Pelli, D. G. (1991) Noise in the visual system may be early. In: M. Landy & J. A. Movshon (Eds.), *Computational Models of Visual Processing* (pp. 147-152). Cambridge: MIT Press.

11

Noise in the Visual System May Be Early

Denis G. Pelli

Most models of visual performance incorporate random variations: noise. Sometimes the model maker injects this randomness as an afterthought, to account for trial-to-trial variations in subjects' responses. Sometimes the noise is an essential part of the model, as in probability summation, where the key idea is the random detection that can occur in any of many independent channels (e.g., Graham, 1977). Recent psychophysical evidence indicates that the main source of noise limiting visual detection is in the proximal stimulus: the photon noise arising from the random nature of light absorption (Banks, Geisler & Bennett, 1987; Cohn, 1976; Krauskopf & Reeves, 1980; Pelli, 1981, 1983, 1990). While noise undoubtedly arises at every stage of visual processing, it is a hallmark of well-engineered systems, and apparently of human vision (at near-threshold contrasts), that the gains of each stage are sufficient to amplify the unavoidable noise of the first stage (i.e., photon noise) so that it dwarfs the additional noises that arise at later stages.[1]

This chapter argues that psychophysical models of the visual system should incorporate noise at the first stage, an equivalent input noise added to the stimulus, rather than injecting an arbitrary noise later. The observer's equivalent input noise is susceptible to direct measurement psychophysically (Pelli, 1981, 1983, 1990), thus removing many degrees of freedom from the specification of the model (Ahumada, 1987; Ahumada & Watson, 1985). Physiological models may usefully represent noise arising at each stage, but with appropriate simplifying assumptions (e.g., linearity) these models will be equivalent to a black-

---

[1] Previous claims that visual detection is photon noise limited (e.g., Rose, 1942, 1948) have been criticized on the grounds that the observer's overall efficiency, from stimulus to decision, is strongly dependent on the experimental conditions (Barlow, 1962). However, the observer's overall efficiency depends not only on the amount of noise in the system. It also depends on the efficiency of the observer's algorithm for reducing the multidimensional noisy data to a single number upon which a detection decision may be made. Our visual system's algorithms are not particularly well matched to the stimuli and tasks that we usually use to study vision, resulting in low efficiency strongly dependent on the stimulus conditions. For example, the observer's overall efficiency is very strongly dependent on disk size (Barlow, 1958; Jones, 1959) but the observer's equivalent input noise is independent of disk size (Pelli, 1981). (See Pelli, 1990.)

Pelli, D. G. (1991) Noise in the visual system may be early. In M. Landy and J. A. Movshon (Eds.), Computational Models of Visual Processing (pp. 147-152), Cambridge: MIT Press.

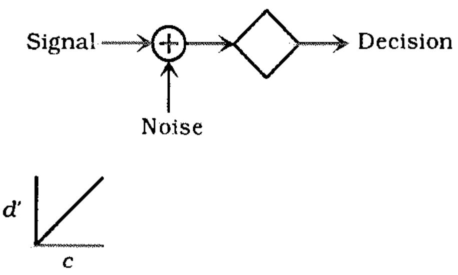

Fig. 11.1
The simplest case: ideal detection of a one-number signal in noise. The psychometric function, in the lower left hand corner, is linear.

box model with only an equivalent input noise (Pelli, 1990; see chapter 7).

The *theory of signal detectability* is a collection of theorems about the detectability of signals in noise (Peterson, Birdsall & Fox, 1954). This naturally formed the starting point for the first attempts to model sensory processes as noise-limited detectors. When considering the task faced by our visual system, it is instructive to begin by asking what would be the ideal way to detect a signal in noise. These sensory models, and many ad hoc modifications of them, have come to be called *signal detection theory*, after Green and Swet's (1974) textbook of that name. Unfortunately, many students of signal detection theory have mistakenly come to think that known human behavior is inconsistent with an early noise. This is because theorems proved for zero-dimensional models—whose stimulus is a single number—have been mistakenly assumed to generalize to the multidimensional case of real vision—whose stimulus is a dynamic image (i.e., many numbers).

