Available online at www.sciencedirect.com
ScienceDirect
Vision Research 48 (2008) 1014–1017
Vision Research
www.elsevier.com/locate/visres

The coordinate frame of pop-out learning

Ayelet McKyton [a,*], Ehud Zohary [a,b]

[a] Neurobiology Department, Life Science Institute, Hebrew University, Jerusalem, Israel
[b] Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem, Israel

Received 24 July 2007; received in revised form 18 October 2007

Abstract

Saccades are ubiquitous in natural vision. One way to generate a coherent representation of a scene across saccades is to produce an extra-retinal coordinate frame (such as a head-based representation). We investigate this issue by behavioral means: Participants learned to detect a 3D-pop-out target in a fixed position. Next, target was relocated in one coordinate frame while maintaining it fixed in the others. Performance was severely affected only when the change in target position occurred in a retinotopic coordinate frame. This further suggests that perceptual learning occurs in retinotopic regions having receptive fields restricted within a hemifield.
© 2008 Elsevier Ltd. All rights reserved.

Keywords: Coordinate system; Retinotopic representation; Craniotopic; Receptive field; Saccade

1. Introduction

It is commonly assumed that the ventral visual pathway is engaged in object recognition, while the dorsal pathway is the one involved in the necessary coordinate transformations that allow visually guided action (Andersen & Buneo, 2002). However, to be able to recognize complex scenes we typically scrutinize the image, making multiple eye movements. Every eye movement results in a different image on the retina, thereby generating a radically different pattern of activity in the retinotopic areas (such as primary visual cortex). Our perception, however, is of a stable scene rather than a jittery one. This suggests that the representation at higher order, object-related areas, may incorporate information about eye position and depend on the location of the objects in *space*. Indeed, recent evidence suggests that the fMRI signal in both LOC (McKyton & Zohary, 2007) and ventral MT (d'Avossa et al., 2007) is influenced by the object's position on the screen more than its position on the retina. This suggests that the neurons' spatial selectivity in the higher order ventral areas may be based on spatiotopic coordinates.

It is has been known for some time that some perceptual tasks show clear position specificity. For instance, it is easier to recognize a previously shown pattern of random dots (Foster & Kahn, 1985; Nazir & O'regan, 1990) or a random checkerboard stimulus (Dill & Fahle, 1997) if it was presented in the same, rather than in a different, location. These experiments, however, could not reveal the *coordinate frame* of this position specificity since the target's new position on the screen was accompanied by a change both in its position on the retina and in its position with respect to the subject's head. In this experiment we manipulate the target position, to generate changes which are exclusive to one frame of reference, and study its effect on perceptual performance. By finding the relevant coordinate frame in which learning occurs, we can indirectly infer the brain regions which are intrinsically involved in perceptual learning of 3D shape detection.

We used a common task called pop-out detection: requiring detection of an odd target in an array of distractors. Typically, while this task is considered effortless, prac-

* Corresponding author.
E-mail address: ayelet.mckyton@mail.huji.ac.il (A. McKyton).

0042-6989/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2008.01.013

tice leads to substantial improvement in performance. However, if the target is always present in a specific location in the array, a change in its location often results in an elevated threshold, closer to the initial performance levels (Ahissar & Hochstein, 1996). This specificity to the target's learned position was taken as evidence that learning occurs at retinotopic visual areas. But since the target's position relative to the screen was also changed, together with its relative position within the array of distractors, it is impossible to infer in what coordinate frame learning occurs.

To pinpoint the relevant coordinate frame in which learning occurs, we trained our participants to detect a target in a fixed position. After this learning phase, we tested performance when the target was relocated in one coordinate frame while maintaining it fixed in the others. We assumed that performance will be severely affected only if the change in target position occurred in the coordinate frame in which the task was originally learned. We take advantage of the putative transformation from a retinotopic coordinate frame in early visual areas (such as V1) to extra-retinal coordinate frame in higher ventral areas (such as LOC) to suggest which brain areas might be involved in 3D shape detection.

