

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL INFORMATION PROCESSING
WHITAKER COLLEGE

A.I. Memo No. 1336
C.B.I.P. Paper No. 68

December 1991

Fast Perceptual Learning in Visual Hyperacuity

Tomaso Poggio, Manfred Fahle and Shimon Edelman

Abstract

In many different spatial discrimination tasks, such as in determining the sign of the offset in a vernier stimulus, the human visual system exhibits hyperacuity-level performance by evaluating spatial relations with the precision of a fraction of a photoreceptor's diameter. We propose that this impressive performance depends in part on a fast learning process that uses relatively few examples and occurs at an early processing stage in the visual pathway. We show that this hypothesis is plausible by demonstrating that it is possible to synthesize, from a small number of examples of a given task, a simple (HyperBF) network that attains the required performance level. We then verify with psychophysical experiments some of the key predictions of our conjecture. In particular, we show that fast stimulus-specific learning indeed takes place in the human visual system and that this learning does not transfer between two slightly different hyperacuity tasks.

© Massachusetts Institute of Technology, 1991

This paper describes research done at the Center for Biological Information Processing, the Department of Brain & Cognitive Sciences, and the Artificial Intelligence Laboratory. This research is sponsored by grants from the Office of Naval Research Cognitive and Neural Sciences Division, the Artificial Intelligence Center of Hughes Aircraft Corporation, the Alfred P. Sloan Foundation, and the National Science Foundation. Support for the A. I. Laboratory's artificial intelligence research is provided by the Advanced Research Projects Agency under Army contract DACA76-85-C-0010, and by ONR contract N00014-85-K-0124. M. Fahle is at the Dept. of Neuroophthalmology, University Eye Clinic, D7400 Tübingen, Germany. S. Edelman is at the Dept. of Applied Mathematics & Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel.

For any given visual competence, it is tempting to conjecture a specific algorithm and a corresponding neural circuitry. It has been often implicitly assumed that this machinery may be hardwired in the brain. This extreme point of view, if taken seriously, may quickly lead to absurd consequences. Consider for instance the many different hyperacuity tasks,¹ some of which are outlined in Figure 1. Computational analysis reveals that the photoreceptor spacing and the low-pass characteristics of the eye’s optics satisfy (in the fovea) the constraints of the sampling theorem.² Thus, the underlying reason for the spectacular performance of human subjects in the hyperacuity tasks is that the signal sampled by the photoreceptors and relayed to the brain contains the information necessary for precise localization of image features. This observation, however, does not constitute an explanation of hyperacuity, since each of a variety of hyperacuity tasks is different and, in principle, would require a different circuit for its solution. Note that the idea of a fine-grid reconstruction of the image in some layer of the cortex² does not address the problem, because it still requires a homunculus looking at the reconstructed image and applying a different routine or circuitry for each specific hyperacuity task.

We propose instead³ that the brain may be able to synthesize – possibly in the cortex – appropriate task-specific modules that receive input from retinotopic cells and learn to solve the task, after a short training phase in which they are exposed to examples of the task. To show the plausibility of our argument, we first describe a model that learns to solve vernier acuity tasks from a few examples. Synthesizing a module from examples for a specific computational task may be often regarded as approximating a multivariate function from sparse data. We have chosen to use for function approximation the HyperBF network technique.⁴ Other schemes, such as the popular Multilayer Perceptrons or more traditional classification techniques,⁵ could probably be used as well. In our model we take the extreme view that the inputs are photoreceptor activities, to demonstrate the plausibility of low-level, or “early” learning. Biologically, it may be more reasonable to assume that the input to the learning stage is provided by the circular center-surround and oriented cells in V1.⁶

