). According to the Reverse Hierarchy Theory, illustrated schematically in Fig. 1, the role of attention in directing perceptual learning is in its determination of the cortical level where learning modification will take place. Selection of an appropriate neural population follows a top-down search tree, seeking a population whose output is discriminative with respect to the task at hand. For most perceptual tasks, several alternative cortical representations along this hierarchy may be employed to reach the correct solution.

The top-down search tree begins at high cortical levels. Consequently early learning results from modifications at high cortical levels where the visual representation is generalized with regard to spatial parameters such as retinal position, while later learning modifies low-level, spatially specific mechanisms. On the other hand, only easy spatial discriminations can be resolved at high levels. When more difficult spatial discriminations are required, the continued search for increased signal-to-noise ratio leads to lower areas with finer spatial representations and yet inducing more spatially restricted learning.

The theory predicts that manipulations affecting the spatial distribution of attention will have the same effect on the spatial distribution of learning. In particular, the following two characteristics are expected:

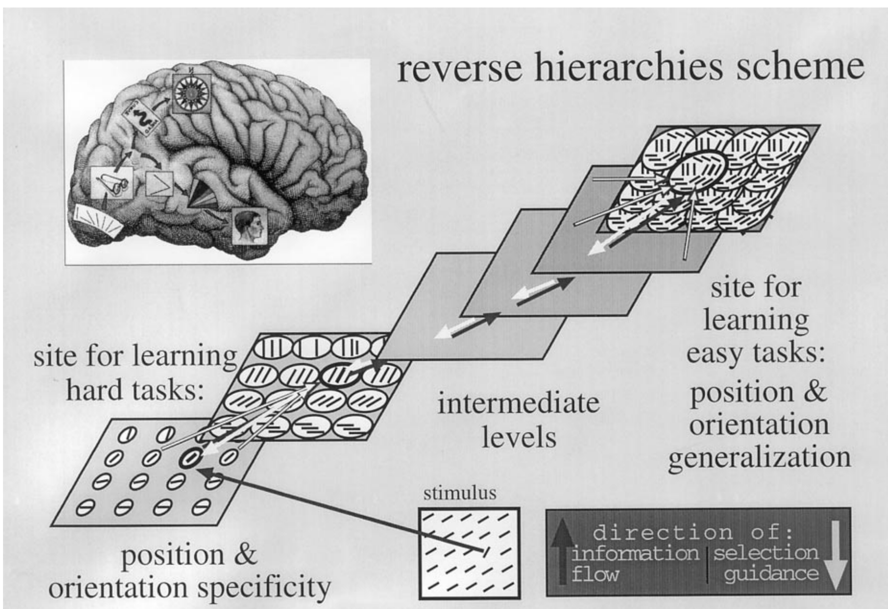

Fig. 1. Reverse hierarchy theory of attention and learning. In this schematic illustration, each circle denotes a neuronal group (e.g. cortical column) and the line(s) inside it the orientation preference. Enhanced lines mark bottom-up stimulus-activated neuron groups and paths interconnecting them. Stimulus is initially encoded by orientation selective neurons, at the lowest level. For example, one group is shown encoding the target element bar. At subsequent levels, there is substantial convergence across position (second level) and orientation (top level; inset depicts ventral and dorsal hierarchies, adapted from Posner and Raichle (1997)). Thus, the actual orientation of the original target signal is gradually less discriminated, and its salience diminished when neurons integrate over many orientations. Instead, convergence refines more abstract features such as orientation difference and spatial organization. Spatial attention and learning are initially directed at the highest, spatially generalizing levels. For easy conditions (e.g. large orientation gradient and long SOA) high levels suffice: even diminished salience leaves some distinction between target presence/absence. In this case, learning occurs at a high level (e.g. by selecting its most informative inputs). If the high level signal is insufficient, high level mechanisms must direct attention and learning to lower levels (downward arrows) to pin-point appropriate mechanisms here, within the sub-domain of their inputs. At lower levels, learning is restricted to trained orientations and positions, reflecting the increased spatial selectivity at this level. For new stimulus parameters, the counter-stream paths from the common top-level mechanism need to be used again to direct learning to different low level sites.

1. Stimulus spatial distribution should determine spatial learning distribution indirectly — via attention and its determination of level of learning. The spatial extent of perceptual learning should depend on the nature of the representation at the level chosen for learning. Thus, when spatial attention is broad, it reflects choice of a high-level site where receptive fields are large. Consequently it may not be necessary to stimulate each location in the field for learning to be effective.
2. Task difficulty (in terms of spatial parameters) should be another determinant of the attended cortical level and hence of the spread of learning effect. The greater the spatial refinement needed, the lower the level of the site chosen for focused attention and learning.

Thus, these two parameters — stimulus distribution and task difficulty, should interact to determine, via attention, the level of learning and the extent of transfer to new locations. These predictions are tested in the present paper.

1.1. Experimental plan

From the perspective of the Reverse Hierarchy Theory, spatial specificity of learning is not the direct result of stimulating only at given positions within the visual field. Rather stimulus-specificity is a top-down determined effect dictated by both the global distribution of the task-relevant stimulus and the difficulty of the task. These factors determine the site chosen for attention and the representation at this site determines the spatial spread of learning effect.

Previous studies did not distinguish between bottom-up and top-down dictated spatial selectivity. Indeed it was found that learning under difficult spatial conditions is confined to the vicinity of the trained retinal area (Eriksen & Eriksen, 1974; LaBerge & Brown, 1986; Berardi & Fiorentini, 1987; Ahissar & Hochstein, 1996a). In the context of the pop-out task (Treisman & Gelade, 1980), learning is restricted to the vicinity of the trained position of the target element (Ahissar & Hochstein, 1996a). However, in previous studies target position and spatial attention overlapped. Thus, these two explanations — direct effect and effect via attentional window, remained as open possibilities. In particular, the effect of shifting target position was studied only in one direction, moving the relevant stimulus away from fixation. If specificity results from the attention-attracting aspect of the target distribution, then no transfer of learning effects will be found in the near-to-far direction when attention is focussed only to locations proximal to fixation. However, it may be difficult not to attend positions between target and fixation. Furthermore, if the relevant attention-directing mechanism is determination of cortical processing level, as suggested by the Reverse Hierarchy Theory, then spread attention may be equivalent to employment of larger receptive fields at higher cortical levels, which generally include central vision. Thus, following practice with distal targets, we may expect transfer to intermediate positions between fixation and the trained locations, or in the area joining a number of eccentric training positions. The alternative bottom-up target-location explanation of specificity would predict that there should not be transfer in either the proximal-to-distal or the distal-to-proximal direction.

We used several alternative target distributions under different degrees of task difficulty during the training period. We then tested detection when target was evenly distributed at all array positions. The spatial distribution of detection performance during this test session forms a two-dimensional performance map. We found that this map matches the one expected if learning follows the spread of attention, rather than affecting only stimulated positions. In addition, the spread of attention and learning depends on task difficulty so that training easier tasks transferred more to new spatial conditions. Task difficulty was increased in several ways including decreased target-distractor orientation difference, shorter processing intervals and, a posteriori, by comparing observers who had an easier (more successful) versus more difficult experience with pop-out training (though given the same physical training conditions). All these tests showed that more difficult conditions induce more focused attention and learning. These findings are consistent with the Reverse Hierarchy Theory whereby attention determines hierarchical level, rather than learning depending on the local physical characteristics of the trained stimulus.

2. Methods

2.1. Stimuli and procedure

Stimuli were arrays of light bar elements (147 cd/m^2) on a dark background (0.2 cd/m^2). The array consisted of 7×7 elements (subtending $4.6 \times 4.6°$) centered at fixation, as illustrated schematically in Fig. 2, top. Each stimulus element subtended $22' \times 1'$. The distance between element centers was 42.6′ ($\pm 4'$ jitter, randomly chosen with uniform probability). In half of the stimulus presentations, all elements had the same orientation (30° or 60° counter-clockwise from horizontal; Fig. 2, top, center). In the other half, one of the elements was a target at a fixed orientation, deviating by 30° from that of the distractor elements (60° or 30°, respectively; Fig. 2, top, left).

