Acknowledgements

SE thanks Yair Weiss for programming assistance and useful discussions. This paper describes research done at the Massachusetts Institute of Technology within the Artificial Intelligence Laboratory and the Center for Biological Information Processing in the Department of Brain and Cognitive Sciences and Whitaker College, and at the Weizmann Institute of Science in the Department of Applied Mathematics and Computer Science. The Center's research is sponsored by grant N00014-88-K-0164 from the Office of Naval Research (ONR), Cognitive and Neural Sciences Division; by the Alfred P. Sloan Foundation; and by National Science Foundation grant IRI-8719392. The Artificial Intelligence Laboratory's research is sponsored by the Advanced Research Projects Agency of the Department of Defense under Army contract DACA76-85-C-0010 and in part by ONR contract N00014-85-K-0124. MF holds a grant from the Deutsche Forschungsgemeinschaft (Heisenberg-Programme).

[14] S. Lin and B. W. Kernighan. An effective heuristic algorithm for the traveling salesman problem. *Operations Research*, 21:498–516, 1973.

[15] T. Nazir and J. K. O'Regan. Some results on translation invariance in the human visual system. *Spatial vision*, 5:81–100, 1990.

[16] T. Poggio. A theory of how the brain might work. *Cold Spring Harbor Symposia on Quantitative Biology*, LV:899–910, 1990.

[17] T. Poggio and S. Edelman. A network that learns to recognize three-dimensional objects. *Nature*, 343:263–266, 1990.

[18] T. Poggio and F. Girosi. A theory of networks for approximation and learning. A.I. Memo No. 1140, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1989.

[19] T. Poggio and F. Girosi. Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247:978–982, 1990.

[20] J. Rimer. Learning to control a robotic arm using Radial Basis Function interpolation, 1991. M.Sc. thesis, Dept. of Applied Mathematics, Weizmann Institute of Science.

[21] E. L. Schwartz. Local and global functional architecture in primate striate cortex: outline of a spatial mapping doctrine for perception. In D. Rose and V. G. Dobson, editors, *Models of the visual cortex*, pages 146–157. Wiley, New York, NY, 1985.

[22] R. Watt and F. W. Campbell. Vernier acuity: interactions between length effects and gaps when orientation effects are eliminated. *Spatial Vision*, 1:31–38, 1985.

[23] G. Westheimer. The spatial sense of the eye. *Invest. Ophthal. Vis. Sci.*, 18:893–912, 1979.

[24] G. Westheimer and S. P. McKee. Visual acuity in the presence of retinal image motion. *Journal of the Optical Society of America*, 65:847–850, 1975.

[25] G. Westheimer and S. P. McKee. Spatial configurations for visual hyperacuity. *Vision Research*, 17:941–947, 1977.

[26] B. Widrow and S. D. Stearns. *Adaptive signal processing*. Prentice Hall, Englewood Cliffs, NJ, 1985.

[27] H. R. Wilson. Responses of spatial mechanisms can explain hyperacuity. *Vision Research*, 26:453–469, 1986.

References

[1] H. B. Barlow. Reconstructing the visual image in space and time. *Nature*, 279:189–190, 1979.

[2] B. Caprile and F. Girosi. A non-deterministic minimization algorithm. A. I. Memo 1254, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, Sept. 1990.

[3] B. Caprile, F. Girosi, and T. Poggio, 1991. in preparation.

[4] C. R. Carlson and R. W. Klopfenstein. Spatial frequency model for hyperacuity. *Journal of the Optical Society of America*, A2:1747–1751, 1985.

[5] F. H. C. Crick, D. C. Marr, and T. Poggio. An information-processing approach to understanding the visual cortex. In F. Schmitt, editor, *The organization of the cerebral cortex*. MIT Press, Cambridge, MA, 1980.

[6] S. Edelman and T. Poggio. Bringing the Grandmother back into the picture: a memory-based view of object recognition. A.I. Memo No. 1181, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1990.

[7] S. Edelman and D. Weinshall. A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics*, 64:209–219, 1991.

[8] M. W. Fahle, S. Edelman, and T. Poggio. Learning of vernier acuity, 1991. in preparation.

[9] E. Hering. Ueber die Grenzen der Sehschaerfe. In *Bericht. Mathem.-Physikal. Klasse Saechs.*, page 16. Ges. Wissenschaften, Leipzig, 1899.

[10] A. Hurlbert and T. Poggio. Synthesizing a color algorithm from examples. *Science*, 239:482–485, 1988.

[11] A. Karni and D. Sagi. Where practice makes perfect in texture discrimination. *Proceedings of the National Academy of Science*, 88:4966–4970, 1991.

[12] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[13] S. A. Klein and D. M. Levi. Hyperacuity thresholds of 1 sec: theoretical predictions and empirical validation. *Journal of the Optical Society of America*, A2:1170–1190, 1985.

(that is, provided there is at least one unit in the network which is sufficiently active, indicating that the new stimulus is close to one of the existing centers). Subsequently, the newly classified stimulus may be designated as an example for incremental learning, increasing the portion of the input space for which information is available.

Learning algorithms: details

The basic mechanism of learning in HyperBF networks is the computation of the optimal set of coefficients $\mathbf{c}_\alpha$ which relate the network's output vector to the vector whose components are the activities of the individual basis function units. Finding the matrix of coefficients amounts to the solution of a linear system, provided that the number of input/output examples is the same as the number of basis functions. If there are more examples than basis functions, the resulting overconstrained system can be solved by pseudoinverse methods. A one-shot method of this type does not appear to be biologically plausible. However, an equivalent result may be achieved, for the case of $\mathbf{c}_\alpha$, by gradient descent that can be implemented through a Hebbian mechanism (see [18]).

In the overconstrained case, repositioning the HyperBF centers $\mathbf{t}_\alpha$ through gradient descent can also improve the module's performance (see also section 2.4). To cite a typical concrete example, we have trained a 20-center network with 50 vernier examples, achieving mean error of 0.67 ± 0.07. After 20 steps of gradient descent, the error dropped to 0.045 ± 0.006.

