**NEUROBIOLOGIAE
EXPERIMENTALIS**

Model for visual localization

Nadejda Bocheva and **Leon Mitrani**

Institute of Physiology, Bulgarian Academy of Sciences,
23 Ac. G. Bonchev St., Sofia 1113, Bulgaria

Abstract. The position of an object in the visual field is a global
characteristic of the object which must be determined shortly after its
appearance. This characteristic is useful for organizing actions towards
the object or for its better examination. A model of visual localization
based on the concept of "image function" is proposed. It predicts that
the centroid of a visual object determines its position. The centroid
could be easily extracted from the maximum in the image function
under certain conditions. Symmetry and size may influence the
accuracy (absolute and relative) in locating visual objects.

INTRODUCTION

For many years now the process of figure-ground segregation has been recognized as fundamental in visual perception. The visual system is highly specialized in determining and locating edges in visual space. It is equipped with mechanisms that analyses light distribution and basing on this analysis it determines effectively the edges between areas of different luminance, colour, and texture. Important as they may be for the process of figure-ground segregation and the perception of contours and form of visual objects, edge determining mechanisms seem not appropriate for providing information about some global characteristics of visual objects, such as overall orientation and position in visual space.

Locating boundaries and contours of a visual form might be thought sufficient for determining the position of that form in space, but there are enough data to dismiss this possibility (Watt and Morgan 1983, Ward et al. 1985). Visual location of an object has to be attributed to some aspect of the visual image of the segregated figure as a whole, irrespective of the local analysis providing information about contour location. Which characteristic of the image represents its position is difficult to deduce only on logical grounds. We tried to find the answer to this question by a series of experiments (unpublished data) where the task of the subjects was to determine the position of tachistoscopically presented patterns. Our results showed that for both convex and concave patterns the point which represents best their position is the centroid of visual image determined by the luminance distribution. If the objects are homogenous material bodies, the centroid coincides with the center of gravity. This characteristic is invariant under transformations of the so-called Euclidean similarity group, which consists of translations, rotations, dilations, reflections and their composites. The centroid characterizes the trajectory of the objects (Proffitt and Cutting 1980).

When an object appears in the visual field, its perceived position is important for quick manipulation with the object, for thorough examination or tracking, for determining object motion characteristics (for moving objects). If the centroid position is numerically calculated this will be very time-consuming, while our results (unpublished data) showed that the position was determined almost immediately. We propose a model of visual localization based on existing ideas of primary information processing and on an analog principle.

BASIC HYPOTHESES

The concept of "image function" suggests that visual processing is based not on the original pattern itself, but on an image obtained through blurring the pattern. When a dot stimulus appears in the visual field, some visual neurons responds with maximal activity, while the evoked activity of the adjacent neurons decreases with the increase of the distance from the most activated cells. The point-spread function can be represented through a Gaussian function G(r):

$$G(r) = M exp\left(\frac{-r^2}{2s^2}\right)$$

where r is the distance in polar coordinates, M is a constant ("amplitude") and s is another constant representing the spread of neural activity. For simplicity we shall accept that M is equal to unity.

We shall also suppose that cell responses are linear in regard to the intensity of stimulation, i.e. there is no interference between the activity caused by different stimuli This leads to the preservation of spatial relations between the different parts of the visual objects in the over-all activity of the cells.

The image function is obtained by a convolution of the input intensity function with a Gaussian function:

$$F(r)=I(r) \otimes G(r)$$

where F is the image function. I is the intensity function, and $\otimes$ is the sign for convolution. If we accept that visual objects consist of a great but finite num-

ber of dot elements, the intensity function could be presented as a sum of delta-functions:

$$I(r) = \sum \delta(r - r_i)$$

Then the image function has the form:

$$F(r) = \sum_i^n G(r - r_i)$$

Our assumptions are typical of a lot of models, connected with the primary levels of visual processing (Olzak and Thomas 1986). In some models the point-spread function is represented by the Laplacian of a Gaussian, or by the difference of two Gaussian. The influence of the choice of the point-spread function will be discussed later. Here we would like to draw attention to the fact that most of these models deal with mechanisms related to the local characteristics of the images (Marr 1982). We tried to show that using similar models it is possible to determine not only local, but also global characteristics of the image, in particular its position.