In the simulated experiments, the learning module was given an array of “photoreceptor” cell activities that corresponded to the input image blurred by the eye’s optics. There were eight “receptors”, positioned randomly on a loose 4×2 grid (see Figure 2). Each of the inputs was calculated by integrating the image over the point spread function of the optics, approximated by a Gaussian of spatial extent $\sigma = 30''$ and time extent $\sigma_t = 0.5$ units. The simulated “retinal” patch had spatial dimensions of $180'' \times 360''$ and a time dimension of 3 units. The 8-component vector of receptor outputs constituted the input to the HyperBF module, which was trained to produce an output of $+1$ for one sense of the input vernier displacement, and -1 for the other over a set of examples of verniers randomly placed relative to the photoreceptor array.⁷ The performance of the module was estimated by measuring the error, defined as the absolute value of the difference between the actual output and the desired output,

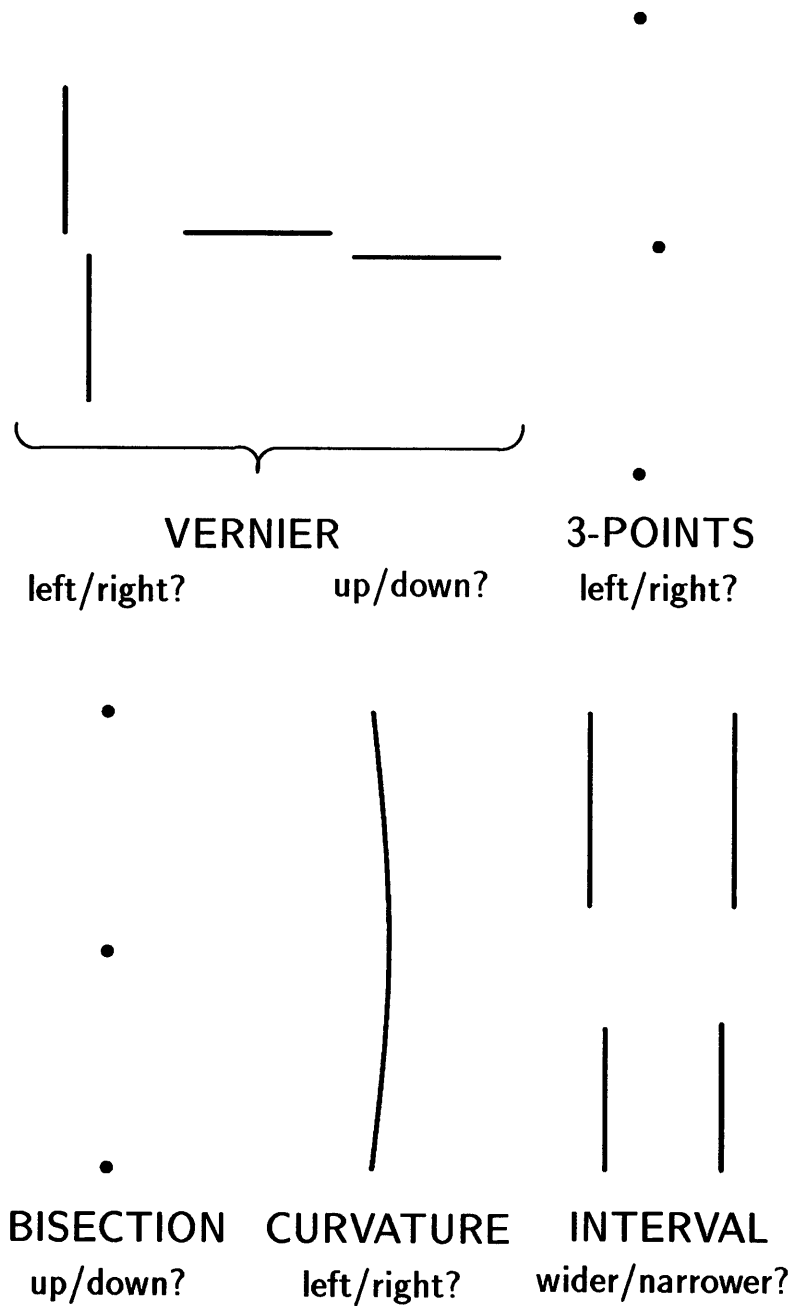


Figure 1: Examples of five tasks in which human subjects perform at hyperacuity levels (that is, exhibit resolution finer than the spacing between individual photoreceptors).

