Title: Characterizing the Effects of Stimulus and Neural Variability on Perceptual Performance

Authors and Affiliations:

Wilson S. Geisler
Center for Perceptual Systems and Department of Psychology
University of Texas at Austin

Johannes Burge
Center for Perceptual Systems
University of Texas at Austin

Melchi M. Michel
Department of Psychology
Rutgers University

Anthony D. D'Antona
Center for Perceptual Systems
University of Texas at Austin

Acknowledgments: Supported by NIH grant EY11747, NSF Grant IIS 1111328 and NIH Training Grant 1T32-EY021462.

In: M.S. Gazzaniga (Ed.) *The Cognitive Neurosciences* V. Cambridge: MIT Press (2014).

Abstract

Perceptual performance is limited by both external and internal factors. External factors include the physical variability of sensory stimuli and the inherent ambiguities that exist in the mapping between the properties of the environment and the properties of stimuli at the sensory organs (natural scene statistics). Internal factors include neural noise and non-random computational inefficiencies. External factors have not received the study they deserve, perhaps because they are difficult to measure and because methods for characterizing them have not been standardized. This chapter describes some of the computational tools used to characterize the effects of external and internal variability on perceptual performance. These tools are based on concepts of Bayesian statistical decision theory, and are illustrated for several basic natural tasks: grouping of contours across occlusions, estimation of binocular disparity, and interpolation of missing pixel luminance values.

Introduction

Evolution pushes sensory and perceptual systems to perform efficiently in those tasks necessary for the organism to survive and reproduce. Nonetheless, even in an organism's natural tasks, perceptual performance can never be perfect. Thus, to understand and predict perceptual performance it is crucial to characterize and understand the many factors that limit performance. These factors include the complexity and variability of the sensory stimuli, as well as many sources of internal variability, ranging from noise in sensory receptor responses, to noise in decision and memory circuits, to noise in motor neuron responses. The aim of this chapter is to describe some of the computational tools used to characterize and understand the effects of external (stimulus) and internal sources of variability on perceptual performance. These tools are based on principles of statistical decision and estimation theory. The computational tools described here are applicable to many perceptual systems, but the examples here are drawn from the vision literature.

The most basic kinds of stimulus variability are irreducible sources of noise that occur in transmission of stimulus information from the environment to the sensory organs. For example, the quantum nature of light causes the number of photopigment molecules activated in a photoreceptor to vary according to the Poisson probability distribution, even when the stimulus is nominally the same (Hecht, Shlaer & Pirenne 1942; Rose 1948; De Vries 1943). Although this source of noise is ubiquitous, there are only a few situations where it is the primary factor limiting performance. These situations consist primarily of simple detection or discrimination tasks where brief, spatially-localized targets are presented in the visual periphery under dark-adapted conditions when the rod photoreceptors are most sensitive (Hecht et al. 1942).

In some laboratory tasks it is possible to avoid all sources of stimulus noise, other than irreducible sources such as photon noise. In such tasks, performance is usually dominated by neural variability, and by limitations in neural computations. Examples of such cases would be simple detection or discrimination tasks with fixed stimuli presented under light-adapted conditions (e.g., detection of a known pattern on a uniform gray background). In other laboratory tasks, and in most natural tasks, additional sources of stimulus variability are also major factors. Examples, of such cases would be detection of targets in pixel-noise backgrounds (Burgess, Jennings, Wagner & Barlow 1981), or estimation of physical properties in the environment such as the depth, shape and reflectance of object surfaces (e.g., see Kersten, Mamassian & Yuille 2004; Geisler 2011).

Most perceptual tasks can be regarded as decision making in the presence of random variability, and hence an appropriate theoretical framework for analyzing perceptual performance is Bayesian statistical decision theory (e.g., see Knill & Richards 1996; Kersten et al. 2004; Geisler 2011). In what follows, we sketch the general Bayesian framework and then discuss several special cases, starting with a discussion of simple detection and discrimination tasks and ending with discussion of optimal estimation in natural scenes.

It is important to note that Bayesian statistical decision theory is used to analyze perceptual performance in two different ways. The first is to derive ideal observer models, which are theoretical devices that perform a perceptual task optimally. An ideal observer usually contains no free parameters and is not meant to be a model of real observers. Rather, its purpose is (i) to help identify task-relevant stimulus properties, (ii) to describe how those properties should be used to perform the task of interest, (iii) to provide a rigorous benchmark against which to compare real perceptual systems, and (iv) to suggest principled hypotheses and models for real performance.

The second way Bayesian statistical decision theory is used is as a framework for modeling perceptual performance. When used in modeling perception, there are generally hypothesized internal (neural or information-processing) mechanisms, which have unknown parameters that are estimated from perceptual performance data.

Bayesian Statistical Decision Theory

Specifying the task. The first step in using Bayesian statistical decision theory is to specify the task. This includes specifying the set of possible stimuli, the set of possible responses, and the goal of the task. Specifying the set of possible stimuli typically requires specifying (i) the ground-truth (distal) stimuli, which reflect the true task-relevant state of the world $\boldsymbol{\omega}$, and (ii) the proximal stimuli, which constitute the input data $\boldsymbol{s}$.[1] For example, in a simple detection-in-noise task, the true state of the world is that a target is either absent $(\omega = a)$ or present $(\omega = b)$ in a noise pattern, and the proximal stimulus is the specific pattern of pixels that would be imaged on the retina given perfect optics. In a typical depth estimation task, the true state of the world is the physical distance of one surface patch from another and the proximal stimulus is the specific pattern of pixels imaged on the two retinas, given perfect optics.

The set of possible responses can be quite complex, but in most perception experiments it is simple. For example, in the detection-in-noise task it would be one of two responses that indicate whether the

[1] In this chapter bold symbols represent multiple-dimensional variables and normal symbols represent single-dimensional variables.

observer judged the target to be absent $(r = a)$ or present $(r = b)$. In the depth estimation task, the response might be an estimate of the number of centimeters in depth separating the surface patches.

Specifying the goal of a task requires specifying the costs and benefits (utility) of each possible response for each possible state of the world: $\gamma(\boldsymbol{r}, \boldsymbol{\omega})$. If the goal in the detection-in-noise task is to be as accurate as possible, then that can be represented by making the utility a positive value *u* when the response is correct $(\gamma(a, a) = \gamma(b, b) = u)$, and –*u* when the response is incorrect $(\gamma(a, b) = \gamma(b, a) = -u)$. As another example, if the goal is to maximize accuracy, while keeping false-positive responses (saying an absent target is present) at some low rate, then that can be represented by assigning a greater cost to false-positive than false-negative responses (i.e., making $\gamma(b, a) < \gamma(a, b)$). In the depth estimation task there are many more possible combinations of response and state of the world, and hence many more possible goals (utility functions). A typical goal would be to minimize the mean squared error, which would be obtained by setting $\gamma(r, \omega) = -(r - \omega)^2$. In an ideal observer model, the utility function (goal) is fully specified. In a perceptual performance model, the utility function is a part of the model and may have free parameters.