In a more realistic situation, the HyperBF module is allowed to improve its performance not only during specially designated training trials, but also throughout the testing stage as in our incremental learning algorithm described in the main text. In our simulation incremental learning was based on a random search method described in section 2.4. Another possibility is to augment the HyperBF module with a Widrow-Hoff learning mechanism (see [26]), in which the coefficients $\mathbf{c}_\alpha$ are modified according to the following formula:

$$\Delta \mathbf{c}^{t+1} = \gamma \mathbf{c}^t \left(\mathbf{f}^t - \hat{\mathbf{f}}^t \right) \mathbf{h}^t$$

where $\mathbf{f}^t$ and $\hat{\mathbf{f}}^t$ are the correct and the estimated output values at trial t, and $\mathbf{h}^t$ is the vector of intermediate-layer values (which are the activities of the basis units). In other words, the coefficients $\mathbf{c}_\alpha$ are modified by an amount proportional to the error made in the current trial. It has been shown [26, 10] that the Widrow-Hoff mechanism is equivalent to an incremental computation of the appropriate pseudoinverse. In our simulations, mean error typically improved by 0.004 per trial for about 100 trials (as found by a linear regression of error on trial number), then became constant. These figures varied with the coefficient γ of the Widrow-Hoff equation.

performed by the subject. More generally, one may apply the same line of reasoning to visual tasks other than hyperacuity, and even to faculties other than vision. Indeed, HyperBF interpolation and related approaches have been recently applied with success to the modeling of human performance in such diverse tasks as three-dimensional object recognition [17, 6, 7] and motor control [20]. Importantly, learning HyperBF interpolation can be implemented in a simple biologically plausible network [19]. It remains to be seen whether the above framework would prove useful in unifying the existing diverse theoretical approaches to the modeling of visual perception, and of brain function in general.

Appendix: Learning modes of the HyperBF scheme

Incremental learning and bootstrapping in the absence of feedback

The HyperBF module must be allowed to improve its performance throughout the testing stage, with or without feedback. This can be achieved by using the algorithm we described in the body of the paper, namely, adding centers when the model's performance is inadequate. Coefficients can be modified and (possibly, on a slower time scale) centers can be moved so as to decrease the discrepancy between the desired value of the response and the correct one. If no feedback is available, it is still possible to estimate performance to provide guidance for learning. Such an estimate can be formed if the new input is not too far away from the existing centers, so that it is likely that the network would classify it correctly (a priori, the reliability of this classification is unknown, however). Imagine that a few examples of the hyperacuity task are given with feedback, that is, accompanied by the correct classification. Subsequently, new stimuli are given without feedback. If these stimuli are sufficiently similar to the original examples, and are classified correctly, they can be incorporated into the network as new centers (i.e., templates), effectively bootstrapping the learning process. Thus, a small modification of the scheme can make it work in the absence of feedback, under certain conditions.[4]

Notice that such incremental learning tasks are not uncommon in psychophysics. In particular, hyperacuity is often tested while adapting the size of the offset to the subject's performance, by slowly decreasing the offsets during testing. Under these conditions, the offset in each trial is never less than half the offset of the previous trial. According to our simulations, the network described earlier can generalize rather well to offsets of half the size (but not to offsets of say, four times the training size). The incremental learning algorithm described in the main text may be extended as follows. At first, in the absence of feedback the network may attempt to classify a new stimulus, provided that the potential classification would be sufficiently reliable

[4] The human subjects, of course, have prior exposure (with feedback) to an extremely wide variety of stimuli, through everyday visual experience.

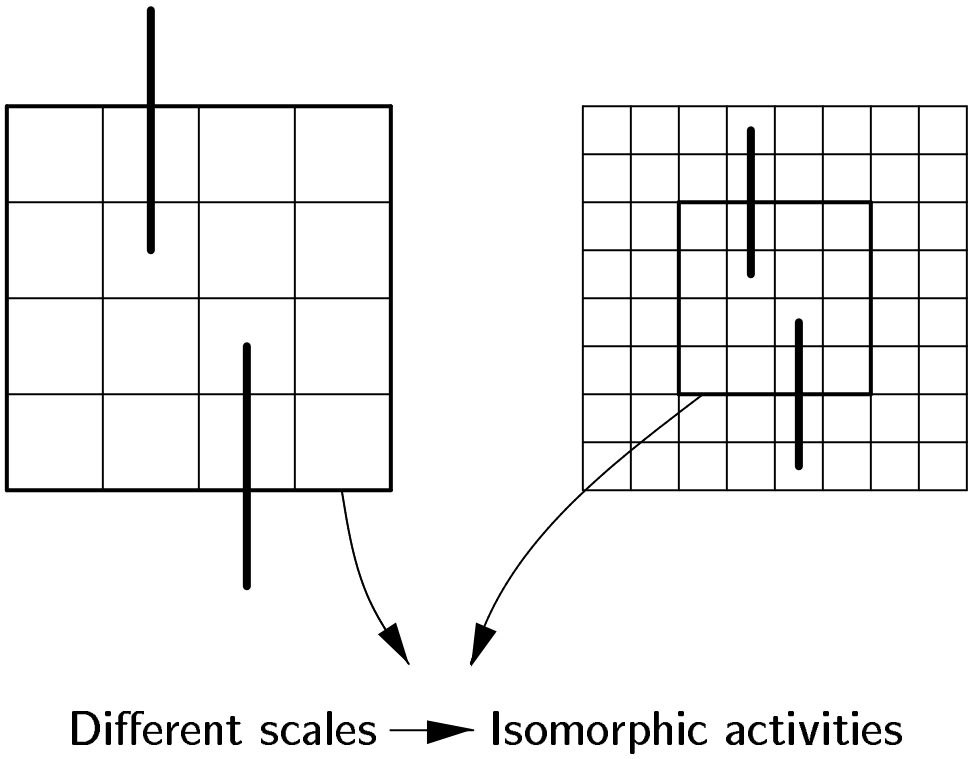

Figure 12: Scale invariance in hyperacuity tasks can be achieved in principle through simultaneous processing of the input at several levels of resolution, corresponding to several overlapping grids of "ganglion" cells of different size and spacing.

with circularly symmetric receptive fields, such as those found in the cortical layer $4C\beta$, or the activities of orientation-selective cells, or a mixture of both). The type of preprocessing in human vision and the associated pseudo-invariances that it may support are experimental issues of great interest. Of course, any lack of transfer of learning would support the simple model. Experimental demonstration of tranfer of learning with respect to translation and scale, would not represent in our view a major problem for the model, though it would require a more complex preprocessing than the one we have simulated. Transfer of learning from one type of stimulus to another (see Figure 1) would be a more serious blow to the spirit of our model and therefore a more critical test of its validity.