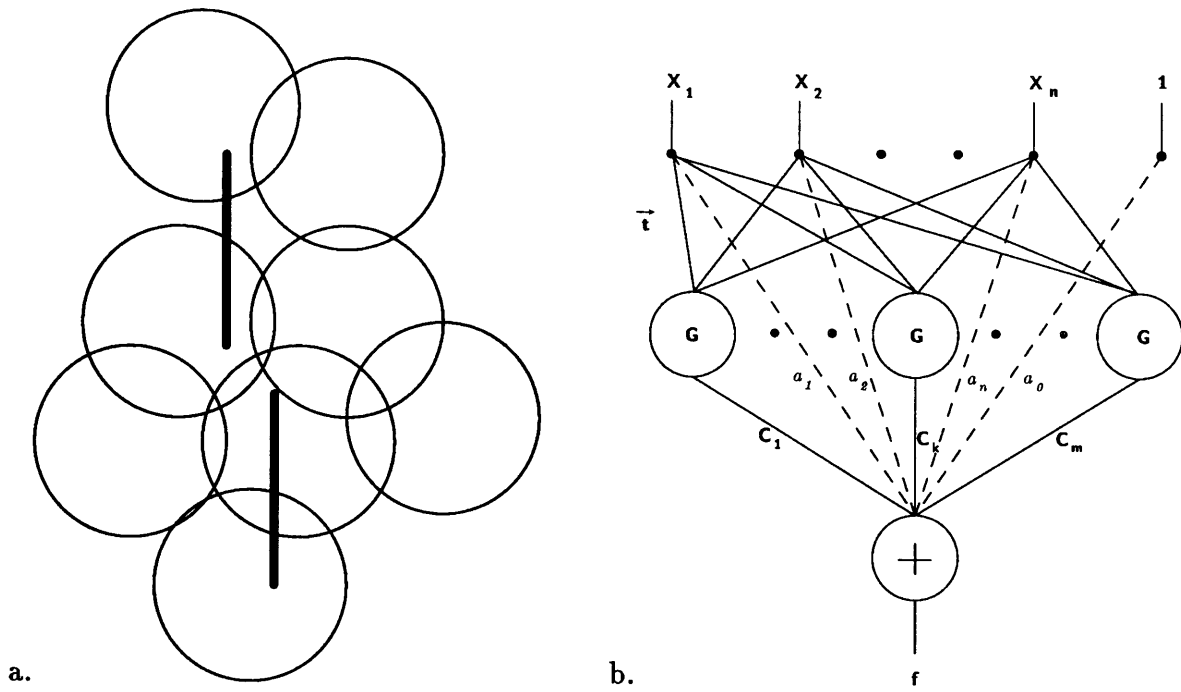


Figure 2: Fig. 2a shows, superimposed on the vernier stimulus, the mosaic of receptive fields of “cells” assumed to provide the input to the HyperBF module shown in (b). Each receptive field is depicted as a circle that refers to the point spread function of the optics. Our simulation is robust with respect to positioning the “cells” at precisely defined locations and to their receptive field properties. The network in (b) is equivalent to equation 1.⁴

which is a good analog of acuity threshold.⁶ Another measure of performance that we have considered is the percentage of correct responses (that is, responses in which the sign of the module’s output agreed with the sign of the vernier displacement, as defined during training).