Ground truth and input stimuli. The second step in using Bayesian statistical decision theory is to specify the statistical relationship between the states of the world and the stimulus. In the most common case, this involves specifying the conditional probability of the different ground-truth states of the world given the input stimuli; this is the posterior probability distribution, $p(\boldsymbol{\omega}|\boldsymbol{s})$. In practice, it is often convenient to first specify the stimulus likelihood distribution $p(\boldsymbol{s}|\boldsymbol{\omega})$ for each possible state of the world, and the prior probability distribution $p(\boldsymbol{\omega})$, and then use Bayes' rule to compute the posterior probability distribution.

Input data. In many applications of Bayesian statistical decision theory there are properties of the perceptual system that are part of the specification of the input to the optimal Bayesian computations. These properties could be either known physical or neural properties that have no free parameters, or models of these properties that have free parameters. For example, these properties might include the optics of the eye, the sampling pattern of the photoreceptors, or the tuning and noise characteristics of retinal ganglion cells. The properties can be represented by a function $\mathbf{g}_{\boldsymbol{\theta}}$ that maps the input stimulus $\mathbf{s}$ into input data $\mathbf{z}$:

$$\mathbf{z} = \mathbf{g}_{\boldsymbol{\theta}}(\mathbf{s}) \tag{1}$$

where $\boldsymbol{\theta}$ represents any free parameters. In other words, this constraint function incorporates the effects of known or assumed properties and its output is the input data to the Bayesian analysis.

Bayes optimal response. Once the task, input data, and posterior distributions are specified, it is possible to write down an expression for the optimal response. Namely, one should pick the response that maximizes the utility (minimizes cost), averaged over the posterior probability of the possible states of the world given the input data:[2]

$$\mathbf{r}_{opt}(\mathbf{z}) = \arg\max_{\mathbf{r}} \left[\sum_{\boldsymbol{\omega}} \gamma(\mathbf{r}, \boldsymbol{\omega})\, p(\boldsymbol{\omega}|\mathbf{z}) \right] \tag{2}$$

Note that $\boldsymbol{z}$ reduces to $\boldsymbol{s}$**,** in the case where the input data are the input stimuli. We define the "ideal observer" for a given task and constraint function, to be the observer that makes responses according to equation (2).

In what follows, we first consider identification tasks (of which detection and discrimination are special cases), then estimation tasks, and finally make some general points about the relative importance of external and internal factors.

Identification Tasks

In an identification task, the observer is required to identify which of *n* possible stimulus categories was presented on a trial. The special cases where there are only two possible categories of stimuli are usually referred to as *detection* or *discrimination* tasks.

In the classic yes-no task, the observer is presented on each trial with one of two randomly chosen stimuli (a or b). The observer is required to report whether the stimulus was a or b. Here we will regard a as the reference stimulus and b as the reference plus signal. The typical goal is to maximize accuracy. In another variant, there are monetary costs and benefits associated with the different stimulus-response outcomes and the goal is to maximize monetary gain. In the two-alternative forced choice (2AFC) task, the observer is presented on each trial both stimulus a and b, either in two temporal intervals or two spatial locations. The temporal or spatial order is randomized, and the observer is required to report whether stimulus a or b was in the first location or interval. Although the yes-no task is more representative of real-world tasks, the 2AFC task is more common in laboratory experiments, because performance tends to be better and response biases smaller than in the yes-no task.

Signal detection theory. Signal detection theory is a special case of Bayesian statistical decision theory that was developed to interpret the behavioral data in detection and discrimination experiments (Tanner & Swets 1954; Green & Swets 1966). The first key assumption of signal detection theory is that on each trial the perceptual system produces a response that is represented by a value of a decision variable ψ. On trials where the stimulus is a, the random values of ψ are described by one probability distribution $p(\psi|a)$, whereas on trials where the stimulus is b, the random values of ψ are described by another probability distribution $p(\psi|b)$, (see Fig. 1A). The second key assumption is that the observer's responses are selected by placing a criterion β along the decision variable axis; if the value of ψ exceeds

[2] Note that arg max $[f(x)]$ is the value of x (the argument) for which $f(x)$ reaches its maximum value.

the criterion then the response is b, if it falls below the criterion the response is a. Two potential criteria are shown in Fig. 1A (i.e., the solid and dashed vertical lines).

There are four possible stimulus-response outcomes in the yes-no task: responding b when the stimulus is b (hit), responding b when the stimulus is a (false alarm), responding a when the stimulus is a (correct rejection) and responding a when the stimulus is b (miss). The proportions of trials that are hits and misses must sum to 1.0, and proportions that are false alarms and correct rejections must sum to 1.0; thus, the data can be summarized by the proportions of hits and false alarms. As is clear from Fig. 1A, these two stimulus-response outcomes are interpreted in signal detection theory as the areas under the two probability distributions to the right of the criterion. The number of standard deviations separating the means of the two distributions represents the observer's sensitivity, and is called d' (d-prime). The bigger the value of d', the greater the potential accuracy of the observer; however, actual performance will also depend on where the criterion is placed.

