Classification Image Weights and Internal Noise Level Estimation

Albert J. Ahumada, Jr.   NASA Ames Research Center, Moffett Field, CA, USA

Abstract

For the case of linear discrimination of two stimuli in white Gaussian noise in the presence of internal noise, a method is described for estimating linear classification weights from the sum of noise images segregated by stimulus and response. The recommended method for combining the two response image for the same stimulus is to difference the average images. Weights are derived for combining images over stimuli and observers. Methods for estimating the level of internal noise are described, especially for the case of repeated presentations of the same noise sample. Simple tests for particular hypotheses about the weights are shown based on observer agreement with a noiseless version of the hypothesis.

Keywords: discrimination, detection, vision

List of symbols in order of appearance

m the number of image components

$\mathbf{s}_s$ s = 0, 1; 1 by m signal vectors

p_s s = 0, 1; probability of signal $\mathbf{s}_s$

$\mathbf{n}$ 1 by m noise vector with components $n(i)$, i = 1, m

$\mathbf{g}$ 1 by m trial stimulus vector with components $g(i)$, i = 1, m

$E[\cdot]$ averaging or expectation operator

$Var[\cdot]$ variance computing operator

σ^2 variance of $n(i)$

$\mathbf{w}$ 1 by m classification vector with components $w(i)$, i = 1, m

β bias of linear classifier

R the observer's response, 0 or 1

^{T} vector transpose operator

$\|.\|$ vector length, $\|\mathbf{w}\| = (\mathbf{w}\,\mathbf{w}^T)^{1/2}$

$Pr\{\}$ probability of enclosed event

$p_{s,R}$ probability of response R given signal $\mathbf{s}_s$, $Pr\{R|\mathbf{s}_s\}$

$\Phi(\cdot)$ cumulative standard normal distribution function

d_0' sensitivity of linear classifier

β_0 shifted bias of linear classifier, $\beta - \mathbf{w}\,\mathbf{s}_0^{\ T}$

$Z(\cdot)$ functional inverse of the cumulative standard normal distribution function, $\Phi^{-1}(\cdot)$

$\mathbf{w}_I$ classification vector $\mathbf{w}$ of the ideal observer

d_I' sensitivity of the ideal observer

ρ the sampling efficiency of $\mathbf{w}$, $\mathbf{w}\,\mathbf{w}_I^T$

$\beta_{0,H}$ random shifted bias of the human observer model

γ^2 variance of β_H

α^2 proportion of external noise in the classification variable, $1/(1+\gamma^2)$

d_H' sensitivity of the human observer model

β_H performance bias of the human observer model

$\phi(\cdot)$ standard normal distribution density function

$\mathbf{n}_{s,R}$ a random noise $\mathbf{n}$ conditional on signal $\mathbf{s}_s$ and detection response R

$\mathbf{a}_{s,R}$ the average of $N_{s,R}$, noises $\mathbf{n}_{s,R}$

$\mathbf{v}_{s,R}$ the expected value of $\mathbf{n}_{s,R}$ when $m = 1$.

x, y, z standard normal variables

U an orthonormal m by m transformation

I the m by m identity transformation

z_i standard normal variables

$N_{s,R}$ the number of presentations of stimulus s that led to response R

N_s the number of presentations of stimulus s, $N_s = N_{s,0} + N_{s,1}$

e the decision contribution from the external noise, replacing $\mathbf{w}\,\mathbf{n}^T$

M_R $R = 0, 1$; the event that an error free model made response R

$p_{M,s,0}$ the probability of event M_0 given that the signal was $\mathbf{s}_s$

$\beta_{M,s}$ the signal-dependent, error-free model criterion, β_0 if $\mathbf{s}_0$, or $\beta_0 - d_0'$ if $\mathbf{s}_1$

1 Historical Introduction

In 1965 a frustrated physiological psychology graduate student was looking for a thesis topic in the auditory research laboratory of E. C. Carterette and M. P. Friedman, the editors-to-be of the Handbook of Perception. They recommended that he tape record the stimulus of traditional tone-in-noise Yes-No detection experiment and analyze the sounds in the four different types of trials to see whether correlates could be found in the stimuli relating to the observer responses. The noise masker was continuous wide band noise and marker tones were recorded on a second track to keep track of the signals presented. The tapes were digitized and analyzed, but the signal-to-noise level at threshold was so low that no trace of the signals could be found in the digitized records. To ensure getting a degree in the foreseeable future the student made several changes in the experiment. To improve the signal-to-noise ratio on the tape, the noise bandwidth was narrowed and the noise was only turned on during the short interval when the signal might be present. To reduce the effects of observer noise, the tape was repeatedly presented to the observer to get average ratings of signal presence. To minimize degrees of freedom in the stimulus measurement, the stimulus was reduced to the energy passed by a filter tuned to the signal tone frequency. This combination of changes allowed the student to find that on signal trials, very narrow filter outputs

correlated best with observer ratings, while on noise trials, wider filter outputs correlated best, contradicting the prediction of single linear filter models for auditory tone detection (Ahumada, 1967).

To gain better control of the masking noise and avoid the limitations of tape recording, Ahumada and Lovell (1971) used computer-generated tones and noises defined by their Fourier component amplitudes and reported linear regressions on the component energies with average observer ratings. These results were essentially auditory classification images that again demonstrated results contrary to simple linear filter theory, in that frequency components were weighted differently on signal trials from noise-only trials and in that negative weights were frequently observed. The results of both experiments seemed to be consistent with models with multiple linear channels that were being non-linearly combined. Ahumada, Marken, and Sandusky (1975) extended the experiment to the combined time and frequency domains with similar results.

Our first visual classification images (Ahumada, 1996) were done to see whether the method we had used in audition could be used to elucidate the features used by observers to accomplish a vernier acuity task. Figure 1 shows a raw classification image and the same image smoothed and quantized so that only weights that are significantly different from zero are colored differently from the gray background. The ideal observer would only have weights on the right side, the side of the line that was either even with or one pixel higher than the left line. Spatial position uncertainty was presumably responsible for the observer needing to compare the two lines and for blurring the image more than optical blurring would predict. Theories that postulate that the response would be determined by the output of a single best-discriminating Gabor-like filter (Findlay, 1973; Foley, 1994)) are not supported by the appearance, but were not tested statistically. Beard and Ahumada (1998) wanted to see whether observer performance was best characterized as orientation discrimination based on an oriented filter outputs or a local position measurement (Waugh, Levi, and Carney, 1993). The question was left unanswered; the linear classification functions obtained from the abutting stimuli were consistent with possible implementations of either theory.
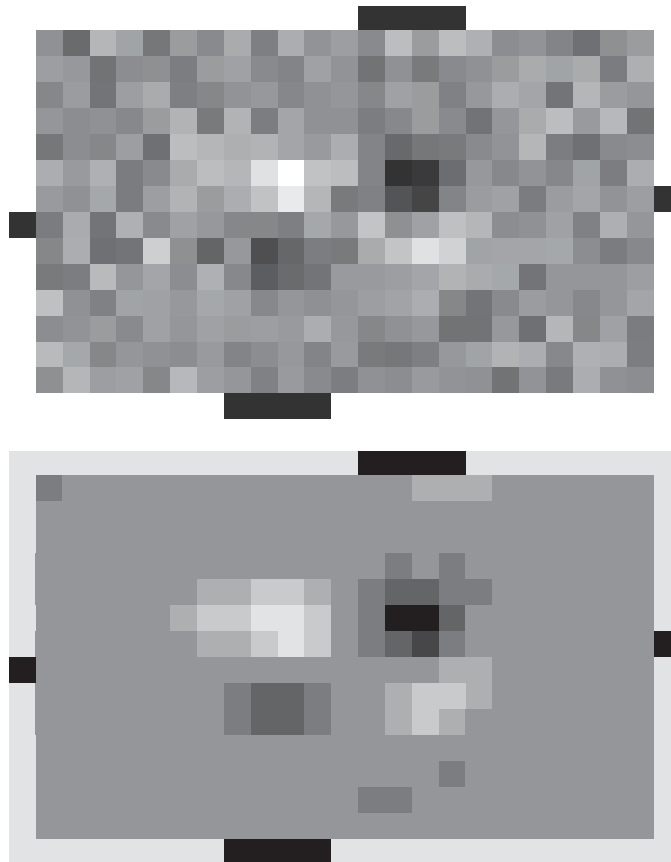

