

# Visual inhomogeneity and reference frames

MIGUEL A. GARCÍA-PÉREZ  
*Universidad Complutense, Madrid, Spain*

(Charles W. Eriksen, *Sponsor*)

The role of visual inhomogeneity in the perceived pointing of equilateral triangles is investigated in this paper. An experiment was performed whose results cannot be accounted for by current hypotheses about the mechanisms that assign reference frames. These results support the view that local and not global reference frames are involved in shape perception. It is feasible that local reference frames are constructed from the high-spatial-frequency information, which, owing to visual inhomogeneity, is exclusively extracted from the immediate neighborhood of the fixation point.

An intriguing fact about visual perception is our impressive ability to recognize shapes regardless of irrelevant variations in size and/or orientation. Marr and Nishihara (1977) set out the basis for studying this ability from an information processing approach. They pointed out that object-centered descriptions are needed if shapes are to be recognized when viewed from any vantage point. The problem is in defining how this viewpoint-independent description is achieved; the existing hypotheses, although they differ in many respects, all agree that canonical *reference frames* must be imposed to describe shape information relative to the frame. It is also assumed that resizing and/or rotating a shape makes the mechanisms that assign the reference frame change the spatial scale and orientation accordingly, in such a way that the structural description of the shape remains the same. Again, the problem is in establishing the processes that determine the reference frame orientation. Understanding how this is done is important not only for gaining knowledge about human visual processing, but also to incorporate flexible recognition capability into computer vision systems.

There are two competing psychological answers to the question of how the orientation of the reference frame is established. On the one hand, Palmer and Bucher (Palmer, 1985; Palmer & Bucher, 1981, 1982) put forward a symmetry hypothesis, which states that one of the axes of the Cartesian reference frame takes the direction of an axis of reflectional symmetry in the whole configuration. On the other hand, Jáñez (1983) proposed that one of the axes of the Cartesian reference frame takes the direction perpendicular to that at which the angular energy distribution (AED) in the stimulus reaches a maximum. Both hypotheses give rise to the same predictions (in a qualita-

tive sense) for all the stimuli that have been considered up to now: when a shape is embedded in a context of other shapes or elements, the target shape seems to be described according to a reference frame whose orientation is consistent with both hypotheses. Experiments by Palmer and Bucher also showed that the strength of the reference frame in directing shape description varies as a function of several parameters in the surroundings of the target shape (e.g., number of elements, separation between them, type of texture, etc.) The symmetry hypothesis cannot account for these variations in reference frame effects, whereas the AED hypothesis can (see Jáñez, 1983, 1986). Furthermore, the AED hypothesis can be applied to any stimulus, whereas the symmetry hypothesis is restricted to stimuli with an axis of reflectional symmetry (see Merino, 1986). If the requirement of reflectional symmetry is relaxed and nearly bilateral symmetry is allowed (Palmer, 1987), both hypotheses lead to the same qualitative predictions for all the stimuli used by either Palmer and Bucher (Palmer, 1985, 1987; Palmer & Bucher, 1981, 1982) or Merino (1986).

Despite their differences, both hypotheses share a common feature: the distal stimulus is the one considered to determine the visual reference frame orientation. However, the reference frame is an internal elaboration that is constructed from the visual representation of the distal stimulus, and it is known that the image to which we have conscious access is an inaccurate copy of the distal stimulus because of visual inhomogeneity: we see the outside world through a window of varying resolution from a maximum value at the fovea to a minimum value at the outermost periphery. Thus, the portion of the stimulus that falls onto the peripheral retina, being internally represented with lower resolution, may not be as relevant as are the foveal portions in determining perceptual effects. Taking visual inhomogeneity into consideration, one might hypothesize that it is not the whole configuration but rather the particular elements around the fixation point that determine the reference frame orientation. The possible influence of visual inhomogeneity on reference frame

---

This work was supported by the Universidad Complutense and by a grant from the Fundación Ramón Areces for a research project entitled "Visión Artificial y Visión Humana: Aplicaciones a la Percepción Visual en Robótica." Address correspondence to Miguel A. García-Pérez, Laboratorio de Visión, Facultad de Psicología, Universidad Complutense, Campus de Somosaguas, 28023 Madrid, Spain.

effects have not been studied thus far. All the experiments referred to in the previous paragraph involved foveal presentations of small configurations, and, under such circumstances, visual processing can be regarded as nearly homogeneous. Nonetheless, there are some trends in the experimental data that may be interpreted as signaling the influence of visual inhomogeneity. For example, the strength of reference frame effects increases asymptotically with increasing number of elements in configuration (see Palmer & Bucher, 1981, Figure 4), as though the influence of each element in a configuration decreases with increasing eccentricity of its location. This asymptotic behavior of reference frame effects cannot be accounted for by Jáñez's (1986) model, probably because he considers homogeneous processing.