3.2 Implications for the modeling of visual performance in general

In view of the results reported above, we conjecture that the modules responsible for hyperacuity-level performance are synthesized in a demand-driven fashion, when the appropriate task is first

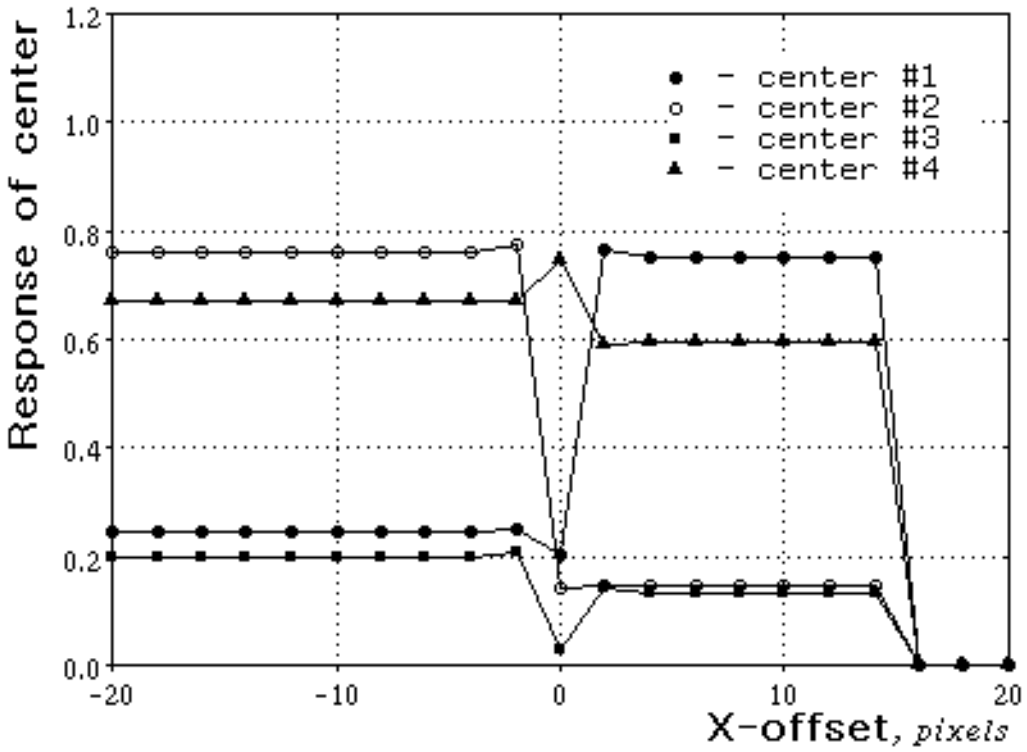

Figure 11: Responses of the four centers acquired during an incremental learning session that consisted of 150 trials vs. the offset of a test vernier presented at a fixed location (see text for interpretation). During learning, the offsets were uniformly distributed between 4 and 12 pixels.

lation, in addition to scale, is another obvious candidate transformation for which invariance could be built in. The bare version of our network, described here, would not generalize from one patch of the retina to another (though this may not be fully necessary; cf. [15, 11]). It seems likely that translation invariance, at least up to a certain extent, should be provided by mechanisms preceding the learning stage. It is possible that preprocessing mechanisms could also provide invariance to the specific stimulus type by computing the equivalent of "place tokens". This would enable the system to generalize automatically (without the need for examples) from, say, line stimuli to, say, dot stimuli. In any case, the input to a learning model such as the one we have outlined should not be raw photoreceptor activities, as in our simulations, but rather pre-processed photoreceptor activities (for instance, the activities of cells

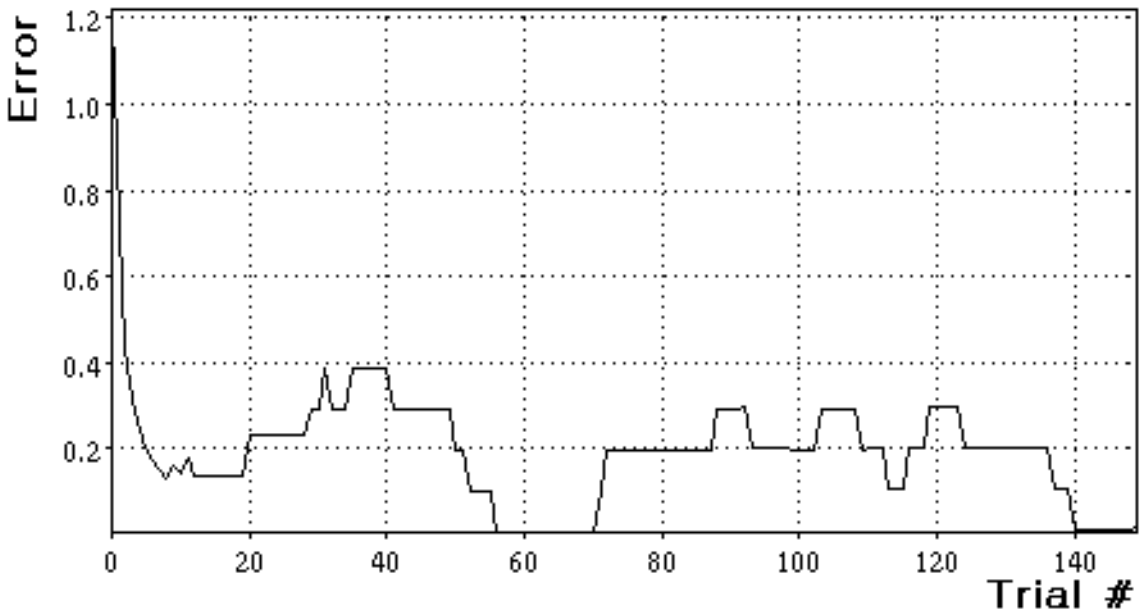

Figure 10: A plot of the output error of the HyperBF module vs. trial number, demonstrating the improvement in vernier acuity over 150 trials by the incremental learning technique described in section 2.4. The number of HyperBF centers in this particular example reached 4 by the 20th trial and remained constant after that. The best linear fit to the data set for the first 20 trials has a slope of -0.020. The slope for the entire 150 trials is -0.001. This illustrates the two learning phases, transient and steady, exhibited by the HyperBF module. The improvement in the performance obtained with human subjects in real psychophysical experiments is also relatively fast in the initial trials and slow but steady thereafter.