Figure 1. A raw classification image (top) and the same image smoothed and quantized (bottom), so that only weights significantly different from zero are colored differently from the gray background. The black squares on the sides show the heights and positions of the fixed line (left) and the variable line offset (right). The dark lines on the top and bottom show the lengths and positions of the lines. The observer was AJA, who ran 1600 trials (Ahumada, 1996).

The first visual classification images were linear combinations of four averaged noise images, one for each of the four stimulus-response categories. For a given stimulus, the average of all the added noises has zero mean, so the sum of noises

from one response class has an expectation equal to the negative of the expectation of the sum of the noises from the other response class, so we knew to combine the two response noise images with opposite sign. It appeared in the initial images that the error images were clearer than the correct response images, so we took the difference of the averages rather than the sums, realizing that this was an arbitrary decision. We also arbitrarily combined the images from the two stimuli with equal weight to get a single overall image. By symmetry this must be the right weighting to use if the observer is making the same number of errors to equal numbers of each kind of stimulus, which was approximately the case. The next section of this paper has an analysis that shows for a simplified theoretical situation, it is possible to show that the averaging is nearly optimal and to find expressions for good weighting functions for the cases of unequal stimulus presentation rates and unsymmetrical response biases. The initial part of the section introduces notation for a standard signal detection experiment as analyzed by Green and Swets (1966).

2 Template estimation for linear classification of two signals in additive white Gaussian noise

2.1 The signals and noise

$\mathbf{s}_0$ and $\mathbf{s}_1$ are 1 by m signal vectors, presented with probabilities p_0 and $p_1 = (1 - p_0)$ for N trials. On each trial a random noise sample vector $\mathbf{n}$ is added to the signal, so the trial stimulus

$\mathbf{g} = \mathbf{s}_0 + \mathbf{n}$

or

$\mathbf{g} = \mathbf{s}_1 + \mathbf{n}$.   (2.1.1)

$\mathbf{n}$ is a 1 by m vector of independent samples of identically distributed Gaussian variables $n(i)$ with

$E[n(i)] = 0$ (2.1.2)

and

$\mathrm{Var}[n(i)] = E[(n(i) - E[n(i)])^2] = \sigma^2,$ (2.1.3)

where $E[\cdot]$ is the averaging or expectation operator and $\mathrm{Var}[\cdot]$ computes the variance. Without loss of generality we can assume that the noise has been normalized by its standard deviation so that

$\sigma^2 = \sigma = 1.$ (2.1.4)

2.2 The linear observer model

The linear observer classifying the noisy signals by responding $R = 0$ or $R = 1$ or would use a vector $\mathbf{w}$ and respond $R = 1$ if and only if

$\mathbf{w}\,\mathbf{g}^\mathsf{T} > \beta,$ (2.2.1)

where β is a response criterion and $^\mathsf{T}$ indicates the matrix transpose operator, so that

$$\mathbf{w}\mathbf{g}^\mathsf{T} = \sum_i^m w(i)g(i)$$

Also, without lack of generality, we will assume that $\mathbf{w}$ has unit length ($\mathbf{w}$ and β have already been divided by the length of $\mathbf{w}$) so that

$\|\mathbf{w}\| = (\mathbf{w}\,\mathbf{w}^\mathsf{T})^{0.5} = 1.$ (2.2.2)

The performance of an observer is characterized by the error rates

$p_{0,1} = \Pr\{R = 1 \mid \mathbf{s}_0\},$

the probability of signal $\mathbf{s}_0$ being followed by response R_1, and

$p_{1,0} = \Pr\{R = 0 \mid \mathbf{s}_1\},$

the probability of signal $\mathbf{s}_1$ being followed by response $R = 0$.
For the linear classifier with vector $\mathbf{w}$ and criterion β, $\mathbf{w} \, \mathbf{n}^T$ is Gaussian with mean zero and unit variance. Hence

$$p_{0,1} = \Pr\{\mathbf{w} \, (\mathbf{s}_0 + \mathbf{n})^T > \beta \} = 1 - \Phi(\beta - \mathbf{w} \, \mathbf{s}_0^T)$$

and

$$p_{1,0} = \Pr\{\mathbf{w} \, (\mathbf{s}_1 + \mathbf{n})^T < \beta \} = \Phi(\beta - \mathbf{w} \, \mathbf{s}_1^T), \qquad (2.2.3)$$

where $\Phi(\cdot)$ is the cumulative standard Gaussian distribution function.
If we define sensitivity and bias parameters

$$d_0' = \mathbf{w} \, (\mathbf{s}_1 - \mathbf{s}_0)^T \qquad (2.2.4)$$

and

$$\beta_0 = \beta - \mathbf{w} \, \mathbf{s}_0^T, \qquad (2.2.5)$$

then the error rates are

$$p_{0,1} = 1 - \Phi(\beta_0) \qquad (2.2.6)$$

and

$$p_{1,0} = \Phi(\beta_0 - d_0'). \qquad (2.2.7)$$

These parameters can be found from the error rates as

$$\beta_0 = Z(1 - p_{0,1}) \qquad (2.2.8)$$

and

$$d_0' = \beta_0 - Z(p_{1,0}), \qquad (2.2.9)$$

where
$$Z(\cdot) = \Phi^{-1}(\cdot)$$
is the functional inverse of the cumulative standard normal distribution function $\Phi(\cdot)$.

2.3 The ideal observer

The ideal observer classifying the noisy signals as $R = 0$ or $R = 1$ would use the linear classifier

$$\mathbf{w}_I = (\mathbf{s}_1 - \mathbf{s}_0) \, / \, \|\mathbf{s}_1 - \mathbf{s}_0\|. \qquad (2.3.1)$$

For the ideal observer,

$$d_I' = \mathbf{w}_I \, (\mathbf{s}_1 - \mathbf{s}_0)^T = \|\mathbf{s}_1 - \mathbf{s}_0\| . \qquad (2.3.2)$$

The efficiency of a non-ideal linear classifier is given by

$$(d_0'/d_I')^2 = (\mathbf{w} \, \mathbf{w}_I^T)^2 = \rho^2, \qquad (2.3.3)$$

the square of the correlation between the actual and the ideal classifier coefficients, sometimes called the sampling efficiency.

