PERGAMON

Vision Research 40 (2000) 3121–3144

Vision Research

www.elsevier.com/locate/visres

Signal detection theory in the 2AFC paradigm: attention, channel uncertainty and probability summation

Christopher W. Tyler *, Chien-Chung Chen

Smith–Kettlewell Eye Research Institute, 2318 Fillmore St., San Francisco, CA 94115, USA

Received 17 March 1999; received in revised form 3 December 1999

Abstract

Neural implementation of classical High-Threshold Theory reveals fundamental flaws in its applicability to realistic neural systems and to the two-alternative forced-choice (2AFC) paradigm. For 2AFC, Signal Detection Theory provides a basis for accurate analysis of the observer's attentional strategy and effective degree of probability summation over attended neural channels. The resulting theory provides substantially different predictions from those of previous approximation analyses. In additive noise, attentional probability summation depends on the attentional model assumed. (1) For an ideal attentional strategy in additive noise, summation proceeds at a diminishing rate from an initial level of fourth-root summation for the first few channels. The maximum improvement asymptotes to about a factor of 4 by a million channels. (2) For a fixed attention field in additive noise, detection is highly inefficient at first and approximates fourth-root summation through the summation range. (3) In physiologically plausible root-multiplicative noise, on the other hand, attentional probability summation mimics a *linear* improvement in sensitivity up to about ten channels, approaching a factor of 1000 by a million channels. (4) Some noise sources, such as noise from eye movements, are fully multiplicative and would prevent threshold determination within their range of effectiveness. Such results may require reappraisal of previous interpretations of detection behavior in the 2AFC paradigm.

Keywords: Psychophysics; Summation; Probability summation; 2AFC; Attention; Uncertainty; Signal detection theory; Additive noise; Multiplicative noise

1. Introduction

A principal function of early human vision is to analyze the spatial structure of images of the visual world. This information is then used to develop a representation of the properties of the objects before us and their layout in 3D space and in time. Despite previous attempts, a valid analytic framework has yet to be applied to the variety of spatial integration phenomena measured in laboratory studies. The analysis provided in this paper will demonstrate the deficiencies in previous approaches and form the basis for a comprehensive analysis of spatial summation based on the tenets of Signal Detection Theory, specifically in the context of detection and discrimination tasks measured by the two-alternative forced-choice (2AFC) paradigm. The analysis is valid for summation in any stimulus domain, but it will be illustrated with specific reference to summation in one and two-dimensional spatial vision.

Detailed analysis of summation behavior requires accurate models of the kinds of summation principles that can operate in psychophysics. The kind of summation performed by physiological receptive fields will be termed physiological summation (whether linear or nonlinear), to distinguish it from probability summation performed on the outputs of a set of decision variables (even though the latter operation must also ultimately be a physiological process in the brain). The primary theoretical analysis will be developed under the assumptions of Signal Detection Theory: that the main source of noise is external, Gaussian and independent of stimulus contrast. The theory also encompasses conditions where threshold is dominated by internal Gaussian noise and other forms of the noise distribution.

* Corresponding author. Fax: + 1-415-3458455.
E-mail address: cwt@ski.org (C.W. Tyler).

0042-6989/00/$ - see front matter
PII: S0042-6989(00)00157-7

Summation is quantified over arrays of processing mechanisms that are equal in sensitivity, although the theory could be extended to arbitrary sets of processes. Extension of the analysis to cases where the internal noise properties are some function of internal signal strength reveals major departures from the behavior with independent noise.

No complete account of summation behavior under the 2AFC paradigm has been published despite its widespread use for several decades. The most extensive published analysis of these issues is by Pelli (1985), which provides the basis for much of the present treatment, although many of our conclusions differ from the approximations derived in that paper. One of Pelli's goals was to show that Weibull High-Threshold Theory could approximate the full predictions of Signal Detection Theory for the 2AFC paradigm. The approximations were valid over a limited range under the assumptions Pelli made, but he did not develop the theory in more general cases. A key assumption was that the human observer is always operating under conditions of high uncertainty. This interpretation seems inherently implausible in practiced observers and we show that there are conditions under which this assumption is violated. Hence the 2AFC predictions need to be developed in accurate and usable form for a full treatment of psychophysical data.

The paper is divided into four main sections. The first section considers the implications of previous analyses of 2AFC probability summation through High Threshold Theory and finds these approaches to be fundamentally flawed in several respects. The second section develops the analysis of 2AFC summation through Signal Detection Theory limited by additive noise (from either external or internal sources). In the third section, the implications of a variety of non-ideal attentional strategies are spelled out for this additive noise case. The final section expands the analysis to cases where the internal noise properties are some multiplicative function of internal signal strength, revealing major departures from the behavior with signal-independent noise.

1.1. Assumptions of the 2AFC analysis

The assumptions of the main 2AFC analysis (Sections 3–5) are generally straightforward. There are also subsidiary issues that arise from considering alternatives to some of the assumptions. These alternatives are noted in brackets (A note on terminology; The term 'distribution' is used here to imply a probability density function, PDF, to which some noise variable conforms, as in 'Gaussian distribution'. The cumulative integral of such a function is termed its 'cumulative distribution function', or CDF).

1. In the 2AFC paradigm, the observer is presented with two defined stimulus events, both containing some background condition, while one also contains a test stimulus to be detected. The observer's task is to indicate which of the two events included the test stimulus.
2. There are sources of noise present in the stimulus events. Any component of the noise that is correlated between the two events forms part of the background from which the test is to be discriminated. We therefore consider 'noise' to include all sources of trial-to-trial variation that are uncorrelated between the stimulus events.
3. The noise is assumed to be white in space and time (for a fixed stimulus level) and Gaussian in its probability density function (PDF). The Gaussian assumption is plausible because of the Central Limit Theorem that the PDF for combinations of non-Gaussian noise is asymptotically Gaussian. If there are many sources of external and internal noise impinging at the decision site, therefore, the resulting noise is most likely to be Gaussian. [Alternatively, the PDF is assumed to take the form of a Poisson noise distribution.] [The Ideal Observer formulation makes the restrictive assumption that there are no noise sources except those present in the stimulus.]
4. The noise is assumed to be additive and independent of the strength of the test stimulus. [Alternatively, the noise variance is assumed to vary as some function of stimulus strength.]
5. Without the noise, the internal signals for each mechanism on which a decision is based are assumed to vary linearly with stimulus strength. [Alternatively, the internal signal is assumed to increase directly with stimulus strength above some level but be limited by a threshold such that the internal signal remains at zero below that level. If the threshold occurs at or above the level of the system noise in the absence of a test stimulus, it is known as a 'high threshold'.]
6. The visual system is assumed to consist of some (large) number of local mechanisms that transmit independent signals concerning the state of the outside world. The mechanisms are independent in the sense that their noise sources are statistically independent.
7. Each local mechanism is assumed to summate linearly over space within some weighting function known as its summation field. The summation may be over signals that are preprocessed for some stimulus attribute (such as orientation) by earlier neural mechanisms. [The Ideal Observer formula-

tion assumes that there is a summation field matching the profile of each stimulus presented.] [The inefficient Ideal Observer formulation assumes that each summation field is incompletely sampled to a similar extent, with the loss of a constant proportion of information for all fields.]

8. The local mechanisms are assumed to draw from the same local noise sources at all sizes of summation fields.
9. The signals from the local mechanisms are assumed to be combined by some nonlinear process known as an 'attention field' that is able to survey the local signals and isolate the largest signal. The implications of several types of control over the size of the attention field are considered. [The Ideal Attention formulation assumes that the attention field matches the stimulus extent, even when the local summation fields do not.]
10. The observer's 2AFC decision is assumed to derive from the larger of the signals from the attention field for the two stimulus events.

2. Problems with High Threshold Theory in the presence of additive noise

This section considers the implications of previous analyses of 2AFC probability summation in relation to High Threshold Theory and finds inherent problems with such approaches in several respects. These flaws indicate that High Threshold Theory does not provide a firm basis for the analysis of attentional integration of neural information in the presence of additive noise. To explain these problems, we first review High-Threshold Theory, but the source references should be consulted for full details.

2.1. Overview of High Threshold Theory

High Threshold Theory (Quick, 1974) is an analysis of the detection of signals that assumes that detection is limited by a noise-free, or fixed, threshold, below which no stimulus information is transmitted (Fig. 1a). The theory gets its name because the threshold is assumed to be high with respect to any noise in the signal arriving at the decision site. The goal of High Threshold Theory is to define the properties of summation over independent channels, which has come to be known as 'probability summation'. In spatial vision, the probability summation hypothesis implies that the mechanism of attention is distributed over many spatial channels rather than being focal, since one cannot monitor many channels without attending to them. It is then assumed that, on every trial, the attention mechanism can select the maximum channel response over the monitored range for use in the detection decision and ignore all other channels. Probability 'summation' is thus a max operator rather than a summing operator in the normal sense, and has generally been considered as the minimal combination rule among independent mechanisms.

The psychometric function Ψ is the theoretical form of the observer's proportion correct in a detection task as a function of stimulus strength. In Quick's (1974) version of High Threshold Theory, the psychometric functions Ψ_i for each individual channel with mean response R_i are given by the Weibull function:

$$\Psi_i = 1 - e^{-(R_i)^\beta},$$

$$\text{where } R_i = f\left(\frac{s}{\alpha_i}\right) \text{ for stimulus strength } s \qquad (1a,b)$$

with f being in general any monotonic function, α_i determining the sensitivity of the ith local mechanism and β controlling the steepness of the psychometric

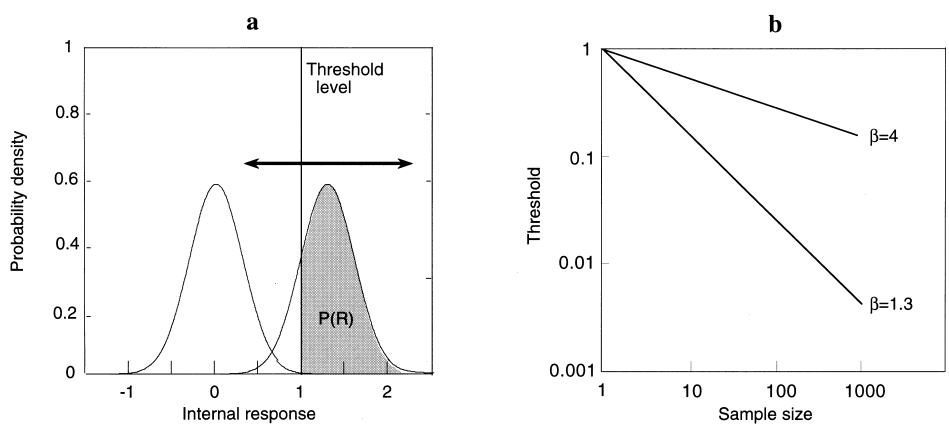

Fig. 1. (a) High-threshold analysis, that noise distribution in the absence of signal (left distribution) lies below some threshold level (R_0). The signal distribution varies in its position as the signal varies (arrow), and passes across the threshold as R increases to reveal some proportion of the signal distribution of correct responses (shaded area). (b) Weibull predictions for probability summation over number of samples (in time, area or any other stimulus parameter) for assumed psychometric exponents of $\beta = 4$ and 1.3 (d' powers of 3.2 and 1; see below).

function. For the remainder of this treatment, we will assume that f is a linear function, such that

$$R_i = \frac{s}{\alpha_i}, \quad s > 0. \tag{1c}$$

How the function behaves for negative s depends on the stimulus domain. For luminance, and for (Michelson) contrast, there are no negative signals, so the function does not exist in the negative region. For other stimulus domains, the negative portion will have to be analyzed according to its particular properties.

The Weibull function is derived from the theory of Failure Analysis and represents the combination of exponentially decaying failure functions. The effect on the overall psychometric function of probability summation over the set of channels (assuming equal sensitivities α_i for the mean responses R_i of the individual channels to the stimulus) is based on the standard statistical formula that probabilities of *not* detecting multiple events should be multiplied together:

$$\Psi = 1 - \mathrm{e}^{-(R)^\beta} = 1 - \prod_n (1 - \Psi_i) = 1 - \prod_n (1 - \mathrm{e}^{-(R_i)^\beta}) = 1 - \mathrm{e}^{-(n^{1/\beta} R_i)^\beta} \tag{2}$$

Psychophysical threshold is estimated by solving Eq. (2) for $\Psi = 0.5$ (This basic version of the theory assumes that the observer's guessing rate is zero). For a criterion level of the output function Ψ, the similarity in form of the first and last expressions makes it clear that the effective mean response over the set of channels is $R = n^{1/\beta} R_i$ (see Robson & Graham, 1979, for details). Thus, as the stimulus extent is increased to sample more of the local mechanisms, the internal response increases in proportion to the βth root of the number of mechanisms sampled by the attention mechanism.

Fig. 1b depicts the degree of probability summation over the number of mechanisms sampled for the typical case of $\beta = 4$ and for the hypothetical case of a linear psychometric function, when $\beta = 1.3$ (Pelli, 1987). Empirically, the exponent β of the Weibull approximation to psychometric data may take values from 1.3 to 6 (Mayer & Tyler, 1986). Under the assumption that f is a linear function (Eq. (1c)), the low value represents the theoretical low limit on the expected slope if the stimulus is present in all the channels that the visual system is monitoring (and there is no phase uncertainty; Pelli, 1985). A high value for β represents a high degree of channel uncertainty. When there is minimal uncertainty, probability summation effects are predicted to be large relative to the possible contrast measurement range (Fig. 1). If $\beta = 1.3$, for example, sensitivity improvement of as much as a factor of 200 is predicted for probability summation over $n = 1000$ equally stimulated channels. Such a result would be predicted by an increase in stimulus diameter by a factor of ~ 30 on homogeneous retina, if n represents the number of local retinal filters). Under such conditions, probability summation could not be dismissed as a minor, near-threshold effect. The generation of such large summation effects from purely attentional processes would cloud the issue of what physiological summation might be taking place because the two effects are of comparable magnitude.