A mask followed each stimulus, as shown in Fig. 2, top, right. The mask was composed of a 7×7 array of asterisk-like elements, located at the grid points of the

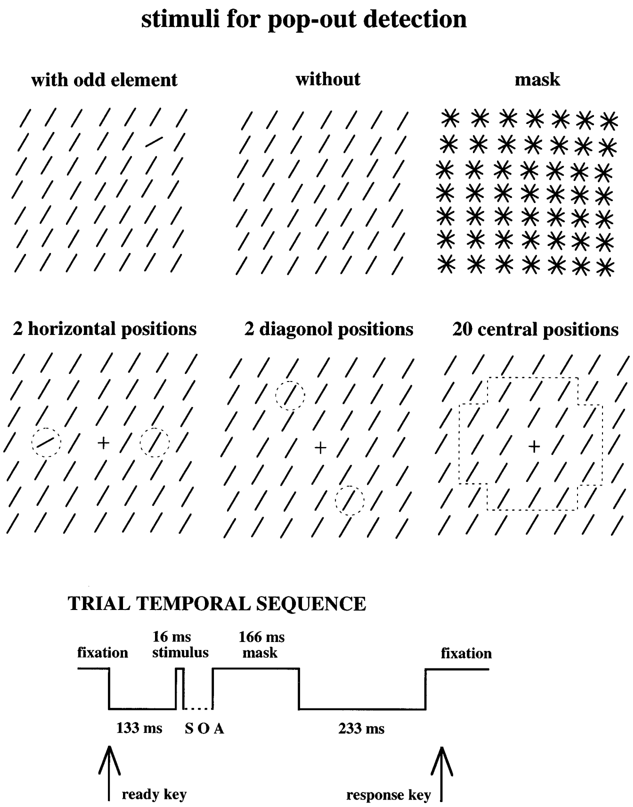

Fig. 2. Stimuli for odd orientation 'pop-out' detection task. Top row: 7×7 element stimulus array with (left) or without (middle) odd element. Following array presentation, a mask stimulus (right) was presented following a variable stimulus-to-mask-onset-asynchrony (SOA). Average distance between array elements was 0.7°, but the element positions were jittered to prevent responses on the basis of unusual blank background areas. Middle row: Illustration of the various target distributions used for this study. (1) Target appeared in any of the 48 array positions (except in the central location where the fixation cross appeared before stimulus presentation). (2) Target appeared in one of two horizontal positions indicated by dashed circles in left diagram. (3) Target appeared in one of two diagonal positions, either two up or down and one left or right, or one up or down and two left or right, as example indicated by circles in middle diagram. (4) Target appeared in any of 20 central positions indicated by dashed line in right illustration. (5) Target appeared always in the same horizontal position indicated by one of the circles in left diagram. We had to use different subject groups for each of these five sets of training tests in order that they be naive at the onset of the training. Dashed circles are shown here to indicate possible target locations and did not appear in tests. Fixation cross appeared before the stimulus array and was replaced by a distractor bar during stimulation, as shown in the top row. Bottom row: Trial temporal sequence. While looking at the fixation cross, subjects pressed the ready key. Then, 133 ms later, the stimulus appeared, followed by the masking stimulus after the variable SOA period. The duration of the stimulus was just 16ms (one monitor frame time), that of the mask, 166 ms. We tested response accuracy (target present or absent) as a function of processing time (SOA), not reaction time. See text and Ahissar and Hochstein (1993, 1996a, 1997a) for more details.

7×7 stimulus lattice ($\pm 4'$ jitter so that element position exactly matched those of the stimulus). Each mask element was a superposition of four lines: the trained target and distractor orientations, and these orientations plus 90° (e.g. 30, 60, 120 and 150°).

The temporal sequence of each trial is shown in Fig. 2, bottom: Each trial started with a fixation cross (a + sign with $22 \times 1'$ lines of intensity 147 cd/m^2). When the observer pressed the ready key, after 133 ms, the stimulus appeared. The stimulus was on for 16 ms. Then, following a variable stimulus onset asynchrony (SOA), the mask was displayed for 166 ms. Finally, following a 233 ms dark period, the fixation point reappeared while the subject pressed a response key. A computer tone confirmed correct responses.

Stimuli were presented in blocks of 20 trials with the same SOA. Each session comprised 70 blocks (1400 trials). Sessions began with a set of nine blocks starting from the longest SOA (183 or 150 ms) and gradually reaching the shortest SOA (16 ms) in an interleaved manner (blocks with SOAs of 183, 133, 100, 66, 33 ms followed by blocks of 150, 116, 83 and 50 ms. Based on performance in these initial blocks, the range of SOAs to be presented next was chosen. The choice was made so that the shortest SOA would be the longest in which the subject still performed near chance level (55% correct) and the longest SOA would be the shortest where the subject already showed near perfect performance (95% correct). Within that chosen range (constrained to include at least three different SOAs), blocks were presented in pseudo-random sequence. Following blocks of presentations with these SOAs, the next range of SOAs was chosen based on performance in these blocks using the above criteria. This procedure was continued until 1400 trials were completed. As a result of this procedure, mean performance was kept around 75% correct, within and throughout sessions.

Stimuli were presented on an HG Trinitron multiscan monitor (Sony, Inc.) or a 5 A Micro-scan monitor (A.D.I., Inc.) running at 60 Hz frame-rate and 1024×1024 pixel resolution, driven by a #9-GXgraphics card (#9 Computer Co., Inc.) in a 486 PC computer. Response keys were '1' (for present) and '0' (for absent) on the numeric keypad of the computer keyboard, followed by the ready key, 'enter', to initiate the next trial.

2.2. Target distribution

Several subject groups were trained with different target distributions:

1. Control: Target presentation location was evenly distributed within the array (data presented in Ahissar & Hochstein, 1996a).
2. Target at one of two horizontal array positions. These positions are circled in the array illustration of Fig. 2, middle row, left. (This group was actually composed of two sub-groups. The first was trained, as were the other groups, with a 30° target-distrac-

tor orientation difference, while the second was trained with a smaller, 16° difference. This latter sub-group was also subsequently tested only with the two-target positions paradigm).

3. Target at one of two diagonal locations, encircled in the array illustration of Fig. 2, middle row, middle. (Thinking of the array as a chess board, target positions are a 'knight's move' from fixation). Subjects were trained either with this pair of opposite locations, or the left-right mirror image pair.
4. Target at one of 20 central positions denoted by the dashed line in Fig. 2, middle row, right.

Following training, all groups were tested on the control paradigm (#1 above), that is, with target presentation location evenly distributed within the array. This served as the crucial test of the effect of training and its spatial spread. For this test session, we present 2-dimensional plots of target detection performance within this map of target locations. On this map we superimpose an outline indication of the limited area of target distribution during the preceding training sessions.

We preferred training with more than one target position because it was very difficult for most subjects to maintain fixation when targets were consistently presented at one side of the fixation point (Ahissar & Hochstein, 1996a). In a pilot study a group of subjects was trained with a single horizontal position, and then tested with all positions. Although learning was not transferred to more distal positions, results were variable across subjects, perhaps due to their different qualities of fixation. This procedure was therefore replaced with two positions symmetrically located with respect to the fixation point.

2.3. Subjects

Fifty-two subjects participated in these experiments. They were 20–27 years old, with normal or corrected-to-normal eyesight. All were naive as to the purposes of the experiment and were reimbursed for participation. Twenty-two subjects learned pop-out with target at all positions; fourteen with target at two horizontal positions (ten with 30° difference; four with 16° difference); six with target at two diagonal positions; ten with target at 20 central locations.

2.4. Analysis

The average session threshold was evaluated by computing the best fit psychometric function of the form: $f(t) = 1 - 0.5 \exp(-(t/\tau)^{\sigma\tau})$ where $f(t)$ is the proportion of correct responses; t, the trial SOA; and σ and τ are free parameters: τ, the threshold SOA at 81.6% correct, and σ, the slope at threshold multiplied by $2e$ (Quick, 1974). The method for determination of the threshold is described elsewhere (Ahissar & Hochstein, 1993, 1996b).

The two-dimensional spatial distribution of the average fraction of correct detection was computed by summing (across subjects), separately for each position, the number of target-present answers among target-present trials, i.e. the fraction of hits (there was no way, of course, of attributing the responses in the target-absent trials to a specific location, so this half of the data had to be discounted). The summation was performed separately for each SOA and the average was then obtained by simple averaging across a group of SOAs (generally 33, 50 and 66 ms). Thus, performance at each SOA was given the same weight, although the number of presentations was typically not precisely equal (see Section 2.1, above and in Ahissar & Hochstein, 1996a). We used at least an entire session to compute this 2-dimensional distribution to reduce variability despite dividing (half) the data among the 48 possible target locations (~ 14 trials/point/subject). To assess the 1-dimensional dependence on eccentricity, data were summed from all array positions with the same eccentricity, regardless of their azimuth.

3. Results

3.1. Spatial distribution of detection: initially and following general training

We first present the detection distribution of a control group trained with target location evenly distributed at all array positions. This detection distribution will serve as a baseline for comparison with detection distributions following training with various target distributions. Threshold was reduced gradually as a function of training experience (session number), as demonstrated in the learning curve of Fig. 3A. Note that inter-subject variability is greater for naive subjects, and is diminished as subjects approach a more common asymptote.

