Bayesian Analysis of Identification Performance in Monkey Visual Cortex: Nonlinear Mechanisms and Stimulus Certainty

W. S. GEISLER,* D. G. ALBRECHT*

The identification performance of single neurons in the primary visual cortex was quantified by measuring how accurately one could know the stimulus based upon the neuron's response. We found that for a typical neuron a response of 10 action potentials, following one brief stimulus presentation, was sufficient to classify the stimulus as belonging to a relatively small region in stimulus space, with a high degree of confidence. The performance was better than that which could be attained through linear summation of excitation and inhibition alone. The results suggest that the enhanced performance is a consequence of two nonlinear mechanisms: contrast gain control and expansive response exponent.

INTRODUCTION

Recent work in the visual cortex of the cat and the monkey has uncovered two nonlinearities. One is a contrast gain control or normalization mechanism (Albrecht & Hamilton, 1982; Sclar & Freeman, 1982; Li & Creutzfeldt, 1984; Albrecht & Geisler, 1991, 1994; Bonds, 1991; Robson, 1991; Geisler & Albrecht, 1992; Heeger, 1992a; Carandini & Heeger, 1994). The other is an expansive response exponent (Albrecht & Hamilton, 1982; Sclar, Maunsell & Lennie, 1990; Albrecht & Geisler, 1991, 1994; Heeger, 1992b; DeAngelis, Ohzawa & Freeman, 1993; McLean & Palmer, 1994). These two nonlinearities may substantially improve identification performance in comparison to that which could be attained through linear summation alone. The potential importance of these nonlinearities motivated us to develop techniques for measuring the identification performance of single neurons, under high degrees of stimulus uncertainty (i.e. when the stimulus is free to vary along one or more dimensions). We show here that because of the nonlinear mechanisms the identification performance of neurons in the primary visual cortex of the macaque monkey is remarkably good.

METHODS

Our techniques for measuring the identification performance of single neurons are based upon Bayesian methods from statistical estimation and decision theory. To apply these methods it is necessary to know both the mean and the variability of the neuron's responses along the stimulus dimensions of interest. Therefore, we measured the means and SDs of responses of individual neurons within the primary visual cortex of macaque monkeys (Macaca fascicularis) to repeated presentations of sine-wave stimuli while varying contrast, spatial frequency, direction of motion, and spatial position. The stimulus duration for an individual trial was 200 msec; the performance of the neuron during this time frame should be roughly comparable to the performance of the neuron during ordinary saccadic inspection of a visual scene. In this paper we report the measurements for spatial frequency and contrast on a sample of 89 neurons.

RESULTS

Figure 1(A) shows the responses of a representative neuron measured as a function of contrast, for three different spatial frequencies. For each spatial frequency the responses increased and then saturated; saturation occurred at the same contrast, but the final response level
differed; the curves fitted to the data all saturate at the same contrast (see Appendix A). Figure 1(B) shows the responses plotted as a function of spatial frequency for three different contrasts. For each contrast the shape of the spatial frequency–response function was the same (i.e. the tuning was invariant), even when the response had saturated. The curves fitted to the data are simply scaled versions of each other, indicating that the dimensions of contrast and spatial frequency are independent or separable (see Appendix A). These response properties indicate a nonlinear gain control mechanism which sets the gain of the neuron based upon the average contrast.

The rapidly accelerating increase in response at low contrast indicates a second nonlinear mechanism: a final expansive nonlinearity, which can be described as a power function with an exponent > 1. For the fitted curves in Fig. 1(A) the exponent was 2.3. One effect of the exponent is to increase responsiveness to optimal stimuli relative to non-optimal stimuli, and hence to increase the selectivity.*

These two nonlinearities should influence identification performance. The contrast gain mechanism prevent non-optimal stimuli from producing a maximum response even if the amplitude is high [see Fig. 1(A)]. Therefore, whenever a near maximum response is observed, subsequent brain mechanisms can be certain that the stimulus was near the optimal for the cell. The response exponent further reduces the possibility that non-optimal stimuli will produce a maximum response. Thus, the two nonlinearities working together should improve the identification performance of individual neurons in the primary visual cortex, particularly under high degrees of stimulus uncertainty, as occurs in the natural environment.

The variability of the neural responses must also influence identification performance. Figure 1(C, D) shows the measured SDs as a function of contrast and spatial frequency. The pattern of results is essentially the same as for the mean responses; e.g. the SDs increase as a function of contrast and then saturate at a final level.