Figure 11.1 illustrates the simplest possible detector. A signal, represented by a single number, is added to noise, a single random number, and the resulting sum is compared to a criterion (in the diamond box) to yield a detection decision, such as "Yes, it's there" or "No, it's not." Absence of the signal would correspond to a value of zero added to the noise. This decision rule, with the appropriate criterion, is the ideal way to decide whether or not the signal is present. Alternatively, in a two-alternative, forced-choice paradigm, the signal and zero would each be presented once, in random order, and the decision stage would choose the interval that produced the larger number. Again, the decision is ideal.

The graph in the lower left hand corner describes the performance of the detector. The horizontal scale is contrast c, in this case equal to the signal. The vertical scale is detectability d', which describes the level of performance. d' is a simple transformation of one of the conventional measures of detection performance, such as proportion correct. d' is defined as the signal-to-noise ratio required by an ideal observer to equal the observed performance of the observer under study (Tanner & Birdsall, 1958). For this simple case the graph is linear, indicating that d' is proportional to the input contrast.

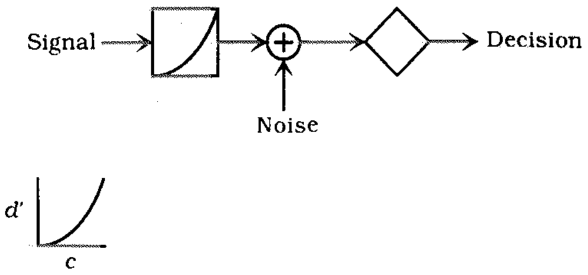

Fig. 11.2
Inserting a nonlinear tranformation before adding the noise results in a nonlinear psychometric function, which is needed to match the human observer. However, this assumes a late noise, which is inconsistent with evidence that visual detection is limited by photon noise.

A linear relationship between d' and contrast is inconsistent with the nearly universal finding that for visual detection of simple patterns d' is a nonlinear accelerating function of contrast, that is, $d' \propto c^2$ (e.g., Nachmias, 1981; Nachmias & Sansbury, 1974). Modelers have ususally resolved this inconsistency by introducing a comparable nonlinearity into the model, as illustrated in figure 11.2. By applying a nonlinear transformation to the signal *before* the noise is added, the signal-to-noise ratio d' becomes a nonlinear function of contrast, as shown in the lower left hand corner of the figure.

The nonlinear transformation introduced in figure 11.2 bent the psychometric function (d' vs. c) appropriately, but it also nullified any claim that the model is ideal. An ideal detector would yield the best possible performance, limited solely by the statistics of the stimulus. In figure 11.2 the noise is assumed to arise inside the model, not to be part of the proximal stimulus, like photon noise.

Figure 11.3 shows what would happen if we were to introduce the nonlinearity *after* the signal and noise are summed. Nothing. It is traditional to call the number that is fed into the decision box the decision variable. The introduction of the monotonic nonlinearity will change the values of the decision variable, but it will preserve

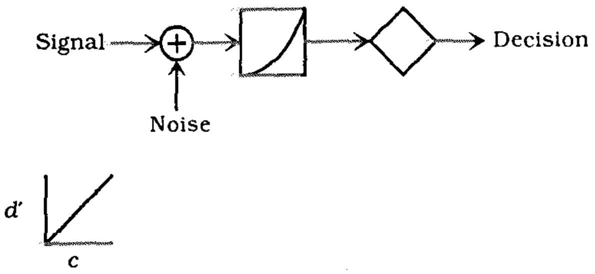

Fig. 11.3
Birdsall's theorem: Inserting the nonlinearity after the noise has no effect.

their ranking (which is greater than which). Since the model observer's decisions all depend solely on comparisons of values of the decision variable, the monotonic transformation will not affect any of the model observer's decisions. (For yes-no or rating-scale tasks, where the observer is comparing the decision variable with a remembered criterion, we are free to think of the criterion as a remembered value of the decision variable.) This immunity to monotonic transformation of the decision variable is sometimes called Birdsall's Theorem (Lasley & Cohn, 1981; Tanner, 1961).