2.2. *High threshold analysis of probability summation assumes non-Gaussian additive noise*

Suppose a signal with intensity s can produce an internal response distribution $D(r;R,\sigma)$, where r represents the dimension of the random internal response variable, with mean R (which is assumed to be a monotonic function of s) and standard deviation σ. Under the assumption of a high threshold, this noise distribution is progressively revealed as the signal intensity moves up beyond the threshold level. Thus, if the noise distribution is additively independent of the mean response, the probability of detecting signal s is the *integral* of the internal signal-plus-noise distribution from the threshold R_θ to infinity. The Weibull formulation of the psychometric function (equation 1a) must correspond to this integral for some particular noise distribution $D_\beta(r - R)$ around the mean signal R,

$$\Psi = 1 - \mathrm{e}^{-R_\beta} = \int_{R_\theta}^{\infty} D_\beta(r - R)\mathrm{d}r \tag{3}$$

where R_θ is the mean internal response level at threshold.

For the assumption that the PDF of the signal + noise distribution generating the Weibull function is of fixed form, $D_\beta(r)$, it can be solved by taking the derivative of both sides of Eq. (3) for each integration limit

$$\beta R^{\beta-1} \mathrm{e}^{-R^\beta} = \lim_{\varepsilon \to \infty} D_\beta(\varepsilon - R) - D_\beta(R_\theta - R) \tag{4a}$$

from which,

$$D_\beta(r) = \beta \rho^{\beta-1} \cdot \mathrm{e}^{-\rho\beta}, \quad \text{with } \rho = R - R_\theta - r, \quad \text{for } r < R - R_\theta \tag{4b}$$

On the assumption that the mean response R is linear with the external stimulus strength s, equation (4) defines the implied PDF that would have generated the Weibull expression for the measured psychometric function. The forms of the psychometric functions and the implied noise distributions $D_\beta(r)$ for values of β from 1.3 to 8 (corresponding to d' exponents from 1 to 6.5; see following sections) are shown in Fig. 2 for a Yes/No experiment (assuming zero false alarm rate). It is evident that the implied noise distributions in the lower panel are generally far from approximating a Gaussian form except in the mid-range of parameter values, the special case where $\beta \approx 4$ (the value for which

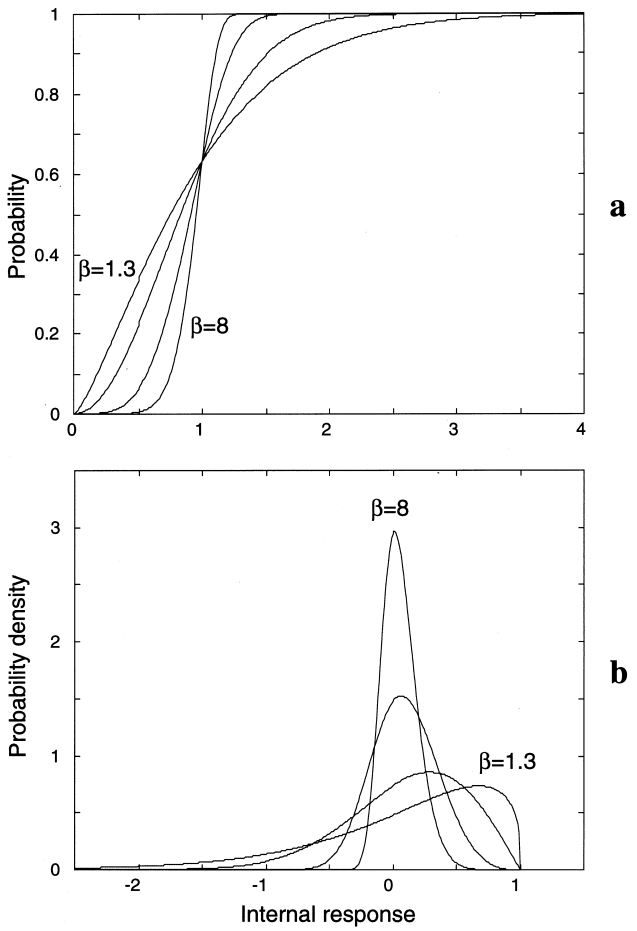

Fig. 2. (a) Psychometric functions with $R_0 = 1$ predicted by High Threshold Theory for values of 1.3, 2, 4 and 8 for the exponent β. Functions are corrected for guessing. (b) Implied noise distribution functions according to equation (4), plotted relative to the mean response (i.e. as the additive noise distribution). Note the marked change in distribution shapes as the exponent varies.

Pelli, 1983, established that the Gaussian is a good approximation). Thus, Weibull analysis is not an accurate theory for the description of systems with a high threshold and Gaussian noise unless the psychometric slope happens to fall at this mid-range value. In practice, empirical slopes have been found to approximate this value in many situations (Robson & Graham, 1981; Williams & Wilson, 1981; Pelli, 1985), but there may be substantial inter-observer differences (Mayer & Tyler, 1986) and large changes in slope under certain circumstances (Tyler, 1997). Thus, there is a need for a comprehensive and accurate theory of probability summation when the assumptions of High Threshold Theory are violated.

2.3. High-threshold probability summation fails for additive noise

High Threshold Theory has been widely used to predict the effects of probability summation, the improvement in sensitivity with an increasing number of chances to detect the presence of a signal. This theory assumes that probability summation occurs because the observer can identify the max of the samples of signal + noise distributions provided by each of n stimulated channels. If we assume a physiological version of the high-threshold system that has Gaussian noise added to the signal, the internal response after probability summation is provided by the distribution of such max values over trials. Note that, for such probability summation to occur, the threshold has to be applied after the max operator. For this analysis, the max operator is assumed to function like an ideal attention mechanism, in that it samples from all of, and only from, the relevant channels.

In general, it is a well-known statistical rule that the cumulative distribution of maximum values for a set of samples $D(r_i)$ from a parent distribution $D(r)$ (where r is the instantaneous internal response) is given by the integral of the parent distribution to the power of the number of values within each sample:

$$\int_{-\infty}^{r} \max_{i=1:n}[D(r_i)]\mathrm{d}r = \left[\int_{-\infty}^{r} D(r')\mathrm{d}r'\right]^n \quad (5)$$

Thus, the expected distribution $M_n(r;R,\sigma)$ of the max of a set of samples is given by taking the derivative of both sides of Eq. (5) with respect to their independent variables:

$$M_n(r;R,\sigma) = \max_{i=1:n}[D(r_i)] = \frac{\mathrm{d}}{\mathrm{d}r}\left[\int_{-\infty}^{r} D(r')\mathrm{d}r'\right]^n \quad (6)$$

The mean R and standard deviation σ parameters in the expression for the max distribution M_n imply that we are deriving the form of the expected function of the resulting probability distribution, which may be characterized by the parameters of its location and spread. It does not imply that these are the only parameters of the distribution (as they would be for a Gaussian distribution), merely that we restrict our consideration to these two parameters.

To obtain the new threshold signal level, the signal can be reduced until the max distribution M_n reaches the original threshold criterion again. The extent to which the signal has to be reduced constitutes the improvement in sensitivity attributable to probability summation on the basis of the max rule. If the noise is assumed to be additive, however, this process creates the fatal problem that, for a large enough number of channels, the mean signal needs to be set to a negative value in order to bring the signal + noise distribution down to threshold. Fig. 3a depicts the case for such summation over 100 channels, where the initial signal is assumed to have a mean of two times the internal threshold level and a σ of 0.67 (so as to provide 75% correct performance at this signal level). The max distribution for 100 channels from Eq. (6) has a mean of about 4.5, giving essentially 100% correct performance. Fig. 3b shows how the signal has to be readjusted to bring the tail of the max distribution down below threshold so as to reattain 75% correct performance. On the linear assumption of Eq. (1c), the level of the mean internal response in the null interval corresponds to an external signal of zero (The internal scale is arbitrary, so we choose zero to represent the mean null level for analytic convenience). Thus, since the internal signal needs to be reduced from 4.5 back to the threshold level of 1.0 (dashed arrow) the external signal represented by the filled arrow has to go substantially *negative* before 75% correct performance is achieved.

The problem is fatal in some domains, such as the amplitude of light, because negative signals do not exist. Other domains, such as contrast, may be defined in such a way that there are negative signals, but the problem reasserts itself because the system contains negative-sensitive elements (e.g. off-center cells) that respond positively to the negative signal. Thus, rather than becoming less detectable by its max value, the signal becomes *more* detectable as the corresponding minimum of the set of samples (at the left-hand tail of the distribution in Fig. 3) passes above the corresponding negative threshold before the max falls below the positive threshold. Once again, therefore, it is impossible to return to the 75% performance level after the max operator has taken effect.

High-threshold analysis is immune to this problem only if the noise on the signal is *multiplicative* with signal strength rather than additive, and hence can be reduced indefinitely by appropriate signal reductions without the signal going negative. Thus, if the noise is purely multiplicative, the max level on the noise distribution may be freely reduced to the threshold level to provide a measure of the threshold sensitivity for the input signal. High-threshold analysis is self-consistent in that the noise implied by the Weibull formulation has the property of being multiplicative. Because this property is rarely made explicit, it should be mentioned that the property follows from Eq. (2), which shows that the Weibull psychometric function has a constant form when plotted on log coordinates, i.e. is scaled in proportion to signal amplitude. This implies that the limiting noise is similarly scaled through the probability summation operator. To reiterate, Fig. 3 goes further in showing that the assumption of additive noise is incompatible with Weibull analysis in general.

Note that, for *mixed* additive and multiplicative noise sources, reducing the signal will tend to reduce the multiplicative noise to the point where additive noise dominates. Since there are always sources of additive noise in any physical signal-detection system (e.g. thermal noise, and quantal noise considered with respect to modulation variables, such as a sinusoidal grating, which keep the mean signal constant), any noise-limited threshold is likely to be limited by its additive component. The only amelioration of this problem is if the high threshold is so high that it sits at or above the level for the *max* of the additive noise from all monitored channels (which one might term an 'ultra-high' threshold). Were it any lower, the negative signal problem would be encountered. Thus, for Weibull analysis to operate, the system must be functioning with thresholds so high as to be quite inefficient, especially considering that the degree of probability summation required by the quantitative application of Uncertainty Theory may be of the order of many thousands or even millions of channels (Pelli, 1985). The Weibull analysis of probability summation is thus implausible in realistic threshold systems.

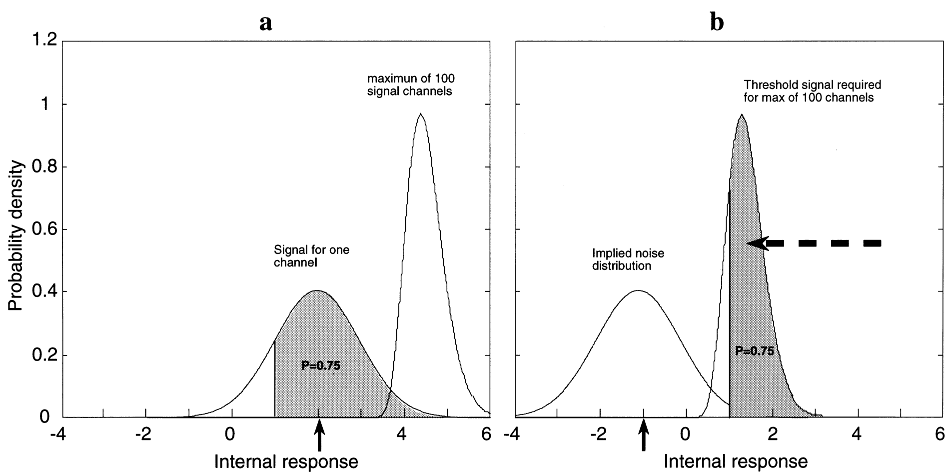

Fig. 3. (a) Threshold signal-plus-noise distribution for 75% correct detection (left distribution, with 75% of the area above the threshold level of 1) together with the distribution of maxes over 100 channels (right distribution). Since the max distribution is effectively all above the threshold level the signal would be detectable close to 100% of the time if probability summation were in operation. (b) Thus, the signal level has to be reduced (dashed arrow) until the max distribution sits at the 75% level above threshold. The problem is that this reduction produces a negative value for the mean signal in each channel (filled arrow), which is likely to be unobtainable in typical vision paradigms.

The conclusions from the analysis of High Threshold Theory are:

1. The high-threshold analysis developed by Quick (1974) implies a reciprocal relationship between the exponent of the psychometric function and the log slope of its probability summation behavior (Fig. 1b). If there are conditions where psychometric functions are empirically found to be shallow (and the noise sources locally independent), steep summation slopes would be predicted. In practice, such conditions have been found to show shallow summation slopes, calling the theoretical framework into question. For suprathreshold masking conditions, a range of studies such as Foley and Legge (1981) and Kersten (1984) report exponents of the d' function close to 1 for the 2AFC paradigm, even when the mask is a noise background that is randomly independent at all locations within the stimulus. Nevertheless, Kersten (1984) showed that summation is negligible under suprathreshold, noise-masked conditions. Both because such near-unity exponents imply strong summation behavior, and because it is hard to conceptualize a threshold operating under 'suprathreshold' conditions, High Threshold Theory cannot be applied to such data. There is thus need for a theory that can be used to analyze suprathreshold discrimination experiments.
2. The form of the Weibull function implies bizarre variations in the noise distribution (Fig. 2b) if it is assumed that the neural noise is additive in the threshold range. Since noise asymptotically Gaussian (such as quantal noise in the light, thermal noise in the photoreceptors or retinal noise in the ganglion-cell outputs), High Threshold Theory is incompatible with plausible assumptions about the noise distribution.
3. Quick's High Threshold analysis through the Weibull function assumes the performance is limited by a high threshold rather than by noise of any kind. However, noise is an unavoidable component of the analysis of the 2AFC paradigm. In order to adapt the high-threshold analysis to the 2AFC paradigm, Pelli (1985) made the assumption that the observer was monitoring a much larger number of channels than were stimulated as a means of obtaining a steep psychometric function that approximated threshold behavior. Thus, Pelli's approximation fails if the stimulus is structured so as to stimulate as many channels as the observer is monitoring, because the predicted psychometric function is then shallow and violates the high-threshold assumptions. No theoretical analysis for these conditions has been published.

3. Signal Detection Theory for the Ideal Observer and its Bayesian approximation

Following from the inadequacies of High Threshold Theory, this section develops the analysis of summation properties in the two-alternative forced-choice (2AFC) task. The analysis is approached by specification of the psychometric function through Signal Detection Theory as limited by additive noise. When the only source of this additive noise is quantum fluctuations (in a contrast detection task), the Signal Detection Theory analysis amounts to a single-channel Ideal Observer model. The implications of an attentional strategy approximating Ideal Observer behavior are also spelled out for this additive noise case.