stimuli if the visual system has a built-in scale invariance mechanism. Several versions of scale-invariant models of early visual processing have been offered in the past (e.g., [21]). For our purpose, a simple scheme, in which invariance is achieved through simultaneous processing of the input at several levels of resolution (corresponding to several overlapping grids of "ganglion" cells of different size and spacing), would suffice. In such a case, the system could be prepared in advance, say, to a reduction in the input scale (up to a certain limit), simply because the small-scale grid would exhibit, after the reduction, a pattern of activity isomorphic to the pattern evoked by the large-scale input in the large-scale grid (see Figure 12).

In general, we expect that the cortex performs suitable pre-processing to provide approximate invariance to certain basic transformations, without the need for explicit learning. Trans-

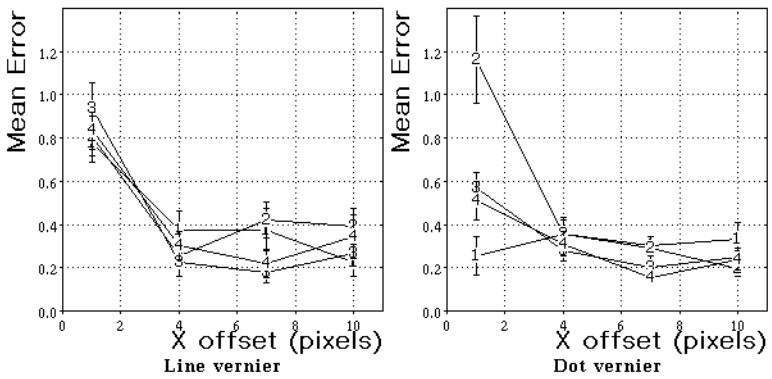

Figure 9: Mean error of the synthesized module vs. X-offset of the vernier stimulus, by Y-offset. *Left:* line vernier stimulus. *Right:* dot vernier stimulus. Note better performance in the latter case for small X-offsets.

3 Conclusions

3.1 Implications for the psychophysics of visual acuity

The skeleton model described in the preceding sections is specific enough to be put to a psychophysical test. One possible way to do so is to test the prediction of the model regarding generalization of performance from a well-practiced to an unfamiliar range of inputs. Consider, for concreteness' sake, the vernier acuity task. If the human visual system relies on a memory-based mechanism such as HyperBF interpolation to solve this problem, a drop in performance (that is, an increase in the error rate) is expected when the range of the stimuli is suddenly changed (e.g., if the verniers are made smaller by a factor of two or more in comparison with their values during training). If regression analysis is used to obtain an estimate of the psychometric function from error rates, such a change in the stimulus range would cause a decrease in the coefficient of determination of the regression, or in related measures of the goodness of fit. Furthermore, the subsequent recovery of performance should be slower if no feedback is provided after the change (even though some learning appears to be possible even without explicit feedback; see the appendix). A similar prediction of poor transfer between conditions holds for a change to a different hyperacuity task (say, from the top left stimulus in Figure 1 to the bottom right one). There are preliminary indications that all these phenomena indeed happen in practice [8].

No problems should be encountered in the transfer of performance to a different range of

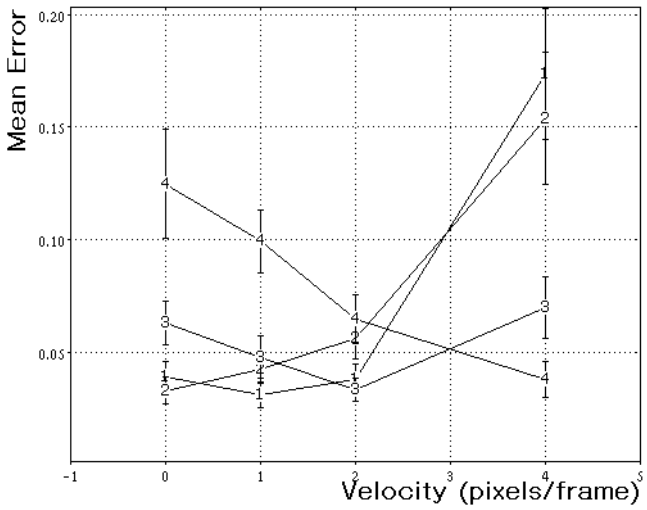

Figure 8: Mean error of the synthesized module vs. velocity of the stimulus during testing. X-offset was between 4 and 8 pixels (12″ to 16″), Y-offset was 1 pixel (3″). The four curves correspond to four values of velocity during training (same set of 4 values as the testing velocities — 0, 1, 2, and 4 pixels/frame, corresponding to 0, 1.5, 3, and 6 ″/frame). In general, performance deteriorates with increased testing velocity, but to a lesser extent if the training velocity was relatively high as well.

resulted in adequate performance was four. The nature of the representation of the input space by the centers in that case is illustrated in Figure 11, which shows the response of each center unit to vernier stimuli centered on the retina and having an X-offset ranging from -20 to 20 pixels.[3] 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. Clearly, 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.

[3]By analogy to single-cell electrophysiology, this illustration may be regarded as a recording of the "receptive fields" of the centers in the space of possible inputs.

