BBS COMENTENTARY ON "REPRESENTATION IS REPRESENTATION OF SIMILARITIES" by
Shimon Edelman

REPRESENTATION IS SPACE-VARIANT

Giorgio Bonmassar Eric L. Schwartz
Dept. of Biomedical Engineering Dept. of Cognitive and Neural Sysems
Boston University

ABSTRACT: *Under shift, caused for example by eye movement, or by relative movement of the subject or object of perception, the cortical representation undergoes very large changes in "size" and "shape". Space-variance of cortical representation rules out models which fundamentally require linear interpolation between shifted patterns (e.g. Shimon Edleman's model), or rigid shift of an invariant retinal stimulus corresponding to shift at the cortex (e.g. the shifter theory of VanEssen). Recently, a computational solution to "quasi-shift" invariance for space-variant mappings has been constructed [Bonmassar and Schwartz, 1997a, Bonmassar and Schwartz, 1997b].*

Shimon Edelman's (SE,from here on) work addresses an important gap in the computational discussion of neural representation, which to date has largely been carried out on a verbal level. His position is that representation is a record of similarities to stored prototypes, rather than direct representation in the form of templates, or feature vectors. Rather than learn all possible prototypes (similarities), a "small" number are stored, with interpolation of new stimuli providing generalization. SE uses a particular form of cluster analysis (multi-dimensional scaling) to effect classification. No neurally plausible means of implementing multi-dimensional scaling in the brain is provided, and no comparison to other similar forms of clustering, or indeed, of statistical pattern recognition in general, is supplied. It seems to us to there is a basic mathematical equivalence between clustering based on "similarities" and clustering based on direct feature vector representation. We will instead focus on the issue of linear interpolation of learned prototypes, which we identify as the key contribution of this model.

Representation in the brain is expressed, we believe, in a wide variety of cortical loci. The majority of cortical visual area's are topographically organized, and we assert that spatial representation in the brain (in the form of topography and columnar spatial patterns) are themselves a form of representation, and one which obviously does not depend on "similarities" between prototypes, but which is an example of direct, template based representation.

The spatial structure of visual stimuli is represented in V-1 as a topographic map, whose fidelity is sufficiently detailed to account for visual acuity. This map is approximated by the complex logarithmic (log-polar) map[Schwartz, 1977]. No other analytic form for V-1 topography has yet been presented, and the approximation of the two dimensional topographic map by the simple one parameter fit in the form of (complex) $\log(z + a)$, with a representing the extent of "foveal" representation (a is roughly 0.5 degrees), is considered to be a "good" approximation by most workers in the field (e.g. see [van Essen et al., 1984, Tootell et al., 1985, Dow et al., 1985]. Recently, a more general conformal map has been numerically generated from 2DG data obtained from primate V-1, and the error bounds for this fit are in the range of 15-20%[Schwartz, 1994]. Although this numerical conformal map has no simple analytic representation, it is similar in its properties to the complex log, and we will use the complex log as a convenient way of modeling the spatial properties of early primate visual representation.

The non-linear spatial structure of V-1 representation poses an unavoidable problem to the basis of SE's model: simple linear "interpolation" between shifted versions of a prototype fails because the human visual representation is strongly space-variant, and both the size and "shape" of the V-1 representation of a stimulus undergoes very large changes. This is demonstrated in figure 1, which shows the behavior of letter's under shifts, or, equivalently, of eye movement. The cortical representation of these shapes is strongly distorted under shift (i.e. eye movement). It is evident that interpolation of the same letter, but with different eye positions, could not possibly work. SE's model would require storage of a large number of eye-position "prototypes", multiplying the combinatorial explosion already present due to the other geometric symmetries. Of course, one can invoke (as does SE) the "deus ex machina" of IT cortex here to somehow unravel this problem, but we know very little about any aspect of trigger feature representation in IT at the present time. There is some evidence that IT trigger features are invariant to size, translation and rotation [Schwartz et al., 1983], but IT receives its representation ultimately from V-1, and therefore inherits the space-variant nature of V-1 representation.

In our lab, we have considerable experience building machine vision systems based on complex logarithmic image representation (reviewed in [Schwartz et al., 1995]). We can confidently state from experience that linear interpolation of view, as used by SE, grossly fails to allow a system built on space-variant design principles (e.g. the human brain) to function. The reader can verify this directly from figure 1.

Recently, we have developed a computational solution to this problem, by devising a new form of Fourier Transform (the exponential chirp transform) which provides quasi-shift invariance, as well as size and rotation invariance that are consistent with the difficulties imposed by V-1 representation [Bonmassar and Schwartz, 1997a, Bonmassar and Schwartz, 1997b]. SE has published a psychophysical study in which perfect translation invariance in human vision is called into question. We can explain this quite simply in terms of "quasi-invariance", which is defined precisely in our papers on the exponential chirp cited above. This discussion indicates a fundamental terminological and conceptual problem in the perceptual literature. Geometric invariance is not possible in a space-variant system. Similarly, there is no (and can be no) "veridicial" representation in the brain, since V-1 discards more than 99.99% of the information available at the level of retinal (optical) image [Rojer and Schwartz, 1990].