3.1. Specification of the psychometric function

The first step to understanding psychometric function in a 2AFC task is to specify the proportion correct of the observer's responses. The 2AFC task typically involves the presentation of two stimulus intervals (or spatial stimulus regions), one of which contains the stimulus to be detected while both contain the background condition from which the stimulus is to be distinguished. The observer's task is to estimate which interval contains the discriminative stimulus. Traditionally, the observer is assumed to exhibit ideal behavior in three ways:

1. to have exact knowledge of the stimulus and to view it with a matched filter, excluding all irrelevant information
2. to be noise-free; performance is limited only by noise in the physical stimulus
3. to respond according to the maximum output of the filter in the two intervals, with no confusion.

When the first assumption is violated by ignorance of the correct filter, the observer may still adopt an ideal *attentional* strategy across a set of filters, to make the best guess as to which is the optimal filter to select on each trial. The second assumption may be violated by the introduction either of early noise before the filter or of late noise at the decision stage. In the case of early noise, the observer's performance will still reflect the form of the Ideal Observer, but at reduced efficiency. In the case of late noise, the threshold will become independent of stimulus extent as long as the late noise dominates other sources of noise.

In Signal Detection Theory (SDT), the proportion correct is conceptualized through an imaginary ROC (receiver operating characteristic) curve of proportion of hits versus proportion of false alarms (Green & Swets, 1966), treating each trial as a separate Yes/No task with a different criterion. Not only is this instantaneous criterion inaccessible, but the 2AFC proportion correct is defined as the area under the ROC curve,

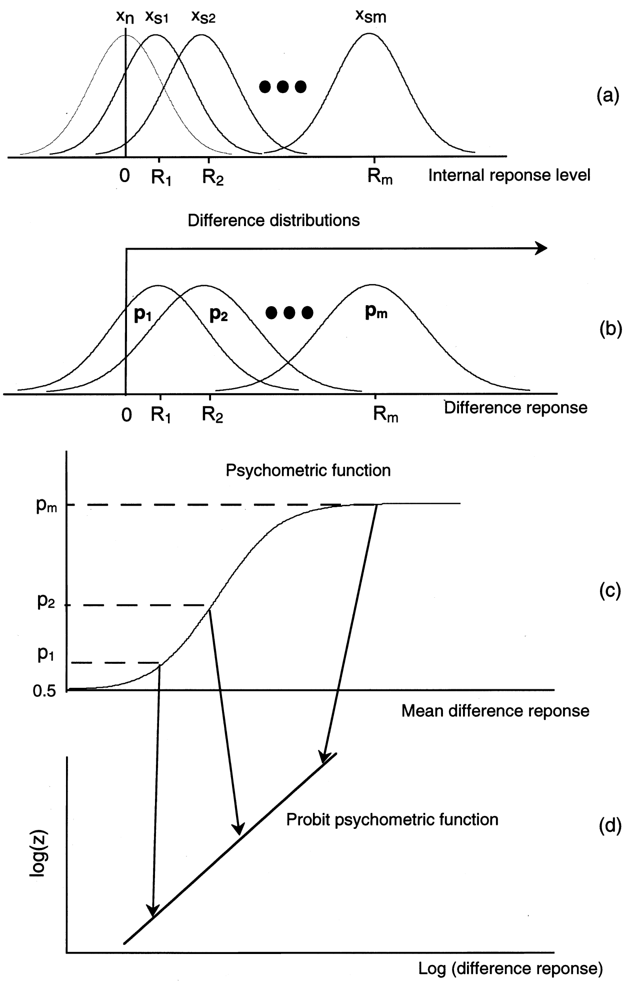

Fig. 4. Derivation of the 2AFC psychometric function. (a) Internal responses for signal levels from zero (distribution x_n) through various mean levels $R_1 - R_m$. (b) Difference distributions between the distributions for each signal level and that for no signal. Proportion correct is derived by integrating areas p_1 to p_m to the right of the vertical line at zero difference indicating the response criterion. (c) Psychometric function derived from plotting the area of each difference distribution lying above the zero criterion from (b). (d) Log (z) transform of the cumulative function to provide a straight line in probit coordinates.

which requires a further level of abstraction from this two-dimensional distribution of signal strength and criterion level (see Macmillan & Creelman, 1993 for details).

The 2AFC task is amenable to a simpler form of analysis based on the difference distribution of the internal responses (MacMillan & Creelman, 1993). On each trial, the observer responds by indicating whichever observation interval produces a larger internal response, which amounts to taking the *difference* between the two internal signals and picking the interval according to the sign of this difference signal (see Fig. 4). The criterion is therefore fixed in this difference space, at a difference of zero (whereas it can range from trial to trial over the whole extent of the internal response distribution). Stated formally:

For signal intensity s, let r_1 be the internal response to the first observation interval, r_2 be the internal response to the second interval and k be the interval the observer chose as the signal interval. (Note that here s serves as a scalar for signal strength and s as an index for the signal interval. Similarly, n is the index for the null interval whereas n defines the number of stimulated channels elsewhere.) We assume that there is a fixed signal level throughout the test interval.

The observer indicates the first interval as the signal interval ($v = 1$) if $r_1 - r_2 > 0$ and indicates interval ($v = 2$) if $r_2 - r_1 > 0$. The response is correct if either $v = 1$ when the signal is the first interval, denoted by $<\mathrm{sn}>$ or $v = 2$ when the signal is the second interval, denoted by $<\mathrm{ns}>$. The proportion correct in terms of the internal difference response δ is:

$$\begin{aligned} p_{\mathrm{corr}}(\delta) &= p(v=1 | <\mathrm{sn}>)^*p(<\mathrm{sn}>) \\ &\quad + p(v=2 | <\mathrm{ns}>)^*p(<\mathrm{ns}>) \\ &= p(r_1 - r_2 > 0 | <\mathrm{sn}>)^*\mathrm{p}(<\mathrm{sn}>) \\ &\quad + p(r_2 - r_1 > 0 | <\mathrm{ns}>)^*p(<\mathrm{ns}>) \end{aligned} \tag{7}$$

If r_n is the internal response to the null interval and r_s is the internal response to the signal interval, Eq. (7) can be rewritten in terms of the psychometric function:

$$\begin{aligned} \Psi(s) &= p(r_s - r_n > 0 | <\mathrm{sn}>)^*P(<\mathrm{sn}>) \\ &\quad + p(r_s - r_n > 0 | <\mathrm{ns}>)^*p(<\mathrm{ns}>) \\ &= p(r_s - r_n > 0) = p(\delta > 0) = \int_0^{\infty} Z_s(\delta;\Delta)\mathrm{d}\delta \end{aligned} \tag{8}$$

where $\delta = r_s - r_n$ is the difference between signal and null interval internal responses and $Z_s(\delta;\Delta)$ is the PDF of the difference distribution for signal strength s (normalized in units of its standard deviation), with mean Δ.

Eq. (8) describes the relation between the proportion correct and the observer's internal responses to signal and null intervals. The psychometric function can be obtained by repeating the computation of Eq. (8) for all relevant signal intensities. Fig. 4 illustrates the relations between the internal responses and psychometric function based on Gaussian additive noise.

The probit transform (Finney, 1952) is the appropriate representation of the psychometric function, on the basis of the additive Gaussian assumption. It normalizes proportion correct to its standard deviation unit (z-score) through the inverse cumulative Gaussian function Φ^{-1}. That is

$$Z_s(\delta;\Delta) = \Phi^{-1}(\Psi(s)) \tag{9a}$$

In the psychophysical literature, the normalized signal $Z_s(\delta;\Delta)$ represents the detectability of the signal at stimulus level s, defined by

$$d' = Z_s(\delta;\Delta) \tag{9b}$$

3.2. Comparison of a Gaussian and a non-Gaussian example

Suppose the system response is dominated by one channel whose internal response distribution $D_N(r)$ in the null interval is added Gaussian noise with expected value 0 and standard deviation σ, denoted as $D_N(r) \approx G(r;0,\sigma)$. At signal intensity s, the internal response in the signal interval will also have a Gaussian distribution but with mean R and standard deviation σ, denoted as $D_R(r) \approx G(r;R,\sigma)$. From the properties of the Gaussian distribution, the difference distribution of $\delta = r_s - r_n$ is another Gaussian distribution with mean R and standard deviation $\sqrt{2}\sigma$. From Eq. (8), the proportion correct is

$$\Psi(s) = 1 - \Phi(\delta,\Delta,\sqrt{\sigma_N + \sigma_R}) = 1 - \Phi(-r;R,\sqrt{2}\sigma_N) = \Phi(r;R,\sqrt{2}\sigma_N) \quad (10)$$

where Φ denotes the Gaussian cumulative distribution function. Eq. (10) is commonly used in fitting the psychometric function to 2AFC data (MacMillan & Creelman, 1993).

In general, however, it is important to avoid the implication that the psychometric function matches the cumulative distribution function of its underlying probability distribution. The match is valid only if the noise is *additive* to the mean internal signal strength R and if its distribution is symmetric (as revealed by Eq. (10)). In general, different signal levels may produce different r_s distributions if the noise is non-additive, and in turn access different $D_R(\delta)$. Thus, the general 2AFC psychometric function would not be a cumulative function of any particular difference distribution. Only when the noise is additive and symmetric (e.g. Gaussian) will the difference distributions at different signal levels all have the same variance and the psychometric function is equivalent to its CDF (e.g. the cumulative Gaussian or *erf*). On the other hand, if the noise distribution is Poisson rather than Gaussian (a common alternative assumption) the noise is no longer additive but varies with the mean level, and also is asymmetric. Thus, the psychometric function derived from Eq. (8) will not exactly match the cumulative distribution function (Fig. 5).

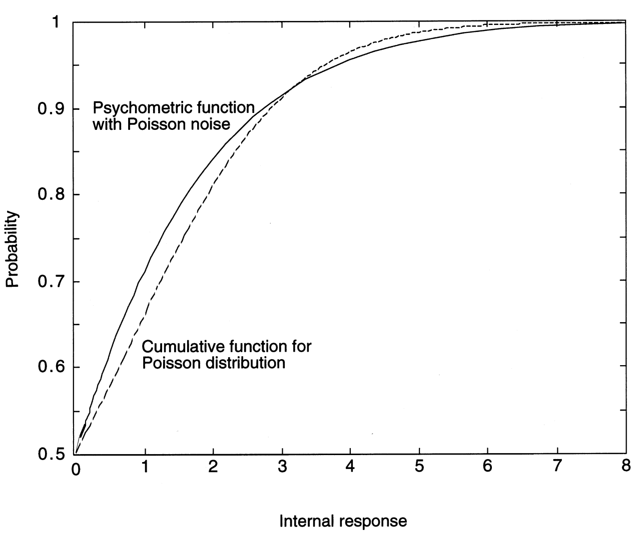

Fig. 5. Theoretical psychometric function for Poisson noise (full curve), showing failure to match the cumulative distribution in this case of non-Gaussian noise (dashed curve).

3.3. Overview of Ideal Observer analysis

The Ideal Observer formalism assumes that the observer has complete knowledge of the stimulus and uses a single matched filter to detect its presence (Wiener, 1949). The Ideal Observer therefore is effectively a Bayesian detector with a prior probability of 1.0 on the matched filter and zero elsewhere. Optimal performance with an ideal filter is assumed to occur with linear summation over the noisy filter inputs sampled by the field. The summation properties of the filters will vary with respect to a large number of stimulus attributes. For simplicity, we consider the case of spatial summation over two-dimensional stimuli $S(x,y)$ varying in one dimension of overall size. This variable size dimension could be the height, the width, the area, or any parameter that is linear with the number of sources of input to each summing field over the domain (x,y). The input for the matched filter is provided by discrete sensors with independent noise sources drawn from the same underlying distribution. When the local regions have identical sources of independent Gaussian noise with standard deviation σ, the summed output of each field is given by summing over the product of the stimulus profile and the matching ideal filter. We can show that the signal-to-noise ratio in such a matched filter is proportional to the square root of the stimulus area.

In general, the response of the matched filter can be approximated as the weighted sum of its responses to the samples

$$\boldsymbol{R} = \sum \boldsymbol{S}(x,y) \cdot \boldsymbol{I}(x,y) \quad (11a)$$

and the signal variance as the weighted sum of the local variances

$$\sigma_R^2 = \sum \boldsymbol{S}(x,y) \cdot \boldsymbol{I}(x,y) \cdot \sigma^2, \quad (11b)$$

Hence

$$\sigma_R = \sigma\left(\sum \boldsymbol{S}(x,y) \cdot \boldsymbol{I}(x,y)\right)^{1/2} \quad (11c)$$

The discriminability of the ith stimulus, d_i', therefore, can be approximated as the reciprocal of the standard deviation times the sampling interval. Appendix A

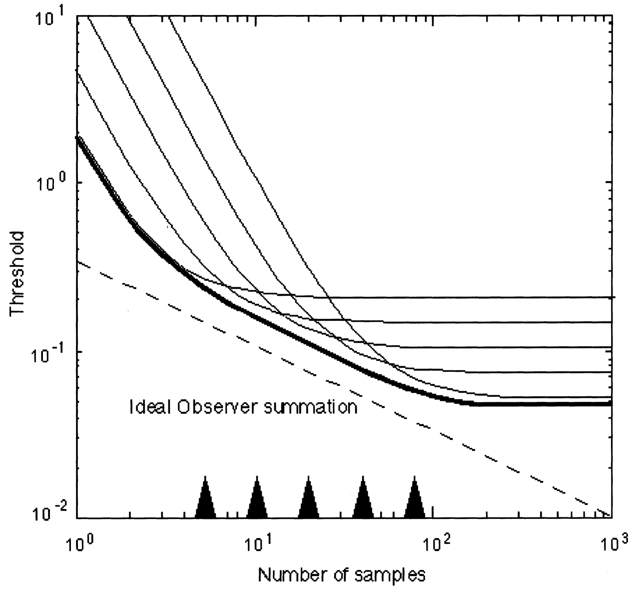

Fig. 6. Physiological implementation of ideal observer behavior. Thin curves: summation behavior for five individual Gaussian filters as the extent of a Gaussian test stimulus is varied (arrowheads indicate filter extents or number of input samples in each filter at half-height). Thick curve: fourth power attentional summation over the individual channels approximates ideal observer summation behavior (a log slope of -0.5, dashed line) in the range where physiological filters are available, with departures below and above. The ideal strategy is to read out from that filter only when it matches the stimulus extent. At the two ends of the range, only one filter dominates detection behavior and hence system performance (thick line) departs from the ideal slope of -0.5 to follow the function for the most sensitive filter in that region.

shows how characterization of the stimulus size in terms of the sampling density within the stimulus envelope allows the discriminability to be expressed in terms of the effective area A_i of the stimulus

$$d'_i = \frac{\boldsymbol{R}_i}{\sigma_{\boldsymbol{R}_i}} \sim \propto \frac{\boldsymbol{A}_i^{1/2}}{\sigma}, \tag{12}$$

showing that ideal discriminability is proportional to the square root of stimulus area.