The initial and final 2-dimensional distributions are plotted in Fig. 3B (left and right, respectively). The distributions are for the entire first session and entire last two sessions, while the dynamics curve of Fig. 3A is by thirds of a session. Both distributions show a large anisotropy: Detection drops off much more steeply vertically compared with horizontally (see also Krose & Julesz, 1989), and more in the lower than the upper hemifield. Training raised detection for all positions (brightening of entire right hand graph compared with the left graph). There was also some expansion of the central region, mainly along the vertical meridian so that asymptotic performance is somewhat less anisotropic. Thus, when trained with target appearing everywhere, learning occurred everywhere. We now ask what will happen when training includes a limited set of target locations. In particular, we ask whether improvement will transfer to all locations, will be limited to trained locations, or will affect some intermediate positions.

3.2. Spatial distribution following training with two horizontal positions

A second group of subjects practiced with target equally likely to be at each of two horizontal positions

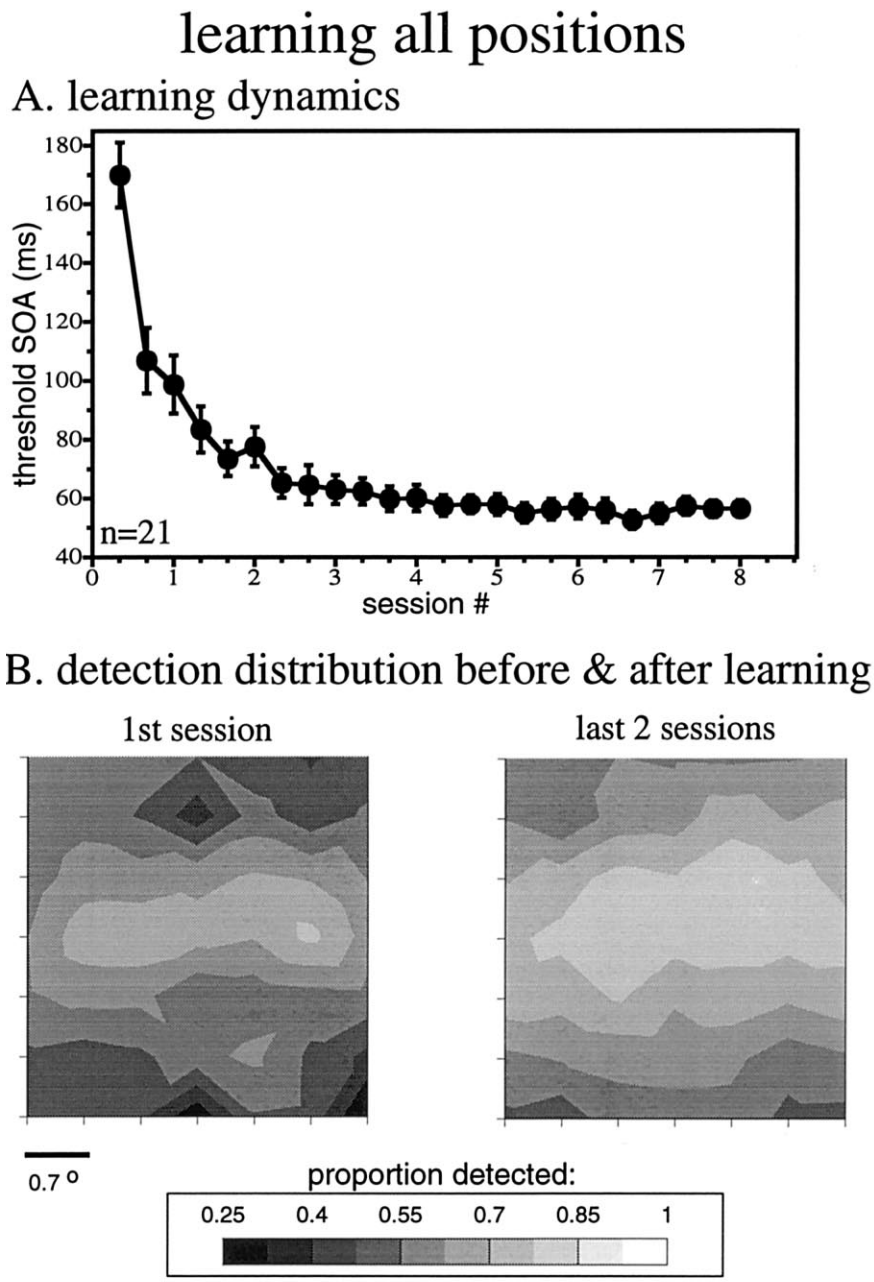

Fig. 3. Effect of training with odd-element appearing at all array positions. (A) Threshold (SOA for achieving 81.6% correct responses; see text) as a function of training session for the group which trained with target appearing in all array positions, by thirds of a session. Twenty one subjects participated in this part of the experiment, and the error bars indicate between-subject standard errors of the mean. Note that in general this variance is greater for the subjects when naive than when they are trained, suggesting that much of the difference between them may be a matter of prior experience or natural training. With substantial training, threshold SOA is reduced by about a factor 3! (B) Two dimensional detection distribution before training (1st session — left diagram) and following substantial training (final two sessions of each subject — right diagram) for the control group that practiced with target appearing at all array positions. Average detection of target when SOA was 33, 50 or 66 ms. Note that detection is better in the center than at the periphery, falling off faster vertically than horizontally. With training, the central area of good performance is brightened (performance improves at the center) and enlarged (performance improves around the center).

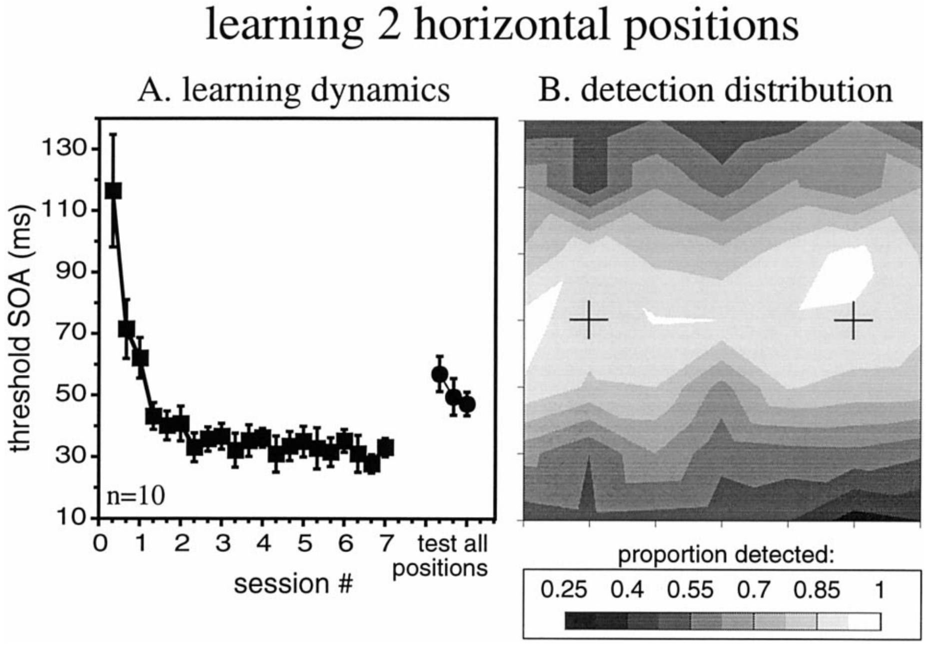

Fig. 4. Spread of effect when training is limited to two horizontal positions. (A) Threshold SOA as a function of training session for the group of subjects who trained with the target appearing at one of two horizontal array positions, by thirds of a session. Note somewhat superior performance (lower thresholds) at start and at asymptote, for this easier task, with less stimulus position uncertainty compared to the control group of Fig. 3. Final three points are for a single final session with target appearing at all positions. Performance is degraded (threshold is raised), but not to original naive level. (B) Two-dimensional detection distribution for subject group who trained with target appearing in one of two horizontal positions. Data for this distribution and those of the following figure are from the final post-training session with target appearing at all positions, averaging detection data for 33 and 50 ms SOA. Note considerable brightening for positions between and around trained locations (indicated by crosses) compared to naive level (Fig. 3B, left), suggesting that subjects attended to this entire region and that attention suffices to induce learning (note that this difference is seen even though the current figure illustrates the average of only two hard SOAs, while Fig. 3B includes an easier SOA).

($\pm$ 1.4° from fixation) as illustrated in Fig. 2, middle-left. The threshold reduction as a function of session number for this group is shown in Fig. 4A. Initial performance is somewhat better than that of the group trained with all positions, and learning is quicker, though learning rate varies among subjects. Better performance might be expected for this easier task with reduced target location uncertainty.