To test whether or not visual inhomogeneity is involved in determining the visual reference frame orientation, an experiment was carried out using triangle stimuli such as that shown in Figure 1. The present formulations of the AED and the symmetry hypotheses give rise to predictions about the orientation of the reference frame that will be imposed on the image that differ from those arising when visual inhomogeneity is considered. The first two hypotheses predict that a single reference frame will be imposed, and, therefore, all triangles should be perceived as pointing in the same direction, regardless of which triangle one is looking at. If, as a consequence of visual inhomogeneity, only the elements around the fixation point are relevant, then the triangles in each radial configuration, when fixated, would be seen to point in the corresponding centrifugal direction. Therefore, if subjects were asked to say whether the particular triangles that are numbered in Figure 1 point up, down, left, or right, and if visual inhomogeneity plays a role in reference frame effects, then their response times to triangles 2, 3, 5, and

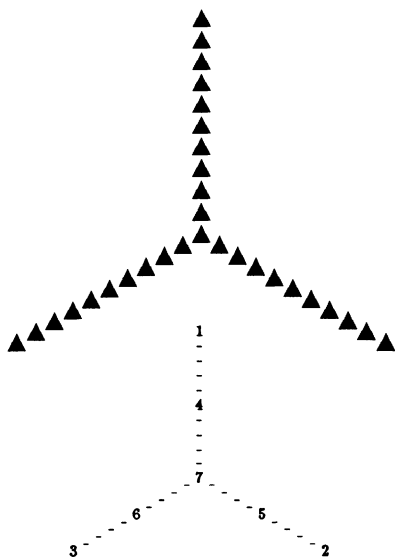


Figure 1. Sample stimulus used in the experiment. In the lower portion of the figure, numbers are assigned to particular triangles for subsequent reference.

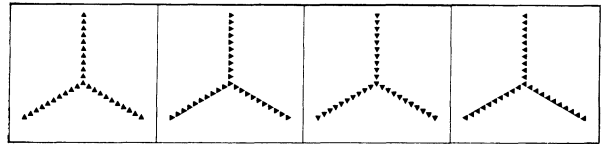


Figure 2. Basic set of four stimuli from which the 16 stimuli that were used in the experiment were generated.

6 would be longer than to triangles 1 and 4 owing to interference and facilitation between the direction of the local configuration and the required response, respectively. If visual inhomogeneity does not play any role and the AED and symmetry hypotheses are valid in their present formulations, then the response times to all seven triangles would be the same. A further prediction may be derived from consideration of visual inhomogeneity: there are more elements in the immediate surroundings of triangle 4 than there are around triangle 1, so there must be a greater facilitatory effect in the former case, which should result in a faster response to triangle 4 than to triangle 1. Similar considerations apply to triangles 2 and 5 on the one hand, and to triangles 3 and 6 on the other, except that the effects in these cases are interfering, and, therefore, the responses to triangles 5 and 6 would be slower than those to triangles 2 and 3. Triangle 7 is embedded in a locally neutral context, and response times to that triangle should be intermediate between those for triangles 1 and 4 and those for triangles 2, 3, 5, and 6.

## METHOD

These predictions were tested with a group of 17 subjects using 16 stimuli derived from the basic set shown in Figure 2. The remaining stimuli were created by rotating those in the basic set by  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$  about triangle 7. A large number of stimuli with triangles pointing in any of the required directions was used to prevent subjects from developing response strategies as the session proceeded. Regardless of these variations, only two dimensions of each stimulus—the bias and its amount—are relevant to the purposes of this experiment, and the experimental data will be analyzed with regard only to them. The *bias* between the orientation of the local configuration and the direction of the required response may be consistent (as in triangles 1 and 4) or inconsistent (as in triangles 2, 3, 5, and 6). The *amount* of bias may be large if there are biasing elements on both sides of the fixated shape (as in triangles 4 to 6) or small if there are biasing elements on only one side (as in triangles 1 to 3). Triangle 7 serves as a control condition since the local configuration in which it is embedded is neutral in all these respects.