Figures 11.1 to 11.3 might seem to imply that human observers must have a late noise. However, these figures apply only to the zero-dimensional case where the signal and noise are each simple numbers. Real vision is a three-dimensional problem: The stimulus is a pattern varying over two spatial dimensions and time. Such a stimulus may be thought of as a continuous function of three dimensions or may be represented by a large three-dimensional array of numbers; it cannot be represented reasonably by a single number. An early visual noise, such as photon noise, has the same high dimensionality.

Figure 11.4 illustrates the multidimensional version of figure 11.3. The large open arrows now represent the transmission of a multidimensional quantity—a movie—and the thin arrows continue to represent the transmission of single numbers. The multidimensional signal and noise are added together, and the multidimensional sum goes to an unspecified nonlinear transformation that somehow reduces it all to a single number, which is the basis for the decision. As indicated by the question mark in the lower left hand corner, such a model could have any psychometric function.

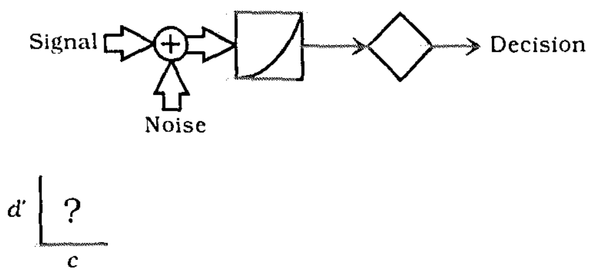

Fig. 11.4
When the signal and noise are multidimensional, as in real vision, then Birdsall's theorem no longer applies, and any psychometric function is possible.

As an example, figure 11.5 illustrates a detector with early noise that nonetheless has a nonlinear psychometric function. This is the uncertainty model, so called because it arose in the context of theory of signal detectability as a nearly optimal way to detect one of M orthogonal signals in white noise (Nolte & Jaarsma, 1967; Pelli, 1985; Peterson, Birdsall & Fox, 1954; also see chapter 10).[2] We assume that there are M possible signals, that the signals are all orthogonal (i.e., have equal energy and zero correlation with each other). There are M filters (i.e., receptive fields), each matched to one of the possible signals. Each filter yields a single number, a measure of the likelihood that a particular signal was present. The decision is based on the largest of these numbers. If M is about 50 then d' will be proportional to squared contrast. Alternatively, if M is 1—no uncertainty—then d' will be proportional to contrast, that is, the psychometric function will be linear. In general, the exponent k in the power law $d' \propto c^k$, increases linearly with log M (Pelli, 1985).

Figure 11.6 shows a more physiologically plausible way to build the uncertainty model of figure 11.5. Each filter is implemented by a cell (e.g., a cortical cell) with the appropriate receptive field and temporal impulse response. Rather than having an explicit maximum oper-

---

[2] Choosing the largest cross-correlation is optimal for identification, and nearly optimal for detection. The ideal detector is considerably more difficult to analyze. The simplifying assumptions of orthogonality of the signals could probably be dropped, provided that M were reduced appropriately, possibly to the number of orthogonal basis functions required to span the stimuli. Of course, in real life one would not expect the signals to match the receptive fields. Pelli (1985) showed that allowing a signal to stimulate multiple filters in the uncertainty model provides a good account for the results of "probability" summation experiments. Also see chapter 10.