Following training with the two target positions, these subjects were tested with target appearing in any array location. Threshold performance for this session is shown in the rightmost points of the learning dynamics graph of Fig. 4A. There was incomplete transfer to this new situation, reflected in a rise in threshold. The spatial distribution of target detection for this group, following training for two horizontal positions, is illustrated in Fig. 4B. It is evident that the entire central elongated cone is brightened compared to not only the initial but also the asymptotic distribution following all-position training (Fig. 3B, left and right, respectively). Detection at positions proximal to fixation is better in this case than following training with all positions, even though target was never presented here during training. This difference was highly significant. For instance, average detection for the eight positions outlining the central square of array positions (at the hardest super-threshold condition, 33 ms SOA) was 0.46 for the group trained with all positions compared with 0.72 for the group trained at two horizontal positions ($P < 0.001$).

We did not monitor eye position so we can not directly refute the interpretation of this result as deriving from subjects' occasionally fixating the expected target positions during training rather than the required fixation position. However, if this were the case and subjects alternated fixation between the right and left target locations, then the target would have occasionally appeared at fixation and occasionally four element positions (2.8°) away, but it never would have been one array position from fixation. Thus, we would have expected more transfer to distal rather than to intermediate horizontal positions. Therefore, this interpretation does not account for the measured spatial distribution of detection.

The advantage of training only two horizontal positions in detecting targets proximal to fixation is most evident for short difficult SOAs. In Fig. 5 we plot detection as a function of eccentricity following two-position and all-position training when testing at all positions. Data are plotted for 33 and 50 ms SOA averaged (Fig. 5A) or separately (Fig. 5B). Note that the slopes for all-position training are shallower. Performance for central positions is consistently better and for peripheral positions consistently worse for two-position training than for all-position training, so that the graphs cross.

We conclude that subjects trained with two positions allocated spatial attention to a continuous central region that extends between the target positions and that this sufficed to induce improvement at the intermediate (interpolated) positions, where the target did not appear. The better performance for these subjects at these central locations must derive from their being trained with an easier task, namely with less target location uncertainty (though tested finally with the same target distribution) where attention could be confined to a smaller area.

3.3. Training with other target distributions

Another group of subjects ($n = 6$) practiced with target equally likely to be at each of two diagonal positions, as illustrated in Fig. 2, middle, middle. Their threshold reduction as a function of session number is shown in Fig. 6A, left. Initial threshold is higher than for the horizontal positions, as expected from the greater detection difficulty for the diagonal positions since they are both more eccentric and lie along the diagonal rather than along the horizontal axis. As shown in Fig. 3, detection decreases more steeply along diagonal and vertical directions than along the horizontal meridian. Still, following practice, a similar asymptotic threshold is achieved.

Fig. 6A, right, illustrates the spatial distribution of detection for this group in the test session (with target appearing at all array locations), following training with target at diagonal locations. The whole central cone is somewhat brightened, with detection at the distal positions, where the targets were actually present, being somewhat better than positions near fixation. The central cone is elongated along the diagonal where targets were trained. The two brightened spots around the trained diagonal target positions show that maximal improvement was attained at the actual distal target positions. However, again, improvement included intermediate positions. Note that in this case, as well, central detection equals or exceeds that attained following general practice (Fig. 3).

Note the large difference between the extent of spatial transfer from the two trained locations to more central positions, seen for two horizontal locations (more transfer) compared to two diagonal locations (less transfer). This difference may be related to the principle stated above that the easier and more spatially confined the task, the more the transfer to new stimulus conditions (Ahissar & Hochstein, 1997a). We also found that the improvement at the diagonal positions did not transfer to new target and distractor orientations (swapping target and distractor orientations; see Ahissar & Hochstein 1996a; Ahissar et al., 1998a).

Another group of subjects ($n = 10$) practiced with target equally likely to be at each of 20 central positions, as illustrated in Fig. 2, middle, right. Their threshold reduction as a function of session number is shown in Fig. 6B, left. Threshold is initially lower than when all positions are equally likely and improvement is quicker; yet asymptotic thresholds are similar.

Fig. 6B right, demonstrates the spatial distribution of detection (as tested in the final session, with target at all locations) for the group trained with target at 20 central locations. The whole central cone is brighter. Yet the outskirts are darker even compared with initial performance of the control group (Fig. 3B, left). Thus, no improvement was obtained beyond the induced central cone of attention.

In summary, if we look at the four detection distributions during the test session with target at all locations (Fig. 3B, right; Fig. 4B; Fig. 6A, right, and B, right) following different training conditions (target appearing at all element locations, at two horizontal locations, at two diagonal locations, or at 20 central locations, re-

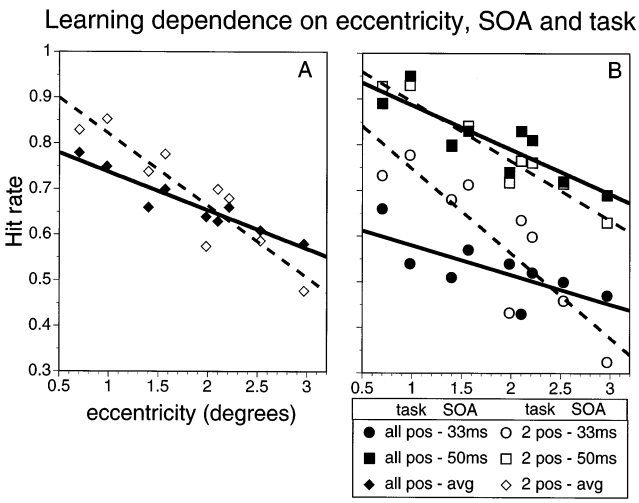

Fig. 5. Detection performance and learning transfer dependence on task difficulty as determined by SOA, target eccentricity and target location distribution. Detection hit rate is shown as a function of eccentricity for two SOAs and two training conditions: target appearing at all locations ('all pos'; filled symbols and full lines; data for 11 better detecting subjects; see Fig. 7A) and for target appearing at two horizontal locations ('2 pos'; open symbols and dashed lines). (A) average performance for 33 and 50 ms SOA. Note that curves cross suggesting that the spread of easy two-position training towards the center comes at the expense of improving performance at the periphery. (B) Separate plots for each of these SOAs. Note that for the easier SOA (50 ms), training everywhere led to better performance at nearly all eccentricities. For the more difficult SOA (33 ms), the performance curves cross: At positions central to the trained positions (where the target never appeared during training), focussed training is better, while for more peripheral locations direct training is required.