The experimental paradigm of interference (see Palmer & Bucher, 1981) was used. The subjects received 10 practice trials before the beginning of the experimental session, which consisted of 112 trials: one presentation of every stimulus for every fixation point. The order of presentation was decided at random. In each trial, a brief tone signaled the presentation of a fixation spot on the screen at the place where the subject had to look; next, the selected stimulus was presented until the subject pressed a button to indicate that he/she had seen the triangle pointing in the appropriate direction; a noise mask was then presented on the screen, and finally the subject indicated the perceived direction of pointing by means of a four-button box. The elapsed time between presentation of the stimulus and pressing of the first button was measured with a resolution of less than 1 msec. Response times were discarded on wrong trials, which were retaken at the end of the session. Viewing was binocular with natural pupils through a translucent gray tunnel that restricted the visible portion of the screen to a circle 27 cm

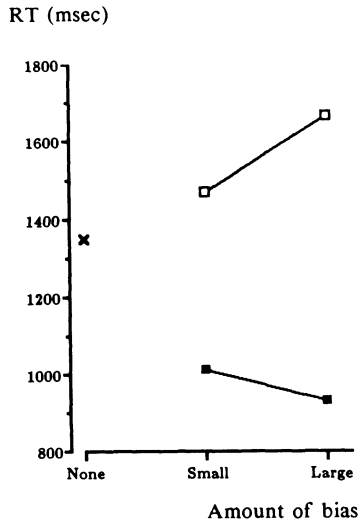


Figure 3. Experimental results. Mean response time (RT, in milliseconds) as a function of stimulus dimensions. Open squares: inconsistent bias. Filled squares: consistent bias. The control condition is represented by a cross.

in diameter. Room lighting was adjusted so that the average luminance of the tunnel matched that of the screen. A head and chin bar was used to control viewing distance, which was set so that the whole configuration subtended 10° with triangle side being approximately .5° in length.

RESULTS

Errors were not analyzed because of their low rate (<2%). Mean response times, as a function of the two relevant stimulus dimensions, are shown in Figure 3, along with the data for the control condition. An analysis of variance confirmed the obvious main effects of bias [ $F(1,1564) = 167.69, p < .00005$ ] and the interaction between bias and its amount [ $F(1,1564) = 9.05, p < .0025$ ]. That is, the response times were not the same to all triangles, and the (facilitatory or interfering) reference frame effects were greater the larger the amount of bias in the immediate neighborhood. Therefore, the predictions derived from the AED and the symmetry hypotheses are disconfirmed, and it seems clear that both hypotheses need to be reformulated to incorporate visual inhomogeneity.