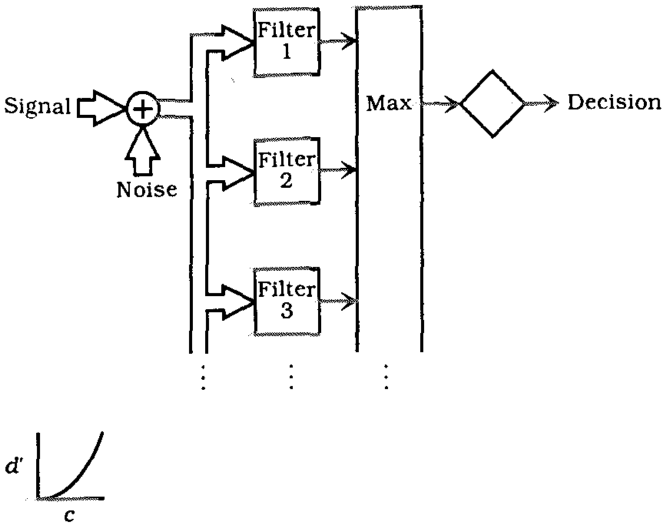

Fig. 11.5
An example: The uncertainty model has an early noise and yet results in a linear psychometric function if $M = 1$ and a nonlinear psychometric function if $M > 1$.

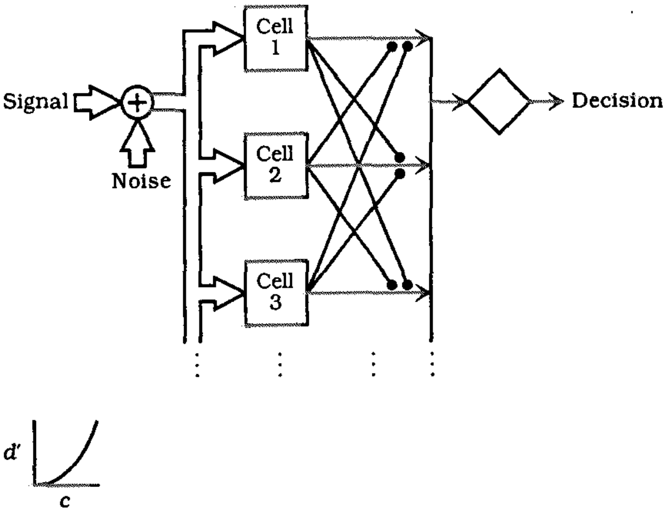

Fig. 11.6
A more physiologically plausible implemention uses mutually inhibitory interconnections to suppress all but the maximum response.

ator, we suppose that each cell sends an output to the final decision stage and that each output receives inhibition from all the other cells. We suppose that the inhibition is such that the output is suppressed unless the signal it is carrying exceeds that in each inhibitory connection. The decision stage will then receive only one signal, the maximum. There is substantial evidence that cortical cells receive input from far beyond their conventional receptive field (Gilbert & Wiesel, 1983, 1989), and are inhibited by a wider range of orientations than will excite them (Burr & Morrone, 1987; Morrone, Burr & Maffei, 1982; Ramoa, Shadlen, Skottun & Freeman, 1986). In the context of this model, that might implement a (nearly) ideal detector of a signal of unknown orientation.[3]

So don't be too quick to inject a late noise in your model. An early noise may be more consistent with what we know.

References

Ahumada, A. J., Jr. (1987). Putting the visual system noise back in the picture. *Journal of the Optical Society of America A, 4,* 2372–2378.

Ahumada, A. J., Jr. & Watson, A. B. (1985). Equivalent-noise model for contrast detection and discrimination. *Journal of the Optical Society of America A, 2,* 1133–1139.

Banks, M. S., Geisler, W. S. & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Research, 27,* 1915–1924.

Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *Journal of Physiology, 141,* 337–350.

Barlow, H. B. (1962). Measurements of the quantum efficiency of discrimination in human scotopic vision. *Journal of Physiology, 160,* 169–188.