The problems associated with space-variance in human vision provide a fatal problem for models which are based on simple linear interpolation (SE's model) or simple "linear shift" to account for the problem of eye movement (the Olshausen-Anderson-VanEssen "shifter theory" referenced by SE as a solution to the problems introduced by eye movement). Linear shift, or linear interpolation, cannot be invoked as a modeling tool in the primate visual system because of the strongly non-linear nature of V-1, and later cortical representation. Linear shift of a cortical pattern does not correspond, in an isomorphic sense, to linear shift of a retinal pattern! Models which require this feature (e.g. the shifter theory , the linear interpolation aspects of SE's model) cannot be correct.

It is a constant source of surprise to us that models which purport to explain biological vision ignore the most basic spatial structure of the visual system. However, it is always useful to be able to falsify models, particularly in fields such as this, in which most models are "not even wrong". For the present time, we can assert with strong confidence, that models which depend fundamentally on the ability to linearly shift or linearly interpolate cortical representations of visual stimuli, and which therefore ignore the space-variant structure of the primate visual system, are, to paraphrase W. Pauli, "even wrong".

References

[Bonmassar and Schwartz, 1997a] Bonmassar, G. and Schwartz, E. (1997a). Lie groups, space-variant fourier analysis and the exponential chirp transform. In *Computer Vision and Pattern Recognition 96*, volume 3 of *CVPR*, pages 229–237.

[Bonmassar and Schwartz, 1997b] Bonmassar, G. and Schwartz, E. L. (1997b). Space-variant fourier analysis: the exponential chirp transform. *IEEE Pattern Analysis and Machine Vision vol. 19 1080-1089*. October 1997.

[Dow et al., 1985] Dow, B., Vautin, R. G., and Bauer, R. (1985). The mapping of visual space onto foveal striate cortex in the macaque monkey. *J. Neuroscience*, 5:890–902.

[Rojer and Schwartz, 1990] Rojer, A. S. and Schwartz, E. L. (1990). Design considerations for a space-variant visual sensor with complex-logarithmic geometry. *10th International Conference on Pattern Recognition, Vol. 2*, pages 278–285.

[Schwartz et al., 1995] Schwartz, E., Greve, D., and Bonmassar, G. (1995). Space-variant active vision: Definition, overview and examples. *Neural Networks*, 8(7/8):1297–1308.

[Schwartz, 1977] Schwartz, E. L. (1977). Spatial mapping in primate sensory projection: analytic structure and relevance to perception. *Biological Cybernetics*, 25:181–194.

[Schwartz, 1994] Schwartz, E. L. (1994). Computational studies of the spatial architecture of primate visual cortex:columns, maps, and protomaps. In Peters, A. and Rocklund, K., editors, *Primary Visual Cortex in Primates*, volume 10 of *Cerebral Cortex*. Plenum Press.

[Schwartz et al., 1983] Schwartz, E. L., Desimone, R., Albright, T., and Gross, C. G. (1983). Shape recognition and inferior temporal neurons. *Proceedings of the National Academy of Sciences*, 80:5776–5778.

[Tootell et al., 1985] Tootell, R. B., Silverman, M. S., Switkes, E., and deValois, R. (1985). Deoxyglucose, retinotopic mapping and the complex log model in striate cortex. *Science*, 227:1066.

[van Essen et al., 1984] van Essen, D. C., Newsome, W. T., and Maunsell, J. H. R. (1984). The visual representation in striate cortex of the macaque monkey: Asymetries, anisotropies, and individual variability. *Vision Research*, 24:429–448.

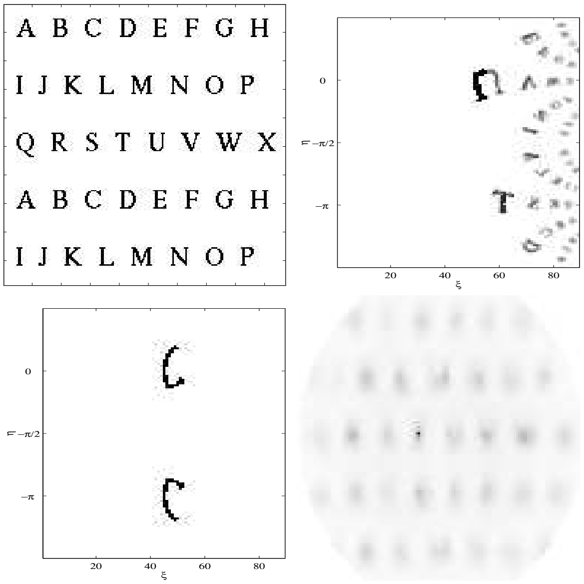

Figure 1: The result of applying the space-variant cross-correlation to an (197 × 194) image of letters with an image of the letter "T" at the fixation point. On the top left is the original image of letters, on the top right is its space-variant representation, on the bottom left is the log-polar (space-variant) image of the letter "T" (split by the "vertical meridian into a "left" and "right" hemisphere segment. These two last space-variant images are used by the ECT (Exponential Chirp Transform) algorithm to compute the space-variant cross-correlation, as shown in the bottom right of the figure. Clearly visible is the sharp peak located in the position of the letter "T" in the original image (cortical) space.