Vision, Hyperacuity

Shimon Edelman[1] Yair Weiss[2,*]

1. Department of Applied Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot 76100, Israel

2. The Interdisciplinary Program, Tel Aviv University, Tel Aviv 69978, Israel

* Present address: Department of Brain and Cognitive Sciences, MIT E25-201, Cambridge, MA 02139, USA.

This article appeared in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, ed., 1009-1011, MIT Press, 1995.

RUNNING HEAD: Visual Hyperacuity

CORRESPONDENCE:
Shimon Edelman
Dept. of Applied Mathematics and Computer Science
The Weizmann Institute of Science
Rehovot 76100
Israel
phone: +972-8-342856
fax: +972-8-344122
email: edelman@wisdom.weizmann.ac.il

1 Introduction

1.1 What is Hyperacuity

The term "visual acuity" refers to the ability of an observer to resolve fine spatial details in a scene. The observer's acuity can be estimated using the familiar eye-chart in an optometrist's office. Alternatively, it can be measured by determining the observer's discrimination threshold in a simple visual task, similar to those shown in figure 1. In the latter case, the acuity estimate is found to be highly dependent on the task itself. While in some tasks (e.g., in telling apart two nearby dots) thresholds are in the range of 30-60 *arcsec*, in other tasks such as the vernier, the threshold may be as low as 5 *arcsec*. A threshold of 5 *arcsec* means that the observer reliably resolves features that are less than 0.02 *mm* at a 1 *m* distance, or the size of a quarter-dollar coin viewed at 17 *km*!

One can better appreciate the astonishing precision of this performance by considering the optical properties of the eye. In the spatially most sensitive region of the retina, the fovea, the diameter of the photoreceptors is in the range of 30-60 *arcsec* and the sizes of the receptive fields of the retinal ganglion cells may be even larger. Thus, humans can resolve detail with an accuracy of better than one fifth of the size of the most sensitive photoreceptor. G. Westheimer coined the term hyperacuity to describe such performance (Westheimer, 1981).

1.2 Why is Hyperacuity Interesting

Visual hyperacuity has been studied by experimentalists for over a hundred years, the first report of vernier hyperacuity dating back to 1892. By now, probably the most interesting point about hyperacuity is that it should not come as a surprise, if considered in the context of the computational neuroscience of the visual system. First, due to the point spread function of the eye's optics, a very small dot of light projected onto the retina may activate as many as 40 different photoreceptors. Thus, there is no *a priori* reason to expect a simple relation between the acuity exhibited by the system and the spacing of adjacent photoreceptors. Second, there is no reason to expect identical thresholds for different tasks, unless it is assumed that the first stage of vision, common to all subsequent processing, amounts to an internal reconstruction of the outside world in some unique and veridical way.

2 Computational Characterization of Hyperacuity

Models of visual information processing can be thought of as lying between two extremes. On the one hand, there is the notion that the goal of the system is to reconstruct the visual scene and display the result for the benefit of a decision module (it is difficult to resist calling this module anything but a homunculus). On the other hand, there is the approach, related to the notion of direct perception, according to which vision is a process of transduction into an abstract feature space followed by classification in this space. Among the models of hyperacuity, one can find instances both of the reconstructionist and of the feature-space approaches.

2.1 Reconstructionist Models

In a classic paper entitled "Reconstructing the visual image in space and time," H. Barlow pointed out that the phenomenon of visual hyperacuity is analogous to our ability to perceive films at the cinema (Barlow, 1979). In a film, shown as a series of discrete frames, we perceive a continuous motion, presumably through a process of temporal interpolation. Likewise, in a hyperacuity task, although the photoreceptors provide a discrete sampling of the retinal image, we perceive detail smaller than the sampling rate, a process of spatial interpolation. Barlow suggested that the first stages of visual processing may be concerned with such interpolation (see also Crick et al., 1981). The output of the spatial interpolation process, he wrote, would be a reconstruction of the retinal image on a finer grid, which may correspond to the granular cells of the cortical layer $4c\beta$ found in the macaque's primary visual area. Such fine-grid reconstruction of the stimulus would then support hyperacuity-level performance by allowing a subsequent stage to base its decision on the activity of individual units, an impossible feat as far as the raw retinal image is concerned.