According to Koenderink (1984) every image could be uniquely embedded in a one-parameter family of derived images with the resolution as a parameter. If these images are obtained through a convolution with Gaussian functions, no spurious details are generated when the resolution diminishes (the resolution is inverse to the spread). With increase in the spread, the number of maxima in the image function decreases.

THE MODEL

We tried to estimate the relation between the spread s of the Gaussian function and the size of an image in order to obtain only one maximum in the image function and to see if this maximum would correspond to some important feature of the object - for instance to its position represented by the centroid.

Each of the constituting functions G(r-r$_i$) could be approximated by a Taylor series. If we discard the members of higher than fifth degree, the image functions becomes:

$$F(r) = F(x,y) = n - \frac{\sum_i^n (x-x_i)^2 + \sum_i^n (y-y_i)^2}{2s^2} +$$

$$+ \frac{\sum_i^n (x-x_i)^4 + \sum_i^n (y-y_i)^4}{8s^4} + \frac{\sum_i^n (x-x_i)^2 (y-y_i)^2}{8s^4}$$

This expression could be transformed in such a way that all sums are substituted by expressions including constants, representing some image characteristics. For instance, the sum :

$$\frac{\sum_i^n (x-x_i)^2}{2s^2}$$

could be written as

$$\frac{n(x-\bar{x})^2}{2s^2} + \frac{ns_x^2}{2s^2}$$

After such a transformation the image function becomes:

$$F(x,y) = n - \frac{n(x-\bar{x})^2}{2s^2} - \frac{n(y-\bar{y})^2}{2s^2} + \frac{n(x-\bar{x})^4}{8s^4} + \frac{n(y-\bar{y})^4}{8s^4} +$$

$$+ \frac{n(x-\bar{x})^2 (y-\bar{y})^2}{4s^4} + \frac{3ns_x^2 (x-\bar{x})^2}{4s^4} + \frac{3ns_y^2(y-\bar{y})^2}{4s^4} +$$

$$+ \frac{ns_x^3 A_x (x-\bar{x})}{2s^4} + \frac{ns_y^3 A_y (x-\bar{x})}{2s^4} + \frac{ns_y^2 (x-\bar{x})^2}{4s^4} +$$

$$+ \frac{ns_x^2 (y-\bar{y})^2}{4s^4} + \frac{nrs_x s_y (x-\bar{x})(y-\bar{y})}{2s^4} + \frac{nm_{12} (x-\bar{x})}{2s^4} +$$

$$+ \frac{nm_{21} (y-\bar{y})}{2s^4} + const$$

where $\bar{x}$, $\bar{y}$ correspond to the coordinates of the centroid of the figure, s$_x$ and s$_y$ are constants, determined by the dimension of the image and the distribution of its elements; they correspond to the standard deviations of the distribution of the dot elements of the pattern, A$_x$ and A$_y$ depend on the symmetry in the distribution of image elements; they are analogous to the coefficients of asymmetry of the distribution of the elements in the image; r is a constant that characterizes the elongation of the image; it is analogous to the coefficient of correlation of the two-dimensional distribution of the coordinates of

the dots, belonging to the image. This constant could be zero if the coordinate system is appropriately chosen. In this coordinate system the constants m_{12} and m_{21} are also zero. We accept that the coordinate system is appropriately chosen. In this coordinate system the constants m_{12} and m_{21} are also zero. We accept that the coordinate system is chosen in such a way that the constants r, m_{12} and m_{21} are equal to zero. In such a coordinate system the values of s_x and s_y are extremal. Then the image function becomes:

