this is just one aspect of temporal vision; taking time seriously appears to be an effort of the magnitude of the four frames model.

Even without new experiments, there is a great deal that might be learned from trying to fit the four frames model to existing bodies of data. Doing this at a crude level has forged the current form of the model. Subsequent efforts are of two different kinds: detailed fitting of small segments of data, and further refinement of the global model. Detailed studies are underway at Rochester on parameter networks in extrastriate cortex and on computational models of specific feature frame and world knowledge formulary computations. These studies, plus commentaries on the present article, will, I hope, lead to an improved and elaborated second version of the four frames model. Commentary on the scientific adequacy of the model and suggestions for alternative formulations would be particularly welcome. At the least, I would hope to direct more attention to the global properties of the visual system, which is often treated as a large number of totally separate problems. The rationale of the whole enterprise is that it is not too early to benefit from more general considerations of the problems of vision and space.

ACKNOWLEDGMENTS

A number of people have made valuable comments on earlier written and oral presentations of the model. Particularly useful were the suggestions of Joanne Albano, Michael Arbib, Dana Ballard, Paul Coleman, Francis Crick, Lydia Hrechanyk, Walter Makous, David Zipser, and the *BBS* reviewers.

The preparation of this paper was supported in part by the Defense Advanced Research Projects Agency Grant No. N00014-82-K-0193 and in part by Defense Advanced Research Projects Agency Grant No. N00014-78-C-0164.

Open Peer Commentary

Commentaries submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Head-centered coordinates and the stable feature frame

Richard A. Andersen

Salk Institute, San Diego, Calif. 92138

Jerome Feldman's ambitious model of higher aspects of visual function brings together several current trends in artificial intelligence and computer vision and is structured in a manner not unlike the organization of visual cortex. I wish to comment on the aspect of the model that addresses the transformation of visual space from the retinal representation of his retinotopic frame to the head-centered coordinate system of his stable feature frame. He proposes the retinotopic frame to be located in striate cortex and the various components of the feature frame to be in extrastriate cortical areas. The cells in the extrastriate fields in this model are spatially invariant, responding to a feature in a specific location in head-centered coordinates regardless of eye position. The mapping from retinal to spatial coordinate frames is accomplished using eye position (gaze) information to direct conjunctive connections between the two representations.

The coordinate frame used by visual cortical fields is best determined by recording the activity of neurons in behaving monkeys, because the eye position aspect of spatial mapping probably requires active gaze by the animal. Little work on the issue of coordinate frames has been done in extrastriate areas with behaving monkeys; however, the retinotopic representations mapped in anesthetized animals and the retinotopic topography of the cortico-cortical connections between many of these areas and between V1 and its extrastriate targets make it unlikely that the responses of cells in these cortical fields deviate much from a retinotopic order (Allman, Baker, Newsome & Petersen 1981; Maunsell & Van Essen 1983; Van Essen 1979). Also, eye position or eye movement signals have not been recorded in many of these areas, nor is it obvious from the anatomy of their inputs how they might receive such signals.

The posterior parietal cortex has long been suspected of playing a pivotal role in spatial perception; lesions in this area in monkeys produce profound defects in localizing ability without apparent defects to other aspects of vision (LaMotte & Acuna 1978). Visual subregions within this area, most notably area 7a, receive inputs from many other extrastriate visual areas (Andersen, Asanuma & Cowan 1982). Many of the visually responsive neurons in this area also carry nonvisual eye position signals (Essick, Andersen & Siegel 1984; Lynch, Mountcastle, Talbot & Yin 1977; Sakata, Shibutani & Kawano 1980). [See also Lynch: "The Functional Organization of Posterior Parietal Cortex" *BBS* 3(4) 1980.] These signals appear to be derived from oculomotor structures in the brainstem that project by way of certain intralaminar thalamic nuclei to 7a (Asanuma, Andersen & Cowan 1982; Schlag-Rey & Schlag 1984). Other 7a neurons are only light sensitive, and still others appear to be light sensitive but change their activity as a function of gaze angle (Andersen & Mountcastle 1983). These latter cells are likely candidates for involvement in mapping from retinal to spatial coordinates. However, their visual receptive fields remain retinotopic and it is only the *magnitude* of their response to identical retinotopic stimuli that varies with eye position (Andersen, Essick & Siegel 1984). Moreover, the response of these neurons to flashes of light can generally be described by the equation $A = G(ex,ey)* R(rx,ry)$, where A is the activity of the neuron, G is a gain factor that is a function of the eye position (ex, ey), and R is the two-dimensional retinal receptive field of the neuron.

These observations suggest three important points pertinent to the target article. First, since area 7a of the posterior parietal cortex receives its visual inputs predominantly from extrastriate visual areas and no spatially invariant responses have been identified in behaving monkeys in area 7a, it is unlikely that the several extrastriate areas projecting to this area are encoding space in other than a retinotopic frame. Second, the change in gain of the visual response of 7a neurons as a function of eye position does tune neurons so that they are most responsive to a particular location in head-centered space, but this tuning is dependent on gaze angle. Thus any spatial information independent of eye position must exist only in populations of neurons in this area or at a position of convergence at an even higher cortical level than the posterior parietal cortex. Third, since it is only the magnitude of the response of cells that varies with gaze angle and not the retinal address of the receptive fields, a neural implementation other than the conjunctive connections approach outlined in the target article must be used for the particular spatial transformation that is presumed to occur in area 7a.

Thus, to fit the model to current knowledge of cortical

function requires that the characteristics of the feature frame be processed within a retinotopic coordinate system by extrastriate visual areas. Conversions to representations that are nonretinal probably begin only at the level of the posterior parietal cortex, and even here head-centered locations independent of eye position can only be represented in the response of populations of neurons. In Feldman's model the environmental frame is proposed to be located in the posterior parietal cortex. It is not unreasonable to assume that the posterior parietal cortex could play a role in the encoding of features in head-centered spatial coordinates and could perform the ego-centered functions assigned to the environmental frame.