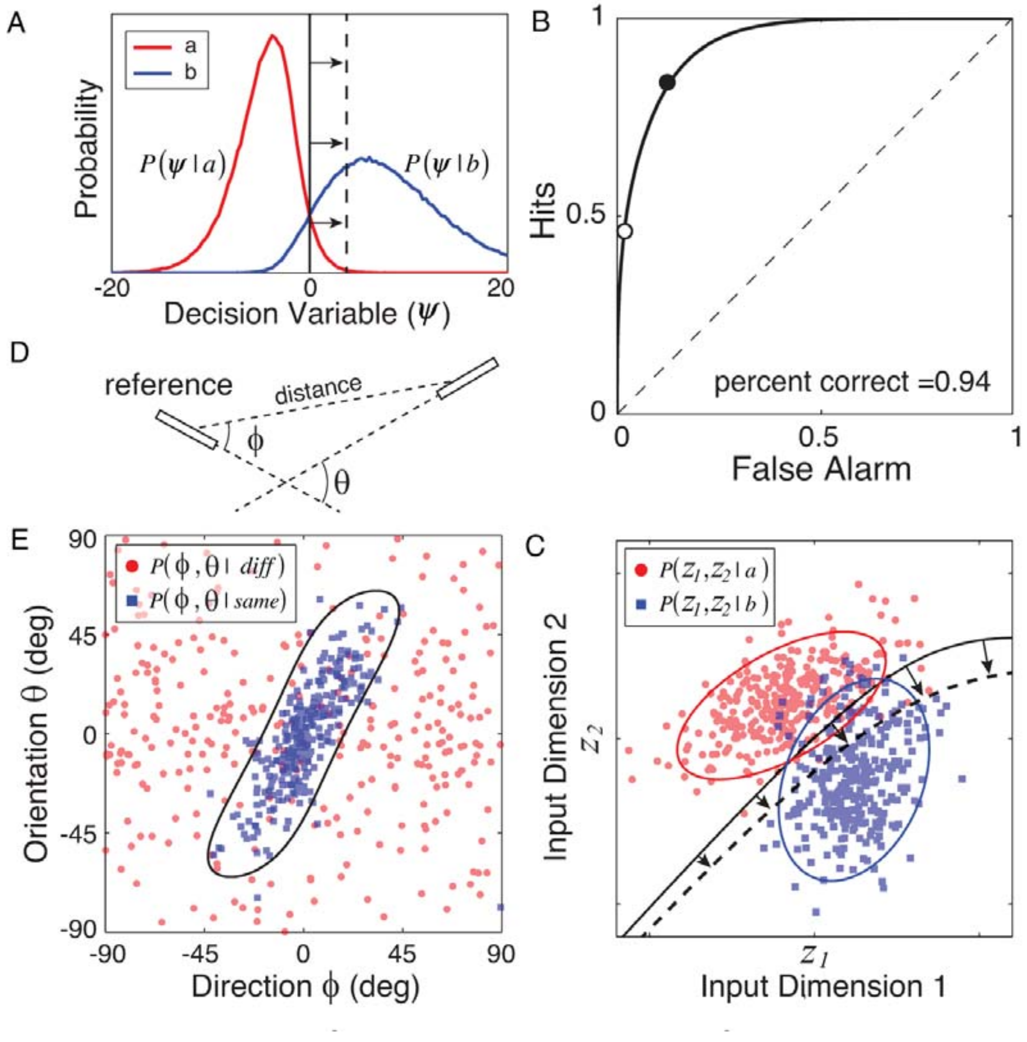

Figure 1. Detection and Discrimination. **A**. The assumptions of signal detection theory (SDT). **B**. The receiver operating characteristic (ROC) for the distributions in A. **C**. Samples from hypothetical Gaussian likelihood

distributions for the two stimulus categories in a yes-no detection task. The log of the ratio of these two likelihood distributions at any point is the decision variable in A (i.e., $\Psi = log[p(z_1, z_2|b)/p(z_1, z_2|a)]$). The solid and dashed contours correspond to the two criteria in A. **D**. Three parameters describing the possible geometrical relationships between two contour elements. **E**. Samples of geometrical relationships for pairs of contour elements belonging to the same physical contour (blue symbols) and belonging to different physical contours (red symbols), for a particular distance between the elements. The black curve shows the optimal decision bound given equal prior probabilities. (Data from Geisler & Perry 2009.)

An important feature of signal detection theory is that it allows estimation of both the sensitivity and the criterion from the proportion of hits and false alarms. For example, if the probability distributions are assumed to be Gaussian and of equal variance, then $d' = \Phi^{-1}(p_h) - \Phi^{-1}(p_{fa})$ and $\beta = \Phi^{-1}(1 - p_{fa}) - d'/2$, where Φ is the cumulative standard normal integral function. These formulas are useful because observers can differ in task performance due to differences in the criterion, even when they are equally sensitive, or vice versa. Once the value of d' is determined it is also possible to calculate what would be the performance of the observer if the decision criterion were placed at some other location (e.g., the optimal location) using the formulas: $p_h = \Phi(d'/2 - \beta)$ and $p_{fa} = \Phi(-d'/2 - \beta)$.

The same logic and equations above hold for the 2AFC task; however, the colorful names hits and false alarms are reserved for the yes-no task. Also, if the responses to the two stimuli are statistically independent, then signal detection theory predicts d' in the 2AFC task to be $\sqrt{2}$ larger than in the yes-no task, for the same stimuli (Green & Swets 1966).

ROC analysis. The effect on performance of changes in the decision criterion, for a given level of sensitivity, can be represented with the receiver operating characteristic (ROC), which plots the proportion of hits as a function of the proportion of false alarms (Fig. 1B). Note that in this plot the value of the decision criterion is implicit; the ROC shows only the locus of hit and false alarm rates for all values of the criterion. The shape of the ROC depends on the shapes of the probability distributions; the ROCs plotted in Fig. 1B are for the distributions in Fig. 1A. The solid and open symbols in Fig. 1B show points on the ROC curve corresponding to the solid and dashed criteria in Fig. 1A, respectively.

The assumptions of signal detection theory can be tested in part by inducing the observer to adopt different decision criteria and then seeing if the observer's hit and false alarm rates fall on the ROC predicted for the measured value of d'. Typically, different decision criteria are induced by either varying the relative probability of stimulus a and b, varying the monetary payoffs for the different kinds of correct and error responses, varying the instructions to observer, or asking the observer to provide a confidence judgment along with each response (Green & Swets 1966). Signal detection theory predicts that in the 2AFC task the ROC curve should be symmetric about the negative diagonal, even when it is not predicted to be symmetric in the yes-no task (e.g., see Green & Swets 1966).

If the probability distributions for the decision variable cross at just one point (as they do for the equal variance Gaussian distributions), and if the two stimuli are equally probable, then the area under the ROC is the maximum percent correct, which is obtained when the criterion is placed where the probability distributions cross (Green & Swets, 1966). In neurophysiology experiments the area under the ROC is frequently used to quantify the "discrimination information" transmitted by individual neurons (e.g., Tolhurst, Movshon & Dean 1983; Britten, Shadlen, Newsome & Movshon 1992). The typical procedure is to record the response of a neuron to multiple presentations of two stimuli, compute the hit and false alarm rate for each value of a criterion along the response axis, and finally compute the area under the resulting ROC. This calculation provides an estimate of the maximum percent correct that could be supported by that neuron alone in a yes-no task, assuming that the all the relevant discrimination information is contained in the total spikes per trial (i.e., no information is in the temporal pattern of responses).

For an ideal observer in a yes-no task, the probability distributions for the optimal decision variable must cross at a single location (see later), and hence the area under the ROC is the maximum percent correct. However, in neurophysiology experiments the neural response (e.g., spike count or spike rate) is typically regarded as the decision variable. The neural response is not guaranteed to represent an optimal decision variable (i.e., a likelihood ratio; see later). Hence, the two probability distributions could cross at more than one location. In this case, the area under the ROC curve will not correspond to the maximum percent correct. A better procedure is to fit the ROC, assuming an appropriate family of distributions (e.g., gamma distributions), compute percent correct, and then do a statistical analysis, for the family of distributions, to correct for bias in the accuracy estimates.