Note, however, that there is a problem with applying this model in practice, since the psychometric function in this model is based on a linear relation between d' and signal strength. This linear relation is violated by most d' measurements, which typically show an exponent of about 2 (e.g. Stromeyer and Klein, 1974). Similarly, translation of this prediction into the Weibull format yields a predicted Weibull exponent of 1.3 in Eq. (2) (whereas most measurements show exponents of 3–4). Extension of the theory to non-ideal attention behavior, which encompasses steeper exponents, is left to the next section. First we consider an approximation to ideal behavior that can be used if the observer knows the set of stimulus types that may be presented in a block of trials, even if the particular stimulus is not known in advance on each trial.

If the task is summation over a range of stimulus sizes, the Ideal Observer model requires a summing receptive field matching every size of stimulus for which the summation behavior is exhibited. A physiological implementation of such behavior is depicted in Fig. 6, where the attention mechanism is assumed to switch to the receptive field size matching the stimulus presented in each condition. This behavior is possible only if the stimuli are presented in blocks of trials, so that the form of the next stimulus on each trial is known. Thus, if human observers exhibit a log-log summation slope of $-1/2$ (dashed curve in Fig. 6) they may be said to manifest Ideal Observer behavior, in the sense of using ideal matched filters to improve in the way an Ideal Observer would, even if the absolute sensitivity is less than predicted for an Ideal Observer (i.e. lower than ideal efficiency). Such (inefficient) Ideal Observer behavior may be taken as evidence that the brain has access to summing fields matching the sizes of all the tested stimuli, either present and selectable by attention as in the central region of Fig. 6, or alternatively as an adaptive mechanism re-forming itself for each new stimulus condition.

If the system has access to only a limited range of summing field sizes, the summation slope should asymptote to -1 for stimulus sizes below that of the smallest summing field size and should asymptote to 0 for sizes above that of the largest summing field, as depicted by the bold curve in Fig. 6. Thus, the form of the summation function in any stimulus domain carries important information about the range of summing field sizes operating in that domain (see Gorea & Tyler, 1997, for an example in the temporal domain and Kersten (1984), for an example in the spatial domain). The model that the brain contains an adaptive filter re-forming itself for each new stimulus condition seems to be incompatible with the occurrence of a limited summation range, for why would such adaptive capability fail at a particular point?

3.4. The concept of probability summation

Probability summation is an option available to a decision mechanism with access to a number of independent signals reflecting the occurrence of a stimulus. The analogy is with a group of human monitors looking out for an approaching plane, for example. The probability of detecting the plane is higher if detection is considered to have occurred when any one of the monitors spots the plane than by relying on a lone observer. In other words, probability summation corresponds to a decision rule in which the group decision is defined by a response from any single member of the group. This decision rule corresponds to defining a detection event when the signal in any one of m monitored channels reaches a criterion level. This decision structure is implemented by applying a max rule to all the channel outputs and defining the detection event when the max reaches some preset criterion level (Pelli, 1985; Kontsevich & Tyler, 1999a). The neural implementation of such a decision rule may be designated as 'attentional summation'.

Although probability summation is often considered as a purely mathematical operation, it is meaningless in the context of the human vision (in a single observer) unless it is mediated by some neural hardware. This raises the issue of whether there are independent neural channels and what is meant by a max operator in a neural system. In terms of detection theory, two channels are considered independent when they are governed by sources of noise that are statistically independent. There is plenty of evidence for a high degree of statistical independence among even neighboring cortical neurons (e.g. Freeman, 1994, 1996; Shadlen & Newsome, 1998), so cortical neurons can be considered to be separate channels for this purpose. What would constitute the probability summation or max operator? It needs to be a neural system receiving signals from an array of channels (or axons) that have statistically independent noise up to that point. It then needs to respond when any of these inputs manifests a signal but not otherwise. Such a neural system would have this property if it would transmit a spike that initiated a detection response on receiving a spike from any one of its inputs. The threshold characteristic of cortical neurons with wide-field input sampling thus provides the requisite hardware for a max operator.

In terms of the detection of signals in additive noise, the optimal strategy is to use a matched filter, to convolve the stimulus input with a linear filter exactly matching the stimulus profile. It is possible to approximate ideal observer strategy by performing probability summation over the full set of filters in the form of the max of the signal-to-noise ratios (Pelli, 1985). This approach may be considered an ideal (or Bayesian) attentional strategy in that the observer knows the set of likely filters to survey on each trial. This strategy will have the effect of isolating the most efficient filter under any condition, and hence mimic ideal observer behavior without requiring prior knowledge of the stimulus. However, implementation of this strategy does require the neural system to have an accurate representation of the noise level, in order to compute the signal-to-noise ratios. Simply taking the max over raw signals will tend to emphasize the noisiest fields. But if it is plausible that the neural system normalizes to the prevailing (long-term) noise level, then a max operator would provide a mechanism for implementing Ideal Observer behavior.

It is common practice to combine the response outputs in neural network models by a Minkowski summation rule:

$$R = \left[\sum_n (R_i^p) \right]^{1/p} \tag{13}$$

where the summing exponent is often set at $p = 4$. Note that such fourth-power summation (thick curve in Fig. 6) produces a completely smooth curve in the range where the filters are present even though in this example the assumed filters are separated by factors of two in size. It is thus possible to approximate Ideal Observer behavior with relatively coarse physiological sampling in a particular domain if there is some way to implement in the cortex the Minkowski summation of Eq. (13) with a high summation exponent.

3.5. Attentional summation in 2AFC experiments does not conform to high threshold analysis, but derives from the σ of the difference distribution

For 2AFC detection using more than one channel, attentional (or 'probability') summation effects should be analyzed through Signal Detection Theory. For a tractable analysis, we assume n stimulated channels of equal sensitivity with additive Gaussian noise. For the full analysis, we will consider the situation where the observing system is monitoring more channels (m) than are being stimulated. The statistical combination rule for attentional summation of the responses over channels is derived again from the maximum value of the set of m monitored channel responses in each stimulus interval (Pelli, 1985; Palmer, Ames, & Lindsey, 1993). For the null stimulus of the pair, which by definition contains only noise, the combined response distribution $M_m(R,\sigma_R)$ is based on the noise-alone distributions in the responses of all m channels. Mathematically, this combined distribution is given in terms of the expected values of the distributions by the derivatives in a similar fashion to Eq. (6), omitting the distribution variables for clarity:

$$M_m(R,\sigma_R) = \max_{i=1:m}[D_N] = \frac{\mathrm{d}}{\mathrm{d}r}\left[\int_{-\infty}^{r} D_N \,\mathrm{d}r'\right]^m = mD_N\left[\int_{-\infty}^{r} D_N \,\mathrm{d}r'\right]^{(m-1)} \tag{14}$$

The two parameters in the expression $M_m(R,\sigma_R)$ for the max distribution imply that we are deriving the form of the expected function of the resulting probability distribution, which may be characterized by the parameters of its location and spread (as for the High Threshold Theory of Eqs. (5) and (6)).

With the inclusion of n signal channels for the signal interval of the stimulus pair, the max must be taken over the maxes of the separate n signal + noise and $m - n$ noise-alone distributions:

$$M_{n,m}(R,\sigma_R) = \max\left[\max_{i=1:n}[D_R(r_i)] \max_{i=n+1:m}[D_N(r_i)]\right]$$

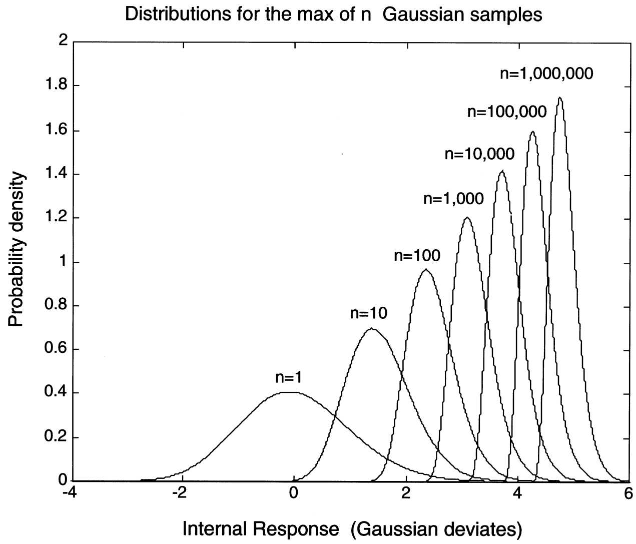

Fig. 7. Max distributions for a Gaussian probability density function for numbers of samples increasing in factors of 10 from $n=1$ to 1 million. Note the decreasing standard deviation and small asymmetry of these max distributions.

$$=\frac{\mathrm{d}}{\mathrm{d}r}\left(\left[\int_{-\infty}^{r} D_R\,\mathrm{d}r'\right]^{n}\cdot\left[\int_{-\infty}^{r} D_N\,\mathrm{d}r'\right]^{m-n}\right) \tag{15}$$

In the general case, Eq. (15) does not simplify in the manner of Eq. (14).

The simplest case of 2AFC attentional summation is the case where $m=n$, so there is no uncertainty as to which of the monitored channels contain the stimulus, and the two distributions differ only in their mean level of internal response. This situation corresponds to assuming that the observer employs an ideal *attention window* that always matches the stimulus extent, so that no unstimulated channels are monitored. Nevertheless, it is assumed that the observer cannot perform ideal *summation* over the stimulus area, but is forced to monitor a set of n local channels to find which gives the max response in any test interval (Pelli, 1985).

Fig. 7 shows the numerical distributions for samples of maxes computed according to the derivation of Eq. (14) for noise alone (or Eq. (15) for signal + noise with $m=n$) in factors of ten from $n=1$ to 1 million channels of equal sensitivity. The σ of these max distributions decreases by a factor of about four (in contrast to the factor of 200 decrease predicted for only 1000 channels under High-Threshold Theory with no uncertainty). In each case, the observer's task is to distinguish between sample stimuli drawn from the max distributions of noise-alone and signal + noise for summation over a given number of channels. Discriminability therefore improves with the reciprocal of the reduction in σ in these max distributions (Fig. 7), as shown in the leftward shift of the d' functions of Fig. 8a. The consequent improvement in sensitivity at the level of $d'=1$ is depicted in Fig. 8b. Because the function in Fig. 8b defines 'ideal' probability summation for the 2AFC paradigm, we provide the values in tabular form in the Appendix for ready reference. Note that the signal + noise max distributions have to be computed by time-intensive numerical integration. We have therefore developed an approximation method (Chen & Tyler, 1999) that captures this function within 1% accuracy. (Pelli, 1985, had also considered this

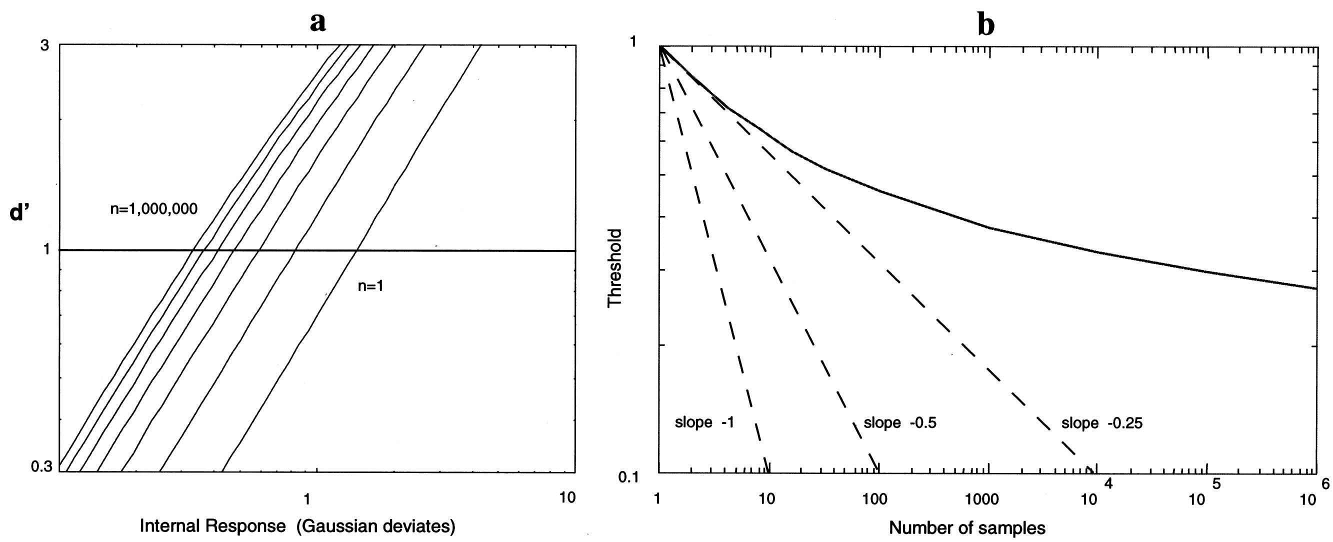

Fig. 8. (a) Theoretical d' functions under 2AFC probability summation assumptions. Note that the exponent (or steepness) is almost invariant with number of equally-sensitive channels monitored from $n=1$ to 1 million (assuming no uncertainty). 2AFC summation behavior is therefore essentially invariant with the d' criterion selected. (b) 2AFC probability summation over six decades on (unequal) double-log coordinates, compared with summation slopes for full summation (-1), for ideal observer summation (-0.5) and for Weibull summation assuming $\beta=4$ (-0.25). Note that the 2AFC summation function is never steeper than a slope of -0.25, and becomes extremely shallow for more than about ten samples.

function and provided an approximation formula that is accurate to within 20%.)