The Shannon-Nyquist sampling theorem states that if the signal is bandlimited, a sampling rate higher than twice the maximal frequency would enable perfect interpolation. Due to the optical point spread function, the retinal image is indeed bandlimited to spatial frequencies less than about 60 cycle/deg, hence a sampling rate of 30 *arcsec* should suffice to reconstruct the retinal image. However, in many experiments the sampling rate does not satisfy the requirements of the sampling theorem. For example, the subjects of Fahle and Poggio (1981) were shown discrete frames of the motion of a vernier target drifting at a constant velocity. Resolving the misalignment of the vernier thus entailed both spatial and temporal interpolation, with the sampling rates $\Delta x, \Delta t$ imposed on the system. Fahle and Poggio found that humans can still solve the vernier task, at thresholds significantly lower than photoreceptor spacing, for a wide range of sampling rates. They concluded that an additional computational constraint is used by the visual system for spatiotemporal interpolation, namely, the assumption of constant velocity, and that this assumption may explain why the system opts for interpolation in the first place. Extending Barlow's suggestion, they proposed that the goal of interpolation may be broader than simple sub-receptor precision, and that it involves some correction in the visual representation by enforcing *continuity* of the visual field in space and time. (see the chapter on Ill-posed Problems in Vision)

2.2 Feature-Space Models

In 1986, H. Wilson showed that the responses of filters, similar in properties to the psychophysically defined spatial frequency channels and to the oriented receptive fields of cells in the primary visual cortex, can serve as the feature space for solving hyperacuity tasks. In Wilson's model, stimuli are represented as points in a multidimensional filter response space. The basic prediction of the model is that the threshold for discriminating two patterns is monotonically related to the Euclidean distance between the local representations of the two patterns, $\|\Delta R\|$. The term "local" here refers to the assumption that not all filter responses contribute to the calculation of ΔR: for each value of orientation and size, only filters overlapping the stimulus, as well as their spatial nearest neighbors, are used.

To test this prediction, Wilson computed the threshold of the model for a wide range of hyperacuity tasks, and obtained a quantitative fit with psychophysical data, with the parameters of the filters determined independently by previous experiments. Wilson's model also provided an intuitive explanation of how widely tuned, discretely spaced filters can support the resolution of small features. The most informative filter, Wilson observed, is not necessarily the one which is most active. For example in the vernier task, the most informative filters are the two oriented filters flanking the stimulus, which are significantly less active than the filter oriented at the same orientation as the stimulus.

A possible architecture for classification in the filter space was suggested by Poggio et al. (1992). In their model, the stimulus is transduced into R^n by a convolution with a bank of widely tuned, circularly symmetric Gaussian filters. Classification, or judging the offset of the vernier, is then carried out using Radial Basis Function approximation (Poggio and Girosi, 1990) in R^n. Poggio et al. found that this model can resolve verniers with offsets significantly smaller than the spacing between the Gaussian filters, again demonstrating the importance of pooling of information between units. Subsequently, it was shown that if Wilson's filters are used as transducers into R^n, then the classification can be performed using a simple perceptron (Weiss et al., 1993).

2.3 Comparison Between Models

It seems that the reconstructionist models of hyperacuity were motivated equally by the desire to explain the performance of the visual system in acuity tasks, and to find a computational rationale for the existence of the dense sheet of cells in layer $4c\beta$ of the macaque primary visual cortex. The appeal of such models was somewhat diminished by the absence of a comparable dense representation area in cats, given the cat's ability to solve the vernier task at a hyperacuity level (Swindale and Cynader, 1986). Specifically, at least in this case, Barlow's fine-grid reconstruction scheme outlined above has no fine grid to rely upon, and an alternative explanation seems to be required. Still, many of the insights on spatio-temporal interpolation remain valid (e.g., the utility of interpolation in avoiding motion smear and in filling in gaps in the retinal input due to local occlusion), even when the interpolation is assumed somehow to be mixed with the decision procedure, rather than merely leading to an explicit reconstruction of the input on a finer grid.