Ideal observer for identification. If the goal in an identification task is to maximize the percentage of correct identifications, then equation (2) reduces to the maximum *a posteriori* (MAP) rule:

$$r_{opt} = \arg\max_{a_i} \left[p\left(\mathbf{z} \middle| a_i\right) p\left(a_i\right) \right] \qquad (3)$$

Further, in the case of just two categories of stimuli ($a_1 = a, a_2 = b$), equation (2) reduces to

$$\text{respond } b \text{ if } \frac{p\left(\mathbf{z}|b\right)}{p\left(\mathbf{z}|a\right)} > \frac{p(a)}{p(b)}\text{; otherwise, respond } a \qquad (4)$$

The left side of the inequality is the likelihood ratio and the right side is the prior-probability ratio (often called the "prior odds"). Figure 1C illustrates this decision rule for an example where the input data are two-dimensional, $\mathbf{z} = (z_1, z_2)$, and Gaussian, with different means and covariance matrices. The ellipses show iso-likelihood contours, and the symbols show random samples from the two Gaussians. Equation (4) says that if the two categories occur with equal probability, then the response should be b when the likelihood ratio exceeds 1.0. The solid black curve shows the locus of points where the likelihood ratio equals 1.0, and thus any input $\mathbf{z}$ below and to the right of the solid curve should be assigned response b. If the ratio of the priors is greater than 1.0, then the decision boundary should shift. For example, the dashed curve shows the locus of points where the likelihood ratio is 44. If the ratio of the priors is 44, then any input $\mathbf{z}$ below and to the right of the dashed curve should be assigned response b. The same logic applies to input data of arbitrary dimensionality and to arbitrary likelihood distributions. It also applies to arbitrary numbers of categories, except that the boundaries now define regions for each of the possible responses.

How is this analysis of optimal identification related to signal detection theory? To see the connection, note that the ideal decision rule is the likelihood ratio (the left side of equation 4). Furthermore, note that the specific decision is unchanged by any strictly monotonic transformation of the two sides of the inequality in equation (4). In other words, for an ideal observer, the decision axis can be any monotonic transformation of the likelihood ratio. For Gaussian (and many other) distributions, a useful monotonic transformation is the logarithm (although any monotonic transformation is valid). Applying this transformation to both sides of equation (4) we obtain the decision variable $\psi = \log[p(\boldsymbol{z}|b)/p(\boldsymbol{z}|a)]$ and the criterion $\beta = \log[p(a)/p(b)]$. When the stimulus on a trial is from category a, then the decision variable will have a distribution $p(\psi|a)$ (red curve in Fig. 1A). When the actual stimulus is b, then the decision variable will have a distribution $p(\psi|b)$ (blue curve in Fig. 1A). Note that for the ideal observer's decision variable, it is impossible for the two probability distributions on the decision axis to cross at more than one point.

We see then that signal detection theory is consistent with ideal observer theory. This fact provides one rationale for the assumptions of signal detection theory. However, signal detection theory is more general in that an observer's decision variable need not be a monotonic transformation of the likelihood ratio, and the criterion need not be optimally placed. In other words, signal detection theory assumes an arbitrary one-dimensional decision variable and criterion.

To obtain quantitative predictions for the ideal observer in an identification task, it is necessary to specify the prior probability of the different stimulus categories and the likelihood of the input data for each of the categories. Specifying the likelihoods and priors can be very difficult. The most common approach is to constrain the stimuli so that it is practical to derive or compute the likelihoods and priors. One way of constraining stimuli is to create them by sampling from probability distributions specified by the experimenter. This is what is done in many perception experiments. For example, stimulus a might be a sample of Gaussian noise and stimulus b a sample of Gaussian noise with a fixed added target. In this case, the decision variable and the criterion can be easily computed, making it straightforward to calculate or simulate ideal performance (e.g., Burgess et al. 1981).

Working with natural stimuli is more difficult because their statistical structure is complex and generally unknown. One approach is to restrict what aspects of the natural stimuli are presented in an experiment. By considering only certain aspects of natural stimuli, it can become practical to measure the relevant probability distributions and compute ideal observer performance. For example, consider a task where the observer is presented with just two contour elements (short line segments) at the boundary of an occluding surface, and must decide whether the elements belong to the same or different physical contours (Geisler & Perry 2009). For a given occluder width (distance), two parameters describe the geometrical relationship between two contour elements: the direction of one element from the other φ, and the orientation difference between to the two elements θ (see Fig. 1D). The blue symbols in Fig. 1E show samples from the actual distribution in natural images of direction and orientation difference when the contour elements belong to the same contour; the red symbols show samples when the elements belong to different contours. The solid curve shows the ideal decision bound when the goal is to maximize accuracy. Specifically, the ideal observer should report that the contour elements are on the same contour if the direction and orientation difference fall inside the boundary; otherwise, the observer should report that the elements are from different contours. The performance of this ideal observer can be determined by applying this decision rule to test stimuli that contain two contour elements (taken from natural scenes) separated by an occluder. These same test stimuli can be presented to human observers. In this case, human and ideal performance is nearly identical, implying that the human visual system accurately applies the decision boundary in Fig. 1E (Geisler & Perry 2009).

We reiterate that ideal observers are not meant to be models of real performance. Rather, they provide a rigorous benchmark against which to compare real performance and a principled starting point for developing models for real performance. For example, the classic signal detection model for interpreting performance in detection and discrimination experiments is motivated by the computational principles of the ideal observer. Also, there are many examples in the perception literature where human performance is found to parallel that of the ideal observer, showing that modest modifications of ideal observers (or heuristic approximations to the ideal observer) can serve as plausible and testable models real performance (for review see Geisler 2011).

Estimation Tasks

In estimation tasks, there is some physically ordered dimension along which stimuli fall and the observer is required to estimate the value along that dimension. The distinction between estimation and identification is not a sharp one, because one can regard estimation as identification with a large number of categories. The primary distinction is captured in the utility (cost/benefit) function. Generally, in the estimation task, the closer the estimate to the true value, the better. On the other hand, in many identification tasks all errors are equally costly; for example, if the task is to identify a criminal from a police lineup of otherwise innocent people, then all errors would be equally bad. Estimation tasks are very common under natural conditions, but in laboratory settings they are less common than identification tasks.

The typical method for measuring estimation performance is similar to that for measuring discrimination performance. On each trial a variable test stimulus is presented and the observer is required to respond whether it is greater or less than some standard along the stimulus dimension of interest (e.g., color, depth, size, shape, etc.). Often the standard is another stimulus, but in some tasks the standard may be an internal reference. For example, in a slant estimation task, the observer may be required to respond whether a test stimulus is right-side-back from frontoparallel (an internal standard). Data are typically plotted as psychometric functions, and the estimate is taken to be the point of subjective equality (PSE)—the value of the variable stimulus where the observer reports that the test is greater than the standard with probability 0.5.