$$F(x,y) = n - \frac{n(x-\bar{x})^2}{2s^2} - \frac{n(y-\bar{y})^2}{2s^2} + \frac{n(x-\bar{x})^4}{8s^4} + \frac{n(y-\bar{y})^4}{8s^4} +$$
$$+ \frac{n(x-\bar{x})^2(y-\bar{y})^2}{4s^4} + \frac{3ns_x^2(x-\bar{x})^2}{4s^4} + \frac{3ns_y^2(y-\bar{y})^2}{4s^4} +$$
$$+ \frac{ns_x^3 A_x(x-\bar{x})}{2s^4} + \frac{ns_y^3 A_y(y-\bar{y})}{2s^4} + \frac{ns_y^2(x-\bar{x})^2}{4s^4} +$$
$$+ \frac{ns_x^2(y-\bar{y})^2}{4s^4} + const$$

The image function will have an extremum in the points for which its first derivatives are zero:

$$F_x' = \frac{-2ns^2(x-\bar{x})^2 + n(x-\bar{x})^3 + n(x-\bar{x})^2(y-\bar{y})}{2s^4} +$$
$$+ \frac{3ns_x^2(x-\bar{x}) + ns_y^2(x-\bar{x}) + ns_x^3 A_x(x-\bar{x})}{2s^4} = 0$$

$$F_y' = \frac{-2ns^2(y-\bar{y})^2 + n(y-\bar{y})^3 + n(x-\bar{x})(y-\bar{y})^2}{2s^4} +$$
$$+ \frac{3ns_y^2(y-\bar{y}) + ns_x^2(y-\bar{y}) + ns_y^3 A_y(y-\bar{y})}{2s^4} = 0$$

If the dot distribution has central symmetry, these expressions will be equal to zero at the point, corresponding to the centroid of the image. However, it is difficult to solve these equations in the general case. We accept that they are equal to zero at the point with coordinates $\bar{x} + d_1$, $\bar{y} + d_2$, where d_1 and d_2 represent the deviation of the extremum from the centroid of the image. The deviation of the extremum from the centroid depends on the presence of central symmetry of the dot distribution and on the size of the image when there is no central symmetry.

The constraint that a function of two variables has extremum is met by a positive Hessian

$$F_{xx}'' F_{yy}'' - F_{xy}''^2$$

This is fulfilled when

$$s^2 > \max\left[s_x^2 + 3s_y^2 + d_1^2 + 3d_2^2, \; s_x^2 + s_y^2 + 3d_1^2 + 3d_2^2\right]$$

However, the necessary condition for having a maximum is that $F_{xx}'' + F_{yy}''$ is negative at the point of the extremum. This leads to the restriction for the size of the spread:

$$s^2 > s_x^2 + s_y^2 + d_1^2 + d_2^2$$

When the distributions of the coordinates of image dot elements have a central symmetry this expression is simplified, because in this case the deviations of the maximum from the image centroid are zero. The condition for having maximum in this case is:

$$s^2 > s_x^2 + s_y^2 \quad (1)$$

It is evident that if s is large, the deviations of the maximum from the centroid of the image have a negligible effect.

A particular case is that of two dot stimuli. In this case the configuration of dots has central symmetry and A_x and A_y are zero. So the maximum of the image function will coincide with the centroid $\bar{x}$, $\bar{y}$ of the image, where $\bar{x}=(x_1+x_2)/2$, $\bar{y}=(y_1+y_2)/2$ if the relation (1) between the spread and the size of the configuration of two dots holds. For the one-dimensional case this relation is:

$$s^2 > s_x^2 = \frac{\left(x_1 - \frac{1}{2}(x_1+x_2)\right)^2 + \left(x_2 - \frac{1}{2}(x_1+x_2)\right)^2}{2}$$

or s should be greater than half of the distance between the two dots. This requirement is used in different models of visual processing (Smith and Vos 1986, Grossberg and Rudd 1989, Mates et al. 1990).