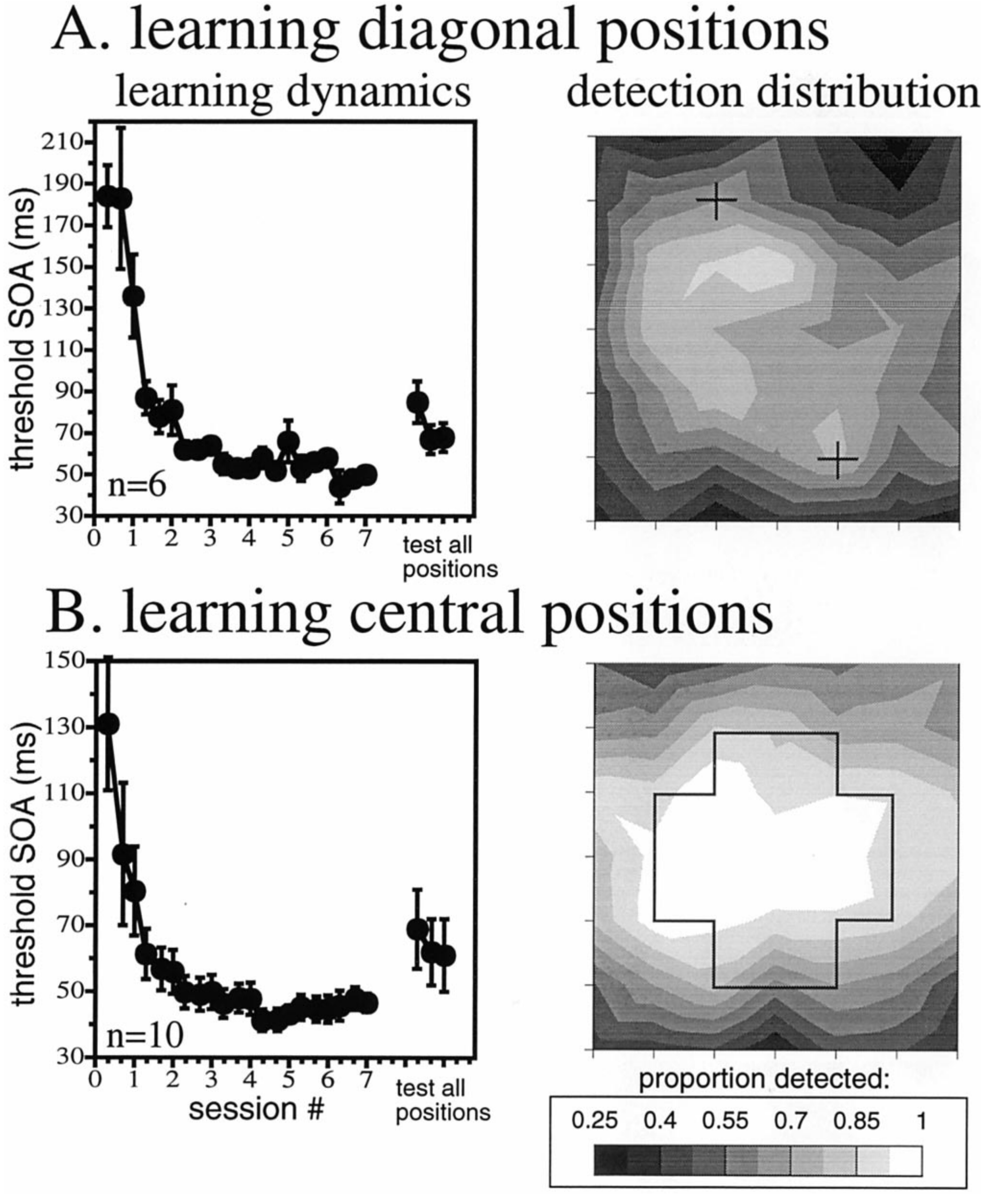

Fig. 6. Effects of training with other target distributions. (A) Learning curve (threshold SOA versus session number-left) and 2-dimensional detection distribution for post-training test (with target appearing at all positions – right) for group which trained with target appearing at one of two diagonal positions (as indicated by crosses). Note training is not as effective for these more difficult positions, nor is the spread of learning so great, as with the two horizontal positions of Fig. 4. Best performance is now near the trained positions, not in the central region between them. Still, there is a spread of learning and performance is considerably better than for naive subjects (compare Fig. 3B, left). (B) Learning curve and detection distribution following training with target appearing in any one of 20 central positions. Performance is enhanced for entire central region, with a very sharp decay at the upper and lower peripheral positions.

spectively), we see four different performance distributions. These distributions reflect the different training conditions, i.e. the different target location distributions during training. However, as a rule, the performance distributions following training demonstrate considerable improvement of performance not only at the locations that were actively trained but rather they always include learning at much of the central region — where attention was directed but target did not appear. Thus, we conclude that attention suffices to improve performance even at positions where the target never appears.

The difference between the degrees of spread for two horizontal and two diagonal locations, together with the different degrees of specificity previously found for training under easier and more difficult conditions (Ahissar & Hochstein, 1997a) suggests that spatial spread may be linked to task difficulty. This conjecture is tested in the following section.

3.4. Interaction between spatial distribution and difficulty

We asked whether the difficulty of the task, in terms of available processing time (SOA) and individual subject skill would affect learning distribution. In particular, whether the diameter of learning will decrease with decreasing SOA, and whether there will be a different spread of training effect for subjects who demonstrate more compared with less skill for the feature search task.

To separate subjects who are better or worse at odd-element detection, we looked at the threshold of detection during the first training session for all subjects trained with target at all locations (the control group). There was a natural division between the 11 better-detecting subjects (with starting thresholds below 115 ms) and ten worse-detecting subjects (with thresholds above 125 ms; the threshold for one subject never declined to below 100 ms so his data were dropped from the averages and figures). The learning curves for these two groups are shown in Fig. 7A. Performances of the two groups remain different throughout the eight session training period and they appear to approach different asymptotes. Note that, as we reported previously, there is generally much greater individual variation before training than after (Fig. 3, above; see also review by Ahissar et al., 1998b).

The dependences of the training effect on target eccentricity, SOA and subject skill are shown in Figs. 7 and 8. In Fig. 8 we present the 2-dimensional target-location contour-plot of performance for the 1st, 2nd and (averaged) final two sessions for the two groups of subjects. These are displayed for two SOAs, chosen to show the greatest degree of learning effect for each group, respectively. In Fig. 7 the learning effect (change in hit rate), is plotted as a function of eccentricity, averaging over azimuth. Analysis of these plots reveals a number of important trends.

There is indeed a difference between the subject groups at all target locations. Comparing the 2-D plots of performance at 50 ms SOA, we see that the more skilled subjects perform better than the less skilled subjects from the first to the last session, and for all eccentricities and target locations. In fact, performance is consistently about one 16 ms step faster in all cases. Performance for the better subjects at 50 ms SOA is about the same as performance for the less skilled subjects for an SOA of 66 ms, for each of the training sessions. This similarity is also found by comparing performance of the more skilled subjects at 33 ms with that of the less skilled at 50 ms.

Learning spread depends on task difficulty. For each set of subjects, the curves plotting learning versus eccentricity (Fig. 7B) are rotated clockwise with decreas-

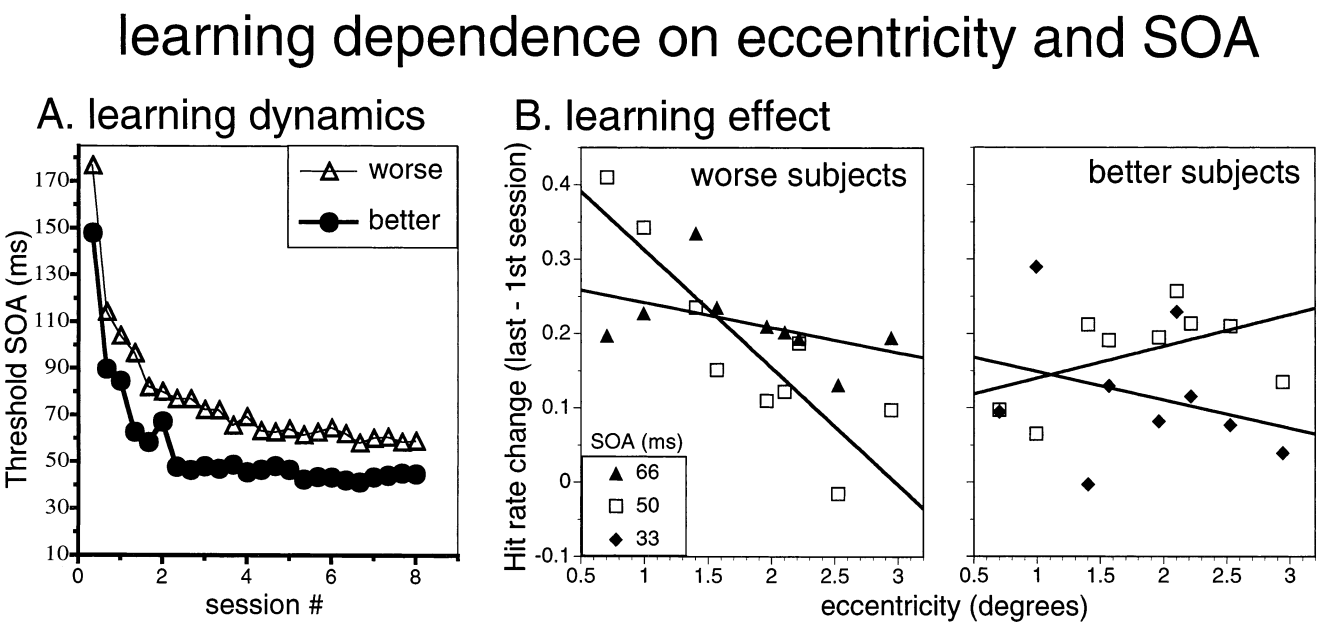

Fig. 7. Learning dependence on eccentricity, SOA and subject detecting ability. We divided the subjects of the control group of Fig. 3 into two sub-groups on the basis of their performance on the first training session ('better' and 'worse' detectors) and compare their further performance in a number of ways: (A) Learning dynamics. Threshold as a function of training session for the two sub-groups (by thirds of a session). Subjects who are better detectors begin and remain better throughout the training. This performance threshold was used to divide the subjects into two groups for the other analyses of this and the following figures. (B) Learning effects. The change in hit rate from first to last sessions is shown as a function of eccentricity for each subject group (left and right graphs) and for two SOAs. Note that the more difficult the detection (in terms of SOA or subject group) the more the learning effect is concentrated at the central part of the visual field.