2.4 A noisy human observer model

Human observers classify the same images different ways on different presentations. This is modeled here by assuming that the observer's criterion $\beta_{0,H}$ (corresponding to β_0) is a normally distributed random variable with

$$E[\beta_{0,H}] = \beta_0 \qquad (2.4.1)$$

and

$$\text{Var}[\beta_{0,H}] = \gamma^2 \tag{2.4.2}$$

independent of the noise **n**. It does not matter whether the variability is added to the criterion or the classification function value. Since the noiseless criterion β_0 was defined as the criterion for a variable with unit variance, the parameter $1 + \gamma^2$ can be interpreted as the total variance of the classification variable and

$$\alpha^2 = 1/(1+\gamma^2) \tag{2.4.3}$$

as the proportion of variance in the classification variable that is from the external noise **n**. The error probabilities are now

$$p_{0,1} = \Pr\{\mathbf{w}\,\mathbf{n}^T > \beta_{0,H}\}$$

$$= \Pr\{(\mathbf{w}\,\mathbf{n}^T - (\beta_{0,H} - \beta_0))/(1+\gamma^2)^{1/2} > \alpha\beta_0\}$$

$$= 1 - \Phi(\alpha\beta_0) \tag{2.4.4}$$

and

$$p_{1,0} = \Phi(\alpha(\beta_0 - d_0')). \tag{2.4.5}$$

If we define the observer's sensitivity and biases as

$$d_H' = \alpha d_0' \tag{2.4.6}$$

and

$$\beta_H = \alpha\beta_0, \tag{2.4.7}$$

then we can compute these parameters from the human model observer error rates as

$$\beta_H = Z(1 - p_{0,1}) \tag{2.4.8}$$

and

$$d_H' = \beta_H - Z(p_{1,0}). \tag{2.4.9}$$

The efficiency of the human observer model is

$$(d_H'/d_I')^2 = \alpha^2 \rho^2. \tag{2.4.10}$$

Since $\rho^2 \leq 1$, a lower bound for alpha is given by

$$\alpha \geq d_H'/d_I', \tag{2.4.11}$$

and an upper bound for γ^2 is given by

$$\gamma^2 \leq (d_I'/d_H')^2 - 1. \tag{2.4.12}$$

These bounds are reached when **w** is $\mathbf{w}_I$ and the inefficiency is only the result of the internal or criterion noise.

2.5 The classification images

The classification image components are the four average noises $\mathbf{a}_{s,R}$ the averages of the noises **n** for the trials segregated by signal $\mathbf{s}_s$ and detection response R. We would like to find the mean and the variance of the pixels of $\mathbf{a}_{s,R}$ as a function of the parameters ($\mathbf{s}_1$, $\mathbf{s}_0$, **w**, β_0, and γ or α).

2.5.1 The single pixel case

In the single pixel (m=1) case we are trying to find the mean of a single Gaussian variable **n** that has been truncated by a random criterion β_H. Let $\mathbf{n}_{s,R}$ be the truncated variable when s was the stimulus and R was the response and

$$v_{s,R} = E[\mathbf{n}_{s,R}]. \tag{2.5.1.1}$$

Then since $\|\mathbf{w}\| = 1$, $w = \pm 1$. We can assume without loss of generality $\mathbf{s}_1$ is greater than $\mathbf{s}_0$, and the sign of w is set to maximize correctness, so that $w = 1$. Hence,

$$\mathbf{w}\,\mathbf{n}^{\mathsf{T}} < \beta_{0,H}$$

if and only if

$$n < \beta_{0,H}. \tag{2.5.1.2}$$

So in the case that $s = R = 0$,

$$v_{0,0} = E[n_{0,0}] = E[n \mid n < \beta_{0,H}]$$

and for the other cases

$$v_{0,1} = E[n_{0,1}] = E[n \mid n > \beta_{0,H}]$$

$$v_{1,0} = E[n_{1,0}] = E[n \mid n < \beta_{0,H} - d_0']$$

$$v_{1,1} = E[n_{1,1}] = E[n \mid n > \beta_{0,H} - d_0'] \ . \tag{2.5.1.3}$$

2.5.2 Single pixel, no noise

Consider now the single pixel case when there is no noise in the criterion ($\beta_{0,H} = \beta_0$).

$$
\begin{aligned}
E\left[n_{0,0}\right] &= E\left[n \mid n < \beta_0\right] \\
&= \left(\int_{-\infty}^{\beta_0} z\,\phi(z)\,dz\right) \Big/ \Phi(\beta_0) \\
&= \phi(\beta_0) \Big/ \Phi(\beta_0)
\end{aligned}
\tag{2.5.2.1}
$$

where $\phi(z)$ is the standard normal density function and the integration of $z\,\exp(-z^2/2)$ is enabled by the variable substitution
$t = -z^2/2.$
Similarly,

$$E[n_{0,1}] = E[n \mid n > \beta_0]$$

$$= \left(\int_{\beta_0}^{\infty} z\,\phi(z)\,dz\right) \Big/ \left(1 - \Phi(\beta_0)\right)$$

$$= \phi(\beta_0) / (1 - \Phi(\beta_0)). \tag{2.5.2.2}$$

2.5.3 Single pixel, noisy criterion

The Gaussian criterion case can be reduced to the fixed criterion case by a change of variables. Let z be the standard Gaussian used to form the criterion $\beta_{0,H}$, so that

$$\beta_{0,H} = \gamma z + \beta_0 \ . \tag{2.5.3.1}$$

Then

$$v_{0,0} = E[n_{0,0}] = E[n \mid n < \beta_{0,H}]$$

$$= E[n \mid n < \gamma z + \beta_0]. \tag{2.5.3.2}$$

If we let

$$x = \alpha\,(n - \gamma z) \tag{2.5.3.3}$$

and

$$y = \alpha\,(\gamma n + z) \tag{2.5.3.4}$$

The new variables x and y are independent ($E[x\,y] = 0$), standard ($E[x] = E[y] = 0$, $Var[x] = Var[y] = 1$) Gaussian variables, such that

$$n = \alpha\,(x + \gamma\,y) \qquad\qquad (2.5.3.5)$$

and

$$n < \gamma\,z + \beta_0$$

if and only if

$$x < \alpha\beta_0. \qquad\qquad (2.5.3.6)$$

So

$$v_{0,0} = E[n_{0,0}] = E[n \mid n < \gamma\,z + \beta_0]$$
$$= E[\alpha\,(x + \gamma\,y) \mid x < \alpha\beta_0]$$
$$= \alpha\,E[\,x \mid x < \alpha\beta_0\,]$$
$$= -\,\alpha\phi\,(\alpha\beta_0)\,/\,\Phi(\alpha\beta_0)$$
$$= -\,\alpha\phi\,(\beta_H)\,/\,\Phi(\beta_H)$$

$$= -\,\alpha\phi\,(Z(p_{0,0}))/p_{0,0}\,. \qquad\qquad (2.5.3.7)$$

The effect of the criterion noise on $v_{0,0}$ is to reduce it by the factor α.