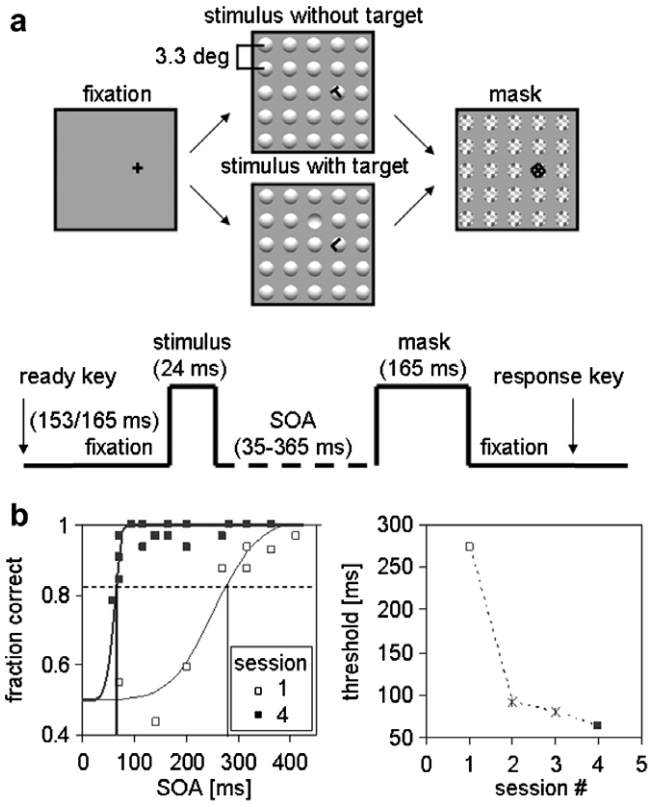

Fig. 1. Experimental paradigm and threshold calculation method. (a) Illustration of the stimuli used in the experiments and the trial temporal sequence. (b) Left: results of one subject in the initial (empty squares) and fourth learning sessions (filled squares). The data were fit by a psychometric curve. The threshold was calculated as the SOA required for 81.6% correct performance (dashed line). Right: a learning curve showing changes in the threshold across sessions.

2. Methods

2.1. Subjects

Eleven subjects participated in these experiments. Subjects were 18–29 years old, naïve and with normal or corrected to normal eyesight. Experiments were undertaken with the understanding and written consent of each subject.

2.2. Stimuli and procedure

The stimuli were 5×5 arrays of shape-from-shading spheres (see Fig. 1a). In half of the trials all elements were identical, seen as illuminated from above and perceived as convex spheres. In the other half, one of the elements (the target presented at a fixed location) was seen as illuminated from below and perceived as an odd concave sphere among convex ones. When the stimulus array appeared, the fixation cross was replaced with the letter T or L. This letter could appear in each of the four different oblique orientations. A mask followed each stimulus after a variable stimulus onset asynchrony. The mask was composed of a 5×5 array of disks positioned at the same location as the original array of elements. Each disk was composed of fragments from the two different spheres, the distractor and the target. The mask also included a stimulus composed of Ts and Ls in all possible orientations at their original location.

The temporal sequence of each stimulus presentation was as follows (Fig. 1b): each trial started with a fixation cross. The subject was required to respond by pressing the ready key, leading to the stimulus array appearance after 153 or 165 ms. The stimulus was on for 24 ms. Following a variable delay from stimulus onset (the SOA), a mask appeared for 165 ms. The subject pressed one response key to indicate whether a T or an L was present and whether a pop-out target appeared. Subjects were instructed to respond "S" if both T and the pop-out target were present; "L" if an L and the pop-out target were present; "X" if they saw a T without the pop-out target; and "<" if they saw an L without the pop-out target. The fixation point changed its color to green or red to indicate a correct response (in both tasks) or an incorrect response (in at least one of the tasks), respectively. Subjects rested their head on a chin-rest facing the middle of the screen.

Stimuli were presented in blocks of 32 trials using the same SOA. There were 6 blocks within a set, each with a different SOA (i.e. 192 trials). A session included 4 sets (i.e. 768 trials). In each set, the order of block presentation was always from easy to difficult trials (i.e. from the longest SOA to the shortest one). The first set on each day was a standard one, spanning the whole dynamic range (from 365 to 35 ms). The range of the SOAs to be presented in each set was chosen according to the subject's performance on the previous set (to optimally span the changing dynamic range due to learning).