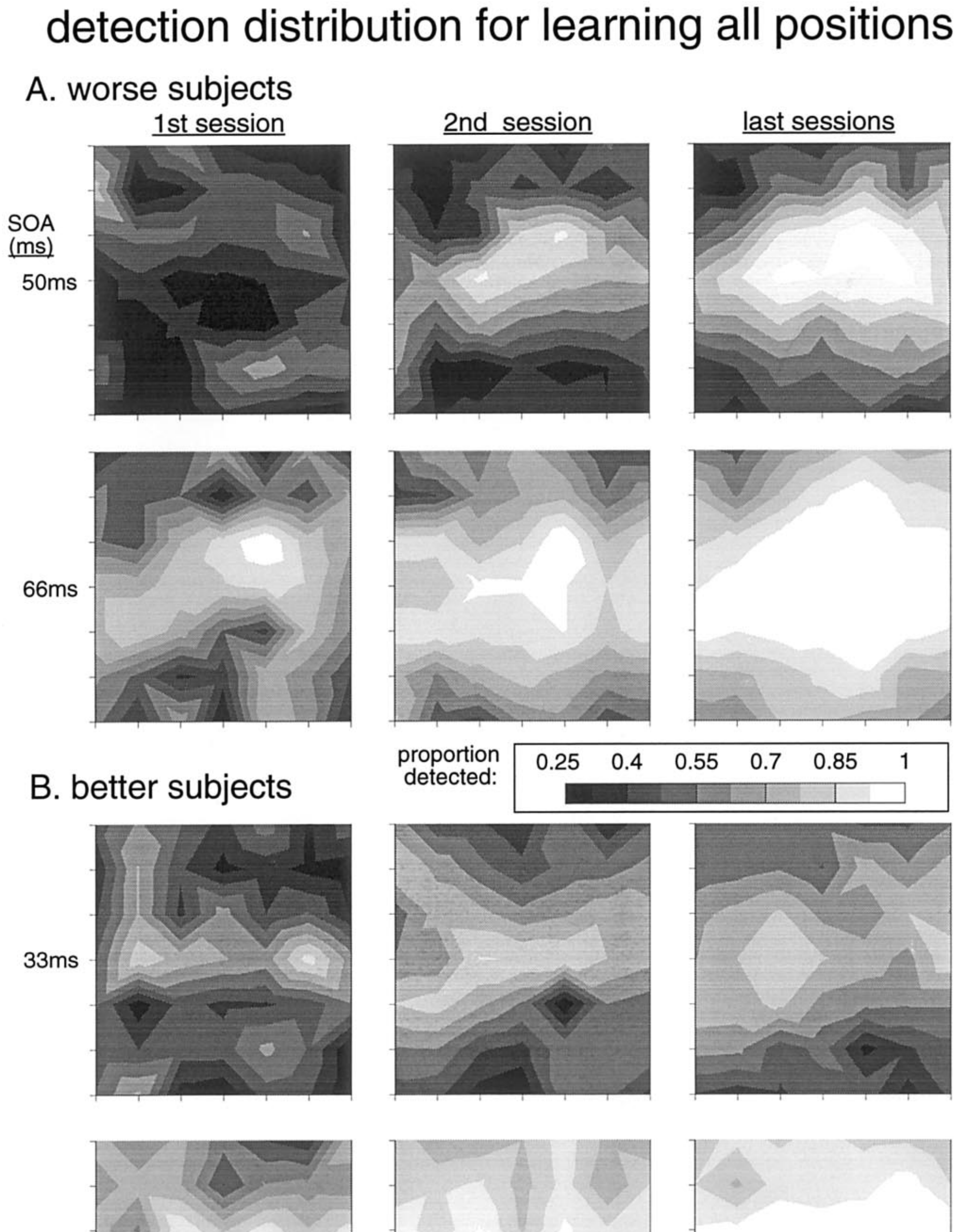

Fig. 8. Detection distributions for better and worse detecting subjects. Detection distributions are presented for the 1st session (left), 2nd session (middle) and last two sessions (right) separately for the 11 better detecting subjects (A) and the ten worse detecting subjects (B). Subject classification for this and the following figure was done on the basis of threshold performance (Fig. 7A). Detection distributions are separated also by SOA, with data for two SOAs (indicated on the left) shown for each subject group. The distributions show gradual learning (brightening) with improvement first at the center and for longer SOAs and then also at the periphery and for shorter SOAs. Note that the better detecting subjects are about 16ms faster than the other subjects.

ing SOA. This means that for more difficult task conditions, learning is concentrated at (easier) more central locations.

In parallel, better detecting subjects seem more able to spread their attentional window. This is evident in the slopes of the curves of Fig. 7 and also in the 2-D detection plots of Fig. 8. The plots for better detectors have much less steep gradients than do those of the poorer detectors. For example, compare the detection distributions for the final sessions of the better subjects at 33 ms SOA with those of the worse detecting subjects at 50 ms SOA (Fig. 8 right column, 3rd and top plots, respectively). Detection falls off much more steeply for the poorer detectors, even for the longer SOA. Thus, the trade-off between personal ability and available processing time may not be complete.

3.5. Spatial distribution with horizontal positions using a smaller orientation gradient

We asked whether the transfer dependence on task difficulty extends to detection difficulty determined by the target/distractor orientation difference. In particular, we wished to know if under more difficult conditions it would be the intermediate positions or the actual positions where target was presented that would be more affected by the extra difficulty introduced. We reduced the target/distractor difference from 30° to 16° for another group of subjects, using the same two-position horizontal distribution studied above. As demonstrated in Fig. 9, we found that even for the more difficult condition there was still nearly complete transfer to central locations (except at the briefest SOA). There was substantial transfer even to more distal locations for long SOAs, while, for brief and intermediate SOAs there was less and less transfer in this direction.

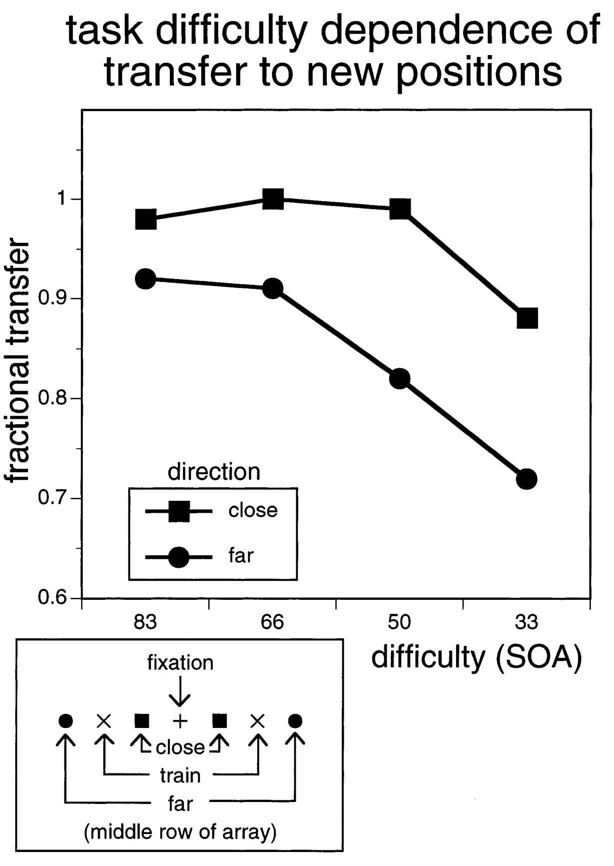

Fig. 9. Transfer with two easy positions and difficult 16° target-distractor orientation difference. Transfer of training effect to more central (squares) and more peripheral (circles) locations for training at two horizontal positions with a difficult discrimination task (target to distractor orientation difference of 16°). Fractional transfer is defined as the amount of training effect at the test positions (detection at test–initial detection) compared to training effect at trained locations (final–initial detection). Initial detection is assumed same for all three pairs of positions, since data are available only for the trained positions; this may somewhat under (over)-estimate the specificity at positions proximal (peripheral) to fixation. Note nearly complete transfer to near locations at nearly all SOAs and decreasing transfer to further locations with decreasing SOA.

Thus, performance increase is not uniform and across the board. For all tested parameters, the easier the task, i.e. the longer the SOA, the greater the orientation difference, or the more central the target location, the greater is the transfer of learning effects. The level of difficulty or ease that dictates the spread learning is subject to individual differences.