The above considerations and the analogy between the constants A_x and A_y to the coefficients of asymmetry suggests that the deviation of the maximum from the centroid of the image is towards the greater density of image elements.

Examples of the image function for symmetrical and asymmetrical patterns of four dots are presented on Figs. 1 and 2. From the figures it is evi-
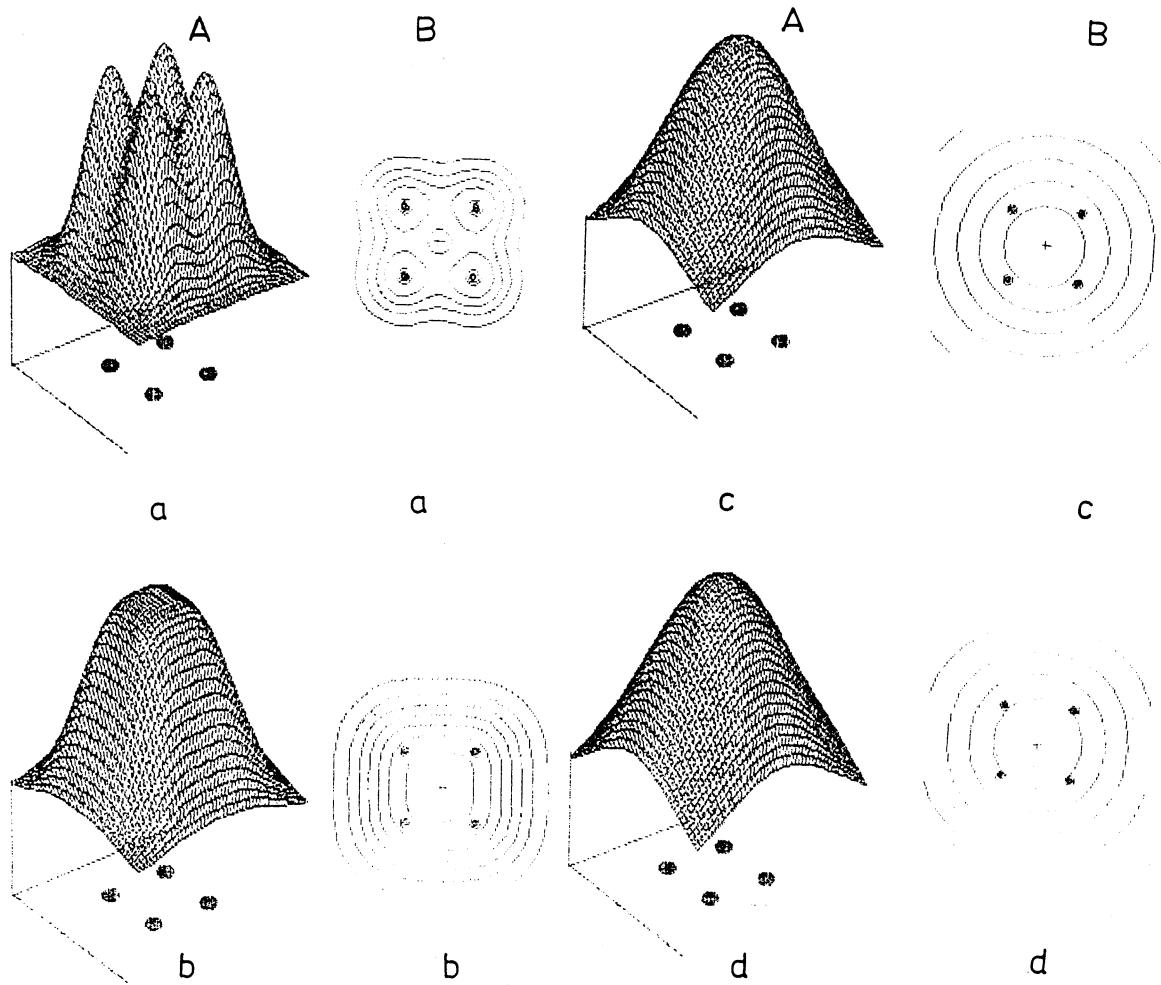
Fig. 1. *A*, Three-dimensional representation of the image function for a pattern of four dots determining a square (the size of the dots is enlarged in figure). The functions are obtained with different spreads (s). *B*, Isolines of the simulated image function for different spreads in an appropriately rotated coordinate system.