Could three frames suffice?

Roger A. Browse[a,b] and Brian E. Butler[b]

[a]Department of Computing and Information Science and [b]Department of Psychology, Queen's University, Kingston, Ontario, Canada K7L 3N6

Feldman's treatment of visual perception involves a competence theory of the representations and interactions that permit visual recognition, and a performance theory of the implementation of visual capabilities using a form of parallel computation. This treatment provides an excellent language to discuss the possible organization of visual perception. In this commentary we concentrate on the competence aspect of Feldman's model and find some inconsistencies with experimental data. Some relatively simple alterations to the model can lessen or eliminate these problems.

Criticisms of the model. Feldman's model is unclear about the extent to which indexing takes place on the basis of feature information obtained peripherally. He suggests at some points that indexing can take place only for foveated objects, but at other places he seems to say that a crude form of indexing is possible in nonfoveated areas. Behavioral studies show that peripheral object recognition does occur. Anstis (1974) demonstrated that the ability to identify letters may be held constant across the visual field by increasing letter size with eccentricity. Peripheral indexing may differ from foveal indexing only in its requirement for information.

Any complete theory of visual perception must specify methods for mediation among the different levels of resolution of information available across the retina. Feldman's model accomplishes this mediation through the formation of concordances among feature values in the stable feature frame. But if indexing is assumed for all resolutions (peripherally and foveally) then sophisticated interactions will be necessary at the level of ongoing interpretations to deal with phenomena such as interference between local and global identities (Navon 1977).

Feldman recognizes that it is unreasonable to encode all possible feature vectors for all locations in the stable feature frame. His alternative is to maintain spatial coherence for pairs of property values and to use these pairs for indexing. A visual element's requirement for a particular pair of feature values is encoded through connection to a unit that represents the appearance of that pair anywhere in the stable feature frame (Figure 12). Feldman claims a relation between the cross talk predicted from this configuration and the illusory conjunctions of features reported by Treisman (1982). Actually, indexing from feature pairs predicts that single features can be misplaced only between objects that share other feature values (as in Figure 12). Experimental evidence does not bear out this prediction, and it has been shown that sharing features does not influence the occurrence of illusory conjunctions (Treisman & Schmidt 1982: experiment V). The idea of indexing from feature pairs also directly contradicts a more basic finding. If features were paired before indexing, as Feldman suggests, then identification requiring a conjunction of features (say a blue circle) could be accomplished in parallel in a background of distractors, each of which has only one feature in common with the target (say blue squares and red circles). Experimentation has shown performance in such tasks to be more consistent with a serial search of the items (Treisman & Gelade 1980).

Finally, Feldman's model is unclear in the positioning of the stable feature frame's coordinate system relative to the retinal frame. It appears that initially (on the "first fixation") the centers of these coordinate systems coincide. Then, with a saccade to the upper left (Figure 10), the retinal frame is positioned on a corner of the stable feature frame. The question is: What will happen if the next fixation is also to the upper left. It appears that the stable feature frame will have to be recomputed at that point. This would mean that for two successive fixation shifts of exactly the same magnitude and direction, different operations of coordinate mapping would be required.

Alterations to the model. Some relatively simple modifications that are consistent with the purpose and intent of Feldman's model can solve the problems noted here. Our first modification is to divide the task of world knowledge formulary indexing. We assume a low resolution indexing that is available in parallel at all locations of an extended retinal frame, including the center of fixation. This indexing leads only to generalized concepts in the world knowledge formulary, such as "automobile" or the even more general "vehicle." The quality of features available for this indexing will vary across the retinal frame, but for most locations there will be very few values for each feature dimension. For example, in the far periphery, orientation might have only four values. With this restricted indexing based on reduced feature value distinctions, it may be possible to formulate all possible indexing feature vectors for each location.

It is assumed that grouping operations within retinotopic feature arrays provide the units for which low resolution indexing may proceed. The generalized elements of the world knowledge formulary may be accessed on the basis of simple rotation and scale invariant encodings for alternative perspectives, as in Feldman's model. Depending on circumstance and task requirements, these objects identified throught low resolution indexing may be interpreted at a satisfactory level of detail. For example, in crossing the street one may wish to confirm only the existence of automobiles, not the particular models and years. However, some of the these objects may not be at an adequate level of specificity, so the system must take further action. Thus we propose a more powerful indexing capability, which may be applied sequentially to one location in the retinal frame at a time. Through the application of this central indexing, very specific object instances may be identified, and so indicated in the world knowledge formulary (Butler 1978). This capability could be applied at the fovea, at parafoveal locations, or, though not normally, at quite distant peripheral locations. The locations to be considered for this sequential indexing will be based on the groupings within single feature maps, but coincidence of groups across maps may also be determined. Any location that is indexed in low resolution may be a candidate for more specific interpretation. If it is determined that the quality of feature information is adequate to support specialized object indexing, then the unique indexing capability will be shifted. The shifting of this form of spatial attention can be accomplished using a mechanism similar to the operations that update gaze location in Feldman's model; but it requires the equivalent of the mapping to a single location in the stable feature frame.

The mapping of features to objects (generalized or specific) can be accomplished using set intersections in an inverted data structure for which each feature value indicates sets of all objects that specify its presence (Browse 1982; Fahlman 1979). As such, a direct counterpart exists in localist connectionism.

If the shifting of this indexing capability should fail to provide the required level of specificity, it may be that the detail of the features is not adequate to support specialized indexing, and a change in fixation may be required. Under normal circum-