4. Summary

In the first part of this study we found that spatial attention is both necessary and sufficient for inducing learning of odd element detection, in agreement with the prediction of the reverse hierarchy theory. Several target distributions were used for the feature search task to dissociate spatial attention from target distribution. When targets were presented at two positions around (but not adjacent to) fixation, detection improved both at the target positions and at intermediate positions within the region of focal attention — even though the target was never presented at these positions. However, no improvement was found at positions farther from target positions, which were not attended. From these findings we conclude that attention is not only necessary but also sufficient for perceptual learning. Note that in accord with previous results attention must not only be spatially focused, but must also by selectively tuned to the aspect of the stimulus that is relevant for the trained task. Thus, training one task did not affect a second task (which was not performed), and which depended on a different attribute of the same set of stimuli (Shiu & Pashler, 1992; Ahissar & Hochstein, 1993; see also Treisman, 1992; Harris & Fahle, 1998).

In the second part of the study we examined the effect of task difficulty in determining the spread of spatial attention and consequently, the spatial distribution of learning. Several manipulations were studied, including the distribution of target positions, the target-distractor orientation difference and the time interval available for processing (determined by stimulus-to-mask onset asynchrony, SOA). In addition, we compared the performances of subjects who had greater or lesser difficulty in detecting the odd element. These various comparisons show a consistent pattern, namely, when the task becomes more difficult, the window of attention shrinks, and learning becomes more localized.

In terms of the reverse hierarchy theory presented in the Introduction, these findings mean that attention determines the cortical level of processing used for task performance and learning. The level chosen is that which satisfies the spatial selectivity requirements of the target distribution and the difficulty of the task conditions as determined by its various parameters. In turn, the choice of cortical level determines the extent of transfer of training effects.

5. Discussion

We can now predict the width of the attentional window under given task conditions, for cases of brief stimulus presentation. From the center of the stimulus, the focus of attention spreads out to accommodate potentially relevant stimuli (a process which is relevant across a time constant of at least several trials). If the task is too difficult to be performed with spread attention, the attentional focus is narrowed. Thus, these two factors — target distribution and task difficulty, both adapted to increase successful detection, may have opposing effects. Understanding the relative role of each factor in determining the spread of spatial attention is important not only from a theoretical perspective. It also has direct implications to the design of optimal training procedures that need to be custom-made to the desired performance goals.

5.1. Assessing the width of the attentional window

Using performance as a measure of the spread of attention is complicated for the following reason. Cortical representation is affected by the cortical magnification factor. Since the physical size of the array elements in our display was fixed, the cortical size and consequently salience, was larger for positions near fixation. By using several target distributions we manipulated the spread of attention without affecting (at least initially) the cortical representation. Thus, the effect on detection of these manipulations reflects the induced window of spatial attention. Since attention spread determines learning distribution, we were able to examine learning and thereby assess the factors determining the spatial extent of attention.

Several of our results show that, consistent with previous reports (e.g. Eriksen & Yeh, 1985; Eriksen & St. James, 1986), as difficulty increases, the width of spatial attention decreases. First, we found that in different subject groups and under different training conditions, for shorter (more difficult) SOAs, subjects show greater improvement near fixation than at larger eccentricities. This negative correlation between the extent of learning and eccentricity can not be explained by subjects' inability to improve at the periphery. In fact, previous studies indicate that once attended, the prospects of improvement at larger eccentricities are higher than near fixation (Johnson & Leibowitz, 1979; see review by Ahissar & Hochstein, 1998). There is more to improve at the periphery, perhaps because attentional mechanisms have not been exploited there in past experience (Ahissar & Hochstein, 1997a).

Second, the detection decrease as a function of eccentricity is steeper for poor detectors than for better detecting subjects (and it is unlikely that they have different cortical magnification factors). This difference in distribution is evident following the same learning procedure, and using the same uniform target distribution. Poorer detectors seem to focus attention more steeply around fixation and thus are very poor at the periphery even under conditions in which they manage to detect better near fixation (Figs. 7 and 8).

These learning distributions arise from shrinking the span of attention in those cases when a large window of attention can not sustain successful task performance. When target distribution is uniform across the entire array, it is beneficial to narrow the window of attention and focus on potentially detectable targets near fixation, even at the cost of missing peripheral targets.

5.2. The site of learning easy, focused cases

The above discussion covers the cases when targets are broadly distributed, yet the focus of attention is narrowed due to difficulty. What will be the size of the attentional window when targets are easy to detect and are near fixation (e.g. one position, nearest to fixation)? Being near fixation, they do not require a broad window of attention, yet they are also easy, allowing a broad window of attention from the perspective of signal-to-noise ratio. According to the Reverse Hierarchy Theory, high-level areas may be accessed first and should suffice, even though a relatively narrow window of attention matches the target distribution. One possible outcome of such a scenario with opposing factors is that there will be large inter-subject variability in the spatial width of attention and learning, as subjects weight the factors differently. An alternative is that a small attentional window and large signal-to-noise ratio is implemented at a higher area, and learning transfer will reflect high level tuning properties. In the present study, we did not attempt to determine experimentally whether a narrow window of attention can be implemented at high levels when spatial constraints are conflicting in this manner.

5.3. The site of learning difficult targets at eccentric positions

Another interesting condition is when a difficult target can appear at one of a few distal positions but never appears at fixation. In this case, shrinking the attention window around fixation — the typical strategy for hard conditions, would be the worst strategy. It would be more useful to split the focus of attention to two (or more) foci, or, if this is not possible, to focus on one of the several plausible positions. Can subjects adopt this useful strategy or are they automatically forced into shrinking attention around the center?

To assess the effects of such a condition, we trained subjects with a smaller target-distractor orientation difference. We reasoned that if subjects had to shrink their window of attention, maximal learning would affect positions near fixation, between targets. In this case, testing with more proximal positions than those trained would reveal full if not greater transfer of the learning effects. Indeed some transfer towards the center was evident (Fig. 9), suggesting that discarding the center entirely was impossible. However, transfer was partial and SOA dependent (with least transfer for briefest SOAs). Thus, in this case, more difficult conditions do not lead to shrinking around fixation.

Yet, previous studies indicate that for brief presentations in which target and distractors are displayed simultaneously, splitting the focus of attention is impossible. Kramer and Hahn (1995; see also Castiello & Umilta, 1992) found that with prolonged viewing subjects may indeed be able to split their attention to two locations. However, when the presentation is brief and has a sudden onset, then attention is automatically directed to the salience centroid and encompasses the entire region including the important items in the scene. For our conditions of brief presentation and rapid onset, we would expect attention to be centered and spread out to the target presentation locations, but perhaps not further. Thus, under difficult conditions, subjects probably focus their attention on one of the two alternative sites (even though they are still physically fixating). Since the detection at each of these two positions was similar, their focus changes and does not stay at one position. The time constant of this shift can not be deduced from these results.

5.4. Relation to previous mappings of attention to the visual hierarchy

Our view of the spatial cone of attention as a function of attended cortical level is novel in its current formulation. However, previous studies introduced similar views from other perspectives. Nakayama (1991), taking the point of view of an 'iconic bottle-neck', suggested the idea of a hierarchy in the form of a pyramid. The basic assumption was that the visual system has a limited capacity for iconic working memory, perhaps in the order of a thousand pixels. A scale has to be chosen — either a low-resolution view where few elements (top of the pyramid) 'cover' a large area or high resolution where covering the same area requires many elements (bottom of pyramid), each with a small aperture. Attention chooses the level. We now suggest that this level is related to cortical hierarchy.

Tsotsos (1995) also suggested a hierarchical view of attention from the perspective of the computational benefit of a search tree. Using a hierarchical search tree minimizes the number of steps needed to search the appropriate representation. Ullman (1995), also from a computational point of view, stressed the benefits of top-down computation reaching low-levels. Recently, Treisman (1996, 1998) (see also Edelman & Duvdevani-Bar, 1997) suggested that fine spatial selection is implemented in the primary visual area.

The idea of attention as a mechanism to refine crude pre-attentive spatial encoding has been suggested by Cohen and Ivry (1989, 1991). Tsal, Meiran and Lamy (1995) extended their view to other dimensions apart from spatial position (e.g. color and shape) and showed its consistency with many experimental results. Recently, the effects of spatial attention in increasing spatial resolution were directly demonstrated by Yeshurun and Carrasco (1998).

The view of these studies regarding the role of attention in fine selection is consistent with ours. We add two new aspects: the dynamics and the implications for learning. The level chosen is determined not by the complexity of the task as verbally described by the experimenter, but rather by the compatibility of the internal representation to the task at hand. Thus, different levels will be chosen for similar conditions depending on relative difficulty. We now further specify factors that determine the extent of spatial attention, in particular target distribution and task difficulty. The first is set by stimulus parameters, and as such may be related to an automatic bottom-up attention mechanism. The latter may only be computed with respect to an internal goal, and as such is also related to top-down attentional mechanisms. These two types of attention may be related to previously described automatic bottom-up driven attention (e.g. Maljkovic & Nakayama, 1994, 1996) compared with top-down attentional control (Mackeben & Nakayama, 1993).