The HyperBF module learned to solve the vernier task at a hyperacuity level from a few examples.⁸ The time course of the learning, illustrated in Figure 3a, shows that the output classification error rate came within 10% of its asymptotic value after just five examples.⁹ All in all, the model replicated⁶ several findings in the psychophysics of spatial acuity: (a) hyperacuity-level performance, (b) improvement in the threshold with increasing length of the two segments comprising the vernier stimulus;¹ (c) deterioration of performance with increasing orientation difference between training and testing trials;¹⁰ (d) high performance for moving verniers;¹ and (e) performance at a similar level for another hyperacuity task, the three-point bisection, after learning from suitable examples.¹¹

The model’s success demonstrates the plausibility of the hypothesis that learning of hyperacuity tasks takes place early in the visual pathway. A more critical test is provided by the predictions that learning of a hyperacuity task should be fast (see Figure 3a) and may not transfer even to a slightly different hyperacuity task (Figure 3b shows that the HyperBF model indeed exhibits no transfer of learning between vertical and horizontal verniers).⁶ We set out to verify experimentally these predictions for human hyperacuity performance. The results of the psychophysical experiments have borne out the predictions of the model. First, the vernier threshold and the error rate in naive subjects improved quickly over a few tens of trials (Figure 4a). Second, the subjects exhibited no transfer of learning between the vertical vernier and the horizontal vernier tasks or vice versa (Figure 4b). In additional experiments (not shown), there was no significant interocular transfer of learning, and little transfer from a position 10° up in the visual field to a position of similar eccentricity down in the visual field (or vice versa).¹²

Our findings pertaining to fast stimulus-specific learning can be viewed in a wider perspective that encompasses the issue of perceptual learning in general. A prominent example is provided by the work of Fiorentini and Berardi¹³ who demonstrated stimulus-specific learning effects in the discrimination of mixed spatial frequency gratings that suggested the involvement of an early-stage mechanism. Similar to our case, they found that learning did not transfer between different orientations of the grating. They also found that there was interocular transfer of learning but little transfer across retinal locations. Karni and Sagi¹⁴ recently described a texture discrimination task in which the subjects showed stimulus-specific learning effect that did show interocular transfer but did not transfer either across orientations or across positions. Other similar instances of specific perceptual learning had been reported even earlier.¹⁵ Plasticity early in the visual pathway has been demonstrated experimentally¹⁶ and could provide the adaptive mechanisms required by a module of the HyperBF type.