Thus, the complete analysis of 2AFC attentional summation over channels of equal sensitivity shows that the Ideal Attention operator provides dramatically different 'probability summation' behavior than that implied by Pelli's (1985) high-uncertainty approximation to High-Threshold Theory. At its steepest, this 2AFC function exhibits a slope of only about -0.25 (from one to four samples) and soon produces negligible summation for larger numbers of samples. The key reason for the difference between this prediction and that for the Weibull approximation is that the tails in the Gaussian distribution fall much more rapidly than exponential tail of the Weibull distribution. A summation mechanism that focuses on the information in this tail region will necessarily give different results for the two distributions. Justification for the ubiquity of the Gaussian distribution is discussed in the assumptions section of Section 1.

Consider the practical implication of the summation function of Fig. 8b. For most reported psychophysical tasks, the smallest stimulus might plausibly stimulate many local mechanisms. The expected starting point for a probability summation prediction would then be some way down this curve, say at the 10^2 level, beyond which little improvement is evident. Under the ideal attention assumption, the only way to achieve summation exponents even close to the reported values of around -0.25 (Watson, 1979; Robson & Graham, 1981; Williams & Wilson, 1981; Pelli, 1985) would be to assume that attention can be focused onto a single neural channel for the smallest stimulus in the series.

A major prediction of the High Threshold theory of probability summation is that the summation exponent can be predicted from the empirical exponent of the psychometric function measured during the summation experiment (Quick, 1974). This prediction has been borne out in several studies (Watson, 1979; Robson & Graham, 1981; Williams & Wilson, 1981; Pelli, 1985), but the result may be coincidental because none have *varied* the psychometric exponent to determine whether the summation exponent varies as predicted. Nevertheless, this analysis shows that the extent of 2AFC attentional summation varies even where the exponent of the psychometric function is invariant at a value close to one (Fig. 8a) and provides a much smaller improvement in sensitivity than is predicted by High-Threshold analysis for conditions yielding shallow exponents (Fig. 8b). Even the early part of the 2AFC attentional summation slope is never steeper than -0.25 (although it must be said that this corresponds to a value commonly assumed for the Weibull exponent, β). Studies that have assumed such a slope, therefore, would seem to have a valid estimate of the probability summation effects as long as the number of elements of equal sensitivity that they are summing remains less than about four. The analysis of Fig. 8 could therefore be regarded as validating the use of Minkowski summation with an exponent of 4 as long as the number of channels remains small and the other assumptions of the analysis are met.

Conversely, there is a major situation in which the summation slope remains unaffected while the psychometric steepness varies. This behavior can occur when the observing system monitors more channels than are being stimulated. This situation is conventionally described as the system having uncertainty as to which channels are being stimulated and is the topic of the next section.

4. Signal Detection Theory with channel uncertainty (and additive noise)

This section develops the implications of a variety of non-ideal attentional strategies for 2AFC in the additive noise case.

4.1. Channel uncertainty effects and their elimination by rescaling

Channel Uncertainty Theory is an elaboration of Signal Detection Theory in which the number of neural channels m monitored in the brain is greater than the number of channels n stimulated (by ratio $M = m/n$) (derived formally in Pelli, 1985). The level of uncertainty would then be defined as $\log_{10} M$ (assumed to be 0 up to this point in the treatment). (An equivalent theory of attentional distraction among the m channels, even where the observer is certain which channel is being attended, has been developed by Kontsevich & Tyler, 1999a.) For the present analysis, we assume that only one channel is being stimulated and that the decision is mediated by attention to successively larger numbers of channels in a non-ideal attentional strategy. Such behavior has been offered as an explanation for the relatively steep psychometric functions that are often measured in practice (Pelli, 1985; Kontsevich & Tyler, 1999a). The full lines in Fig. 9a show the d' functions obtained through the 2AFC derivation of Eqs. (13) and (14) for the certain condition (monitoring only one stimulated channel) and uncertain conditions (in which from 10 up to one million channels are monitored, with only one stimulated). The d' functions get progressively steeper in this operating range as channel uncertainty increases. The dashed lines in Fig. 9a show an analytic approximation to these d' functions that was fitted over the full set within the range from $d' = 0.5$ to 2 (i.e. within the practical measurement range). The approximation is a power function whose log slope U (straight dashed lines) is related simply to uncertainty ($\log_{10} M$) by the expression:

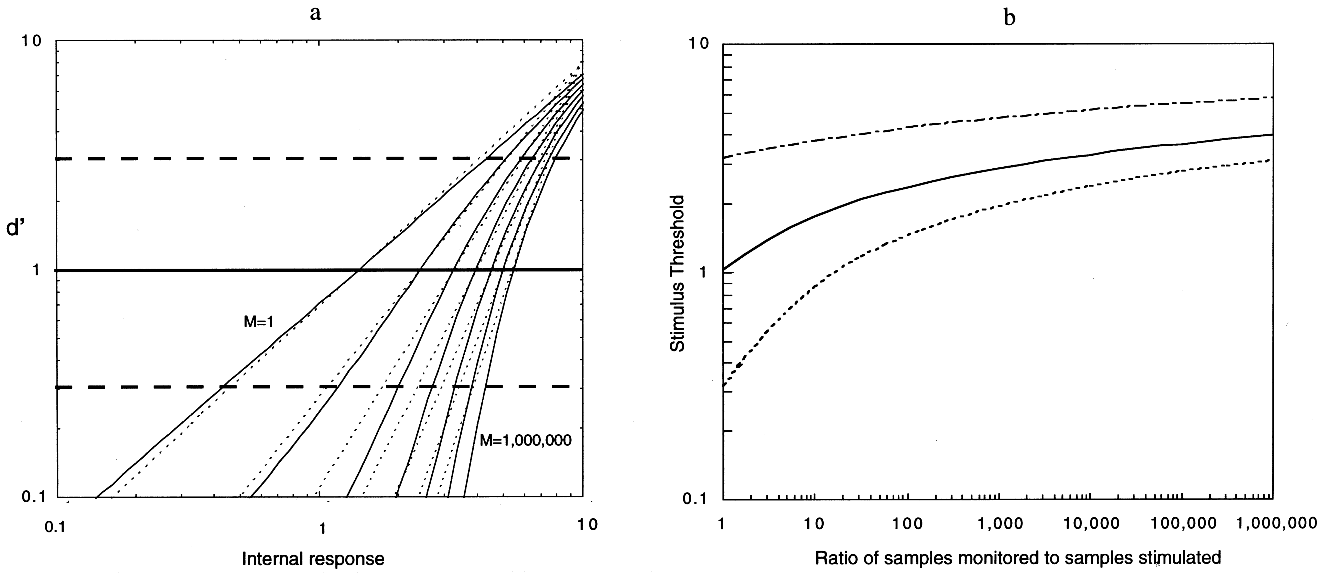

Fig. 9. (a) Log–log d' functions under various degrees of channel uncertainty ($M = 1–1\,000\,000$ in factors of ten running left to right) when one channel is stimulated. Dotted lines represent the least squares fit of equation (16) within the readily measurable range of $-0.5 < \log d' < 0.5$ (horizontal dashed lines). (b). Summation behavior with an attention window of increasing extent, due to increase of channel uncertainty as stimulus size increases over the array of local filter channels, at log d' levels of -0.5, 0 and 0.5 (three curves corresponding to the horizontal criterion lines in a). Threshold rises gradually at first as number of monitored samples is increased, then shows little further effect.

$$d' = A(s/s_0)^U, \quad \text{with } U = C + B\log_{10}M \qquad (16a,b)$$

where M is the ratio of monitored to stimulated channels and A, B, C and s_0 take values of 7.9862, 0.4468, 1.0779 and 9.5414.

The point of presenting this analytic approximation is that it allows reverse inference of the level of uncertainty from the log slope of the psychometric function, fitted to the data as a straight lines on double-logarithmic d' coordinates. Pelli (1985) had provided a similar approximation to a Monte Carlo simulation of the theoretical curves that we derive analytically, but his approximation was formulated in terms of a Weibull analysis and consequently appeared to emphasize the lower range of d' values, which are unmeasurable in practice. Our reanalysis focuses on the most accessible range of the psychometric function, that between log d' values of -0.5 and 0.5 (or percent correct values between about 60 and 90%). Fitting in this range generates fits at high levels of uncertainty that are substantially shallower than Pelli's. One can use our fitted function to derive the inferred uncertainty directly from equation (16), within the accuracy of the slope determination (Empirically, slopes may be determined with an accuracy of about 0.1 log units in 300 trials using the efficient Bayesian maximum likelihood algorithm proposed by Kontsevich & Tyler, 1999b; cf. Cobo-Lewis, 1997. This accuracy would imply a practical resolution of about 6 discriminable slopes in the slope range from 1 to 4).

The inverse equations for sensitivity at the criterion level of $d' = 1$ are straightforward:

$$M = 10^{(U-C)/B} \quad \text{and} \quad s = s_0 A^{-1/U} \qquad (17a,b)$$

Equivalently, channel uncertainty effects may be removed by extrapolating the measured slope of the log d' function up to $d' = 8$, then extrapolating back down a slope of $U = C$ to provide an estimate of the sensitivity that would have been obtained with no channel uncertainty. The extrapolation back to the level of $d' = 1$ may be approximated by dividing the measured threshold by a value of 8. This simplified procedure allows compensation of channel uncertainty effects with minimal computation, merely from knowledge of the log d' slope. For complete accuracy, the computed d' functions as depicted in Fig. 9 may be used to model of the psychometric function with no approximation. If a threshold estimate is required to be more accurate than the proposed approximation formula, the data for the psychometric function may be fitted over the family of computed d' functions to refine the compensation for channel uncertainty.

Of course, removing channel uncertainty does not imply eliminating measurement error in the estimates, only eliminating the *bias* in the threshold estimate introduced by channel uncertainty. The adjusted threshold estimates are no less variable, but threshold changes due to varying uncertainty levels are eliminated. In situations where the channel uncertainty remains constant across conditions, such bias reduction is not needed. But in cases where it may vary, such as summation functions over any stimulus domain, it is critical to partition the threshold variations between the underlying sensitivity variations and the effects of probability summation, as described in the following sections.

4.2. 2AFC attentional summation with a fixed attention window

The previous section considered the general case of estimating the degree of uncertainty from the psychometric function. With this analysis in hand, we may evaluate the particular case of the effect on threshold of varying *stimulus* extent with a *fixed* attention window. For studies that do not expend the effort required to measure the psychometric slope, it is important to have a model of the effects of uncertainty under plausible assumptions. Clearly, if the attention window can be matched to the stimulus extent, the uncertainty (or opportunities for distraction, see Kontsevich & Tyler, 1999a) will remain constant at zero and have no effect on the measured summation function. However, in this case the slope of the psychometric function should be low (assuming a linear transducer), which is known to be invalid in many situations.

Contrary to Robson and Graham's (1981) claim for this situation, 2AFC spatial probability summation effects with a fixed attention window are not proportional to $1/\beta$ (β being the exponent of the Weibull approximation, equation (1)). Such summation effects are controlled by the *change* in the exponent as uncertainty is reduced by increasing stimulus area (Fig. 10a); as seen Fig. 10b, the summation effects at $d'=1$ approximate a log slope of $-1/4$ over most of the range of ratios of stimulated to monitored samples. This result may be considered a justification for the widespread use of 4th power Minkowski rule to approximate probability summation. It is a quite different analysis from that developed by Williams and Wilson (1983), Robson and Graham (1981) and even Pelli (1985), since those analyses all assumed a fixed form of the log psychometric function. In contrast, the shape varies substantially in the fixed-attention-window model of Fig. 10a. Nevertheless, it may correspond to a plausible set of assumptions, so tabular values for the example depicted in Fig. 10 are provided in Table 1. The fixed-attention-window model is the main theoretical alternative to the probability summation effects of the ideal attention window of Fig. 8.

Thus, the curve of 2AFC attentional (or 'probability') summation in double-log coordinates may have either a concave or an approximately linear form according to whether the attention window is assumed to match the stimulus extent (Fig. 8) or to remain fixed (Fig. 10). The two forms are empirically distinguishable from threshold measurements alone. Note that, to provide the fourth-root approximation, the fixed attention window must be at least as large as the largest stimulus, and detection efficiency will necessarily be extremely low for the smallest stimuli. Because summation is only probabilistic within this large attention field, efficiency will still be low for stimuli filling the attention field. Thus, the assumption of the fourth-

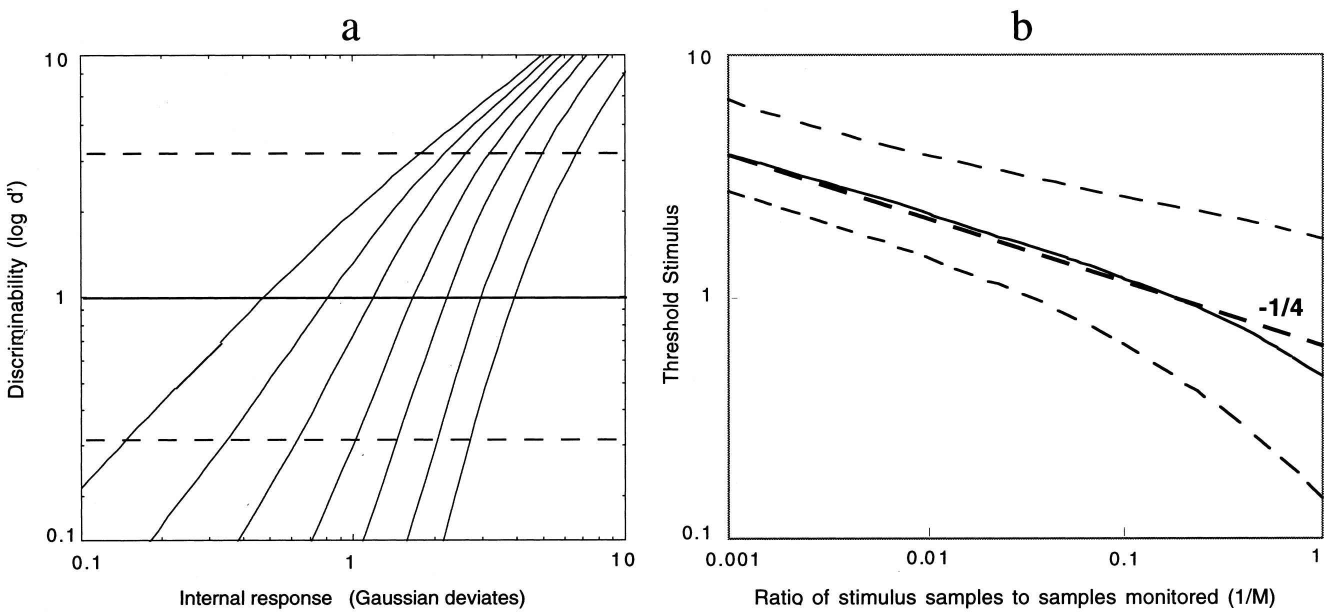

Fig. 10. Probability summation for varying numbers of samples within a fixed attention window (assumed here to allow a maximum of 1000 samples). (a) Psychometric functions in log d' versus log stimulus strength. Note similarity in shape to those in Fig. 9 but with extra shifts at high uncertainties. (b) Summation as a function of ratio of number of samples to total number monitored, at the three d' criteria indicated by the horizontal lines in (a). Thick dashed line in (b) depicts a slope of $-1/4$, which provides a good approximation to fixed-window probability summation over most of the computed range.