Similarly,

$$v_{0,1} = E[n_{0,1}] = E[n \mid n > \beta_{0,H}]$$
$$= E[\alpha\,(x + \gamma\,y) \mid x > \alpha\beta_0]$$
$$= \alpha\,E[\,x \mid x > \alpha\beta_0\,]$$
$$= \alpha\phi\,(\alpha\beta_0)\,/\,(1 - \Phi(\alpha\beta_0))$$
$$= \alpha\phi\,(\beta_H)\,/\,(1 - \Phi(\beta_H))$$

$$= \alpha\phi\,(Z(p_{0,0}))\,/\,(1 - p_{0,0})$$

$$= \alpha\phi(Z(p_{0,1}))\,/\,p_{0,1}, \qquad\qquad (2.5.3.8)$$

since $p_{0,1} = 1 - p_{0,0}$ and

$$Z(p) = -Z(1 - p). \qquad\qquad (2.5.3.9)$$

If false alarms are less frequent than correct rejections ($p_{0,1} < p_{0,0}$), then

$$|v_{0,1}| > |v_{0,0}|, \qquad\qquad (2.5.3.10)$$

a larger absolute expected value on a false alarm than a correct rejection trial. The signal case is the same with the criterion changed to $\beta_{0,H} - d_0'$ so that

$$v_{1,0} = E[n_{1,0}] = E[n \mid n < \beta_{0,H} - d_0']$$

$$= \alpha\phi\,(Z(p_{1,0}))/p_{1,0}\,. \qquad\qquad (2.5.3.11)$$

and

$$v_{1,1} = E[n_{1,1}] = E[n \mid n > \beta_{0,H} - d_0']$$

$$= \alpha\phi\,(Z(p_{1,1}))/p_{1,1}\,. \qquad\qquad (2.5.3.12)$$

Again, if misses are less frequent than hits ($p_{1,0} < p_{1,1}$), then

$$|v_{1,0}| > |v_{1,1}|, \qquad\qquad (2.5.3.13)$$

a larger absolute expected value on a miss than a hit trial. Regardless of the signal, the expected value only depends on the response proportion and the criterion variability.

2.5.4 The multiple pixel case

Another independent variable transformation allows the single pixel case result to solve the multiple pixel case. Let us first examine the means and variances of the pixels of $\mathbf{n}_{0,0}$. For any vector $\mathbf{w}$ of unit length, it is possible to construct an orthonormal transformation $\mathbf{U}$ whose first row is $\mathbf{w}$, that is

$$\mathbf{U} = (\mathbf{w}^T\ \mathbf{w}_2^T\ \dots\ \mathbf{w}_m^T)^T\ , \tag{2.5.4.1}$$

such that

$$\mathbf{U}^T\,\mathbf{U} = \mathbf{I}, \tag{2.5.4.2}$$

where $\mathbf{I}$ is the identity transformation (the transpose of $\mathbf{U}$ is its inverse).

When this transformation is applied to $\mathbf{n}^T$ we get a new noise $\mathbf{U}\,\mathbf{n}^T$ whose distribution is the same as that of $\mathbf{n}^T$, but whose first pixel is $\mathbf{w}\,\mathbf{n}^T$. On an $\mathbf{s}_0$ trial, a noise vector $\mathbf{n}$ will be classified as $\mathbf{n}_{0,0}$ if and only if the first pixel of $\mathbf{U}\,\mathbf{n}^T$,

$$z_1 = \mathbf{w}\,\mathbf{n}^T < \beta_{0,H}. \tag{2.5.4.3}$$

The rest of the pixels $(z_2, \dots, z_m)$ of $\mathbf{U}\,\mathbf{n}^T$ are independent standard Gaussian variables (with mean zero and variance 1).

$$E[\mathbf{n}_{0,0}^T] = E[\ \mathbf{n}^T \mid \mathbf{w}\,\mathbf{n}^T < \beta_{0,H}]$$
$$= E[\mathbf{U}^T\,\mathbf{U}\ \mathbf{n}^T \mid \mathbf{w}\,\mathbf{n}^T < \beta_{0,H}]$$
$$= E[\mathbf{U}^T\,(z_1, z_2, \dots, z_m)^T \mid z_1 < \beta_{0,H}]$$
$$= \mathbf{U}^T\,E[(z_1, z_2, \dots, z_m)^T \mid z_1 < \beta_{0,H}]$$
$$= \mathbf{U}^T\,(v_{0,0}, 0, \dots, 0)^T$$

$$= v_{0,0}\,\mathbf{w}^T. \tag{2.5.4.4}$$

Using similar argument for the other cases, leads to the general result that

$$E[\mathbf{n}_{s,R}] = v_{s,R}\,\mathbf{w}\ . \tag{2.5.4.5}$$

The mean of a classified noise is proportional to the classifying vector $\mathbf{w}$. The variance of individual elements of $\mathbf{n}_{s,R}$

$$\mathrm{Var}\big(\mathbf{n}_{s,R}(i)\big) = \sum_{j=2}^{m} \mathbf{U}_{j,i}^2\,\mathrm{Var}\big(z_j\big) + w_i^2\,\mathrm{Var}\big(z_1 \big| s_s R\big)$$

$$= (1 - \mathbf{w}_i^2) + \mathbf{w}_i^2\,\mathrm{Var}[z_1 \mid \mathbf{s}_s\,R]\ , \tag{2.5.4.6}$$

which is bounded by $\mathrm{Var}[z_1 \mid \mathbf{s}_s\,R]$ and one. Truncation of a Gaussian can only decrease the variance, so

$$\mathrm{Var}[\mathbf{n}_{s,R}(i)] < 1. \tag{2.5.4.7}$$

Since $\|\mathbf{w}\| = 1$, if there are very many significant weights in $\mathbf{w}$ they will have to be small, so that

$$\mathrm{Var}[\mathbf{n}_{s,R}(i)] \sim 1. \tag{2.5.4.8}$$

Let $\mathbf{a}_{s,R}$ be the average value of a number $N_{s,R}$ of $\mathbf{n}_{s,R}$. Any combination of the form

$$\mathbf{w}_{est} = k_{0,1}\,\mathbf{a}_{0,1} - k_{0,0}\,\mathbf{a}_{0,0} + k_{1,1}\,\mathbf{a}_{1,1} - k_{1,0}\,\mathbf{a}_{1,0}, \tag{2.5.4.9}$$

with positive weights $k_{s,r}$ will be an estimate of $\mathbf{w}$ times a positive constant.

2.5.5 Combining the categorized noises

If we have two independent estimates b and c of the same quantity (having the same expected value, $E(b)=E(c)$) with variances σ_b^2 and σ_c^2, the linear combination of the two estimates with the same expected value and the smallest variance is

$$(b / \sigma_b^2 + c / \sigma_c^2) / (1 / \sigma_b^2 + 1 / \sigma_c^2). \tag{2.5.5.1}$$

That is, each estimate should be weighted inversely by its variance.