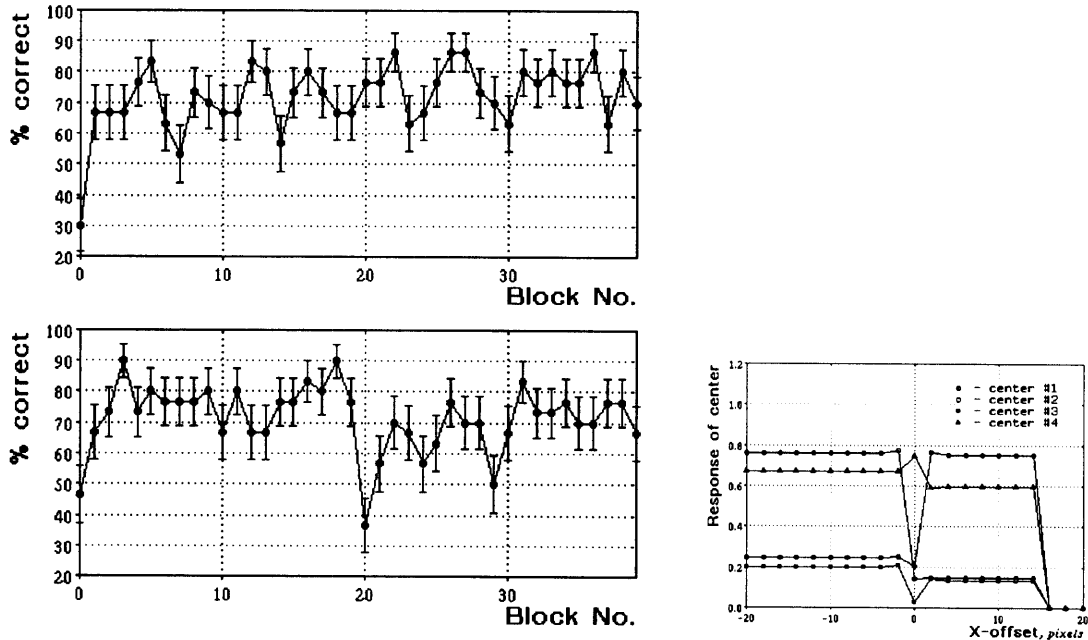


Figure 3: (a) shows the time course of learning by a HyperBF module given the input shown in Fig. 2a (vertical verniers appearing at random positions, with random offsets in a certain range). Each block in this simulation consists of just one trial; the ordinate shows percentage of correct responses in each block (mean and ± 1 standard error over 30 simulation runs). (b) shows the effect of changing stimulus orientation from vertical to horizontal at block 20: there is no transfer of learning, as expected, since the examples used by the network correspond to very different patterns of activation of the photoreceptors in the two cases. Feedback was provided in these simulations (but is not strictly required by the learning algorithm). (c) shows responses of the four HyperBF centers (acquired during an incremental learning session that consisted of 150 trials) vs. the offset of a vertical vernier presented at a fixed location. During learning, the offsets were uniformly distributed between 4 and 12 pixels. The response was tested with vertical verniers shown at the same location and having an offset ranging from -20 to 20 pixels. This illustration may be regarded as a recording of the receptive fields of the centers in the space of possible inputs. Of the four centers, one responded strongly to positive offsets and weakly to negative ones, another one preferred negative offsets, and the other two had no clear preference for any offset sign. An appropriate response representing the sign of the offset may be formed at the output level of the HyperBF module, using the responses of the sign-selective centers.