Burr, D. C. & Morrone, M. C. (1987). Inhibitory interactions in the human vision system revealed in pattern-evoked potentials. *Journal of Physiology, 389,* 1–21.

Cohn, T. E. (1976). Quantum fluctuations limit foveal vision. *Vision Research, 16,* 573–579.

Gilbert, C. D. & Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *Journal of Neuroscience, 3,* 1116–1133.

---

[3] This physiological model is only meant to illustrate the availability of the kind of hardware needed to implement this kind of decision rule, not to advocate this particular realization. In fact, uncertainty in orientation would generate only a modest M, perhaps 8, not enough to account for the steepness of human psychometric functions. Furthermore, neurometric functions have been measured for cortical cells in the cat and monkey, and they too are much shallower than human psychometric functions (Tolhurst, Movshon & Dean, 1983). Relkin and Pelli (1987) found a similar discrepancy in psychometric/neurometric slopes for auditory thresholds.

Gilbert, C. D. & Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience, 9,* 2432–2442.

Graham, N. (1977). Visual detection of aperiodic spatial stimuli by probability summation among narrow band channels. *Vision Research, 17,* 637–652.

Green, D. M. & Swets, J. A. (1974). *Signal detection theory and psychophysics.* Huntington, NY: Krieger.

Jones, R. C. (1959). Quantum efficiency of human vision. *Journal of the Optical Society of America, 49,* 645–653.

Krauskopf, J. & Reeves, A. (1980). Measurement of the effect of photon noise on detection. *Vision Research, 20,* 193–196.

Lasley, D. J. & Cohn, T. E. (1981). Why luminance discrimination may be better than detection. *Vision Research, 21,* 273–278.

Morrone, M. C., Burr, D. C. & Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proceedings of the Royal Society of London [Biology], 216,* 335–354.

Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision Research, 21,* 215–223.

Nachmias, J. & Sansbury, R. V. (1974). Grating contrast: Discrimination may be better than detection. *Vision Research, 14,* 1039–1042.

Nolte, L. W. & Jaarsma, D. (1967). More on the detection of one of *M* orthogonal signals. *Journal of the Acoustical Society of America, 41,* 497–505.

Pelli, D. G. (1981). *Effects of visual noise.* Ph. D. thesis. Cambridge University, Cambridge, England.

Pelli, D. G. (1983). The spatiotemporal spectrum of the equivalent noise of human vision. *Investigative Ophthalmology and Visual Science (Supplement), 4,* 46.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, 2,* 1508–1532.

Pelli, D. G. (1990). The quantum efficiency of human vision. In C. Blakemore (Ed.), *Vision: Coding and efficiency.* Cambridge, England: Cambridge University Press.

Peterson, W. W., Birdsall, T. G. & Fox, W. C. (1954). Theory of signal detectability. *Transactions of the IRE PGIT, 4,* 171–212.

Ramoa, A. S., Shadlen, M., Skottun, B. C. & Freeman, R. D. (1986). A comparison of inhibition in orientation and spatial frequency selectivity of cat visual cortex. *Nature, 321,* 237–239.

Relkin, E. M. & Pelli, D. G. (1987). Probe tone thresholds in the auditory nerve measured by two-interval forced-choice procedures. *Journal of the Acoustical Society of America, 82,* 1679–1691.

Rose, A. (1942). The relative sensitivities of television pickup tubes, photographic film, and the human eye. *Proceedings of the IRE, 30,* 293–300.

Rose, A. (1948). The sensitivity performance of the human eye on an absolute scale. *Journal of the Optical Society of America, 38,* 196–208.

Tanner, W. P., Jr. (1961). Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences, 89,* 752–765.

Tanner, W. P., Jr. & Birdsall, T. G. (1958). Definitions of d' and η as psychophysical measures. *Journal of the Acoustical Society of America, 30,* 922–928.

Tolhurst, D. J., Movshon, J. A. & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research, 23,* 775–785.