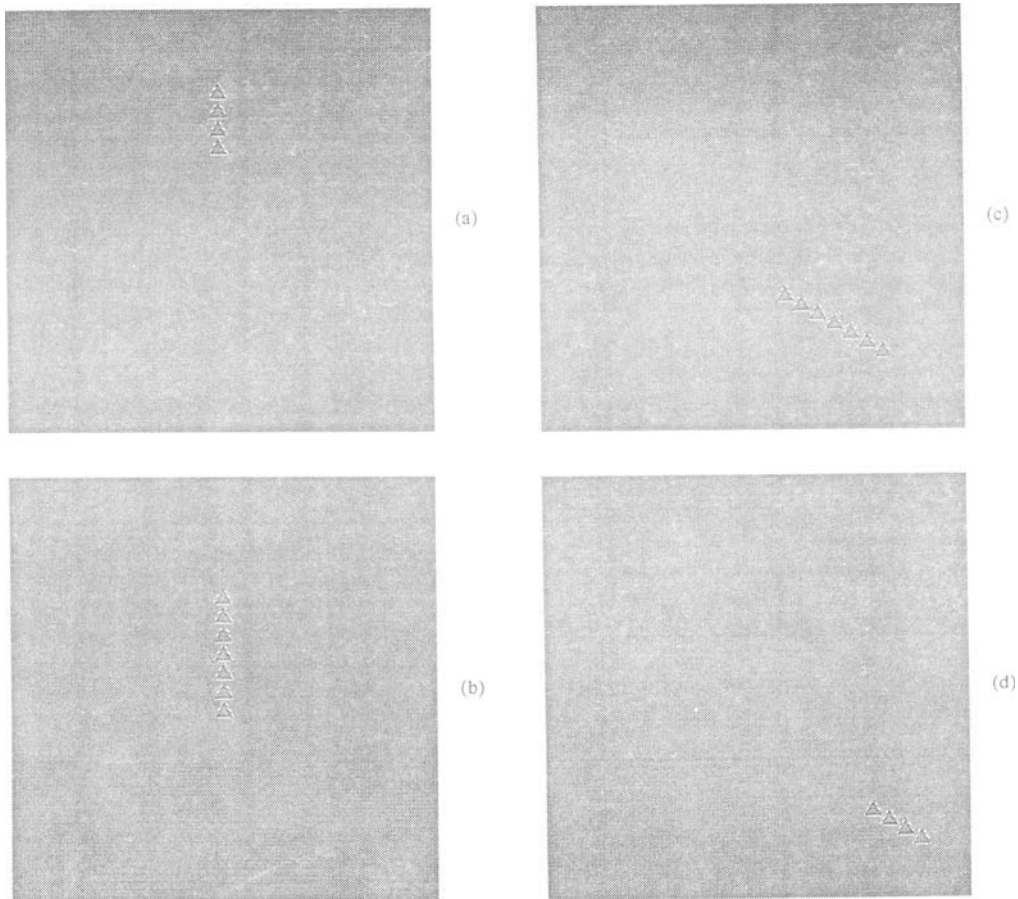


Figure 4. High-spatial-frequency information available after fixating different triangles of Figure 1 under the experimental conditions. See text for details.

## DISCUSSION

It might be said that the above results argue in favor of local reference frames and have nothing to do with visual inhomogeneity. However, it has been shown elsewhere (García-Pérez, in press) that local analyses based on high-spatial-frequency information are possible only with the intervention of visual inhomogeneity, which restricts the available high-spatial-frequency information to that present in a small area around the fixation point. Figure 4 makes the point. Parts a, b, c, and d, respectively, show the high-spatial-frequency information that is available when triangles 1, 4, 5, and 2 of Figure 1 are fixated under the experimental conditions.<sup>1</sup> Clearly, visual inhomogeneity allows local orientation analysis to be performed in the high-spatial-frequency range. The only features of the original stimulus that remain present in this range of frequencies are those around the fixation point, and this fact can be easily related to the data in Figure 3 if the parametric results of Palmer and Bucher (1981) concerning the effects of number of elements in configuration are borne in mind. One way of interpreting these results of Palmer and Bucher is that the strength of reference frame effects, especially for inconsistent configurations, increases with the number of elements that are present within a distance of approximately  $1.5^\circ$  from the fixation point.<sup>2</sup> This is approximately the radius of the circular area for visibility of suprathreshold high-spatial-frequency information found by García-Pérez (in press) and used to obtain the images in Figure 4. Therefore, these experimental data can be explained as a consequence of visual inhomogeneity if high-spatial-frequency information is principally concerned in determining the reference frame orientation.

This emphasis on high spatial frequencies is in contrast to the original proposal by Jáñez (1983), who, based on earlier work by Ginsburg (1978), assigned the main role to low spatial frequencies. Nonetheless, there are at least two observations that argue in favor of a dominant role of high spatial frequencies. First, low spatial frequencies are analyzed by our visual system all over the visual field (García-Pérez, in press), and, hence, the asymptotic increase in reference frame effects with increasing area of stimulation cannot be simply related to the energy in the low-spatial-frequency range. Second, and more important, reference frame effects also appear in the absence of low spatial frequencies (Palmer, Kube, & Kruschke, 1987).

It may well be that things are not this simple and no single spatial-frequency band can be said to determine the orientation of the reference frame. In fact, Jáñez's (1983) model includes a weighting function that describes the relative contribution of every spatial-frequency band. However, he arbitrarily gave null weights to all bands except those centered one and two octaves below the object's fundamental frequency, although the high spatial frequencies also give a good account of the data (L. Jáñez personal communication; see also García-Pérez, 1987b). Furthermore, empirical estimates of the weighting function also show that all spatial-frequency bands seem to play a role in reference frame effects (García-Pérez, 1985).

The recourse to high spatial frequencies allows us to explain some other parametric data on reference frame effects. For example, Palmer and Bucher (1982, Experiment 5) found that a ground texture consisting of a square-wave grating of 2.7 cpd biases the perceived pointing of a central triangle to an extent that depends on the proximity between the triangle and the textured background. Interestingly, when that distance reaches  $1.3^\circ$ , no significant bias appears (see Palmer & Bucher, 1982, Figure 8). As noted above, this is approximately the region at which high-spatial-frequency information is visible, and, hence, the textured background lies almost completely outside the area at which this kind of biasing factor is effective. Also, Merino (1986) used similar textured backgrounds of noise filtered in spatial frequency and orientation and of varying spatial frequencies. He found that the strength of reference frame effects on the perceived pointing of a central triangle diminishes as the center frequency of the noise band gets higher. Since the visual field area for visibility of spatial-frequency information is smaller the higher the spatial frequency (García-Pérez, in press), the amount of bias is in fact smaller the higher the spatial frequency.