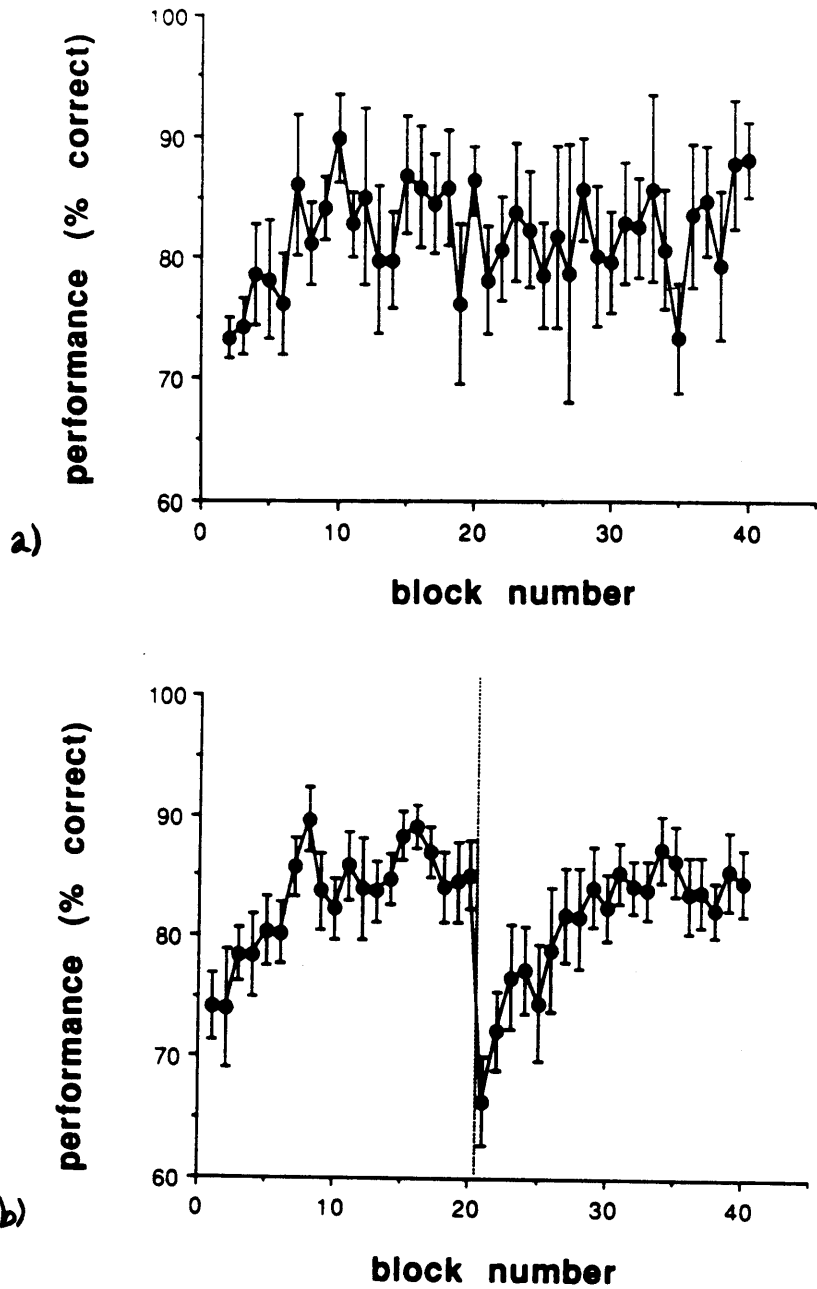


Figure 4: Psychophysical experiments corresponding to the simulations of Fig. 3: (a) shows the time course of learning in a vernier task. A fast initial component is clear. Data are means of six subjects; vertical bars represent standard errors. Each block consisted of 40 trials. We found similar learning effects with verniers consisting of three dots rather than of two lines. (b) shows the effect of switching from vertical to horizontal verniers (or vice versa) after block 20. Averaged results of 12 subjects; six started with horizontal verniers, the others started with vertical verniers. There is no transfer of learning. In these experiments feedback was provided to the subjects.

Our computational and psychophysical results support the conjecture that the modules responsible for hyperacuity-level performance are synthesized early in the visual pathway in a demand-driven fashion, when the appropriate task is first performed by the subject. Related evidence regarding perceptual learning mentioned above suggests that the same line of reasoning can be applied to visual tasks other than hyperacuity, and even to faculties other than vision.^{3,17} Importantly, learning HyperBF interpolation can be implemented in a simple biologically plausible network.^{3,4} The proposal that much of the information processing in the brain is performed by mechanisms related to the HyperBF modules acting as *enhanced look-up tables* may bridge apparently conflicting paradigms, such as Gibson's *immediate perception* and Marr's *representational theory*, since appropriately encoded icons or "snapshots" of the world appear to allow the synthesis of computational mechanisms effectively equivalent to vision algorithms for tasks ranging from hyperacuity to object recognition.¹⁸

References and Notes

1. G. Westheimer and S. P. McKee, *Vision Research*, 17:941–947, 1977.
2. F. H. C. Crick, D. C. Marr, and T. Poggio, In F. Schmitt, editor, *The organization of the cerebral cortex*. MIT Press, Cambridge, MA, 1980. H. B. Barlow, *Nature*, 279:189–190, 1979.
3. T. Poggio, in: *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. LV, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 899-910, 1990.
4. T. Poggio, and F. Girosi, *A.I. Memo No. 1140* (Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, 1989), T. Poggio, and F. Girosi, *Proc. IEEE* **78**, 1481-1497, 1990. The HyperBF network scheme for the approximation of smooth functions has the form

$$f^*(\mathbf{x}) = \sum_{\alpha=1}^n c_{\alpha} G(\|\mathbf{x} - \mathbf{t}_{\alpha}\|_W^2) + p(\mathbf{x}) \quad (1)$$

where the parameters \mathbf{t}_{α} that correspond to the centers of appropriate basis functions G (such as the Gaussian), and the coefficients c_{α} are unknown, and are in general much fewer than the data points ($n \leq N$). The norm is a *weighted norm*

$$\|\mathbf{x} - \mathbf{t}_{\alpha}\|_W^2 = (\mathbf{x} - \mathbf{t}_{\alpha})^T W^T W (\mathbf{x} - \mathbf{t}_{\alpha}) \quad (2)$$

where W is an unknown square matrix and the superscript T indicates the transpose.