3 The Neural Basis of Hyperacuity

What neurobiological mechanisms in the brain support hyperacuity-level performance? To obtain a comprehensive answer to this question, one must relate data from anatomy and physiology (e.g., dendritic tree shapes and functional tuning curves of neurons) to behavioral findings (e.g., thresholds) measured by psychophysicists.

Parker and Hawken (1985) suggested that this relation can be elucidated by measuring the psychometric function of a single unit, called the "neurometric function." Rather than averaging the firing rate of a neuron over many presentations, this function captures the ability of a neuron to respond correctly to a single stimulus presentation. A response is defined to be correct if it agrees with the unit's behavior in the long run. For example, a neuron tuned to vertical bars is said to respond correctly in an orientation discrimination trial if its firing rate exceeded a certain value, given that the stimulus in that trial was vertical. Once the neurometric curve (a plot of the percentage of correct responses vs. the difference between the two patterns) is plotted, the behavioral threshold of the neuron can be determined by standard regression methods such as those used in psychophysics.

Applying this technique to neurons in the visual cortex of an anesthetized monkey, Parker and Hawken found that some neurons had thresholds significantly below photoreceptor spacing. Motivated by Barlow's suggestion that hyperacuity is linked to non-oriented neurons in cortical layer $4c\beta$, they compared the performance of these neurons to orientationally selective cells. They found that the non-oriented neurons performed worse in localization tasks. Using a similar technique, Swindale and Cynader (1986) measured the vernier threshold of cells with oriented receptive fields in cortical area 17 of the cat. They found cells with vernier thresholds as low as 6 *arcmin*.

The significance of these experimental results depends on the assumption that the threshold of a single cell limits the behavioral threshold of the organism. If an organism indeed can only perform as well as its best-tuned single neurons, then the finding of neurons with exceptionally low thresholds would constitute a step towards unraveling the neural basis of hyperacuity. If, on the other hand, information sufficient for solving the task at a hyperacuity level is available in the coarse-coded population response of only marginally selective units, then the response of each such unit does not tell us much about the emergent capability of the entire system.

A computational study by Snippe and Koenderink (1992) suggests that the population coding scenario may be more relevant to the understanding of hyperacuity than the single-unit one. Snippe and Koenderink determined the discrimination threshold of an ideal observer given the activities in an array of noise-perturbed receptors, which were assumed to have a Gaussian response profile, with an unknown width. These authors derived a formula for the threshold as a function of the width of the Gaussian, the distance between two neighboring receptors, and the functional dependence between the noise and signal in each receptor. They found that the threshold as compared to the tuning width can be made arbitrarily small if the individual receptors show high fidelity, or if there is considerable overlap between the receptors. Thus, an ensemble of cells with large *single-cell* thresholds may in fact support hyperacuity-level *behavioral* thresholds.

4 Perceptual Learning in Hyperacuity

Human performance in a vernier task improves with practice (that is, the threshold decreases), by as much as 40% over 2,000 trials (McKee and Westheimer, 1978). As in many other cases of perceptual learning, improvement is specific for stimulus parameters such as orientation and the eye to which the stimulus is presented. Interestingly, learning does not require that the subjects be informed as to the correctness of their responses.

Poggio et al. (1992) proposed that improvement in performance with practice may result from synthesizing a perceptual module for the task by acquiring examples of the input-output relationship, and by subsequent optimization of the use of the stored examples with practice. They demonstrated that a model based on this principle can learn to solve hyperacuity tasks starting with a "tabula rasa" state and continuously improving its performance. The basic prediction of their model is that generalization (supported by conservative interpolation among the stored examples), should be limited to stimuli that are similar to the previously seen ones, as indeed is observed in psychophysical experiments.