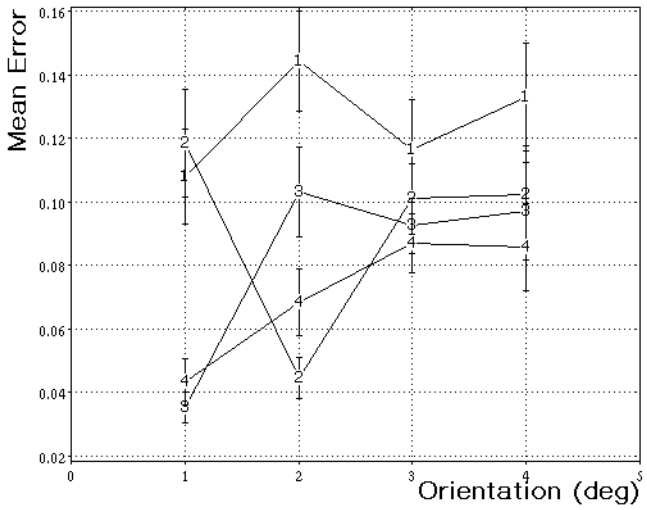

Figure 7: Mean error of the synthesized module vs. orientation of the stimulus (shown along the abscissa as the lower bound of a 1-octave range, in degrees), by stimulus length. X-offset was between 4 and 8 pixels (12″ to 24″), Y-offset was 1 pixel (3″). The four curves correspond to four values of stimulus length, from 10 to 40 pixels (30″ to 80″). In general, performance is seen to deteriorate with increased orientation range.

The algorithm for adjusting the positions of the existing centers was as follows. For each new input, the system made between 10 and 100 random changes in the value of a randomly chosen coordinate of a center (the amplitude of the change was about ten percent of coordinate value). After each change the error for that particular input was recalculated. If the new error was lower (and, with a small probability, if the error increased), the change was incorporated into the system, otherwise the change was reversed (cf. [2]).[2] If at any stage during the simulated experiment the current input was too distant from any of the existing HyperBF centers, that input was adjoined to the model as a new center (cf. learning by example acquisition in the CLF model of object recognition [7]; see also [18]). The time course of the performance of the resulting algorithm that combined adjustment of existing centers with recruitment of new centers is shown in Figure 10.

The number of centers eventually incorporated into the HyperBF module varied with the parameter that governed the acquisition of new centers. The smallest number of centers that

[2]The probability of keeping a change that led to a higher error could be decreased with time, as in the simulated annealing approach to optimization [12] (this feature, however, appeared to be unnecessary for our purposes).

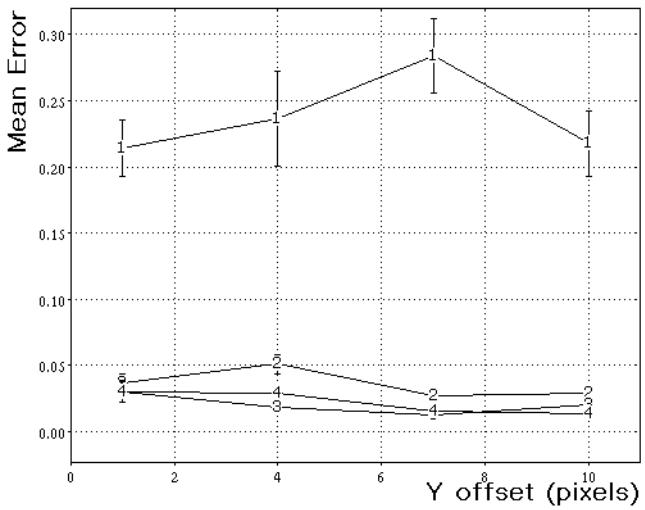

Figure 6: Mean error of the synthesized module vs. Y-offset of the vernier stimulus, by X-offset. The four curves correspond to four ranges of X-offset ($1-2$, $4-8$, $7-14$, $10-20$ pixels, corresponding to $1.5''-3''$, $6''-12''$, $10.5''-21''$, and $15''-30''$). Once the X-offset is high enough to guarantee good performance (curves 2, 3 and 4), increasing the Y-offset improves the performance level, as it does in human subjects.

2.4 Replication of the decrease of vernier threshold with practice

A major characteristic of human performance in hyperacuity tasks is the gradual and constant improvement of the threshold, which continues, albeit at a slow rate, after ten thousand trials ([8]; see the appendix). We have replicated this phenomenon by endowing the model with a dual incremental learning mechanism (see also [3]). 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). In the appendix we discuss how the incremental learning algorithm can be naturally extended to work even without explicit feedback, in certain tasks.

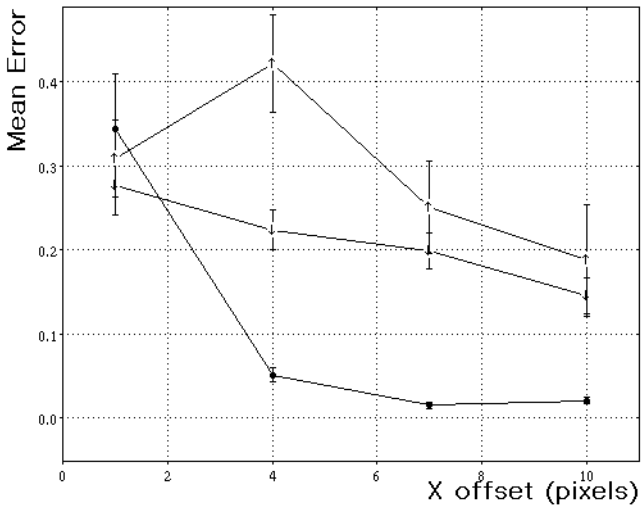

Figure 5: Mean error of the synthesized module vs. X-offset of the vernier stimulus. The module was trained to output 1 for left offset and -1 for right offset. Consequently, error of 0.1 corresponds to high performance (bars in this and other figures denote ± 1 standard error of the mean). The values of X-offset along the abscissa are the lower bounds of an octave range (e.g., 4 pixels means that the offsets were uniformly distributed between 4 and 8 pixels; in all our simulations the scale was 10 pixels to $30''$). The three curves correspond to three training/testing combinations. In the first one ($\bullet$), the same X-offset range was used both for training and testing. In the other two combinations ($\downarrow$ and $\uparrow$), the testing range was one-half and twice as large as the training range, respectively. Note that the smallest X-offset that still yielded high performance (mean error smaller than 0.05) is much smaller than the photoreceptor spacing ($6''$, compared to about $30''$).

Another experiment 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 comparison between the two vernier tasks appears in Figure 9. The better performance of the HyperBF module in the dot vernier task for small X-offsets parallels a recent surprising finding with human subjects (M. Fahle, personal communication).