The network of Fig. 2b corresponds exactly to equation 1. Its interpretation is the following. The centers of the basis functions are similar to prototypes (see Fig. 3c), since they are points in the multidimensional input space. Each unit computes a (weighted) distance of the inputs from its center and applies to it the radial function. In the case of the Gaussian, a unit will be the most active when the input exactly matches its center. The output of the network is a linear superposition of the activities of all the basis functions, plus direct, weighted connections from the inputs (the linear terms of $p(\mathbf{x})$) and from a constant input (the constant term). Notice that in the limit case of the basis functions approximating delta functions, the system becomes equivalent to a look-up table holding the examples.

The parameters $\mathbf{c}, \mathbf{t}, \mathbf{W}$ are searched for during learning by minimizing an error functional defined as

$$H[f^*] = H_{\mathbf{c}, \mathbf{t}, \mathbf{W}} = \sum_{i=1}^N (\Delta_i)^2,$$

where

$$\Delta_i \equiv y_i - f^*(\mathbf{x}) = y_i - \sum_{\alpha=1}^n c_{\alpha} G(\|\mathbf{x}_i - \mathbf{t}_{\alpha}\|_{\mathbf{W}}^2).$$

Thus learning in the HyperBF network corresponds to finding values of parameters that minimize H . Iterative methods of the gradient descent type can be used for the minimization of H . An even simpler method that does not require calculation of derivatives is to look for random changes (controlled in appropriate ways) in the parameter values that reduce the error. In the simulations described in this paper the model was endowed with a dual incremental learning mechanism. First, when the model's performance on a new input was markedly inadequate (in comparison with recent history), that input was adjoined to the model as an additional center (prototype). This happened mainly in the initial trials, with the number of centers eventually reaching an asymptote that depended on the nature of the task and on the parameters that affected the decision to add new centers. The performance of the model during these first trials improved quickly, then stabilized as the number of centers asymptoted. Second, further gradual improvement in the performance was obtained by letting the model carry out a local random search in the space of existing HyperBF center coordinates. This search was guided by feedback given to the model (that is, by indicating whether the response at each trial was correct). Details of the learning algorithms, including an extension of the incremental learning algorithm to a situation in which no explicit feedback is available, can be found in Poggio, Fahle & Edelman, *MIT AI Memo 1271*, 1991 and Weiss, Edelman, Fahle & Poggio *Weizmann CS-TR 21*, 1991.

5. A description of MLP and the backpropagation technique used for learning is in D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Nature* **323**, 533 (1986). An overview of some of the classical techniques can be found in S. Omohundro, *Complex Systems* **1**, 273 (1987) and in R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, N.Y., 1973. Relations between MLP networks and HyperBF networks are mentioned in Ref. 4 and studied in Maruyama, Girosi and Poggio, *A.I. Memo No. 1291* (Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, 1991).
6. T. Poggio, M. Fahle and S. Edelman *A.I. Memo No. 1271* (Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, 1991), Poggio,