Ideal observer for estimation. A typical goal in an estimation task is to minimize the mean squared error between the estimate and the true value. This is the MMSE estimate given by

$$\hat{\omega}_{opt} = \arg\min_{\hat{\omega}} \left[\sum_{\omega} \left(\omega - \hat{\omega}\right)^2 p\left(\omega \middle| \mathbf{z}\right) \right] = \sum_{\boldsymbol{\omega}} \omega p\left(\omega \middle| \mathbf{z}\right) = E\left(\omega \middle| \mathbf{z}\right) \tag{5}$$

In other words, the optimal estimate is simply the mean of the posterior probability distribution. Using Bayes rule to expand $p\left(\omega \middle| \mathbf{z}\right)$ in equation 5, the optimal estimate can also be expressed in terms of the likelihood and prior probability distributions:

$$\hat{\omega}_{opt} = \sum_{\boldsymbol{\omega}} \omega \frac{p\left(\mathbf{z} \middle| \omega\right) p\left(\omega\right)}{\sum_{\mathbf{z}} p\left(\mathbf{z} \middle| \omega\right) p\left(\omega\right)} \tag{6}$$

Although minimizing some measure of the deviation from the true value is the intuitive goal for most estimation tasks, it is not uncommon for researchers to consider the MAP estimate (which penalizes all errors equally; c.f., equation 3):

$$\hat{\omega}_{opt} = \arg\max_{\omega} p\left(\omega \middle| \mathbf{z}\right) = \arg\max_{\omega} \left[p\left(\mathbf{z} \middle| \omega\right) p\left(\omega\right) \right] \tag{7}$$

The reason for this choice is that sometimes the MAP estimate is easier to compute, and if the posterior distributions are unimodal and not skewed, then the MAP and MMSE estimates are the same.

As in identification experiments, the difficult step in generating ideal observer predictions is specifying the likelihood and prior distributions (or equivalently the posterior distributions). For both laboratory and natural stimuli this generally requires constraining the stimuli in some way. We briefly describe two approaches that can be applied to tasks with natural stimuli.

Both approaches begin by constraining the amount of data in the input. To be concrete, suppose that the task is to estimate some state of the world (e.g., depth) at each location in the retinal image. The input ***z*** to the visual system is the entire image, which for natural stimuli is far too big and complex to allow specification of the likelihood or posterior distributions. Thus, it is typical to restrict the input to some small neighborhood or context, ***c***, over the location where the estimate is to be made. This is reasonable because, in many cases, image correlations drop rapidly with distance (e.g., Deriugin 1956; Field 1987). Also, in experiments on real observers it may be possible to use stimuli restricted to the local context so that the ideal observer is appropriate for the stimuli tested on real observers.

The first approach is to make an assumption about the parametric form of the likelihood distributions. A common (but sometimes unverified) assumption is that the likelihood distributions are Gaussian,

$$p\left(\mathbf{c} \middle| \omega\right) = gauss\left(\mathbf{c}; \boldsymbol{\mu}_{\omega}, \boldsymbol{\Sigma}_{\omega}\right) \tag{8}$$

where $\boldsymbol{\mu}_\omega$ and $\boldsymbol{\Sigma}_\omega$ are a mean vector and covariance matrix that depend on the specific state of the world (e.g., depth). This assumption implies that the probability distribution of the context vector $p(\boldsymbol{c})$ is a mixture of Gaussian distributions with weights given by the prior probabilities, a form of Gaussian mixture model (GMM): $p(\boldsymbol{c}) = \sum_\omega p(\boldsymbol{c}|\omega)p(\omega)$. In applying this approach to natural stimuli, the mean vectors and covariance matrices can be measured (learned) from a large set of contexts taken from natural images, for each state of the world. Empirical measurements of natural stimuli also allow researchers to verify whether their assumptions about the parametric form of the likelihood are valid (e.g., Burge & Geisler 2011; 2014). If the size of the context vector is n, then the number of parameters that must be estimated for each possible state of the world is $n\,(n+1)/2 + n$. This number is small enough to make it practical to measure all parameters for moderate context sizes. Once the means and covariance matrices are measured, equations (6) or (7) can be used to compute the ideal observer's estimates.

The second approach makes no assumptions about the parametric form of the likelihood or prior distributions, but instead makes the analysis tractable by considering only small context sizes (e.g., Geisler & Perry 2011). One version of this approach, parallel conditional means (PCM), involves measuring separately the mean of the posterior probability distribution for all context values for two or more contexts in the input data (gray squares in Fig. 2A). These means can be measured directly by computing sample means from training data, and do not require measuring (or modeling) the posterior probability distributions. These means specify estimation functions that map context values into optimal (MMSE) estimates: $\hat{\omega}_1 = E(\omega|\boldsymbol{c}_1)$, $\hat{\omega}_2 = E(\omega|\boldsymbol{c}_2)$. Once these functions are measured the final estimate is obtained by combining the estimates, typically by weighting the estimates by their relative reliability (e.g., Oruc, Maloney & Landy 2003). Another version of this approach, recursive conditional means (RCM), involves first measuring the mean of the posterior probability distribution for one context $\boldsymbol{c}_1$ in the input data $\mathbf{z}$ (gray squares in the first box in Fig. 2B). Again the optimal estimate is $\hat{\omega}_1 = E(\omega|\boldsymbol{c}_1)$. The recursive step is to define a second-level context $\boldsymbol{c}_2$ that includes one or more values of the first-level estimates (gray squares in third box in Fig. 2B), and then directly measure the mean of the posterior distribution for all possible values of the variables in this second-level context. The result is a second estimation function that maps second-level context values into optimal estimates: $\hat{\omega}_2 = E(\omega|\boldsymbol{c}_2)$. This process can be repeated to obtain a series of n estimation functions; the value of n is determined by when performance reaches asymptote. The final estimate is obtained by applying the n estimation functions sequentially.

Which version performs best depends on the particular task. Both versions require having enough training data to estimate the mean of ω for each possible pattern of context values. For example, if the values of the context variables range from 0 to 255, then the context size is limited to three or four variables, because more variables would require an impractical amount of training data. However, the context size does grow (in effect) at each step. In the recursive case, the effective context grows because the context for a higher level estimate contains estimates that were obtained using the contexts at lower levels (in Fig. 2B, the context for the second estimate can effectively include the light gray pixels). Note that it is also possible to apply Gaussian mixture models recursively.

Another distinction between these two approaches is that the GMM observer is “generative”, in the sense that the GMM parameters specify the joint distribution of the context and true values. (The term “generative” refers to the fact that if the joint distribution is specified, then it is possible to generate random samples from the distribution.) On the other hand, the RCM observer is “discriminative”, in the sense that it provides optimal estimates, but does not specify the joint distribution of the context and true values (McLachlan 1992; Vapnik 1998). The distinction between generative and discriminative is separate from the distinction between parametric and non-parametric. For example, direct non-parametric measurements of higher-order moments beyond the conditional mean may allow generation of random samples from the joint distribution. Alternatively, parametric models such as multiple linear regression produce estimates, but cannot generate random samples from the joint distribution. In perceptual systems, a potential advantage of representing the joint distribution is that it may be possible to switch utility/cost functions without needing to learn whole new estimation functions.