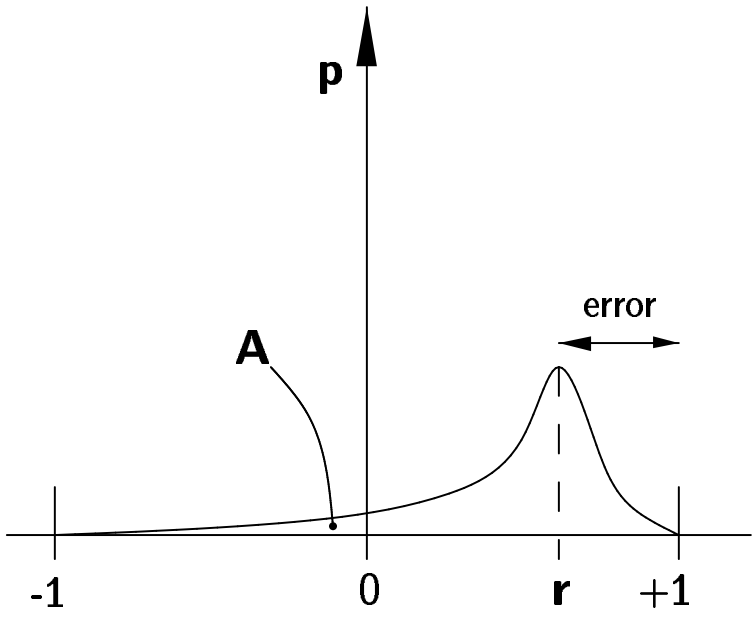

Figure 4: The relationship between the performance index used in the simulations — the absolute output *error* of the HyperBF module — and the acuity threshold. The probability density of the output is shown as a distribution centered at some $r > 0$, whose tail extends across 0 to the other half of the ± 1 range of possible values. The area A under the tail of the distribution indicates the probability of erroneous response, given the statistics represented by the mean and standard deviation of *error* (the two parameters we have measured in the simulations). The acuity threshold, in turn, can be related to the probability of erroneous response through probit analysis.

Importantly, the hyperacuity-level performance was independent of the precise location of the receptors. At the same time, different quasi-random receptor mosaics yielded thresholds that sometimes varied by as much as a factor of two. A similar range of hyperacuity thresholds is observed in human subjects with full acuity and perfectly normal eyes.

2.3 Comparison among different perceptual tasks

The next experiment compared the performance of the HyperBF module in two tasks: line vernier and three-point bisection. The stimulus in the bisection task consisted of three dots, arranged in a vertical line, at an approximately even spacing. The task is 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 as easily as it did in the line vernier case.

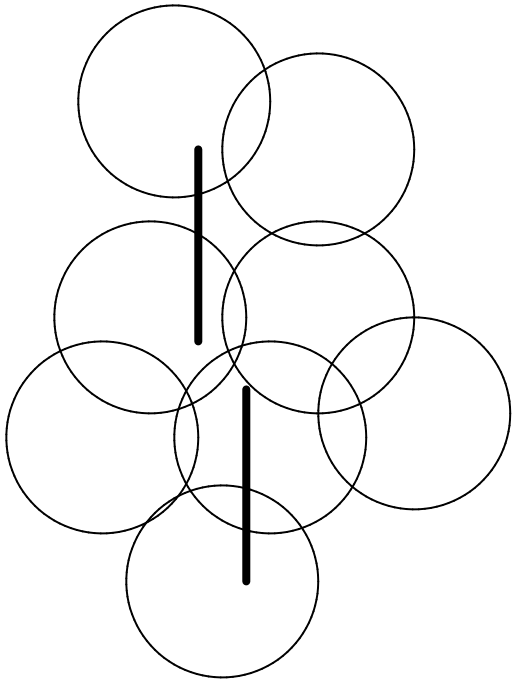

Figure 3: An illustration of the vernier acuity task: the subject has to tell whether the upper bar is to the left or to the right of the lower one. Human subjects (and the HyperBF simulation) perform this task at hyperacuity levels, that is, the minimum discernible horizontal displacement of the two bars is much smaller than the average distance between adjacent photoreceptors. The photoreceptor mosaic is shown superimposed on the stimulus. Each cone is shown as a circle that represents the Gaussian spread of a point source shining at the corresponding retinal location. This spread is due to the low-pass characteristics of the optics of the eye. Our simulation does not require positioning the "receptors" at precisely defined locations.

- The threshold improved with increasing vertical separation of the two segments comprising the vernier stimulus ([25]; Figure 6). We note that in human subjects this improvement reverts with further increase in the vertical separation; this phenomenon was also replicated by the model.

- The threshold deteriorated with increasing orientation difference between training and testing trials. This deterioration was more pronounced for shorter stimuli ([22]; Figure 7).

- Performance remained at hyperacuity levels when the stimuli moved across the retina, and was the highest when the velocity of the stimulus translation was the same during training and testing ([24]; Figure 8).

2 Example: simulated experiments in hyperacuity

In the preceding section we have proposed a new general approach to the modeling of perceptual function, whose central tenet is that the perceiver learns to solve the task at hand in a demand-driven fashion. According to the proposed approach, cortical modules that perform learning from examples rely on a computational process that may be regarded as multivariate function approximation from sparse data. The present section provides a rather detailed example of an application of this approach to the class of spatial discrimination tasks illustrated in Figure 1. Accounts of application of multivariate function approximation to a wide variety of learning tasks, ranging from object recognition to robot control may be found elsewhere [17, 20]. We proceed to describe a series of simulated psychophysical experiments, in which a simple version of a HyperBF module has been trained to perform several different hyperacuity tasks.

2.1 Simulation details

The input to the module was an array of "photoreceptors" whose activity corresponded to the input image, blurred by the eye's optics.[1] There were eight receptors, positioned randomly on a loose 4×2 grid (see Figure 3). Each of the receptors calculated its response by integrating the input over a Gaussian-shaped region of the "retina", with two space dimensions ($\sigma = 30''$) and one time dimension ($\sigma = 0.5$ units). The space dimensions of the "retina" were $180'' \times 360''$, while the time dimension had an extent of ± 1 unit. 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.

The performance of the module was estimated by measuring the absolute error, that is, the difference between the actual output (which could be any number between -1 and $+1$; for a proof see [6]) and the desired output (± 1; see Figure 4). Without going into the details, we point out that the absolute output error is a good analog of acuity threshold, since the two are related monotonically.