- Edelman and Fahle, 1992 (*Computer Vision, Graphics and Image Processing B*, in press); Weiss, Edelman, Fahle and Poggio, *CS-TR 91-21*, (Dept. of Applied Mathematics and Computer Science, Weizmann Institute), 1991. The simulation results were robust with respect to all parameters, including the number of inputs.
7. We have also experimented with a different version of the HyperBF model, in which orientation-selective receptive fields similar to those of simple cells in V1 played the role of the basis functions. See Weiss et al. 1991.
 8. Hyperacuity-level performance was independent of the precise location of the receptors. At the same time, different quasi-random receptor mosaics yielded different thresholds, sometimes by as much as a factor of two. A similar range of hyperacuity thresholds is observed in human subjects, even at full acuity and perfectly normal eyes.
 9. The model exhibited learning also on a longer time scale (see note 4), similar to the slow long-term learning component found in human subjects (M. Fahle, S. Edelman & T. Poggio, *Vision Research*, submitted).
 10. R. Watt and F. W. Campbell, *Spatial Vision* **1**, 31-38 (1985).
 11. The stimulus in the bisection task consists of three dots, arranged in a vertical line, at an approximately even spacing. The subject has to determine whether the middle dot is above or below the midpoint of the segment formed by the other two dots. The HyperBF module learned this hyperacuity task just as easily as it did in the line vernier case.⁶ Another simulation made a comparison between the line vernier task and a similar one in which each of the line segments has been replaced by two dots (situated at its endpoints). The network learned this task, as it did previously in the line vernier and the bisection cases. The better performance of the HyperBF module in the dot vernier task for small offsets parallels a recent surprising finding with human subjects (M. Fahle, personal communication).
 12. In a recent study, R. Bennett and G. Westheimer (*Percept. Psychophys.* **49**, 541-546, 1991), found surprisingly little learning of thresholds in three-dot alignment and grating discrimination. Their experiments used transfer of training across stimulus range to probe for learning, hiding possible effects of fast learning that may have happened in the baseline session (p. 544). Interestingly, the lack of transfer across stimulus range in these experiments is consistent with our notion of experience-based learning.
 13. A. Fiorentini and N. Berardi, *Nature* **287**, 453-454 (1981).
 14. A. Karni and D. Sagi, *Proc. Natl. Acad. Sci.* **88**, 4966-4970 (1991).

15. K. Ball & R. Sekuler, *Science* **218**, 697-698, (1987); S.P. McKee & G. Westheimer (1978) *Percept. & Psychophys.* **24**, 258-262 (1978); V. S. Ramachandran & O. Braddick, *Perception* **2**, 371 (1973).
16. Y. Fregnac, D. Schulz, S. Thorpe & E. Bienenstock, *Nature* **333**, 367-370, (1988).
17. T. Poggio and S. Edelman, *Nature* **343**, 263 (1990); S. Edelman and T. Poggio, *A.I. Memo No. 1181*, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1990, S. Edelman and D. Weinshall, *Biological Cybernetics* **64**, 209 (1991).
18. S. Edelman, D. Reisfeld & Y. Yeshurun, *CS-TR 91-20*, Dept. of Applied Mathematics and Computer Science, Weizmann Institute, 1991; R. Brunelli & T. Poggio, *Proc. 12th Intl. Joint Conf. on Artif. Intell.*, 1278-1284, Sydney, Australia, 1991; S. Edelman & T. Poggio, *A.I. Memo No. 1181*, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1990; R. Brunelli & T. Poggio *I.R.S.T. TechReport 9110-04*, I.R.S.T., Trento, Italy, 1991.
19. *We are grateful to F. Crick, F. Girosi, A. Hurlbert and Y. Weiss for useful discussions and suggestions. This report describes research done within the Artificial Intelligence Laboratory and the Center for Biological Information Processing in the Department of Brain and Cognitive Sciences, MIT and at the Department of Applied Mathematics and Computer Science, Weizmann Institute of Science. Support for this research is provided by a grant from ONR, Cognitive and Neural Sciences Division, and by the Artificial Intelligence Center of Hughes Aircraft Corporation as well as by the Deutsche Forschungsgemeinschaft (Heisenberg-Programme). Support for the A.I. Laboratory's artificial intelligence research is provided by the Advanced Research Projects Agency of the Department of Defense. TP is supported by the Uncas and Helen Whitaker chair.*