To obtain a minimum variance estimate of $w(i)$ from a sample with $N_{s,R}$ approximately independent samples of each type, the individual estimates of $w(i)$, $\mathbf{n}_{s,R}(i)/ v_{s,R}$ should be weighted inversely by their variances, which are approximately

$$\text{Var}[\mathbf{n}_{s,R}(i)/v_{s,R}] = 1/v_{s,R}^2, \tag{2.5.5.2}$$

so we should weight each $\mathbf{n}_{s,R}(i)$ by $v_{s,R}$. Since the weights do not depend on i , we can then just weight $\mathbf{n}_{s,R}$ by $v_{s,R}$. A good un-normalized estimate of the classifier $\mathbf{w}$ is thus given by the $v_{s,R}$ weighted sums $N_{s,R} \; \mathbf{a}_{s,R}$.

$$\mathbf{w}_G = v_{0,1} N_{0,1} \, \mathbf{a}_{0,1} - v_{0,0} N_{0,0} \, \mathbf{a}_{0,0}$$

$$+ v_{1,1} N_{1,1} \, \mathbf{a}_{1,1} - v_{1,0} N_{1,0} \, \mathbf{a}_{1,0} . \tag{2.5.5.3}$$

If we replace $p_{s,R}$ in $v_{s,R}$ by $p_{s,R} = N_{s,R}/ N_s$, where $N_s = N_{s,0} + N_{s,1}$, and take advantage of the fact that

$$\phi (Z(p)) = \phi (Z(1{-}p)), \tag{2.5.5.4}$$

we obtain

$$\mathbf{w}_G = \alpha [N_0 \, \phi (Z(p_{0,1})) (\mathbf{a}_{0,1} - \mathbf{a}_{0,0})$$

$$+ N_1 \, \phi (Z(p_{1,0})) (\mathbf{a}_{1,1} - \mathbf{a}_{1,0})]. \tag{2.5.5.5}$$

The more frequent stimulus should be given more weight and, since for $p < 0.5$, $\phi (Z(p))$ increases monotonically, the more error prone stimulus should be given more weight. These weights take into account all the parameters assumed to determine the observer's performance. If both stimuli are equally frequent and the error rates are equal, the formula is proportional to

$$\mathbf{w}_{ave} = \mathbf{a}_{0,1} - \mathbf{a}_{0,0} + \mathbf{a}_{1,1} - \mathbf{a}_{1,0} , \tag{2.5.5.6}$$

the combination rule originally used by Ahumada and Beard (Ahumada, 1996; Ahumada & Beard, 1998, 1999; Beard & Ahumada, 1997, 1998, 2000).

In the next section we add a third subscript to refer to the observer. The good weighting scheme for combining average classification images $\mathbf{a}_{s,R,O}$ over responses, stimuli, and observers can be described as a sequential process. To combine over responses, just take the difference,

$$\mathbf{w}_{G,s,O} = \mathbf{a}_{s,1,O} - \mathbf{a}_{s,0,O}. \tag{2.5.5.7}$$

To then combine images for different stimuli, weight by factors involving the relative frequencies of the stimuli and the extremeness of the error proportions for the stimuli.

$$\mathbf{w}_{G,O} = k_{1,O} \, \mathbf{w}_{G,1,O} + k_{0,O} \, \mathbf{w}_{G,0,O}. \tag{2.5.5.8}$$

where

$$k_{s,O} = N_{s,O} \exp(- Z(p_{s,r,O})^2/2) . \tag{2.5.5.9}$$

If estimates are to be combined over M observers, they need to be weighted by the square root of the observer's proportion of decision variance due to the external noise, $\alpha^2 = 1/(1 + \gamma^2)$, and the number of trials run by the observer (which is already included here in the $N_{s,O}$).

$$\mathbf{w}_G = \alpha_1 \, \mathbf{w}_{G,1} + \alpha_2 \, \mathbf{w}_{G,2} + ... + \alpha_M \, \mathbf{w}_{G,M} . \tag{2.5.5.10}$$

3 Measuring the internal noise

3.1 Response agreement with the same external noise sample

Estimates of the internal noise (α or γ) are needed in order to use the above formula (2.5.5.10) to combine estimates over observers. Internal noise estimates are also needed to compute the variance of an estimate of $\mathbf{w}$ to plan the number of

trials that need to be run (see Equation $(2.5.5.2)$). For this section we are using the model of section 2.4, relaxing the linearity assumptions about the classification function and the external noise to the assumption that the term

$$e = \mathbf{w}\,\mathbf{n}^T$$

is a standard Gaussian.

The subscripts i and j are added to indicate two separate trials. The response agreement probability for a particular signal and response with the same noise is denoted by

$$p_{A,s,R} = \Pr\{R_i = R_j = R \mid \mathbf{s}_{s,i}, \mathbf{s}_{s,j}, \mathbf{n}_i = \mathbf{n}_j\}. \tag{3.1.1}$$

To obtain this probability we will compute it conditional on the value of e and then average over the possible values of e.

Conditional on the value of e, for $s = 0$, the probability of a response $R = 0$ is given by

$$\Pr\{R_i = 0 \mid \mathbf{s}_{0,i}, e\} = \Pr\{e < \beta_{0,H} \mid e\}$$

$$= \Pr\{e < \gamma z_0 + \beta_0 \mid e\}$$

$$= \Pr\{(e - \beta_0)/\gamma < z_0 \mid e\}$$

$$= \Phi((\beta_0 - e)/\gamma). \tag{3.1.2}$$

For another response to the same signal and the same noise the criterion variability is independent so the probability of two $R = 0$ responses is given by

$$\Pr\{R_i = R_j = 0 \mid \mathbf{s}_{0,i}, \mathbf{s}_{0,j}, e\} = \Pr\{e < \beta_{0,H} \mid e\}^2$$

$$= \Phi((\beta_0 - e)/\gamma)^2.$$

$$= \Phi((\beta_H / \alpha - e)/\gamma)^2.$$

$$= \Phi((\beta_H (1 + \gamma^2)^{0.5} - e)/\gamma)^2. \tag{3.1.3}$$

The probability of two correct responses $R_i = R_j = 0$ to the same noise is then

$$p_{A,0,0} = \int_{-\infty}^{\infty} \Phi\left(\frac{\left(\beta_H \sqrt{1+\gamma^2}\right) - e}{\gamma}\right)^2 \phi(e)\,de$$

and the probability of two incorrect responses $R_i = R_j = 1$ to the same noise is

$$p_{A,0,1} = \int_{-\infty}^{\infty} \left(1 - \Phi\left(\frac{\left(\beta_H \sqrt{1+\gamma^2}\right) - e}{\gamma}\right)\right)^2 \phi(e)\,de$$

The equations for arbitrary s can be written

$$p_{A,s,0} = \int_{-\infty}^{\infty} \Phi\left(\frac{\left(Z(p_{s,0})\sqrt{1+\gamma^2}\right) - e}{\gamma}\right)^2 \phi(e)\,de$$

and

$$p_{A,s,1} = \int_{-\infty}^{\infty} \left(1 - \Phi\left(\frac{\left(Z(p_{s,0})\sqrt{1+\gamma^2}\right) - e}{\gamma}\right)\right)^2 \phi(e)\,de \tag{3.1.7}$$

Either of these equations can be used to solve for an estimate of γ (the same estimate results by using either one).