Table 1
2AFC max summation effects in relative threshold values[a]

Number of samples	Additive noise			Multiplicative noise
	Ideal attention window	Fixed attention window	Increasing attention window	Ideal attention window
1	1.0000	3.9311	1.0000	1.0000
2	0.8235	3.2978	1.2235	0.4522
3	0.7442	2.9836	1.3524	0.2896
4	0.6963	2.7823	1.4421	0.2143
5	0.6632	2.6354	1.5104	0.1713
6	0.6385	2.5202	1.5653	0.1437
7	0.6191	2.4278	1.6107	0.1244
8	0.6032	2.3503	1.6500	0.1103
9	0.5900	2.2823	1.6845	0.0995
10	0.5787	2.2245	1.7142	0.0910
20	0.5149	1.8695	1.9046	0.0530
30	0.4847	1.6827	2.0110	0.0402
40	0.4658	1.5566	2.0826	0.0336
50	0.4523	1.4626	2.1377	0.0294
60	0.4421	1.3889	2.1826	0.0265
70	0.4339	1.3278	2.2183	0.0244
80	0.4271	1.2763	2.2492	0.0227
90	0.4213	1.2313	2.2762	0.0213
100	0.4163	1.1924	2.3005	0.0203
200	0.3871	0.9475	2.4538	0.0147
300	0.3725	0.8158	2.5398	0.0124
400	0.3629	0.7275	2.6003	0.0111
500	0.3560	0.6620	2.6441	0.0102
1000	0.3367	0.4762	2.7793	0.0080
2000	0.3201		2.9078	0.0064
5000	0.3015		3.0720	0.0049
10 000	0.2893		3.1868	0.0041
20 000	0.2785		3.3005	0.0035
50 000	0.2658		3.4437	0.0029
100 000	0.2573		3.5475	0.0025
200 000	0.2495		3.6524	0.0022
500 000	0.2403		3.7793	0.0019
1 000 000	0.2339		3.8743	0.0017

[a] Results for fixed attention window assume a window size of 1000 samples.

power approximation to attentional summation carries the implication that the neural system is operating at low efficiency, and is not applicable to situations where efficient detection performance is demonstrable.

4.3. Two-component summation and channel analysis

A classic case in both spatial and color vision is the summation for the detection of two stimulus components as their intensities are varied relative to each other. The results of this paradigm are plotted on a dual axis plot of the contrast threshold for the pair of components when combined in a variety of ratios (Guth, 1967; Graham & Nachmias, 1971; Stromeyer & Klein, 1974; Yager, Kramer, Shaw, & Graham, 1984). These references should be consulted for the theoretical development, but various outcomes are summarized in Fig. 11a. If the two components are detected entirely independently (by a noiseless max rule), the detection contour forms a square corner (independent channels); if they are added linearly in a single mechanism limited by late noise (linear summation), the detection contour is a negative oblique line; if they are combined linearly but detection is limited by independent sources of Gaussian noise in the two channels, the detection contour is a circular arc (squaring).

Two-component summation is an important case for channel analysis in general, because it represents the combination rule between adjacent channels for detection by sets of channels in any domain. Channel summation is often modeled as a fourth-power Minkowski rule (or pth norm) for combination over channels (Graham & Nachmias, 1971; Stromeyer & Klein, 1974; Williams & Wilson, 1981; Wilson, McFarlane, & Phillips, 1983; Yager et al., 1984). The justification for this rule is usually expressed in terms of Weibull analysis, which we show to be on shaky grounds, but the situation may be reanalyzed for the 2AFC paradigm with Gaussian noise.

The strict 2AFC probability summation prediction is based on the case where the system takes the max on every trial, for two channels with additive Gaussian noise, as the relative signal strength is varied between the channels (Fig. 12). The analysis of this situation is essentially an uncertainty analysis because the observer always monitors both channels as the stimulation progresses from one channel alone through both together to the other alone. Even at the extremes, where only one channel is stimulated, the joint signal always has to exceed the max of the noises in both channels. The applicable distributions are plotted in Fig. 12 in terms of both the max response over the two channels in each test interval and the difference response δ between the signals in each pair of 2AFC intervals. The response distributions for the two intervals are set so that the detectability for an individual channel falls at the 75% correct position. The 2AFC uncertainty prediction may be developed for a full range of component ratios, as is shown by the full curve in Fig. 11a. This prediction is part-way between the linear and the square-law summation rule of linear summation over sources with independent additive noise. In fact, it is well-described by a pth norm (Minkowski summation rule) with a power of 1.5 in the case of linear d' functions (i.e. involving no additional uncertainty about the stimulus properties).

Although most reported cases of two-component summation under the 2AFC paradigm show less complete summation than this probability summation prediction, their analysis requires consideration of the free parameter of the slope of the psychometric function, which is rarely specified in published studies of two-component summation. When the slope is steep, one interpretation is that there is much additional uncertainty, i.e. the observer is monitoring many more channels of whatever kind than are being stimulated (Fig. 9a). Quantitatively speaking, most 2AFC studies report the exponent of the d' function to be close to 2 in the fovea (e.g. Stromeyer & Klein, 1974), which implies a ratio of monitored to stimulated channels of $M = 116$ (with $n = 1$). As can be seen in Fig. 11b, this power of 2 assumption produces a curve matching a Minkowski exponent between 3 and 4.

Another commonly reported value of the d' exponent is 3 (approximating reports of the Weibull β from 3.5 to 4). This slope requires a channel monitoring ratio of $M = 20\,000$, but this large increase generates a curve that is sharper than that for the exponent of 4. Beyond this range, double-precision computation was no longer capable of computing the required max distributions for the Gaussian function, but we could use the analytic approximation in the form of the Poisson distribution that we developed for this purpose (Chen & Tyler, 1999). The resulting curves for d' exponents of 4 and 6 show that, again, there is very little change in the shape of the curve. Thus, one can conclude that there is no

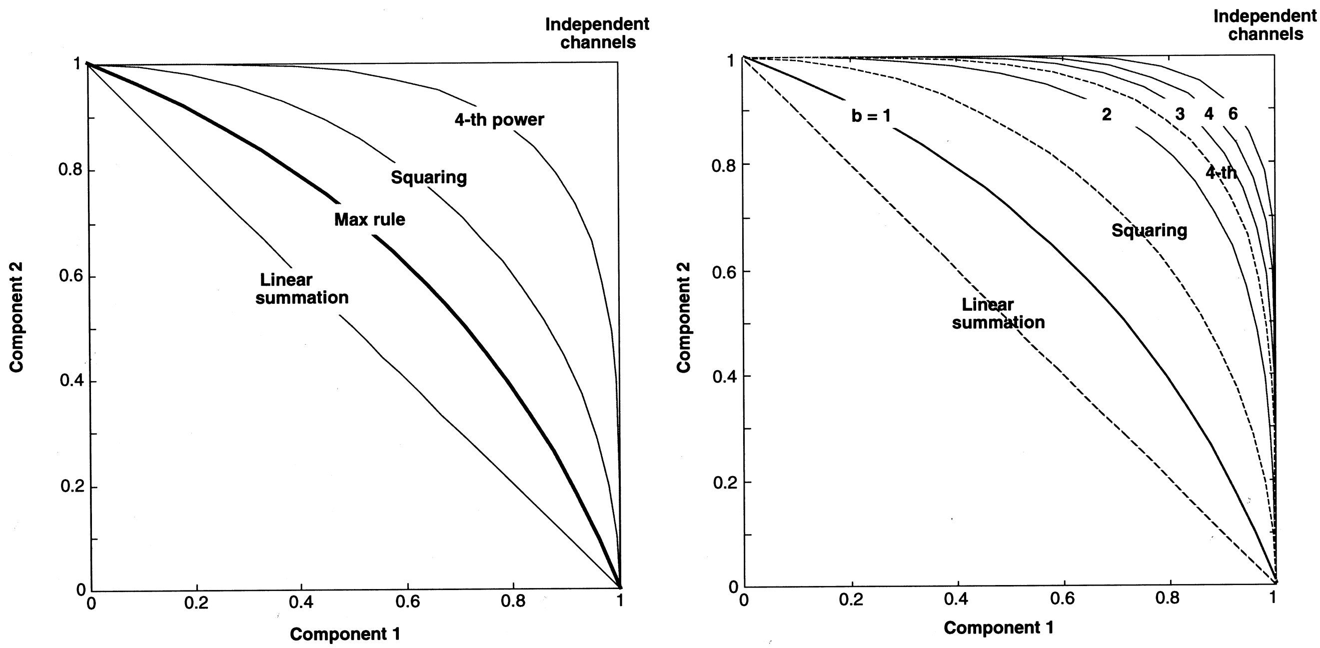

Fig. 11. (a) Theoretical functions for two-component summation under various combination rules, assuming linearity of the psychometric functions. Thin curves: Minkowski summation with exponents of one, two and four and independent channels. Thick curve: max rule. (b) Thin curves show theoretical functions for two-component summation assuming accelerating psychometric functions with the levels of uncertainty required to approximate psychometric slopes of 2, 3, 4 and 6. Dashed curves show Minkowski summation with exponents of 1 (linear), 2 (squaring) and 4 for comparison. Note that even a steep slope of six departs substantially from the corner prediction for independent channels, so that one can expect to determine accurately whether two channels are fully independent or subject to some kind of combination rule.

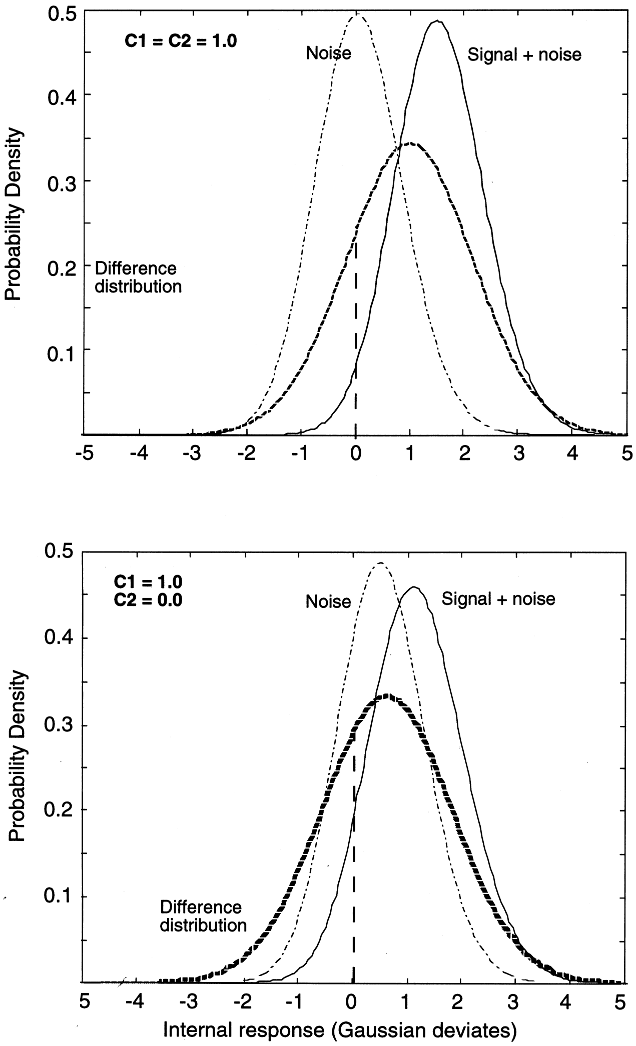

Fig. 12. Derivation of the 2AFC probability summation prediction for two-component summation. Dash-dotted curves: max distribution for the two channels for the noise-alone interval; full curve: max distribution for the two channels for the signal + noise interval; heavy dashed curve: distribution of the differences between the intervals. (a) Both channels equally stimulated. (b) Only one channel stimulated.

plausible degree of steepness of the psychometric function that will push the curve to the corner of the box if probability summation is operating. Finally, it should be noted that the effect of uncertainty in increasing the exponent is essentially equivalent to the same change in exponent from an accelerating threshold nonlinearity.

In conclusion, the analysis of the two-component summation paradigm in terms of the Minkowski summation rule (Eq. (13)) provides an adequate approximation to the 2AFC behavior in additive Gaussian noise, as long as the Minkowski exponent is not misinterpreted according to High-Threshold Theory.

5. Effects of multiplicative noise

5.1. Multiplicative noise makes the psychometric function shallower

Instead of the classical assumption of additive noise, analysis of the noise in cortical neurons suggests that it may have a multiplicative component, with σ_R in Eq. (10) increasing according to some function of the strength of the mean signal R (Tolhurst et al., 1981). In a general expression, the total noise in the signal may be expressed by a power relation:

$$\sigma_R \propto kR^q + \sigma_N \tag{18}$$

The additive constant σ_N represents some irreducible level of noise that is present even when there is no signal, when the multiplicative component kR^q will fall to zero. Such additive noise is a physiological requisite because no real system is noise-free.

Other than scaling the noise according to Eq. (18), the multiplicative analysis employs the identical analytic structure of Signal Detection Theory developed in Section 3, but the results are very different. The presence of multiplicative noise radically alters the expected shape of the psychometric function derived by inserting Eq. (18) into Eq. (15), in both the absence and the presence of channel uncertainty. Even in the absence of uncertainty, the log-log steepness of the psychometric function changes according to the rate of increase of noise with stimulus intensity. If the exponent q of this rate of increase is 0.5, as in Poisson noise (which governs the quantal fluctuations of light, for example), the psychometric slope for a single channel goes to about 0.5 (Fig. 13a, bold curve), a striking deviation from what is seen empirically in psychophysical measurements. (Eq. (18) assumes that the noise distribution is Gaussian rather than strictly Poisson, a good approximation for high mean levels of quantal events.) Note that a slope of $q = 0.5$ represents a tremendously shallow increase of d' with stimulus strength, implying that the measurable range of the psychometric function extends over as much as two log units, the entire visible contrast range.