The model suggested by Poggio et al. requires feedback for parameter update. Weiss et al. (1993) developed a model that operates without feedback, in which learning is achieved not by synthesizing a new perceptual module, but rather by improving the performance of an existing one, based on a stimulus-driven, feedback-independent amplification of unit responses. They proposed a simple feedback-independent rule for synaptic modification, and showed that this rule can improve the signal to noise ratio in an existing network module, as a result of repeated exposure to the stimulus. The psychophysical data on the time course of learning indicates that it can be divided into an initial fast and a subsequent slow and prolonged components. This finding suggests that at least two different learning strategies may coexist in the visual system, perhaps corresponding to the two rules outlined above.

5 Discussion

To conclude, we point out that the two most intriguing aspects of hyperacuity (the surprisingly high resolution exhibited by a "sloppily built" system, and the dependence of the resolution on the exact nature of the task) may be found in many biological systems, both in vision and in other natural computation domains, ranging from bat echolocation to jamming avoidance response of electric fish (Altes, 1988). The study of this phenomenon indicates that issues vitally important for the understanding of how the brain works — representation, processing, and learning — are frequently best addressed with the simplest of stimuli.

References

Altes, R. A. (1988). Ubiquity of hyperacuity. *J. Acoust. Soc. Am.*, 85:943–952.

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

Crick, F. H. C., Marr, D. C., and Poggio, T. (1981). An information-processing approach to understanding the visual cortex. In Schmitt, F., editor, *The organization of the cerebral cortex.* MIT Press, Cambridge, MA.

Fahle, M. W. and Poggio, T. (1981). Visual hyperacuity: spatiotemporal interpolation in human vision. *Proceedings of the Royal Society of London B*, 213:451–477.

McKee, S. P. and Westheimer, G. (1978). Improvement in vernier acuity with practice. *Perception and Psychophysics*, 24:258–262.

Parker, A. J. and Hawken, M. J. (1985). Capabilities of monkey cortical cells in spatial resolution tasks. *Journal of the Optical Society of America*, 2:1101–1114.

Poggio, T., Fahle, M., and Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, 256:1018–1021.

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

Snippe, H. P. and Koenderink, J. J. (1992). Discrimination thresholds for channel-coded systems. *Biological Cybernetics*, 66:543–551.

Swindale, N. V. and Cynader, M. S. (1986). Vernier acuity of neurones in cat visual cortex. *Nature*, 319:591–593.

Weiss, Y., Edelman, S., and Fahle, M. (1993). Models of perceptual learning in vernier hyperacuity. *Neural Computation*, 5:695–718.

Westheimer, G. (1981). Visual hyperacuity. *Prog. Sensory Physiol.*, 1:1–37.

Legend for figure 1.

Three examples of stimuli (vernier, chevron, and bisection) used for testing spatial acuity of the visual system. The task is to tell the sense of the offset of some parts of the stimulus with respect to the others.

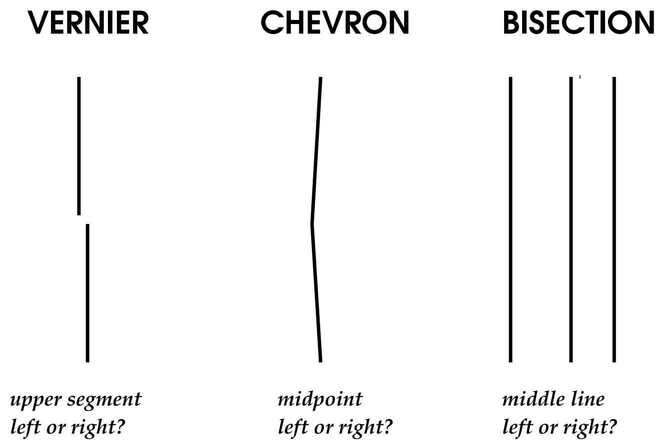
VERNIER
CHEVRON
BISECTION
upper segment
left or right?
midpoint
left or right?
middle line
left or right?

Figure 1