2.2 Replicating the basic psychophysics of vernier discrimination

The HyperBF module coupled to the input mechanism described above successfully replicated, after a training phase typically consisting of about 50 examples, the following four basic findings of the psychophysics of hyperacuity in human subjects:

- The equivalent acuity threshold was significantly lower than the spacing of the receptors in the simulated retina ([9, 23]; Figure 5).

[1] The identification of the input units with photoreceptors is done here for simplicity of exposition and does not imply an attempt to model the neurophysiological aspects of hyperacuity perception.

approximation of smooth functions was proposed recently under the name of HyperBF networks [19]. Detailed descriptions of this scheme, its theoretical underpinnings and its performance can be found in [19, 16, 18, 2]. Within the HyperBF scheme, a multivariate function is expanded in terms of basis functions, with parameter values that are found or "learned" from the data, i.e., the examples. The expansion has the form

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

where the parameters $\mathbf{t}_\alpha$ that correspond to the centers of basis functions, and the coefficients c_α are unknown, and are in general much fewer than the data points ($n \leq N$). The norm $\|\cdot\|$ is a weighted norm

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

where W is an unknown square matrix and the superscript T indicates the transpose. In the simple case of diagonal W the diagonal elements w_i assign a specific weight to each input coordinate, determining in fact the units of measure and the importance of each feature [19]. Equation 1 can be implemented by the network of Figure 2. The parameters $\mathbf{c}, \mathbf{t}, \mathbf{W}$ are searched for during learning by minimizing the error functional defined as

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

where

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

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 looks for random changes (controlled in appropriate ways) in the parameter values that reduce the error (cf. [14, 2]). The interpretation of the network of Figure 2 is as follows. The centers of the basis functions, which are points in the multidimensional input space, may be considered prototypical inputs for which the desired response is known. 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 basis function, 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 when the bases are delta functions, the system becomes equivalent to a look-up table holding the examples.

after a quick training phase in which it is exposed to examples of the task. In most psychophysical experiments, subjects are actually shown several examples of the task before testing takes place. Hyperacuity tests, in particular, require a significant training period in order to achieve good performance (thresholds typically decrease by a factor of two to four during the first several hundreds of stimulus presentations [25]; on the other hand, some subjects have thresholds of $10''$ or less upon the first testing). A broad prediction of our conjecture is that almost any psychophysical task that is initially tractable can be performed better after suitable training, provided that the necessary information is available in the stimulus. Computational support for this conjecture, outlined below, as well as the results of recent psychophysical experiments [8], supplies new evidence regarding the purposive aspect of human vision, in which the capability for solving any given task emerges only if the need for it is dictated by the environment.

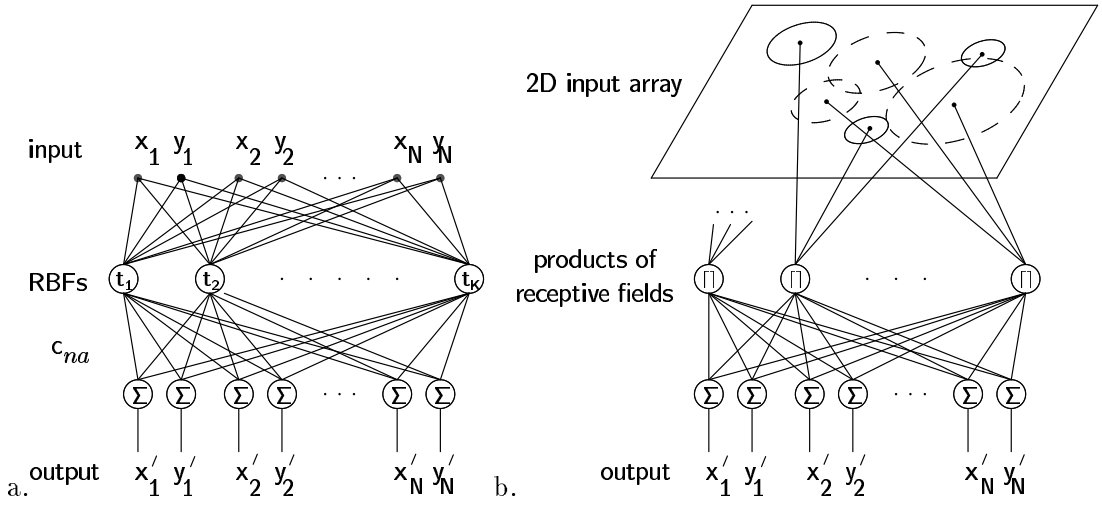

Figure 2: (a) A network representation of approximation by Hyper Basis Functions. (b) shows an equivalent interpretation of (a) for the case of Gaussian radial basis functions. Gaussian functions can be synthesized as the product of two-dimensional Gaussian receptive fields operating on retinotopic maps of features. The solid circles in the image plane represent the 2D Gaussians associated with the first radial basis function, which represents the first view of the object. The dashed circles represent the 2D receptive fields that synthesize the Gaussian radial function associated with another view. The Gaussian receptive fields transduce positions of features, represented implicitly as activity in a retinotopic array, and their product "computes" the radial function without the need of calculating norms and exponentials explicitly.

Synthesizing a module for a specific task from sample input/output pairs may be often regarded as approximating a multivariate function from sparse data. An efficient scheme for the