$$a, \; s = \frac{1}{4}\left(s_x^2 + s_y^2\right) \qquad b, \; s = \frac{1}{2}\left(s_x^2 + s_y^2\right) \qquad c, \; s = \left(s_x^2 + s_y^2\right) \qquad d, \; s = 1.5\left(s_x^2 + s_y^2\right)$$

dent that the maximum shifts toward the centroid with increase of spread for asymmetrical configurations, while for symmetrical it coincides with the centroid. It could also be seen that for the asymmetrical pattern though we obtained one maximum in cases c and d of Fig. 2 the isolines are still not elliptical which could be attributed to the fact that for asymmetrical patterns the constraint for the spread contains not only the constants, corresponding to the size of the patterns, but also constants determined by the symmetry.

DISCUSSION

We demonstrated that the position of every object in the visual field could be quickly determined through the maximum in the image function if one accepts the point-spread function to be Gaussian with a spread greater than a constant, determined by object dimensions. It is reasonable to assume that the values of the spreads (which could be regarded as the size of the receptive fields) in the visual system are limited. Thus for very large objects one
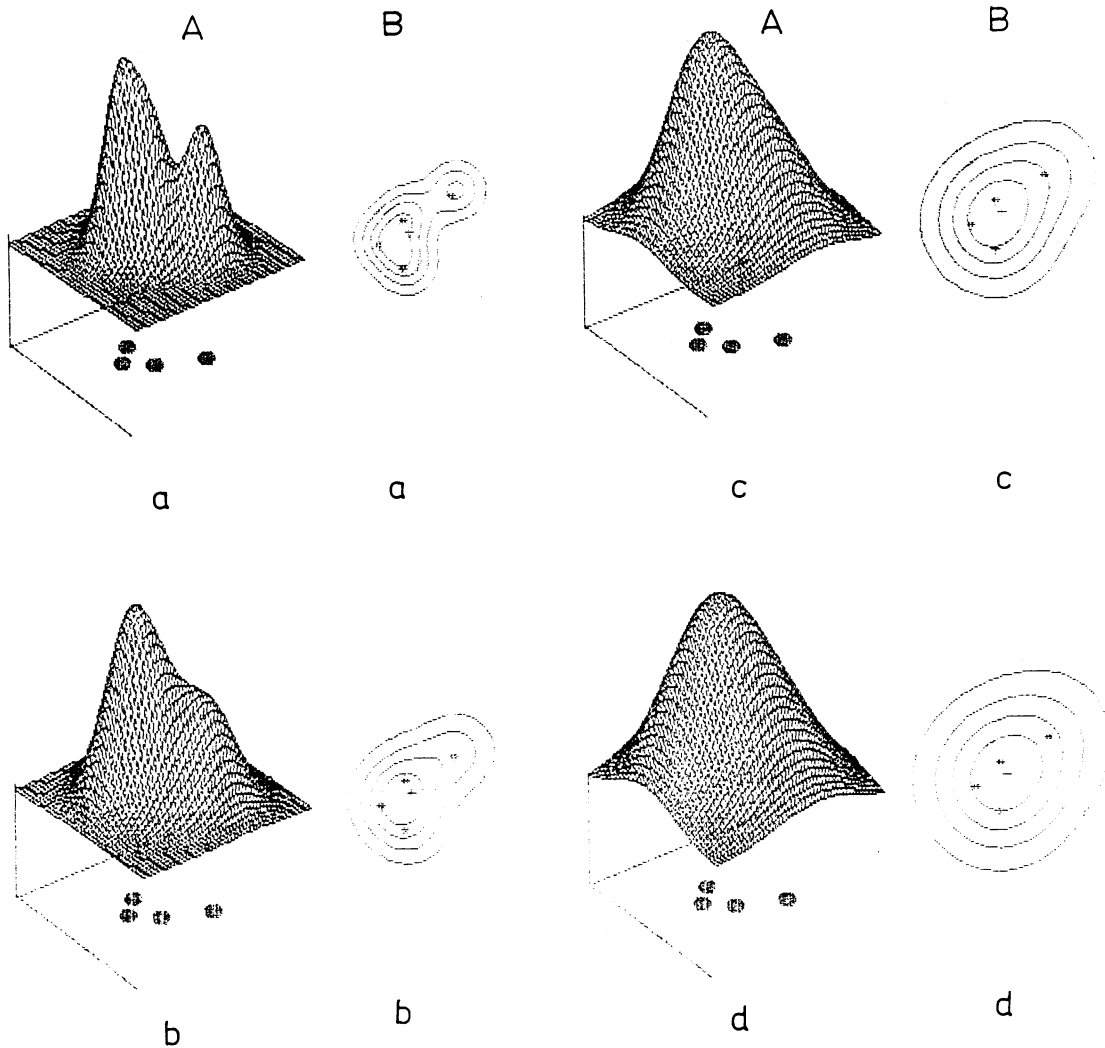
Fig. 2. _A_, Three-dimensional representation of the image function of four asymmetrically positioned dots. Simulation as in Fig. 1_B_. Isolines of the simulated image function for different spreads in an appropriately rotated coordinate system.