The results reported in this paper do not definitely rule out the symmetry hypothesis or the AED hypothesis. The symmetry hypothesis can be maintained by adding emphasis on local symmetry. The AED hypothesis can be maintained if visual channels are regarded as spatially limited in the manner described by García-Pérez (in press).

In summary, then, these results can be interpreted as proof that visual inhomogeneity is involved in determining the reference frame orientation with the likely intervention of high-spatial-frequency information. The influence of the variations of processing across the visual field also have been shown to be relevant to perceptual phenomena such as visual illusions and multistability (García-Pérez, 1987a), and these results will be reported in separate papers.

## REFERENCES

- BURT, P. J. (1981). Fast filter transforms for image processing. *Computer Graphics & Image Processing*, **16**, 20-51.
- GARCÍA-PÉREZ, M. A. (1985). Reference frame effects, perceptual axes and spatial-frequency channels in vision. *Informes de Psicología*, **4**, 91-107.
- GARCÍA-PÉREZ, M. A. (1987a). *Procesamiento visual anisoplanático*. Unpublished doctoral dissertation, Facultad de Psicología, Universidad Complutense, Madrid.
- GARCÍA-PÉREZ, M. A. (1987b, May). *Visual inhomogeneity and reference frame effects*. Poster presented at the Conferencia sobre Percepción del Objeto: Estructura y Procesos, Madrid.
- GARCÍA-PÉREZ, M. A. (in press). *Space-variant visual processing: Spatially limited visual channels*. *Spacial Vision*.
- GINSBURG, A. P. (1978). *Visual information processing based on spatial filters constrained by biological data*. Unpublished doctoral dissertation, Cambridge University.
- JÁÑEZ, L. (1983). *Stimulus control of visual reference frame orientation: Quantitative theory*. *Informes de Psicología*, **2**, 133-147.
- JÁÑEZ, L. (1986). *Stimulus control of perceptual axes orientation: Quantitative theory*. Unpublished manuscript.
- MARR, D., & NISHIHARA, H. K. (1977). *Representation and recognition of the spatial organization of three dimensional shapes* (MIT-AI Lab Memo 416). Cambridge, MA: Massachusetts Institute of Technology, Artificial Intelligence Laboratory.
- MERINO, J. M. (1986). *Efectos del marco de referencia visual: Comparación experimental de las teorías basadas en la simetría y en la distribución angular de energía*. Unpublished doctoral dissertation, Facultad de Psicología, Universidad Complutense, Madrid.
- PALMER, S. E. (1985). The role of symmetry in shape perception. *Acta Psychologica*, **59**, 67-90.
- PALMER, S. E. (1987, May). *Reference frames in perception of shape and orientation*. Paper given at the Conferencia sobre Percepción del Objeto: Estructura y Procesos, Madrid.
- PALMER, S. E., & BUCHER, N. M. (1981). Configural effects in perceived pointing of ambiguous triangles. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 88-114.
- PALMER, S. E., & BUCHER, N. M. (1982). Textural effects in perceived pointing of ambiguous triangles. *Journal of Experimental Psychology: Human Perception & Performance*, **8**, 693-708.
- PALMER, S. E., KUBE, P., & KRUSCHKE, J. (1987). *A test between two theories of a Gestalt effect: Symmetry versus low spatial frequency channels*. Manuscript in preparation.

## NOTES

1. These images have been obtained as follows. First, the original image of Figure 1 was generated in a  $1,000 \times 1,000$  pixel format, which was subsequently regarded as subtending  $10^\circ$  in each direction so that 100 pixels represent  $1^\circ$  of visual angle and, therefore, the highest spatial frequency in the image is 50 cpd. Next, the image was high-pass filtered using Burt's (1981) hierarchical discrete correlation. Each image in Figure 4 was obtained by multiplying that filtered image by a circular window with blurred edges, centered at the fixation point (different for each image) and with the radius experimentally found by García-Pérez (1988) as defining the spatial extent of the highest spatial-frequency visual channel in suprathreshold vision.

2. This interpretation is based on the facts that (1) the asymptotic response time was reached when five triangles were aligned and (2) each individual triangle subtended about  $.5^\circ$  of visual angle (see Palmer & Bucher, 1981, Figure 4).