2.3. Experimental procedure

The subjects first practiced the task for four days (one session a day) using fixed conditions, in which the array, the target and the fixation point positions were set (Fig. 2a "learned"). Following this initial learning stage, generalization was tested using various experimental manipulations on different days. Object-based translation was achieved by relocating the distractors and keeping the target and the fixation point in the same position as in the "learned" state (Fig. 2a "object"). Head-based translation (Fig. 2a; "head"), tested during a separate session, was attained by shifting the entire array (target, distractors and fixation point) with respect to the screen (and therefore also with respect to the subject's head). Finally, we generated a retinotopic trans-

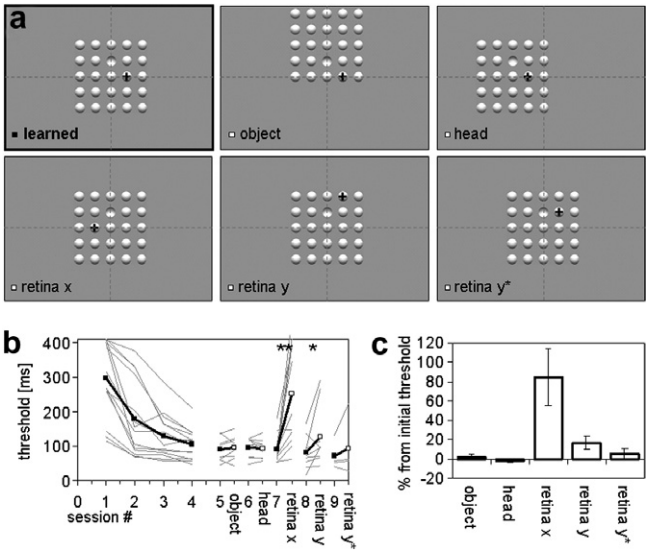

Fig. 2. Experimental design and results. (a) The stimulus configuration during the learning phase (“learned”) and the various test phases. Screen center is marked by dashed lines. (b) Learning curves for each subject (gray) and the population average (black). *P*-values were calculated based on comparison between the learned task and the new task performance on the same day, using a non-parametric one-tailed Wilcoxon matched-pairs signed-ranks test. *P*-values for “object”, “head”, “retina x”, “retina y1” and “retina y*” are 0.21, 0.25, 0.0005, 0.027 and 0.31; $n = 9, 9, 11, 9, 4$, respectively. (c) Performance level, normalized according to each individual’s initial threshold and final threshold at the day of testing. Error bars represent SEM.

lation condition by relocating only the fixation point on the screen either in the horizontal (Fig. 2a “retina x”) or in the vertical (Fig. 2a “retina y” and “retina y*”) axis.

The order of the various tests was randomized across subjects. Each of these testing sessions consisted of 4 sets. In 2 sets (384 trials) the position of the whole array, or the target, or the fixation point was altered in the same way. The other 2 sets served as internal controls and were identical to those presented in the learning phase (designed to monitor baseline performance on the same day). The order of the sets within a session differed between subjects. To assure that the results were not due to some left–right visual field asymmetry, the position of the stimulus array on the screen was randomized among subjects such that half the subjects saw a mirror image of the stimuli shown to the other half.

2.4. Data analysis

The percentage of correct responses was measured as a function of SOA. The average performance of the day’s sessions was evaluated by computing the best fit psychometric function of the form: $f = 1 - 0.5 \cdot \exp - (t/\tau)^{\sigma \cdot \tau}$, where f is the proportion of correct responses, t is the trial SOA, τ and σ are free parameters: τ the threshold SOA at 81.6% correct, and σ is the slope at the threshold multiplied by $2e$. We used the threshold τ as a measure of subjects’ performance. We included only trials in which subjects correctly identified the foveal T/L targets.

3. Results

Our participants were trained to detect an odd-element sphere within an array of spheres, at a specific location (Fig. 1). For each session, the percent of correct answers was measured as a function of the SOA and the psychophysical threshold was evaluated by fitting the behavioral data with a psychometric function (see example in Fig. 1b). Typically, detection thresholds improved dramatically across sessions in all subjects until reaching asymptotic performance (Fig. 2b; sessions #1–4). After this learning phase, we tested the degree of learning transfer by changing the target position only in one specific coordinate frame at a time. Each such manipulation was preceded by a session using the originally learned configuration to compare performances on the same day.

3.1. Testing for object-based coordinates

To investigate if learning occurred in object-based coordinates, the target’s position relative to the distractors was changed while maintaining its retinal and head-based position (Fig. 2a “object”). This manipulation was performed by changing only the location of the distractors while keeping the fixation point and target position fixed on the screen. This resulted in individual performance thresholds similar to those of the control session (Fig. 2b and c; compare sessions “object” and “5”), suggesting that this task is not learned using object-based coordinates.

3.2. Testing for head-based coordinates

Next, to test whether the 3D target detection task is carried out in head-based coordinates, we shifted the entire array, together with the fixation point, to a different location on the screen (Fig. 2a; “head”). This manipulation kept the target in the same retinotopic position and in the same location within the array but changed its position relative to the subject’s head. Here too, performance was similar to that in the well-trained condition (Fig. 2b and c; compare sessions “6” and “head”), suggesting that this pop-out task is invariant to head-based position.