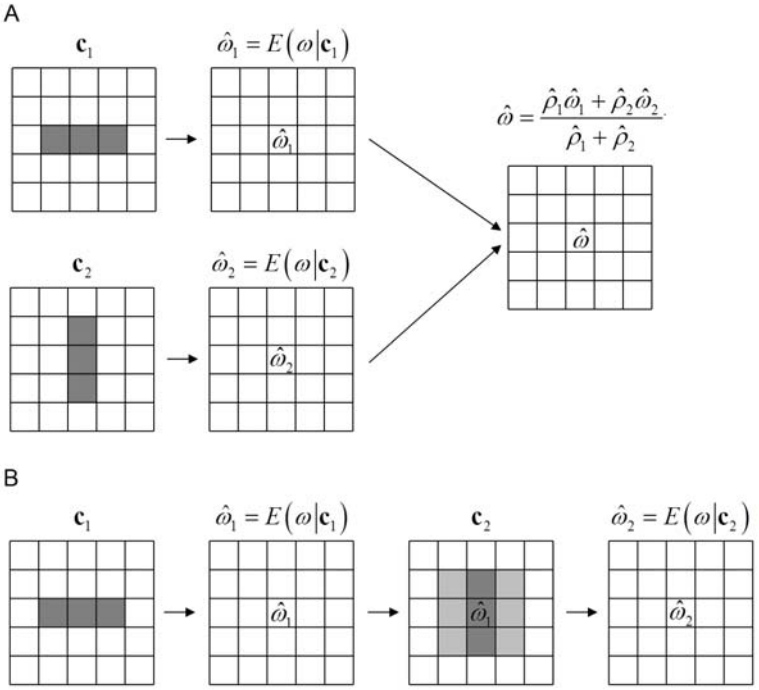
A
$\mathbf{c}_1$
$\hat{\omega}_1 = E(\omega|\mathbf{c}_1)$
$\hat{\omega}_1$
$\hat{\omega} = \frac{\hat{\rho}_1\hat{\omega}_1 + \hat{\rho}_2\hat{\omega}_2}{\hat{\rho}_1 + \hat{\rho}_2}$
$\hat{\omega}$
$\mathbf{c}_2$
$\hat{\omega}_2 = E(\omega|\mathbf{c}_2)$
$\hat{\omega}_2$
B
$\mathbf{c}_1$
$\hat{\omega}_1 = E(\omega|\mathbf{c}_1)$
$\hat{\omega}_1$
$\mathbf{c}_2$
$\hat{\omega}_1$
$\hat{\omega}_2 = E(\omega|\mathbf{c}_2)$
$\hat{\omega}_2$

Figure. 2. Simple non-parametric minimum mean squared error (MMSE) estimates. **A**. Parallel conditional means (PCM) **B**. Recursive conditional means (RCM).

We illustrate the two approaches with two examples: (i) disparity estimation, which underlies binocular depth perception, and (ii) missing pixel estimation, which is a simple form of image interpolation (amodal completion).

Disparity estimation. To illustrate the first approach, consider the task of estimating horizontal disparity from the images formed in the left and right eye when binocularly viewing a small patch of natural scene. In this example, the context consists of eight variables, where each variable is the dot product of a different vertically oriented binocular receptive field with the retinal images in the two eyes. These eight receptive fields were found (by a separate analysis) to be the most useful vertical receptive fields for disparity estimation given the optics of human eyes and the properties of natural stereo images (Burge & Geisler, 2014). The symbols in Fig. 3B show joint responses of the first two binocular units (Fig. 3A) to randomly selected contrast-normalized natural image patches, for a range of horizontal disparities (-15 to 15 minutes of arc). As can be seen, the likelihood distributions are roughly Gaussian in shape (solid curves are 95% volume contours) with mean vectors that change little with disparity and covariance matrices that change rather dramatically. This pattern holds for all pairs of variables, and for disparities intermediate to those shown in Fig. 3B, strongly suggesting that equations (7) and (8) should give near-ideal performance. Fig. 3C shows that the optimal estimates (for a separate set of test patches) are unbiased and that the confidence intervals grow with the magnitude of disparity.

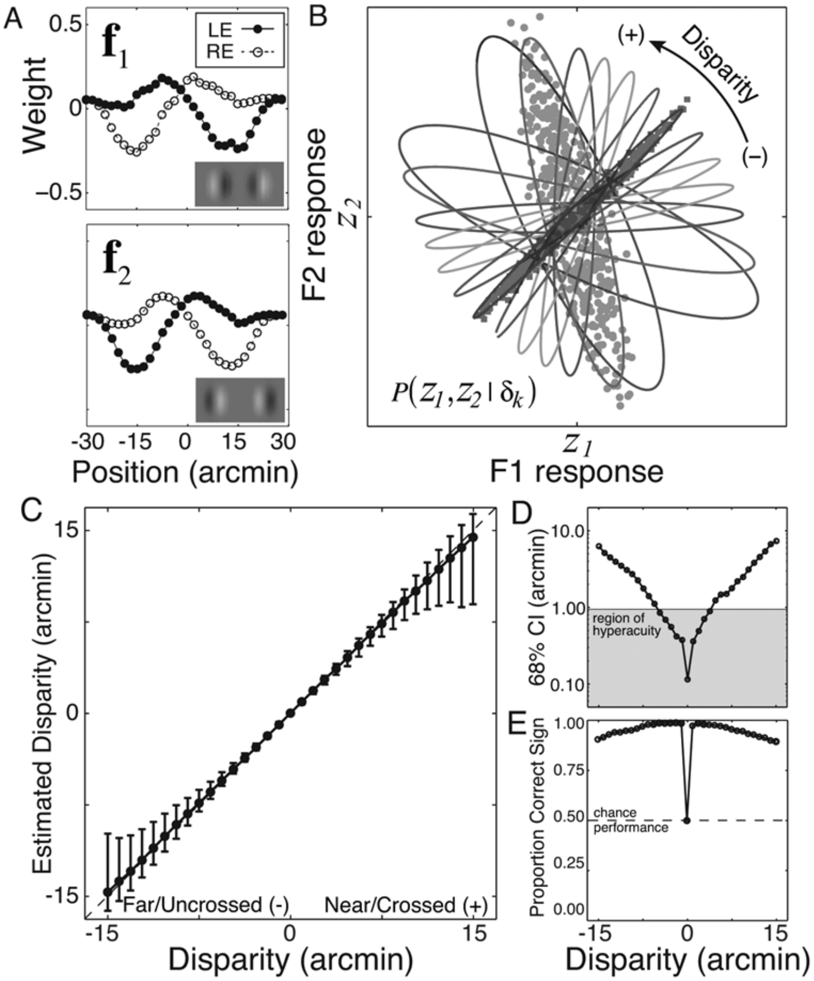

Figure 3. Disparity estimation. **A**. Two of the eight vertically oriented binocular receptive fields (filters) optimal for disparity estimation. **B**. Likelihood distributions of first two filter responses for disparities ranging from -15 min to 15 min. Symbols show responses to individual natural image patches for two disparities. Contours indicate 95% of the volume of Gaussian distributions fit to the joint responses. **C**. Estimation performance on random natural image test patches, for ideal Gaussian mixture model (GMM) using all eight optimal filter responses. Symbols are mean estimates and error bars represent 68% confidence intervals. **D**. Confidence intervals of estimates as

function of disparity. **E**. Proportion correct estimation of disparity sign (crossed vs. uncrossed) as a function of disparity. (From Burge & Geisler 2014.)