$$a, \ s = \frac{1}{4}\left(s_x^2 + s_y^2\right) \qquad b, \ s = \frac{1}{2}\left(s_x^2 + s_y^2\right) \qquad c, \ s = \left(s_x^2 + s_y^2\right) \qquad d, \ s = 1.5\left(s_x^2 + s_y^2\right)$$

could expect an image function with more than one maximum. The position of such an object should not be determined unambigously. It is possible that for large objects there occurs grouping, determined by the maxima in the image function.

When the object dimensions decrease, the activity of the cells with largest spreads (receptive fields) could be represented by an image function with one maximum, but the number of such cells will be not large enough and the accuracy of localization will be not high with respect to both absolute deviation

from the centroid and to relative accuracy (variability of the responses).

With a further decrease of object dimensions the requirements for one maximum in the image function will be fulfilled for a large number of cells with different spreads. This should allow the position of the object to be determined by the cells with optimal spread. Hence with some limits changes in the size of the object will not lead to significant changes in the absolute accuracy of localization, represented by its centroid. Changes in the relative accuracy could

only be due to less expressed maximum in the image function for larger spreads.

It might also be supposed that the proposed model for visual localization is applicable only a short time after the stimulus appearance. This suggestion seems to be reasonable, considering the reorganization of the cell receptive fields. The latter is taken into account when modeling primary visual processing either by the involvement of functions, changing with time (Wörgöter and Holt 1991) or by the assumption about switching from cells with larger receptive fields to cells with smaller ones (Watt 1988). The reorganization in the structure of the receptive fields could be due to the interaction between the cells or to an influence from cells at a higher level. According to Podvigin et al. (1986) in the first 100 ms after the appearance of a stimulus the activity of the cells in the laterate geniculate body is characterizes by excitation only. After this time inhibitory zones are formed. The size of the receptive fields after stimulus appearance is big (8-10° for the laterate geniculate body). The initial activity of these cells could be described by Gaussian functions and the appearance of inhibitory zones could be represented through point-spread functions that contain minima. Such functions are more sensitive to the peculiarities of the image. It is for this reason that most models of visual processing for revealing details involved functions, containing minima - like the Laplacian of a Gaussian, difference of two Gaussians with unequal spreads, Gabor functions. etc. The existence of minima in the point-spread function does not interfere with our considerations, but the range of image sizes for which the image function will have only one maximum will be narrower.