3.2 Estimating observer noise using a model classifier

Ahumada and Beard (1998) also derived other estimates for γ based on the assumption that e comes from a known model (and is distributed as a Gaussian) and that Gaussian internal noise is added by the observer. This allows falsification of the model if the estimates of γ are not consistent. The model they tested was a particular parametric linear filter estimated by Barth, Beard, and Ahumada (1999), but any model can be tested using their scheme if the performance of the noiseless model can be computed for the same noises presented to the observer. The noiseless model's performance index, d_0', is needed as is the trial by trial agreement of the observer and the model.

One estimate of γ is based on the ratio of the performance of the model, d_0', to that of the observer, d_H'. From above we have

$$d_H' = \alpha d_0' \, ,$$

so

$$\alpha = d_H'/d_0' \, . \tag{3.2.1}$$

and

$$\gamma = ((d_0'/d_H'))^2 - 1)^{0.5}. \tag{3.2.2}$$

Another γ estimate comes from the trial by trial agreement between the observer and the model. For a given value of e, the event M_0 that the model will respond that the signal was s_0 when it actually was s_s will occur if e is less than the criterion for the model,

$$\beta_{M,s} = Z(p_{M,s,0}), \tag{3.2.3}$$

which is β_0 if $s = 0$ or $\beta_0 -- d_0'$ if $s = 1$. The probability that the observer and the model agree on this response will be

$$P\{R = 0, M_0\} = \int_{-\infty}^{\beta_{m,s}} \Phi\left(\frac{\left(\beta_H \sqrt{1+\gamma^2}\right) - e}{\gamma}\right) \phi(e) de \tag{3.2.4}$$

Note that the other possibilities can be computed from this and simpler ones, from the relationships

$$\Pr\{R_0, M_0\} + \Pr\{R_0, M_1\} = \Pr\{R_0\} \, , \tag{3.2.5}$$

$$\Pr\{R_0, M_0\} + \Pr\{R_1, M_0\} = \Pr\{M_0\} \, , \tag{3.2.6}$$

and

$$\Pr\{R_1, M_1\} + \Pr\{R_0, M_1\} = \Pr\{M_1\} \, . \tag{3.2.7}$$

The question arises as to which alternative models could or could not be rejected by this test. One reader hypothesized that any template with the same sampling efficiency (relative to the ideal observer) could not be rejected. Within the current framework, this is true for tests comparing the noise estimate based on observer self-agreement of Section 3.1 with the estimate based on the ratio of observer performance to model performance (3.2.1). However, responses from any other template of the same efficiency would correlate worse with the observer responses than the correct template, leading to a larger estimate of internal noise. Thus only the correct template can lead to agreement among all three noise estimates.

3.3 Example of Model Testing based on Internal Noise Estimates

This section illustrates model testing based on noise estimates using the data of observer CS from Ahumada and Beard (1999). The observer was detecting the presence or absence of a Gabor signal (2 cpd or 16 cpd) in noise. Figure 2 shows that the classification images for 2 cpd signal present and signal absent and for 16 cpd signal present resembled the signal, but no image emerged for the 16 cpd signal absent, a result consistent with position or phase uncertainty.
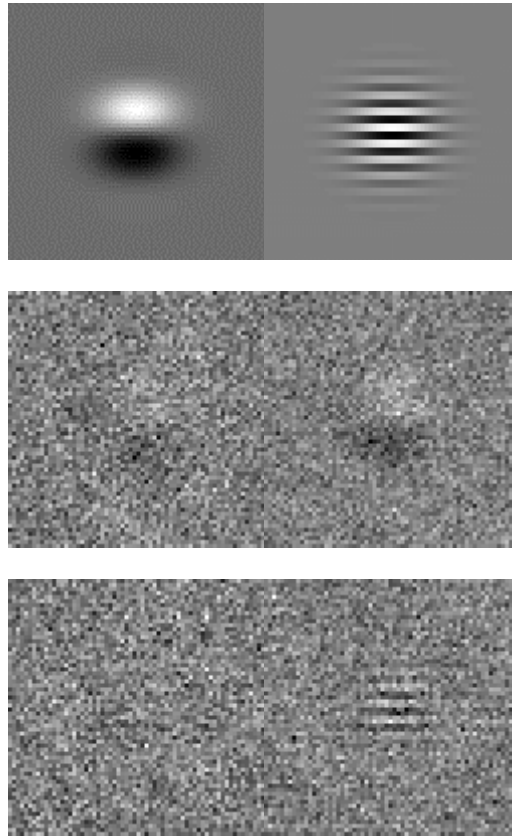
Figure 2. (Top) Gabor signals with spatial frequencies of 2 cpd (left) and 16 cpd (right). (Center) Classification images for the 2 cpd target for no-signal trials (left) and signal trials (right). (Bottom) Classification images for the 16 cpd target for no-signal trials (left) and signal trials (right).smoothed and quantized (bottom). The observer was CS, who ran 5900 trials at 2 cpd and 8000 trials at 16 cpd (Ahumada and Beard, 1999).

Figure 3 shows estimates of the proportion of external noise α^2 for this observer. Within each block of 100 trials, each stimulus was repeated twice. The circles show estimates based on the agreement of these two responses to the same stimulus. The data and calculation details are shown in the Appendix. Error bars for the self-agreement estimates were computed by generating 95% confidence limits for the proportion of self-agreement and then computing the corresponding noise proportion. For both 2 and 16 cpd the agreement estimate of α^2 for signal trials is greater than that for no-signal trials. The confidence intervals overlap slightly, but the difference is significant because the data are independent. This difference rejects the linear model for both spatial frequencies, the result being consistent with position uncertainty combined with internal noise reducing response agreement for no-signal conditions relative to signal conditions at both spatial frequencies. Using the noise analysis rather than the actual agreement proportions provides compensation for different criterion positions in the two cases.