Fig. 3D and 3E show, in agreement with human psychophysics, that the growth in the confidence interval is approximately exponential with disparity (Blakemore 1970; McKee, Levi & Bowne 1990), and that the proportion of disparity sign confusions decreases rapidly at small disparities, is minimal at intermediate disparities and decreases gradually at large disparities (Landers & Cormack 1997). Thus, an ideal (GMM) observer for disparity estimation in natural images shows that human performance tracks the information available in the retinal images and provides a principled starting point for developing models of human disparity estimation under natural conditions.

Missing-pixel estimation. To illustrate the second approach, consider the task of estimating the gray level of missing pixels in 8-bit (0-255 gray level) calibrated natural images (D'Antona, Perry & Geisler 2013; task modified from Kersten 1987). The left side of Figure 4A shows a large patch of natural image with a missing center pixel; the right side shows an enlargement of the center 5x5 neighborhood around the missing pixel. PCM (Fig. 2A) was applied using two contexts: the four pixels left and right of the center pixel, and the four pixels above and below the center pixel. From a large set of natural image training patches (on the order 10^{10}), the mean of the center pixel is computed for each combination of the four context values. These conditional means (which are a smooth function of the context values) were used to obtain two estimates that were then combined.

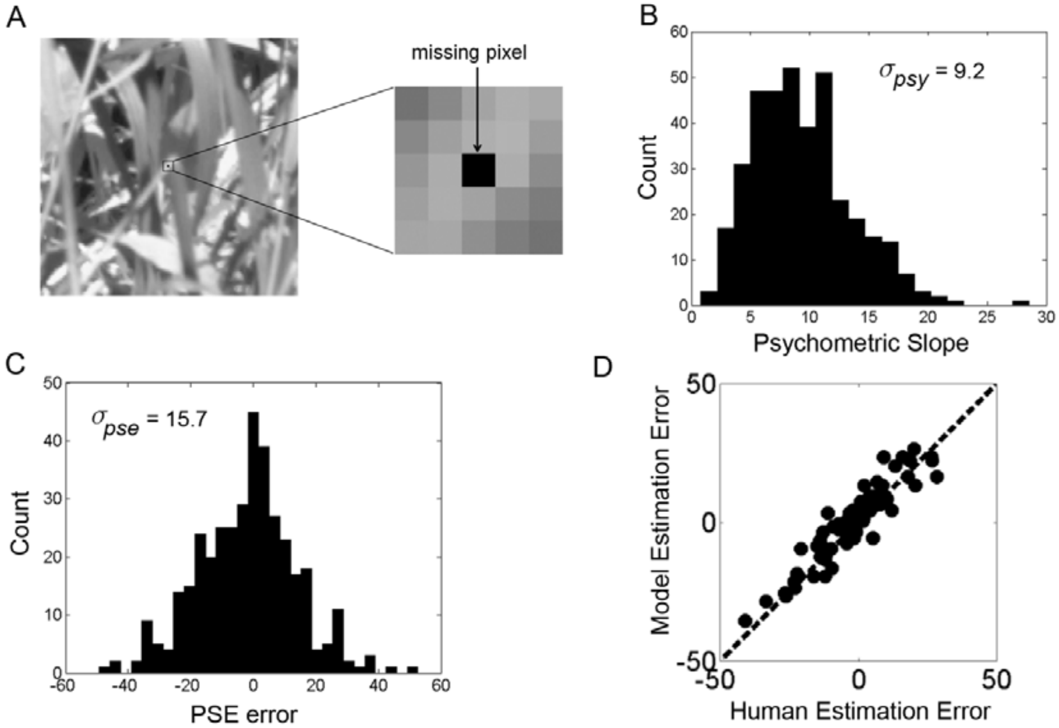

Figure 4. Missing pixel estimation task. **A**. Example test image (calibrated 8-bit gray scale). **B**. Distribution of estimation errors (in gray-level) of the point of subject equality (PSE) for 3 human observers on 62 test patches. **C**. Distribution of psychometric function slopes (standard deviation values) obtained by fitting psychometric data with

a cumulative Gaussian. **D**. Estimation of error of a PCM observer trained on natural images plotted as function the estimation error of three human observers, for 62 test patches. If PCM and human errors were identical the points would fall on the positive diagonal.

For test patches of natural image, the mean squared error of the estimates of the PCM observer is approximately 90 (SD of error = 9 gray steps). Analysis shows that the most useful pixels for this task are the four neighboring horizontal and vertical pixels (other pixels provide much less information), and hence the performance of the PCM observer is likely close to the true optimum (Geisler & Perry 2011). The mean squared error of the estimates of the GMM observer that uses simultaneously the four pixels in the horizontal direction and the four pixels in vertical directions (8-dimensional Gaussian distributions) is approximately 140. Thus, for this task the PCM observer is considerably closer to ideal.

The mean squared error of human estimates on exactly the same test patches is approximately 246 (Fig. 4C), and thus humans are well below optimal in this task. However, because of luminance gain control and center-surround mechanisms in the retina, the output of the retina is probably better described as a contrast image rather than a luminance image (a contrast image is obtained from a luminance image by subtracting and then dividing by the local mean luminance at each pixel location). Interestingly, humans match the performance of a PCM observer trained on contrast images. Figure 4D shows, for 62 representative test patches, the estimation error of the contrast PCM observer plotted against the human estimation error. The contrast PCM observer does a good job of predicting the specific errors made by humans for arbitrary natural image patches.

Relative Influence of External and Internal Factors

Efficiency. An observer's performance is generally limited by both external factors (variability and ambiguity of the inputs) and by internal factors (neural, decision, and motor noise, and non-random computational inefficiencies). For the purpose of estimating the relative influence of external and internal factors, the combined effect of all the internal factors can be regarded as a level of internal noise, which can be estimated by calculating how much the external (stimulus) variability must be scaled up for the performance of the ideal observer to match that of the organism. This scale factor κ is closely related to the definition of efficiency, η, in signal detection theory: $\eta = d'^2_{real}/d'^2_{ideal}$ (Tanner & Birdsall 1958). In a detection task with a fixed signal in Gaussian noise, κ is simply the inverse of the efficiency ($\kappa = 1/\eta$).

As a more general example, consider an ideal observer in an identification task, where each category is represented by a Gaussian distribution. In this case, κ is the scale factor on the covariance matrices that brings the ideal performance down to real performance. If the external variability must be scaled by a factor of κ, then the effective internal variability equals $\kappa - 1$ times the external variability. In other words, if the value of κ is near 1.0, then the internal noise is near zero and external factors dominate performance; if the value of κ is large, then internal factors dominate performance.

Whether external or internal factors dominate performance is highly task dependent. For detection of targets in fixed backgrounds, the only external variability is photon noise (the ideal observer is limited only by the signal's energy and photon noise) and the value of κ is large (typically greater than 10) showing that internal factors dominate (e.g., Geisler 1989). For detection in high contrast pixel noise the values of κ can be quite a bit smaller (sometimes less than 2), showing that external factors dominate or at least play a major role in limiting performance (e.g., Burgess et al. 1981).