5.2. Dramatic probability summation with multiplicative noise

If we evaluate the effects on sensitivity of taking the max of n equally-sensitive channels in the presence of root-multiplicative Gaussian noise, the results are also profound (Fig. 13b). Summation over the first ten channels actually exceeds the amount expected for linear summation in additive noise (see tabular values specified for this condition in Table 1). This result seems counterintuitive, but it arises because any decrease in the signal provides a concomitant decrease in the accompanying noise level. Since the effect of probability summation is to allow a small decrease in the signal to begin with, the multiplicative reduction in the noise provides a further enhancement of the signal/noise ratio, resulting in the net improvement depicted in Fig. 13b.

This dramatic degree of probability summation under root-multiplicative noise conditions has powerful implications for neural processing, which seems to be governed by this type of principle throughout the cortex (Tolhurst, Movshon, & Thompson, 1981; Tolhurst, Movshon, & Dean, 1983; Vogels, Spileers, & Orban, 1989). Shadlen and Newsome (1998) point out that the multiplicative behavior makes the signals at individual neurons so noisy that they cannot account for the discriminative behavior of the animal as a whole, even if the neuron's response is optimal for the local stimulation employed. They estimate that the activity must be integrated over 50–100 neurons to account for the observed behavior, implying that the signal/noise ratio of the optimal neuron is about a log unit below the required level.

However, the plot in Fig. 13b implies that a different strategy is available under root-multiplicative noise conditions. Instead of integrating the activity of 100 neurons, and losing the potential specificity available from the elements of that assemblage, the cortex could monitor the activity of just ten *relevant* neurons. Taking the *max* of the ten responses gives the required boost of a factor of 10 in signal/noise ratio, equivalent to *summing* over 100 neurons. Thus, a much smaller pool is required for the same gain in detectability, if the brain is capable of implementing a max rule. Such implementation seems plausible because it is the core operation of an attentional process, for which there is much behavioral and increasing neurophysiological evidence. In fact, a simple neural threshold has the effect of implementing a max rule in a psychophysical task where the stimulus is reduced until the last response of the most sensitive neuron carries it. (The detailed effects of a hard threshold on 2AFC performance, which are beyond the scope of the present treatment, are discussed in Kontsevich & Tyler, 1999c.)

5.3. Disambiguating multiplicative noise and uncertainty

Inclusion of channel uncertainty in the case where the noise is somewhat multiplicative has similar effect to the case of additive noise (see Fig. 9a), except that the entire fan of uncertainty functions is rotated to become shallower. Fig. 14 plots sample psychometric functions for square-root-multiplicative noise (the case of $p = 0.5$ in Eq. (18)). At first sight, it might seem that this result implies that the empirical effects of the noise multiplier and uncertainty would be hard to disentangle. However, notice that the steepening effects of uncertainty in Fig. 14 are much reduced at high levels of d'. Thus, the steepness at high d' (say, above the level of $d' = 2$), are highly diagnostic of the degree to which the noise is multiplicative. If the fitted slope in this region of the unmasked psychometric function is 1 or above, as in Fig. 9a, the implication is that the major component of the noise is additive. A high d' slope significantly less than 1 (Fig. 14), on the other hand, is strong evidence for multiplicative noise operating in the near-threshold region.

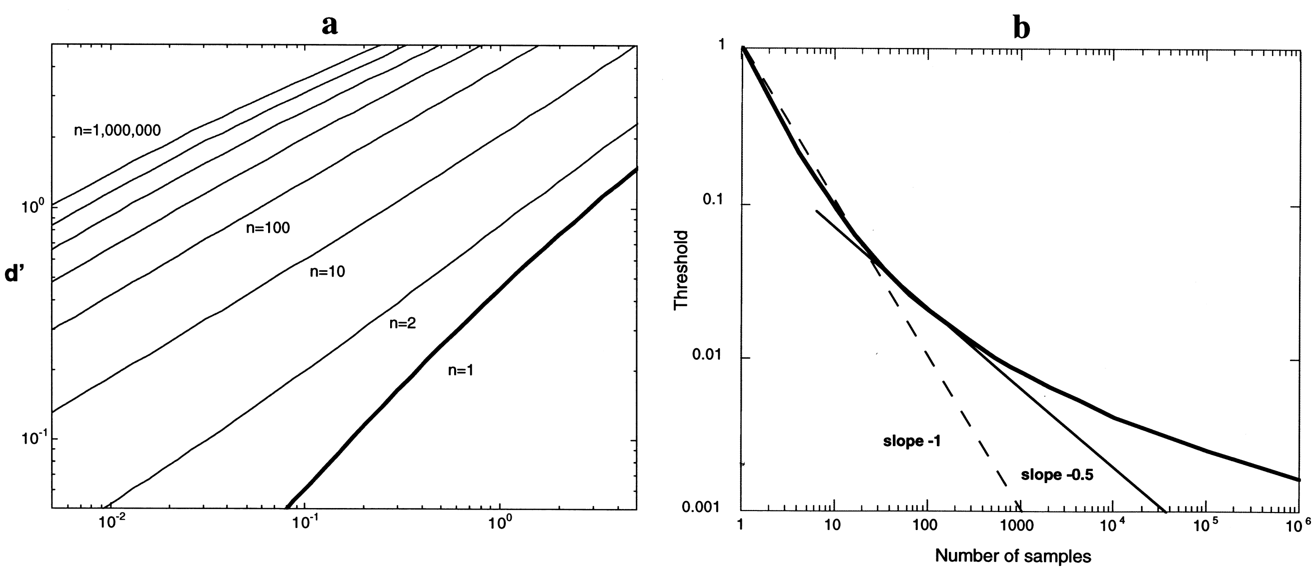

Fig. 13. Effects on probability summation of assuming square-root multiplicative noise according to Eq. (18) with $p = 0.5$. (a) Shallower d' functions. (b) Dramatically enhanced summation behavior for threshold stimulation at the criterion of $d' = 1$ that is even supralinear for small numbers (Upper and lower criterion levels are omitted for clarity).

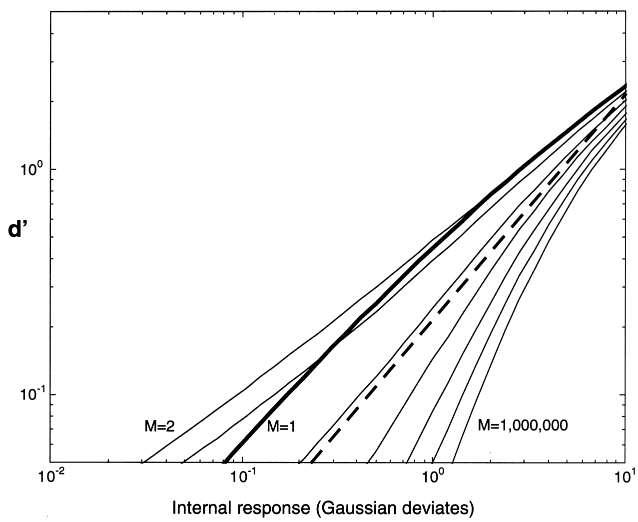

Fig. 14. Effects of uncertainty on the steepness of the d' function with the root-multiplicative noise assumption, with various values of M. Unlabeled curves have monitoring ratios in factors of ten from ten to 100 000. Note that a high level of uncertainty is required before the fitted slope approaches 1.

There is a curious crossover in the functions in Fig. 14 at low d', where the curve for no uncertainty actually shows a slightly higher threshold than the curves for low uncertainty. This result may seem counterintuitive, but it arises from the necessary assumption that there is an additive component to the noise (Eq. (18)), which tends to reduce sensitivity as the multiplicative component approaches the level of the additive component at low contrasts. As uncertainty increases, it de-emphasizes the role of the additive noise, effectively increasing the sensitivity. Without this additive component, the log slope fitted to the psychometric function would be at 0.5 even with no uncertainty. However, there must always be a noise component that is additive with respect to contrast due to the existence of quantal noise in the stimulus and thermal noise in the receptors. Because these curves are governed by two free parameters, the particular summation function at some level on the curves is not of canonical interest, and summation functions are therefore not plotted for this case. If such multiplicative noise is implicated in detection behavior, the role of additive and multiplicative noise components must be estimated by measurement of full psychometric functions.

5.4. Fully multiplicative noise introduces psychometric saturation

A more extreme form of multiplicative noise is the case where the noise σ_R is directly proportional to the stimulus strength ($q = 1$ in Eq. (18)). Direct proportionality is not implausible, as such a form occurs in the case of noise due to eye movement fluctuations over the image being viewed. Whatever the distribution of eye movements, they will introduce some level of fluctuation in the response of a local filter viewing any kind of contrast stimulus. The resulting fluctuation is a form of noise that is necessarily in direct proportion to the stimulus contrast (assuming that the eye movements are independent of contrast). This property of direct proportionality may be shown analytically in terms of the temporal waveform of the signal fluctuation of the output of each linear filter $I_{k_i}(x,y)$ responding some stimulus $S(x,y)$, such as a sinusoidal grating, projected on to the moving retina.

$$\begin{aligned} r_i(t) &= I_{k_i}(x,y) \otimes s \cdot S(x - \Delta x(t), y - (\Delta y(t))) \\ &= s \cdot [I_{k_i}(x,y) \otimes S(x - \Delta x(t), y - (\Delta y(t)))] \end{aligned} \quad (19)$$

where $\otimes$ is the convolution operator, $\Delta x(t)$, $\Delta y(t)$ is the retinal shift over time and s is the scaling constant of stimulus strength.

Thus, for a given filter and eye-movement sequence, the filter output $r_i(t)$ is directly proportional to the contrast of the stimulus, because convolution is a linear operation. We may treat the response of the filter derived from such eye movements as a noise source by determining its standard deviation σ_E computed over some temporal window $t_1{:}t_2$ according to

$$\sigma_E = \left[\int_{t_1}^{t_2} \left(r_i(t) - \int_{t_1}^{t_2} r_i(t) \mathrm{d}t \right)^2 \mathrm{d}t \right]^{0.5} \propto r_i(t) \propto s \quad (20)$$

The standard deviation of this source of noise is thus directly proportional to the contrast of the background display. Such proportional noise will tend to overtake other sources of noise that do not increase so rapidly with contrast, and will therefore tend to dominate at high contrast. We are not aware of any previous consideration of such noise.

The effect of proportional multiplicative noise ($q = 1$ in Eq. (18)) on the form of the psychometric functions is shown in Fig. 15. The fitted slopes become even shallower than in the case of square-root noise (Fig. 14) when the signal rises out of the additive noise regime, where the slopes approximate unity. As stimulus strength increases, the effect is to make the functions asymptote to a constant d' level, with no further improvement in sensitivity at high stimulus strengths. This horizontal asymptote thus becomes a conspicuous signature of the presence of full multiplicative noise. Such behavior has rarely been seen in psychometric functions for contrast detection (e.g. it is not evident in the high-contrast study of Foley & Legge, 1981), suggesting that this type of multiplicative noise is not a usual feature of contrast detection tasks. However, it is not clear that previous workers have designed their studies for careful evaluation of this high d' region of the psychometric function, so there is room for further evaluation of particular situations of interest before the case may be considered to be settled. For example, although such noise might not be expected in a simple detection task, it might plausibly be found in a difficult discrimination task where the contrast threshold is being measured as a function of a slight spatial difference between two stimuli, with long-duration presentations allowing eye-movement-generated noise in the pedestal stimulus to become a significant factor limiting discrimination performance.

For summation over increasing numbers of mechanisms, the curves of Fig. 15 show that the additive noise regime (approximating a slope of 1) tends to dominate the domain of measurable (and computable) range of d' functions. As a result, derivation of a summation curve for this case is relatively meaningless because its form would depend on the exact ratio of additive to multiplicative noise assumed. When in the domain dominated by the fully multiplicative noise (horizontal leg of curves), summation is indeterminate because reduction of the signal would be accompanied by a proportionate reduction of the noise, and signal/noise ratio (discriminability) would be maintained at a constant level. In the full-multiplicative noise regime, therefore, discriminability is insensitive to the signal level, and threshold cannot be determined. Only when the signal level is finally reduced into the domain dominated by additive noise (the left-hand region of Fig. 15) would summation revert toward the form depicted in Fig. 8b.

6. Conclusion

Psychophysical measures of summation are widely used as indexes of underlying integrative mechanisms in visual processing. The preceding analysis provides a rigorous approach to the universe of such mechanisms, detailing the properties of physiological summation, ideal observer summation and attentional summation in the 2AFC detection paradigm, for situations of both additive and multiplicative noise limiting the detection task. The key difference between these three types of summation is the type of attention process accessing an array of filters. If attention accesses a single filter, the physiological summation within that filter predominates; if attention can switch among filters matching each stimulus, ideal observer behavior occurs; if attention can access the max response of an array of filters, the result has been described as probability summation, but we favor the term 'attentional summation' for this case.

If the noise limiting detection behavior is additive, ideal observer summation proceeds with a log slope of $-1/2$ (i.e. as long as there are channels available matching the extent of each stimulus). For sets of local channels of equal sensitivity, ideal attentional summation approximates the fourth root of the number of channels up to about four channels, then proceeds at a diminishing rate thereafter. The maximum improvement asymptotes to about a factor of 4 up to a million channels. The log slope of the psychometric function remains virtually invariant at 1 in this regime. If attention is paid to a fixed (large) number of channels, attentional summation operates under a regime of varying channel uncertainty. In this fixed-attention paradigm, attentional summation approximates a power of $-1/4$ over a wide range, although the log slope of the psychometric function varies.

If the noise in the detection task is multiplicative with some positive power of stimulus contrast, the psychometric function for d' becomes much shallower than for additive noise, with the d' exponent approximating the multiplicative power at high d' levels, even in the presence of channel uncertainty. The consequent increase in sensitivity attributable to attentional summation is, correspondingly, radically increased to levels matching full physiological summation behavior. The interpretation of summation curves therefore depends critically on the form of the d' function measured in the same paradigm.