The crosses in Figure 3 show predictions of α^2 from a comparison of detection performance with that of the linear observer using the signal as the template, where the proportion of observer variance from external noise α^2 is then given by the square of the ratio of the observer d_H' to the model d_0',

$$\alpha^2 = (d_H'/d_0')^2 \ . \tag{3.3.1}$$

For both spatial frequencies the performance estimate of α^2 is well below the observer self-agreement estimates, rejecting the hypothesis that the observer is using the signal as a linear template, but is noisy.
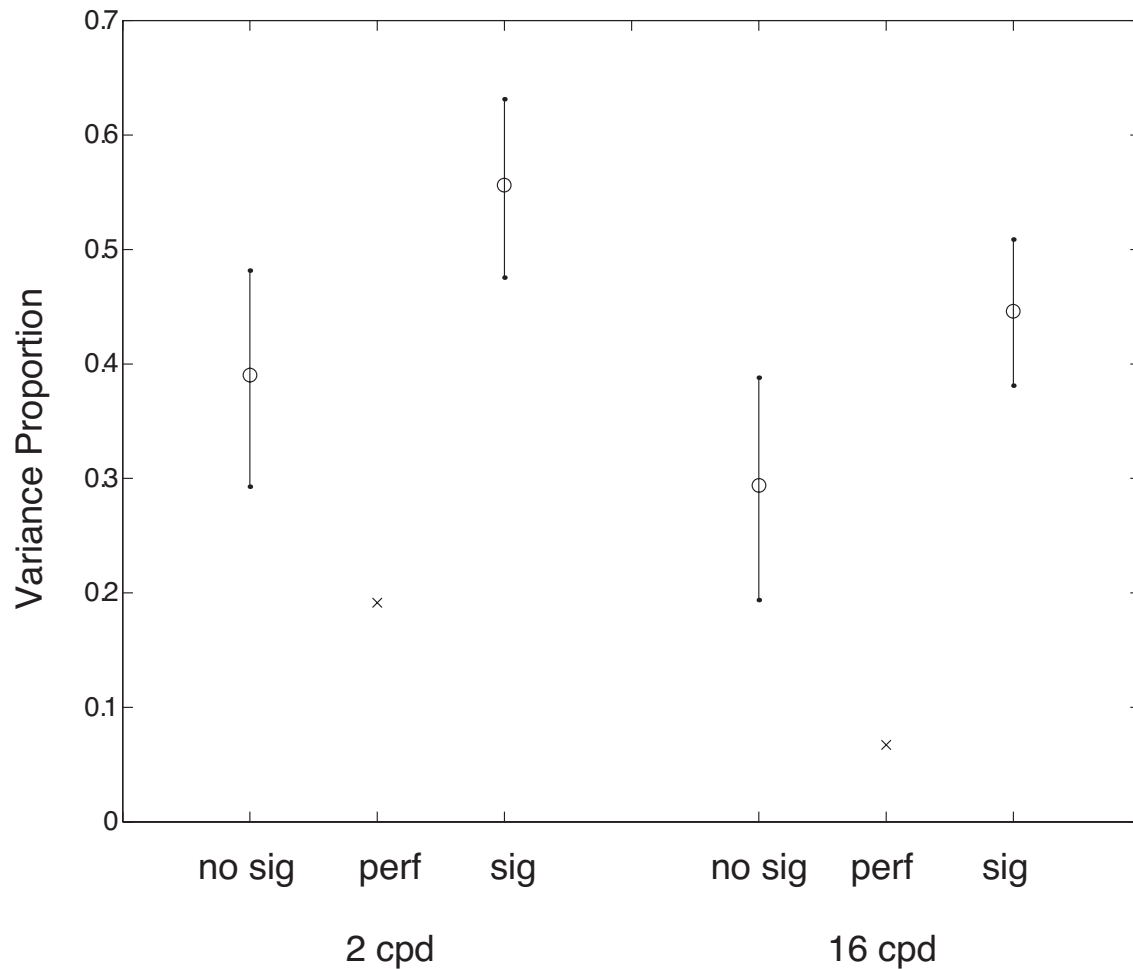
Figure 3. Estimated proportions of variance due to external noise, α^2, from observer self agreement (circles) and from relative detection performance (crosses) based on the data of observer CS from Ahumada and Beard (1999).

4 Discussion and Conclusions

4.1 Classification Function Estimation

The combining rules derived in the first part of the paper agree with those derived by Murray, Bennett, and Sekuler (2002) for the same case of linear discrimination of two stimuli in white Gaussian noise in the presence of internal noise. Their derivation assumes that both the external and internal noise are white, but since the internal noise only participates after its linear combination with the observer template, its covariance matrix does not matter. Abbey, Eckstein, and Bochud (1999) and Abbey and Eckstein (2000; 2001a; 2001b; 2002) consider the 2 alternative forced choice situation with no alternative bias, a special case of the above analysis. They also derive formulas for the case of non-white external noise, which theoretically reverts to the white noise analysis after a pre-whitening filter is applied. Solomon (2002) also has a derivation of the expected value of a classification image based on the same transformation argument presented above.

4.2 Internal Noise Estimation

In the auditory studies (Ahumada, 1967; Ahumada and Lovell, 1971; Ahumada, Marken, and Sandusky, 1975), multiple observations of the same stimuli provided estimates of internal noise. The observer ratings were regarded as approximately continuous variables and standard parametric analyses provided the estimates. Burgess and Colborne (1988) solved for both the probability of observer response agreement on two repetitions of the same 2AFC stimulus and the probability of a correct response as a function of detectability (d0') and internal noise (γ) for the unbiased observer (their Equations (4)-(6)), obtaining results similar to those in Section 3.1. To get estimates of γ (their k), they plotted these two

variables as a function of d_0' with γ as a parameter and found the γ curve that the data points fell on (their Figure 2). Richards and Zhu (1994) solved for the response correlation on two repetitions of the same trial using the model of the above Section 2 (their Theorem 3). They report this correlation as a function of both external and internal noise for the unbiased case (their Table II), pointing out that it is just a function of γ. Their results are also similar to those in Section 3.1, but they look different because the random variables are integrated in reverse order and the squared terms are recast as the variance of a dichotomous variable. Ahumada and Beard (1998) presented a method for estimating γ for the same general situation as Richards and Zhu (1994) using the same agreement measure as Burgess and Colborne (1988). Since the overall agreement measure and the response correlation measure are functions of one of the agreement measures of Equations (3.1.6) or (3.1.6) given the hit or false alarm rates, it suffices to use one of these simpler measures to solve for γ.

4.3 Design Considerations

Beard and Ahumada (1998) arbitrarily tried to adjust the detection parameters so that the percent correct would be near 75% and wanted their observers to have roughly the same number of errors for both stimuli, which were presented with equal probability. The 75% value was considered to be a compromise between trying to get as many errors as possible for the image and keep the task easy enough so that the observer could maintain a stable template (Beard and Ahumada, 1999).

The relation of the weights to the error rates shows that more errors are desirable if the internal noise is held constant (2.5.5.5). They increased the difficulty in two ways, decreasing stimulus duration and increasing the external noise level. Within the present model, lowering performance without increasing external noise can only be the result of a less efficient template or more internal noise, so it was not a good idea for improving the quality of the template estimate. Fortunately, it had little effect and the trials went by faster. Increasing the level of external noise when the observers were performing very well was successful, and did not degrade observer efficiency as compared with the ideal linear observer, but increased noise when the observer was already at 75% led to decreased observer efficiency and probably did not improve the quality of the classification images.

References

C.K. Abbey, M.P. Eckstein (2000) Estimates of human-observer templates for simple detection tasks in correlated noise, Proc. SPIE (E.A. Krupinski, Ed.) 3981, 70-77.

C.K. Abbey, M.P. Eckstein (2001a) Maximum-likelihood and maximum a-posteriori estimates of human-observer templates, Proc. SPIE (E.A. Krupinski and D.P. Chakraborty, Eds.) 4325.