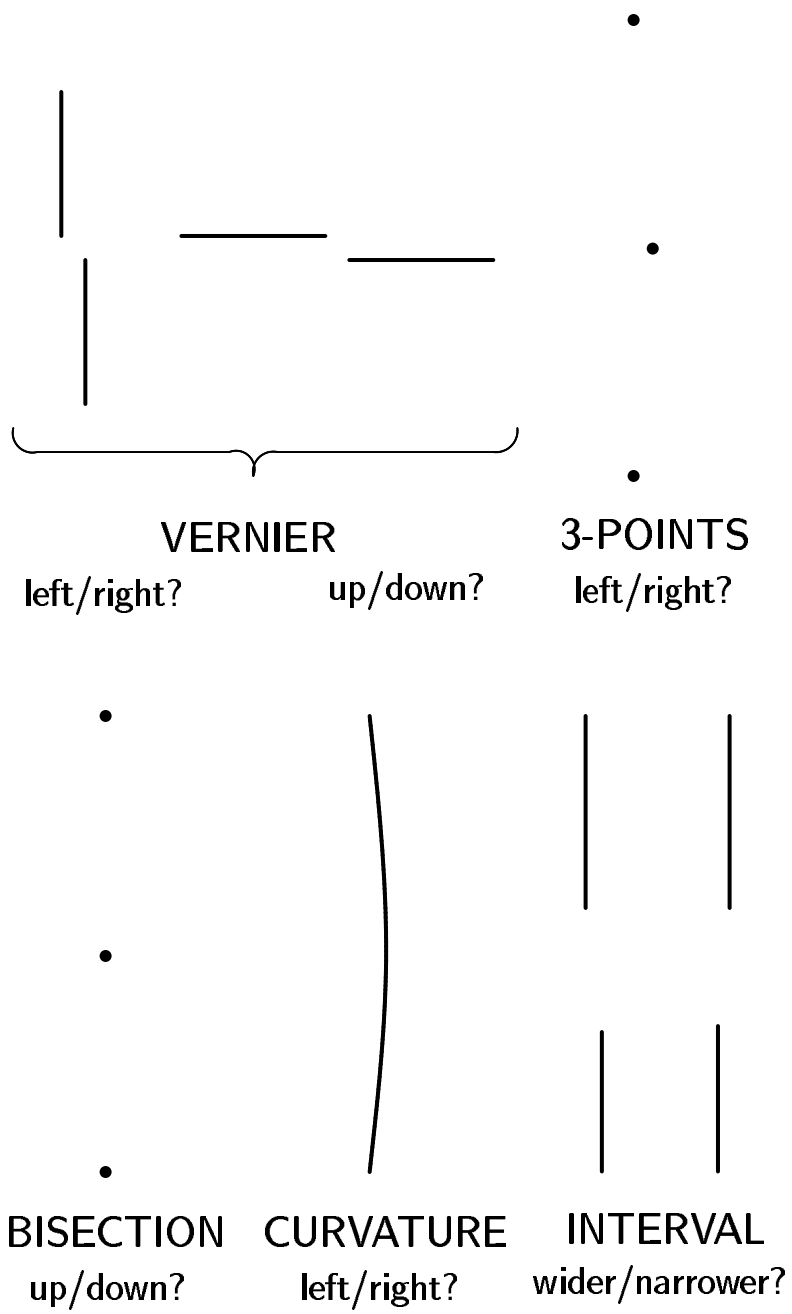

Figure 1: Illustration of five of the many different tasks in which human subjects perform at hyperacuity levels, that is, exhibit resolution finer than the spacing between individual photoreceptors.

1 A general framework for psychophysical modeling

The human visual system excels not only in the difficult tasks it faces routinely, such as recognition and navigation, but also in other, rather esoteric and no less difficult tasks, whose counterpart in the natural environment is not readily apparent. Figure 1 illustrates several such tasks that test different kinds of visual acuity. In all these tasks the visual system exhibits a hyperacuity-level performance (that is, the effective resolution it achieves is better than the best possible two-point acuity, dictated by the spacing of the photoreceptors in the fovea, the most densely sampled region in the visual field).

The impressive performance of the human visual system in the hyperacuity tasks indicates that the information necessary for solving each of these tasks is present in the stimulus and in principle can be recovered through appropriate processing [5]. The easiest approach to the computational modeling of hyperacuity processing, and of any other visual competence, is to conjecture a specific algorithm and a corresponding neural circuitry for each given task. This approach, if taken seriously, quickly leads to the absurd conclusion that the visual system had a hardwired mechanism ready to process any stimulus configuration the psychophysicists came up with in the last hundred years.

A more plausible alternative to the task-specific explanation postulates a common mechanism underlying the ability of the visual system to solve the different tasks. The possibility that such a mechanism exists has been mentioned in the last decade by several researchers. Consider again the example of acuity tasks, which we adopt as a test case throughout this paper. It has been shown that, in principle, spatial mechanisms that account for grating resolution can support hyperacuity-level performance [13, 4, 27]. Furthermore, some of the hyperacuity tasks can be solved by detecting "secondary" cues such as luminance difference (as in the bisection task) or orientation (as in the detection of vertical vernier stimuli).

The accounts offered in the past for hyperacuity performance suffer, however, from two limitations. First, the detailed structure of the neural circuitry postulated by the models tends to be problematic. In particular, the idea of a fine-grid reconstruction of the image in some layer of the cortex [1, 5] is unsatisfactory, because it still requires a homunculus looking at the reconstructed image and applying a different routine for each specific hyperacuity task. Second, it is not clear how the computational mechanisms mentioned in the models (e.g., metric computations in the space of receptive field activities [27]) can be integrated into a more general theory of brain function.

The present paper outlines an account of visual performance that is both biologically plausible and sufficiently general to encompass a variety of stimuli and tasks. We propose [16] that the brain may be able to synthesize, possibly in the cortex, appropriate modules for specific tasks

Learning of visual modules from examples: A framework for understanding adaptive visual performance

Tomaso Poggio[1] Shimon Edelman[2,4] Manfred Fahle[3]

1. Center for Biological Information Processing, Dept. of Brain and Cognitive Sciences, MIT, Cambridge MA, USA

2. Department of Applied Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot 76100, Israel
e-mail: edelman@wisdom.weizmann.ac.il
Tel: +972 – 8 – 342856
Fax: +972 – 8 – 344122

3. Universitäts Augenklinik, Schleichstr. 12, D7400, Tübingen, Germany

4. To whom correspondence should be addressed.

Abstract

Networks that solve specific visual tasks, such as the evaluation of spatial relations with hyperacuity precision, can be easily synthesized from a small set of examples. The present paper describes a series of simulated psychophysical experiments that replicate human performance in hyperacuity tasks. The experiments were conducted with a detailed computational model of perceptual learning, based on HyperBF interpolation. The success of the simulations provides a new angle on the purposive aspect of human vision, in which the capability for solving any given task emerges only if the need for it is dictated by the environment. We conjecture that almost any tractable psychophysical task can be performed better after suitable training, provided the necessary information is available in the stimulus.