Noise from eye movements and similar sources of retinal jitter will generate contrast noise whose amplitude is likely to increase linearly with contrast. When such noise becomes the dominant source of noise, it will preclude the measurement of sensitivities because signal/noise ratio becomes independent of signal level.

The analysis shows that a full treatment of 2AFC summation is necessary because previous approaches are poor approximations in many situations (though generally adequate for the cases for which they were designed). It is hoped that the new level of precision of

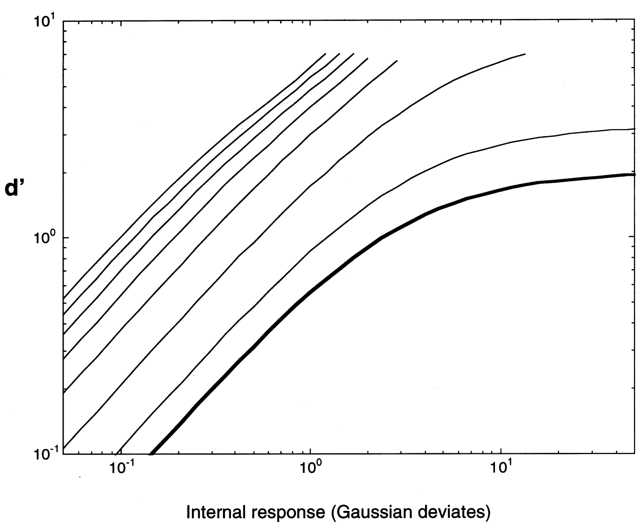

Fig. 15. Saturating d' functions obtained assuming linearly multiplicative noise according to equation (18) with $p = 1$. Thick curve: one mechanism, next curve to left: two mechanisms, successive leftward curves from ten to 1 000 ,000 in factors of ten.

these summation analyses will clarify many issues that have remained clouded up until now, and will inspire further rigorous evaluation of the detailed make-up of human spatial vision. At least one effort in this direction is under way under the rubric of the Modelfest project, a joint effort to collect psychophysical data across many laboratories and provide them on the Internet (at www.neurometrics.com/projects/Modelfest/IndexModelfest) as a publicly available testbed for models of spatial vision. A description of the philosophy and progress of the Modelfest project has been published (Carney et al., 1999). It is to be hoped that all candidate models will adhere to the summation principles developed in the present treatment, or provide empirical tests for deviations from its underlying assumptions.

Acknowledgements

Our thanks to Lenny Kontsevich for formulating Appendix A and to Marius Usher, Tim Meese and two anonymous reviewers for detailed comments on the mathematical exposition. Supported by NEI grant R01 EY 7890.

Appendix A. Derivation of the Square-Root Law of ideal summation

Leonid L. Kontsevich

Suppose that the generic stimulus profile (with a unity area A = 1) is provided by a function $S(x,y)$ and the inputs for the matched filter are provided by a square grid of the samples (x_j,y_k) (where $x_j = \varepsilon j$, $y_k = \varepsilon k$, with $j,k \in Z$, where Z is the set of all integers, and ε is the grid step). Each input has noise standard deviation σ. In general, the response of the matched filter can be approximated as the weighted sum of its responses to the samples

$$\boldsymbol{R} = \sum_{j,k \in Z} \boldsymbol{S}(x_j,y_k) \cdot \boldsymbol{I}(x_j,y_k) = \sum_{j,k \in \boldsymbol{Z}} \boldsymbol{S}^2(x_j,y_k)$$
$$\approx (1/\varepsilon^2) \int_{-\infty}^{\infty} \boldsymbol{S}^2(x_j,y_k) \mathrm{d}x \; \mathrm{d}y$$

and the signal variance as the weighted sum of the local variances

$$\mathrm{Var}(\boldsymbol{R}) = \sum_{j,k \in \boldsymbol{Z}} (\boldsymbol{S}(x_j,y_k)\sigma)^2 \approx (\sigma^2/\varepsilon^2) \int_{-\infty}^{\infty} \boldsymbol{S}^2(x_j,y_k) \mathrm{d}x \; \mathrm{d}y,$$

because $\lim_{\varepsilon \to 0} \Sigma_{i \in \boldsymbol{Z}} f(\varepsilon_i)\varepsilon = \int_{-\infty}^{\infty} f(x)\mathrm{d}x$ according to the Riemannian definition of the integral. The discriminability of the ith stimulus, d'_i, therefore, can be approximated as the reciprocal of the standard deviation time the sampling interval

$$d'_i = \boldsymbol{R}_i/\sigma_{\boldsymbol{R}_i} \approx 1/\sigma\varepsilon_i.$$

Consider now what would happen to the signal-to-noise ratio if the stimulus is magnified by a factor of a_x along the axis x and a_y along the axis y, i.e. its area is increased by a factor $\boldsymbol{A}_i = \boldsymbol{a}_x\boldsymbol{a}_y$. The stimulus (and the matched filter) profile in this case is given by $S_i(x/a_x, y/a_y)$, and the approximations for the response and its variance correspondingly are

$$\boldsymbol{R}_i(\boldsymbol{a}_x,\boldsymbol{a}_y) = \sum_{j,k \in Z} \boldsymbol{S}_i^2(x_j/\boldsymbol{a}_x, y_k/\boldsymbol{a}_y)$$
$$\approx (\boldsymbol{a}_x\boldsymbol{a}_y/\varepsilon_i^2) \int_{-\infty}^{\infty} \boldsymbol{S}_i^2(x_j,y_k) \mathrm{d}x \; \mathrm{d}y$$

and

$$\mathrm{Var}(\boldsymbol{R}_i(\boldsymbol{a}_x,\boldsymbol{a}_y)) \approx (\boldsymbol{a}_x\boldsymbol{a}_y\sigma^2/\varepsilon_i^2) \int_{-\infty}^{\infty} \boldsymbol{S}_i^2(x_j,y_k) \mathrm{d}x \; \mathrm{d}y.$$

The discriminability then becomes

$$d'_i = \frac{\boldsymbol{R}_i}{\sigma_{\boldsymbol{R}_i}} \approx \frac{(\boldsymbol{a}_x\boldsymbol{a}_y/\varepsilon_i^2) \int_{-\infty}^{\infty} \boldsymbol{S}_i^2(x_j,y_k) \mathrm{d}x \; \mathrm{d}y}{\left((\boldsymbol{a}_x\boldsymbol{a}_y\sigma/\varepsilon_i^2) \int_{-\infty}^{\infty} \boldsymbol{S}_i^2(x_j,y_k) \mathrm{d}x \; \mathrm{d}y \right)^{1/2}}$$
$$= \frac{(\boldsymbol{a}_x\boldsymbol{a}_y)^{1/2}}{\sigma\varepsilon_i} \propto \boldsymbol{A}_i^{1/2},$$

i.e. it is proportional to the square root of stimulus area.

Appendix B. Symbols used in text

s	test stimulus strength (for a spatial periodic target, s would be contrast)
t	time variable
v	test interval choice variable
s	index for signal interval
n	index for null interval
i	index variable over the set of summation field responses
j,k	index variables over a set of inputs over local positions within a summation field
$f()$	an unspecified monotonic function
$r_i(t)$	internal response to the test stimulus in the ith local mechanism over time
r'	dummy integration variable on variable r
x,y	two-dimensional index variables of the local positions of retinal inputs
a_x,a_y	scaling factors in x and y directions
A_i	effective area of the ith summation field
R	mean internal response to the test stimulus at the decision site (or a

	corresponding location parameter in the case of the Weibull formalism)
R_i	mean internal response to the test stimulus in the ith local mechanism (or a corresponding location parameter in the case of the Weibull formalism)
R_θ	internal response threshold below which no response information is transmitted to the decision site
R	Pearson product moment correlation coefficient
δ	difference between internal response variables r_s and r_n
ε_i	sampling distance between local inputs scaled in relation to the width of the ith summation field
$p(v=x\|<g>)$	probability that the observer's response indicates interval x given stimulus state g
σ	standard deviation of the additive noise on each local input
σ_N	standard deviation of the additive noise, i.e. the residual noise when $R=0$
σ_R	standard deviation of all operative noise components for nonzero mean response strength R; will equal σ_N for a system governed by additive noise
σ_E	standard deviation of full-multiplicative noise due to eye movements over a stimulus
d'	discriminability of signal from null response scaled in terms of R/σ_R
β	exponent controlling the psychometric function steepness in the Weibull formulation
b	exponent controlling the steepness of the psychometric function in the Signal Detection formulation
p	exponent for summation of information from independent channels according to the Minkowski rule
q	exponent controlling the multiplicative relation of an internal noise component with internal response strength
s_0	sensitivity constant for d' fit
n	number of local mechanisms stimulated
m	number of local mechanisms monitored by the attention mechanism
M	ratio of m/n
A	scaling constant for uncertainty approximation
B	log scaling constant for uncertainty approximation
C	additive constant for uncertainty approximation
U	uncertainty exponent for uncertainty approximation
$S(x,y)$	profile of stimulus over spatial dimensions x,y
$I_i(x,y)$	weighting function for an ideal filter mediating detection of the ith stimulus, matching the stimulus over spatial dimensions x,y (and all the other properties)
$\Delta x(t),\Delta y(t)$	shift of retinal position of the stimulus in time over spatial dimensions x,y
$\Psi(s)$	cumulative probability distribution of the observer's responses; assumed to be proportional to $\Psi(R)$ in most of the present treatment
$\Psi_i(R_i)$	the cumulative probability distribution that would be obtained if the observer's response were based on the output of the ith local mechanism
$D(r;R,\sigma)$	PDF of some noise source r with mean R and standard deviation σ
$G(r;R,\sigma)$	PDF of a Gaussian noise source r with mean R and standard deviation σ
$\Phi(r;R,\sigma)$	CDF of a Gaussian noise source r with mean R and standard deviation σ
$\Phi^{-1}(r;R,\sigma)$	inverse CDF (or erf) of a Gaussian noise source r with mean R and standard deviation σ
$P(r;R,\sigma)$	PDF of a Poisson noise source r with mean R and standard deviation σ
$D_\beta(r;R)$	PDF of an additive noise source r with mean R generating a Weibull CDF with exponent β
$Z_s(\delta;\Delta)$	normalized PDF of the difference response δ between two test intervals, with mean Δ in units of its standard deviation
$M_n(0,\sigma_N)$	PDF of samples of the max of a noise-alone event over n channels with zero mean response and standard deviation σ_N
$M_{n,m}(R,\sigma_R)$	PDF of samples of the max of a signal-plus-noise event over m channels with (equal) mean responses R and standard deviation σ_R, together with n-m noise-alone channels of zero mean response and standard deviation σ_N

References

Carney, T., Klein, S. A., Tyler, C. W., Silverstein, A. D., Beutter, B., Levi, D., Watson, A. B., Reeves, A. J., Norcia, A. M., Chen, C.-C., Makous, W., & Eckstein, M. P. (1999). The development of an image/threshold database for designing and testing human vision models. *Proceedings of the SPIE*, *3644*, 542–551.

Chen, C.-C., & Tyler, C. W. (1999). Accurate approximation to the extreme order statistics of Gaussian samples. *Communication in Statistics*: *Simulation and Computation*, *28*, 177–188.

Cobo-Lewis, A. B. (1997). An adaptive psychophysical method for subject classification. *Perception & Psychophysics*, *59*, 989–1003.

Finney, D. J. (1952). *Probit analysis: a statistical treatment of the sigmoid response curve* (2nd ed.). Cambridge: Cambridge University Press.

Foley, J. M., & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research*, *21*, 1041–1053.

Gorea, A., & Tyler, C. W. (1997). New look at Bloch's law for contrast. *Journal of the Optical Society of America*, *A14*, 52–61.

Graham, N., & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channels models. *Vision Research*, *11*, 251–259.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Guth, S. L. (1967). Nonadditivity and inhibition among chromatic luminances at threshold. *Vision Research*, *7*, 319–328.

Kersten, D. (1984). Spatial summation in visual noise. *Vision Research*, *24*, 1977–1990.

Kontsevich, L. L., & Tyler, C. W. (1999a). Distraction of attention and the slope of the psychometric function. *Journal of the Optical Society of America*, *A16*, 217–222.

Kontsevich, L. L., & Tyler, C. W. (1999b). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*, 2729–2737.

Kontsevich, L. L., & Tyler, C. W. (1999c). Nonlinearities of near-threshold contrast transduction. *Vision Research*, *39*, 1869–1880.

Macmillan, N. A., & Creelman, C. D. (1993). *Detection theory: a user's guide*. Cambridge, UK: Cambridge University Press.

Mayer, M. J., & Tyler, C. W. (1986). Invariance of the slope of the psychometric function with spatial summation. *Journal of the Optical Society of America*, *A3*, 1166–1172.

Palmer, J., Ames, C. T., & Lindsey, D. Y. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology*: *Human Perception and Performance*, *19*, 108–130.

Pelli, D. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America*, *A2*, 1508–1531.

Pelli, D. (1987). On the relation between summation and facilitation. *Vision Research*, *27*, 119.

Quick, R. F. (1974). A vector-magnitude model of contrast detection. *Kybernetik*, *16*, 65–67.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, *21*, 409–418.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation and information coding. *Journal of Neuroscience*, *18*, 3870–3896.

Stromeyer, C. F. III, & Klein, S. A. (1974). Spatial frequency channels in human vision as asymmetric (edge) mechanisms. *Vision Research*, *14*, 1409–1420.

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.

Tolhurst, D. J, Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research*, *41*, 414–419.

Tyler, C. W. (1997). Why we need to pay attention to psychometric function slopes. *Vision Science and Its Applications*, *Technical Digest*, *1*, 240–243.

Vogels, R., Spileers, W., & Orban, G. A. (1989). The response variability of striate cortical neurons in the behaving monkey. *Experimental Brain Research*, *77*, 432–436.

Watson, A. B. (1979). Probability summation over time. *Vision Research*, *19*, 515–522.

Wiener, N. (1949). *Extrapolation*, *interpolation and smoothing of stationary time series*, *with engineering applications*. Cambridge, MA: MIT Press.

Williams, D. W., & Wilson, H. R. (1983). Spatial frequency adaptation affects spatial probability summation. *Journal of the Optical Society of America*, *73*, 1367–1371.

Wilson, H. R., McFarlane, D. K., & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, *23*, 873–882.

Yager, D., Kramer, P., Shaw, M., & Graham, N. (1984). Detection and identification of spatial frequency: models and data. *Vision Research*, *24*, 1021–1035.