C.K. Abbey, M.P. Eckstein (2001b) Theory for estimating human-observer templates in two-alternative forced-choice experiments, Proc. 17th International Conference on Inferential Processes in Medical Imaging (M.F. Insana and R. Leahy, Eds.) Springer-Verlag, Berlin.

C. Abbey, M. Eckstein (2002) Classification image analysis: Estimation and statistical inference for two-alternative forced choice experiments, Journal of Vision 2(1), 66-78.

C.K. Abbey, M.P. Eckstein, F.O. Bochud (1999) Estimation of human-observer templates for 2 alternative forced choice tasks, Proc. SPIE (E.A. Krupinski, Ed.) 3663:284-295.

Albert J. Ahumada, Jr. (1967) Detection of tones masked by noise: A comparison of human observers with digital-computer-simulated energy detectors of varying bandwidths. Unpublished Ph.D. dissertation (available as Technical Report No. 29, Human Communications Laboratory, Department of Psychology, University of California, Los Angeles).

A.J. Ahumada, Jr. (1996) Perceptual classification images from vernier acuity masked by noise, Perception 26 (ECVP Suppl.), 18 (abstract). HTML

A.J. Ahumada, Jr., B.L. Beard (1998) Response classification images in vernier acuity, Investigative Ophthalmology and Visual Science 39 (4, ARVO Suppl.) S1109 (abstract).

A.J. Ahumada Jr., B.L. Beard (1999) Classification Images for Detection, Investigative Ophthalmology and Visual Science 40 (4, ARVO Suppl.), S572 (abstract). HTML

A.J. Ahumada, Jr., J. Lovell (1971) Stimulus features in signal detection, Journal of the Acoustical Society of America 49, 1751-1756.

A.J. Ahumada, Jr., R. Marken, A. Sandusky (1975) Time and frequency analyses of auditory signal detection, Journal of the Acoustical Society of America 57, 385-390.

E. Barth, B.L. Beard, A.J. Ahumada, Jr. (1999) Nonlinear features in vernier acuity, in Human Vision and Electronic Imaging III, ed. B.E. Rogowitz and T.N. Pappas, Proc. Vol. 3644, pp. 88-96, SPIE, San Jose, CA.

B.L. Beard, A.J. Ahumada, Jr. (1997) Relevant image features for vernier acuity, Perception 26, 38 (abstract).

B.L. Beard, A.J. Ahumada, Jr. (1998) Technique to extract relevant image features for visual tasks, in Human Vision and Electronic Imaging III, ed. B.E. Rogowitz and T.N. Pappas, Proc. Vol., 3299, pp.79-85, SPIE, San Jose, CA. HTML

B.L. Beard, A. J. Ahumada Jr. (1999) Detection in fixed and random noise in foveal and parafoveal vision explained by template learning, Journal of the Optical Society of America A, 16(3), 755-763.

B.L. Beard, A.J. Ahumada (2000) Response Classification Images for Parafoveal Vernier Acuity, Investigative Ophthalmology and Visual Science 41 (4, ARVO Suppl.), S804 (abstract). HTML

A.E. Burgess, B. Colborne (1988) Visual signal detection: IV. Observer inconsistency, JOSA A 5(4), 617-628.

J.M. Findlay (1973) Feature detectors and vernier acuity, Nature 241(5385), 135-137.

J.M. Foley (1994) Human luminance pattern-vision mechanisms: masking experiments require a new model, JOSA A 11 (6), 1710-1719.

D.M. Green, J.A. Swets (1966) Signal detection theory, (Wiley, New York).

R. Murray, P. J. Bennett, A. B. Sekuler (2002) Optimal methods of calculating classification images, in preparation.

V.M. Richards, S. Zhu (1994) Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients. Journal of the Acoustical Society of America 95(1),423-434.

J. A. Solomon (2002) Noise reveals visual mechanisms of detection and discrimination, Journal of Vision, in press.

S.J. Waugh, D.M. Levi, T. Carney (1993) Orientation, masking, and vernier acuity for line targets, Vision Research 33, 1619-1638.

Appendix

```
% data from Ahumada and Beard (1999) web page

% calculations shown only for 2 cpd data

%                S0              S1

%        R0   R1   R2    R0   R1   R2

pCS02 = [911  417  149    180  352   941] ;

pCS16 = [1290  552  115    370  657  1016] ;

pCS = pCS02;

pfa =(pCS(3)+0.5*pCS(2))/...

(pCS(3)+pCS(2)+pCS2(1)) ; 0.2420 ;

pht = (pCS(6)+0.5*pCS(5))/...

(pCS(6)+pCS(5)+pCS(4)) ; 0.7583 ;

dH = znorm(pht)-znorm(pfa) ;1.4006 ;

d0 = 3.2006 ;% stimulus snr

se2d=(dH/d0)^2 ; 0.1915 ;

c = -znorm(pfa) ;  0.6997 ;

nPairs(1)=(pCS(3)+pCS(2)+pCS(1)); 1477

nPairs(2)=(pCS(6)+pCS(5)+pCS(4)); 1473

sameProp(1)=(pCS(1)+pCS(3))/nPairs(1); 0.7177

sameProp(2)=(pCS(4)+pCS(6))/nPairs(2); 0.7610

for s = 1:2

  p=sameProp(s);

  n=nPairs(s);

  confIntProp=1.96*sqrt( (p-p*p)/n);

  confInterval(s,:)=[p, p-confIntProp, p+confIntProp];

end; % for s

[ 0.7177   0.6947   0.7406
```

```
  0.7610   0.7393   0.7828];
% solve for internal noise from psame, c, m
n = 25;% integration approx. parameter
range = 4.0;% ditto
f=inline('(psame-probSame(si,c,m,n,range))^2',...
   'si','psame','c','m','n','range');
x1=0.001;% zero won't work
x2=1;
ms= [0.0 dH];% no signal,  signal
for s=1:2
  m = ms(s);
  for i=1:3
     psame=confInterval(s,i);
     sfinal(s,i)=fminbnd(f, x1,x2, [],psame,c,m,n,range);
  end;% for i
end;% for s
se202 =1-sfinal.*sfinal;
[ 0.3904   0.2931   0.4819
  0.5564   0.4755   0.6313] ;
se216=...
[ 0.2942   0.1938   0.3880
  0.4462   0.3811   0.5087] ;
plot(1,se202(1,1),'ko',[1 1],se202(1,2:3),'k.-',...
     3,se202(2,1),'ko',[3 3],se202(2,2:3),'k.-',...
     2,se2d02,'kx',...
     5,se216(1,1),'ko',[5 5],se216(1,2:3),'k.-',...
     7,se216(2,1),'ko',[7 7],se216(1,2:3),'k.-',...
     6,se2d16,'kx') ;
axis([0 8 0 0.7]);
%%%% probSame.m
function prob=probSame(si,c,m,n,range)
% Prob(same) from si, c, m
```

```
se=sqrt(1.-si*si);
a=(m-c)/si;
b=se/si;
s =[-range:1/n:range];
p=normalcdf(a+b*s);
p=(p.*(1.-p)).*exp(-0.5.*s.*s);
prob=1.0-2.0*sum(p)/(n*sqrt(2*pi));
```