For tasks involving natural stimuli, the variability of the stimuli is often high and hence there are likely to be many cases where external factors dominate. An example is the task described earlier where the observer is presented with two contour elements at the boundary of an occluding surface, and must decide whether the elements belong to the same or different physical contours (Fig. 1D,E). The accuracy of the ideal observer in this task is 87% correct, and is entirely due to external (stimulus) variability. Human performance under exactly the same conditions is 83% correct. The value of κ necessary to degrade ideal to real performance is 1.5, and hence in this task human performance is dominated by external factors. In other words, the human visual system uses a decision rule that closely approximates the solid curve in Fig. 1E and hence has efficiently incorporated the statistics of natural contours.

Fixed-stimulus and across-stimulus variation in performance. Another important distinction is between the variations in behavioral response that occur when an observer is presented with the same fixed stimulus repeatedly, and the variations that occur across different stimuli. Fixed-stimulus variation must be entirely due to internal factors that are varying from presentation to presentation (e.g., sensory neural noise, decision noise, motor noise). On the other hand, variation in response across stimuli must be due either to external factors or to non-random internal factors.

A method for separating the two types of variation in detection-in-noise tasks is the "frozen noise" experiment, where each noise background (or natural stimulus background) is repeated occasionally in the course of the experiment. Fitting the subject's responses with standard signal detection models allows estimation of the relative variance of the two sources of variation. Furthermore, if the pixel-based ideal observer for the task is known, then it is possible to separately estimate the effective variance due to external factors, non-random internal factors, and internal noise (e.g., see Swensson & Judy 1996).

A related simple analysis for estimation experiments is to measure the fixed-stimulus variance from the slopes of the psychometric functions and the across-stimulus variance from the differences between the PSEs and the true values. For example, Fig. 4B shows the distribution of psychometric function slopes in the pixel estimation task. These slope values are standard deviations of the cumulative Gaussian distributions fitted to the psychometric data. If the human observers had no internal variability, they would make the same decision every time the same stimulus was presented and the psychometric functions would be step functions. Thus, the fitted standard deviations estimate all the internal variability, which in this case is equivalent to a pixel noise standard deviation of 9.2 gray steps (variance = 85). On the other hand, Fig. 4C shows the distribution of systematic errors (PSE errors). These errors are largely due to external factors or fixed (non-random) internal factors (the confidence intervals on the PSEs are quite small). The root mean squared PSE error is 15.7 gray steps (variance = 246), which is substantially larger than the internal variability. This result makes the important point that in many natural tasks performance is limited more by external factors and non-random internal inefficiencies than by neural, decision, and motor noise.

Conclusion

This chapter reviewed some tools that are useful for characterizing the external and internal factors that limit perceptual performance. These tools are based on applying the concepts of Bayesian statistical decision theory to the analysis of natural signals, neural responses, and behavioral responses. Application of the Bayesian approach to natural signals can identify task relevant dimensions of information, provide principled hypotheses for neural mechanisms, and determine the limitations on perceptual performance imposed by external factors. Application of the Bayesian approach to neural responses can provide similar insight into task relevant dimensions of neural information and can provide principled hypotheses for subsequent decoding. Application of Bayesian approaches to behavior can provide principled perceptual models and can be used to separate effects on performance due to sensitivity from those due to decision criteria.

References

Blakemore C (1970) The range and scope of binocular depth discrimination in man. *J Physiol (Lond)* 211:599–622.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience, 12*(12), 4745-4765.

Burge J & Geisler WS (2011) Optimal defocus estimation in single natural images. *Proceedings of the National Academy of Sciences*, 108, 16849-16854.

Burge, J. & Geisler W.S. (under review) Optimal disparity estimation in natural stereo-images. *Journal of Vision*.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981) Efficiency of human visual signal discrimination. *Science, 214*, 93-94.

D’Antona A.D., Perry, J.S. & Geisler W.S. (minor revision) Humans make efficient use of natural image statistics when performing spatial interpolation. *Journal of Vision*.

De Vries, H. L. (1943). The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye. *Physica* X(7): 553-564.

Deriugin, N. (1956) The power spectrum and the correlation function of the television signal. *Telecommunications*, 1(7), 1–12.

Field, D.J. (1987) Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12): 2379-2394.

Geisler, W. S. (1989) Sequential ideal-observer analysis of visual discrimination. *Psychological Review,* 96, 267-314.

Geisler WS (2011) Contributions of ideal observer theory to vision research. *Vision Research*, 51, 771-781.

Geisler W.S. & Perry J.S. (2009) Contour statistics in natural images: Grouping across occlusions. *Visual Neuroscience*, 26, 109-121.

Geisler WS & Perry JS (2011) Statistics for optimal point prediction in natural images. *Journal of Vision* 11(12):14, 1–17.

Green, D. M., & Swets, J. A. (1966) *Signal Detection Theory and Psychophysics*. New York: Wiley.

Hecht, S., Shlaer, S., & Pirenne, M. H. (1942) Energy, quanta, and vision. *Journal of General Physiology, 25*, 819-840.

Kersten, D. (1987). Predictability and redundancy of natural images. *J. Opt. Soc. Am. A*, 4(12): 2395-2400.

Kersten, D., Mamassian, P., & Yuille, A. L. (2004) Object perception as Bayesian inference. *Annual Review of Psychology, 55*, 271-304.

Knill, D. C. & Richards, W. (Eds.). (1996) *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.

Landers D.D., Cormack L.K. (1997) Asymmetries and errors in perception of depth from disparity suggest a multicomponent model of disparity processing. *Percept Psychophys* 59:219–231.

McKee S.P., Levi D.M., Bowne S.F. (1990) The imprecision of stereopsis. *Vision Research* 30:1763–1779.

McLachlan, G. J. (1992) Discriminant Analysis and Statistical Pattern Recognition. New York: Wiley

Oruc, I., Maloney, L.T., & Landy, M.S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Res*, 43:2451-2468.

Rose, A. (1948) The sensitivity performance of the human eye on an absolute scale. *Journal of the Optical Society of America* 38(2): 196-208.

Tanner, W.P. & Birdsall, T.G. (1958) Definitions of *d*′ and η as Psychophysical Measures *J. Acoust. Soc. Am.* 30 (10), 922-928.

Tanner, W.P. & Swets, J.A. (1954) A decision-making theory of visual detection. *Psychological Review*, Vol 61(6), 401-409.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in the cat and monkey visual cortex. *Vision Research, 23*(8), 775-785.

Vapnik, V. N. (1998) Statistical Learning Theory. New York: Wiley.

Key Words

Detection, Discrimination, Identification, Estimation, Ideal observer, Bayesian statistical decision theory, Depth perception, Binocular disparity, Perceptual interpolation, Receiver operating characteristic (ROC), Stimulus noise, Neural noise, Utility function, Gaussian mixture model, Recursive conditional means, Parallel conditional means, Bayes rule, Signal detection theory.