A mechanism for visual localization activated at short times after stimulus appearance would allow the use of this information about localization for other functions and activities: directing eye or limb, directing attention towards the object, determining the trajectory of a moving object. Probably it is not incidental that for objects with a size smaller then 5 angular degrees the saccades are towards the centroid, while for bigger objects the direction of the saccades is less determined, and sight is directed towards different parts of the image (Richards and Kaufman 1969, Zetzche et al. 1984). It is known that there is a relation between eye movements and directing attention (Desimone et al. 1985, Posner et al. 1985). The effects of shift of gaze and shift of attention are nearly the same for the visual system, because they cause visual processing to be dominated by a new input. The structure of the receptive fields could change with time in accordance with the task and the level of attention (Moran and Desimone 1985). The perceived trajectory of object motion is also connected with their centroid (Proffitt and Cutting 1980). The role of the centroid in manipulation with objects is obvious. No great accuracy is necessary for all these processes, but they require information about the object position shortly after the stimulus appearance.

REFERENCES

Desimone R., Wessinger M., Thomas L., Shneider W. 1989) Effects of deactivation of lateral pulvinar or superior colliculus on the ability to selectively attend to a visual stimulus. Soc. Neurosci. Abstr. 15: 162.

Grossberg S., Rudd M. (1989) A neural architecture for visual motion perception: Group and element apparent motion. Neutral Networks 2: 421-450.

Koenderink J.J. (1984) The structure of images. Biol. Cybern. 50: 363-367.

Marr D. (1982) Vision. WH Freeman Comp., New York.

Mates J., Lansky P., Yakimoff N. (1990) Models for the perception of orientation in random dot patterns. Biol. Cybern. 63: 71-80.

Moran J., Desimone R. (1985) Selective attention gates visual perception visual processing in the extrastriate cortex. Science 229: 782-784.

Olzak L., Thomas J. (1984) Seeing spatial patterns. In: Handbook of perception and human performance. Vol. 1. (Eds. K. Boff, L. Kaufman and J. Thomas). John Willey and Sons, West Sussex, p. 7-27.

Podvigin N.F., Makarov F.N., Shelepin J.E. (1986) Elementy strukturno-funktsional'noj organizatsii zritel'no-glazovodvigatel,noj sistemy. Nauka, Leningrad, p. 252.

Posner J.I., Choate L.S., Rafal R., Vaugh J. (1985) Inhibition of return: neural mechanisms and function. Cog. Neuropsychol. 2: 211-228.

Proffitt D.R., Cutting J., (1980) Perceiving the centroid of curvilinearly bounded rolling shapes. Percept. Psychophys. 28: 484-487.

Smith J.T.S., Vos P.G. (1986) A model for the perception of curves in dot figure: The role of local salience of "virtual lines". Biol. Cybern. 54: 407-416.

Richards W., Kaufman L. (1969) "Center of gravity" tendencies for fixation and flow patterns. Percept. Psychophys. 5: 81-84.

Ward R., Casco C., Watt P.J. (1985) The location of noisy visual stimuli. Can. J. Psychol. 39: 387-399.

Watt R.J. (1988) Visual processing: Computational, psychophysical and cognitive research. Lawrence Erlbaum Assoc., Hove, Hillsdale.

Watt R.J., Morgan M.J. (1983) Mechanisms responsible for the assesment of visual location: theory and evidence. Vision Res. 23: 97-109.

Wörgötter F., Holt G. (1991) Spatiotemporal mechanisms in receptive fields of visual cortical simple cells: A model J. Neurophysiol. 65: 494-510.

Zetzche G., Deubel H., Elsuer T. (1984) Similar mechanisms of visual processing in perception and oculomotor control. Perception 13: A6a.