3.3. Testing for retinotopic coordinates

Finally, we changed the target position only relative to the retina by moving the fixation point to different locations on the screen while keeping the array (including the target) constant (Fig. 2a; conditions “retina x”, “retina y” and “retina y*”). Moving the fixation point along the horizontal axis by 6.6 degrees shifted the target position to a (mirror symmetric) position in the other hemifield (Fig. 2a; “retina x”). This resulted in a markedly increased threshold (Fig. 2b and c; compare sessions “7” and “retina x”), suggesting that learning is specific to retinal position. However, moving the fixation point along the vertical axis by the same amount had a much smaller effect (Fig. 2b and c; compare session “8” and “retina y”), which disappeared altogether when the fixation point was moved by only 3.3 degrees (Fig. 2b and c; compare sessions “9” and “retina y*”).

4. Discussion

Our interpretation of these results is that the essential element in learning to detect a shape-from-shading pop-out target occurs in areas where the neuronal receptive fields are still retinotopic and restricted to one hemifield. Their receptive field size is typically large enough to include the closer alternate target position ("retina y*") but not the more distant target ("retina y"). A reasonable candidate area is V4. Its neurons' average receptive field size is restricted to the contralateral visual field, spanning about 5 degrees in diameter at the target eccentricity used in this experiment (Gattass, Sousa, & Gross, 1988; Smith, Singh, Williams, & Greenlee, 2001). It has been shown that neurons in monkey V2 are more selective than V1 to similar shape-from-shading pop-out images (Lee, Yang, Romero, & Mumford, 2002). The authors have not tested this in V4 but their gradient of pop-out selectivity is in accordance with our results, suggesting that shape-from-shading pop-out may rely on activity originating in extra-striate areas (such as V4). Indeed, perceptual learning is accompanied by an improvement in the stimulus selectivity of extra-striate neurons in V4 and MT (Yang & Maunsell, 2004; Zohary, Celebrini, Britten, & Newsome, 1994).

Our behavioral paradigm introduces a novel way to elucidate *where* might the changes that result in improved perceptual performance, occur in the brain. For example, it seems reasonable to expect that learning to detect a "simpler" 2D pop-out target among an array of "flat" discs (half white, half black) will also be specific to the *retinal* position, with sensitivity for even smaller translations. That would suggest that learning is based on changes occurring in neurons with *small*er receptive fields earlier in the visual pathway. It seems feasible that perceptual learning in tasks requiring comprehension of the "gist of the scene" (such as the presence or absence of an animal in a natural scene (Thorpe, Fize, & Marlot, 1996), might well-generalize across retinal positions. This would suggest that such learning is based on neuronal changes that occur at higher order object-related visual areas (Ahissar & Hochstein, 2004).

Acknowledgments

We thank Merav Ahissar for insightful comments. This study was funded by the Israel Science Foundation of the Israel Academy of Sciences Grant #8009.

References

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

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences, 8*, 457–464.

Andersen, R. A., & Buneo, C. A. (2002). Intentional maps in posterior parietal cortex. *Annual Review of Neuroscience, 25*, 189–220.

d'Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nature Neuroscience, 10*, 249–255.

Dill, M., & Fahle, M. (1997). The role of visual field position in pattern-discrimination learning. *Proceedings. Biological Sciences, 264*, 1031–1036.

Foster, D. H., & Kahn, J. I. (1985). Internal representations and operations in the visual comparison of transformed patterns: Effects of pattern point-inversion, position symmetry, and separation. *Biological Cybernetics, 51*, 305–312.

Gattass, R., Sousa, A. P., & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *Journal of Neuroscience, 8*, 1831–1845.

Lee, T. S., Yang, C. F., Romero, R. D., & Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nature Neuroscience, 5*, 589–597.

McKyton, A., & Zohary, E. (2007). Beyond retinotopic mapping: The spatial representation of objects in the human lateral occipital complex. *Cerebral Cortex, 17*, 1164–1172.

Nazir, T. A., & O'Regan, J. K. (1990). Some results on translation invariance in the human visual system. *Spatial Vision, 5*, 81–100.

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex, 11*, 1182–1190.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

Yang, T., & Maunsell, J. H. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. *Journal of Neuroscience, 24*, 1617–1626.

Zohary, E., Celebrini, S., Britten, K. H., & Newsome, W. T. (1994). Neuronal plasticity that underlies improvement in perceptual performance. *Science, 263*, 1289–1292.