5.5. Implications for practical applications

The results of this study have direct implications for the design of training procedures. Three main functional conclusions may be drawn:

1. When easy conditions need to be learned, a few exemplars may suffice to induce interpolation and

learning. Attention will immediately spread around the stimulus center and encompass untargeted as well as targeted positions. Under more difficult conditions, reduced transfer is expected.

2. When very difficult conditions have to be learned, training should contain only the easiest examples of these conditions. For achieving best performance at short SOAs it is better to train only with target positions proximal to fixation, since otherwise, attention is drawn out over too large an area, yielding only crude resolution, which is not sufficient for detecting anything with these brief SOAs.

Within any one block of trials, trainers should present only test conditions for which the same attentional spread is appropriate, since the attentional window is determined with respect to average block success. This procedure therefore allows for more appropriate attention and learning. For example, in our case, learning for targets proximal to fixation was not optimal for brief SOAs when these positions were mixed with distal targets.

Acknowledgements

We thank Roni Laiwand for technical assistance in performing these studies and the data analysis. We thank Anne Treisman, Robert Shapley and Howard Hock for many helpful discussions of the work. This study was supported by grants from The Israel Institute for Psychobiology (to MA); the US–Israel Binational Science Foundation (BSF) and the Israel Science Foundation of the Israel Academy of Arts and Sciences (to SH).

References

Ahissar, M., & Hochstein, S. (1993). The role of attention in early perceptual learning. *Proceedings of the National Academy of Science USA*, *90*, 5718–5722.

Ahissar, M., & Hochstein, S. (1996a). Learning pop-out detection: specificity to stimulus characteristics. *Vision Research*, *36*, 3487–3500.

Ahissar, M., & Hochstein, S. (1996b). Perceptual learning transfer over space and orientation. *Investigative Ophthalmology and Visual Science*, *37*, 3182.

Ahissar, M., & Hochstein, S. (1996c). Eureka: one shot viewing enables perceptual learning. *Investigative Ophthalmology and Visual Science*, *37*, 3183.

Ahissar, M., & Hochstein, S. (1997a). Task difficulty and visual hierarchy: reverse hierarchies in sensory processing and perceptual learning. *Nature*, *387*, 401–406.

Ahissar, M., & Hochstein, S. (1997b). Does over-training lead to automaticity? *Investigative Ophthalmology and Visual Science*, *38*, 1123.

Ahissar, M., & Hochstein, S. (1998). Perceptual learning. In V. Walsh, & J. J. Kulikowski, *Perceptual constancies: why things look as they do* (pp. 455–498). Cambridge University Press.

Ahissar, M., & Hochstein, S. (1999). The role of attention in simple perceptual learning. In M. Fahle, & T. Poggio, *Perceptual learning*. MIT Press (in press).

Ahissar, M., Laiwand, R., & Hochstein, S. (1998a). 'Learning attention' cannot be divided between tasks in a dual task paradigm. *Investigative Ophthalmology and Visual Science*, *39*, S628 (#2923).

Ahissar, M., Laiwand, R., Kozminski, G., & Hochstein, S. (1998b). Learning pop-out detection: building representations for conflicting target–distractor relationships. *Vision Research*, *38*, 3095–3107.

Ball, K., & Sekuler, R. (1987). Direction specific improvement in motion perception. *Vision Research*, *27*, 953–965.

Beard, B. L., Levi, D. M., & Reich, L. N. (1995). Perceptual learning in parafoveal vision. *Vision Research*, *35*, 1679–1690.

Berardi, N., & Fiorentini, A. (1987). Interhemispheric transfer of visual information in humans: spatial characteristics. *Journal of Physiology*, *384*, 633–647.

Castiello, U., & Umilta, C. (1992). Splitting focal attention. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 837–848.

Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 650–663.

Cohen, A., & Ivry, R. (1991). Density effects in conjunction search: evidence for a coarse location mechanism of feature integration. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 891–901.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkey. In F. Boller, & J. Grafman, *Handbook of neuropsychology*, vol. 2 (pp. 267–299). Amsterdam: Elsevier.

Edelman, S., & Duvdevani-Bar, S. (1997). A model of visual recognition and categorization. *Philosophical Transactions of the Royal Society of London B, Biological Science*, *352*(1358), 1191–1202.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Perception and Psychophysics*, *16*, 143–149.

Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 583–597.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Perception and Psychophysics*, *40*, 225–240.

Fiorentini, A., & Berardi, N. (1980). Learning in grating waveform discrimination: specificity for orientation and spatial frequency. *Vision Research*, *21*, 1149–1158.

Harris, J. P., & Fahle, M. (1998). The use of different orientation cues in vernier acuity. *Perception and Psychophysics*, *60*, 405–426.

Johnson, C. A., & Leibowitz, H. W. (1979). Practice effects for visual resolution in the periphery. *Perception and Psychophysics*, *24*, 218–224.

Karni, A., & Sagi, D. (1991). Where practice makes perfect: evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences USA*, *88*, 4966–4970.

Kramer, A., & Hahn, S. (1995). Splitting the beam: distribution of attention over non-contiguous regions of the visual field. *Psychological Science*, *6*, 381–386.

Krose, B. J. A., & Julesz, B. (1989). The control and speed of shifts of attention. *Vision Research*, *29*, 1607–1619.

LaBerge, D., & Brown, V. (1986). Variations in the size of the visual field in which targets are presented: an attentional range effect. *Perception and Psychophysics*, *40*, 188–200.

Lawrence, D. H. (1952). The transfer of discrimination along a continuum. *Journal of Comparative and Psychology*, *45*, 511–516.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory and Cognition*, *6*, 657–672.

Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Perception and Psychophysics*, *58*, 977–991.

Mackeben, M., & Nakayama, K. (1993). Express attentional shifts. *Vision Research*, *33*, 85–90.

Mollon, J. D., & Danilova, M. V. (1996). Three remarks on perceptual learning. *Spatial Vision*, *10*(1), 51–58.

Motter, B. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2 and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, *70*, 909–919.

Nakayama, K. (1991). The iconic bottleneck and the tenuous link between early visual processing and perception. In C. Blakemore, *Vision: coding and efficiency* (pp. 411–422). Cambridge: Cambridge University Press.

Pavlov, I. P. (1927). *Conditioned reflexes* (pp. 121–122). Oxford: Oxford University Press.

Poggio, T., Fahle, M., & Edelman, S. (1992a). Fast perceptual learning in visual hyperacuity. *Science*, *256*, 1018–1102.

Poggio, T., Edelman, S., & Fahle, M. (1992b). Learning in visual modules from examples: a framework for understanding adaptive visual performance. *CVGIP*: *Image Understanding*, *56*, 22–30.

Polat, U., & Sagi, D. (1994). Spatial interactions in human vision: from near to far via experience dependent cascades of connections. *Proceedings of National Academy of Science USA*, *91*, 1206–1209.

Posner, M. I., & Raichle, M. E. (1997). *Images of mind*. New York: Freeman.

Quick, R. F. (1974). A vector magnitude model of contrast detection. *Kybernetic*, *16*, 65–67.

Ramachandran, V. S., & Braddick, O. (1973). Orientation specific learning in stereopsis. *Perception*, *2*, 371–376.

Rosenthal, O. (1995). Behavioral effects in V1. Master thesis, Hebrew University, Jerusalem.

Rubin, N., Nakayama, K., & Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory contour perception. *Current Biology*, *7*, 461–467.

Schoups, A., & Orban, G. (1995). Perceptual learning in texture discrimination. How specific is the improvement? *Perception* (*Supplement*), *24*, 83.

Schoups, A. A., Vogels, R., & Orban, G. (1995). Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularity. *Journal of Physiology*, *483*, 797–810.

Shiu, L., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception and Psychophysics*, *52*, 582–588.

Sutherland, N. S., Mackintosh, N. J., & Mackintosh, J. (1963). Simultaneous discrimination training of Octopus and transfer of discrimination along a continuum. *Journal of Comparative and Physiological Psychology*, *56*, 150-156.

Treisman, A. (1992). Perceiving and re-perceiving objects. *American Psychologist*, *47*, 862–875.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–178.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London B*, *Biological Science*, *353*(1373), 1295–1306.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Tsal, Y., Meiran, N., & Lamy, D. (1995). Towards a resolution theory of visual attention. *Visual Cognition*, *2*, 313–330.

Tsotsos, J. K. (1995). Toward a computational model of visual attention. In T. V. Papathomas, C. Chubb, A. Gore, & E. Kowler, *Early vision and beyond* (pp. 207–218). Cambridge MA: MIT Press.

Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bi-directional information flow in the visual cortex. *Cerebral Cortex*, *5*, 1–11.

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, *396*(6706), 72–75.