*It is possible that cortical neurons have both a response threshold and an expansive exponent. A response threshold can have a similar effect on selectivity. However, an expansive response exponent alone provides a good fit to the data, and further, it is consistent with many other cortical cell properties (e.g. Albrecht & Hamilton, 1982; Sclar et al., 1990; Albrecht & Geisler, 1991, 1994; Heeger, 1992b; DeAngelis et al., 1993; McLean & Palmer, 1994).
which varies with spatial frequency. The curves show the SDs predicted from the means, assuming that the variance is proportional to the mean, with a proportionality constant of 1.15 (see Appendix A). In agreement with earlier studies (e.g., Tolhurst, Movshon & Dean, 1983), we found that this simple proportionality rule adequately described the relationship between the mean and the variance for the entire sample of cells (the average value of the proportionality constant was 1.5). Furthermore, the proportionality constant remained the same across stimulus dimensions for a given neuron [e.g., in Fig. 1(C, D), the same constant is seen to hold for contrast and spatial frequency]. In other words, the variance is proportional to the mean response, independent of the stimulus which produced that mean response. (This behavior is characteristic of a multiplicative noise source internal to the neuron and large enough to dominate the noise in the synaptic inputs.*) This simple proportionality rule implies (all other things being equal) that the signal-to-noise ratio, and hence identification performance, will improve as the response increases.

How can one use the means and variances to determine the identification performance of single neurons, under high degrees of stimulus uncertainty (i.e., when the number of possible stimuli is large)? Because traditional methods of measuring identification performance (e.g., Barlow & Levick, 1969; Cohn, Green & Tanner, 1975; Hawken & Parker, 1990; Geisler et al., 1991) are not applicable under conditions of high uncertainty, we developed an appropriate method using Bayesian estimation theory (see also de Ruyter van Steveninck & Bialek, 1988). Specifically, we measured identification performance by determining how accurately one could know the stimulus based upon the neuron's response.

The first step in applying the method was to obtain a certainty function, which gives the probability of each possible stimulus given the observed response, $p(x|R)$. Bayes' formula was used to obtain the certainty function from the means and SDs measured as a function of the stimulus. Bayes' formula expresses the probability of the stimulus given the response, $p(x|R)$, in terms of the probability of the response given the stimulus, $p(R|x)$, and the a priori probability of the stimulus in the environment, $p(x)$:

$$p(x|R) = \frac{p(R|x)p(x)}{\int p(R|x)p(x) \, dx}. \quad (1)$$

The measured means and SDs (e.g., Fig. 1) were used to determine the probability of the response given a stimulus, $p(R|x)$ (see Appendix B). We assumed no knowledge of the stimulus prior to the response; i.e., we assumed that $p(x)$ is constant, in which case $p(x)$ cancels in the above equation. (Incorporating true stimulus probabilities, i.e., natural environmental constraints, should generally narrow the certainty functions, and thus would only improve identification performance.)

The solid curve in Fig. 2(A) shows the certainty function along the dimension of spatial frequency (when contrast was fixed at 30% and only spatial frequency was determined by summation of excitatory and inhibitory inputs. For example, consider two different stimuli that produce the same mean response: (a) an optimal stimulus at low contrast, and (b) a non-optimal stimulus at high contrast. Presumably, stimulus (b) would produce stronger excitatory inputs and stronger inhibitory inputs than stimulus (a). Thus, if the inputs were multiplicative noise sources, then the variability would be greater for stimulus (b) than for stimulus (a). This would be true whether the noise sources were statistically independent or positively correlated. The assertion that the multiplicative noise seen in cortical neurons is unlikely to arise from the summation of excitatory and inhibitory synaptic inputs is consistent with Softky and Koch's (1993) finding that cortical neurons are noisier than expected from the synaptic inputs.

*If the variance is proportional to the mean independent of how the mean is produced, and if the selectivity of cortical neurons is established by summation of excitatory and inhibitory inputs, then it would be surprising if the multiplicative noise arose from the synaptic inputs. For example, consider two different stimuli that produce the same mean response: (a) an optimal stimulus at low contrast, and (b) a non-optimal stimulus at high contrast. Presumably, stimulus (b) would produce stronger excitatory inputs and stronger inhibitory inputs than stimulus (a). Thus, if the inputs were multiplicative noise sources, then the variability would be greater for stimulus (b) than for stimulus (a). This would be true whether the noise sources were statistically independent or positively correlated. The assertion that the multiplicative noise seen in cortical neurons is unlikely to arise from the summation of excitatory and inhibitory inputs is consistent with Softky and Koch's (1993) finding that cortical neurons are noisier than expected from the synaptic inputs.
contrasts are spread out over a considerable range.

The spatial frequency–response function measured at 30% contrast [from Fig. 1(B)] is superimposed for comparison. As can be seen, the most probable stimuli are confined to a relatively narrow region around the peak frequency; thus, a single response of 11 spikes is sufficient to identify the stimulus as belonging to that region. Identification performance, in general, is related to the width of the certainty function—the narrower the function, the more precisely the neuron can identify the stimulus.

The second step in applying the method was to use the certainty function to determine a maximum-likelihood confidence region in stimulus space: the smallest region to which the stimulus could be classified with some criterion level of accuracy (see Appendix B). For example, the 95% confidence region is defined as the set of most likely stimuli for which the probabilities sum to 0.95. A subsequent brain mechanism could be 95% certain that the stimulus belongs to this region. The maximum-likelihood confidence region quantifies identification performance; the smaller the region the better the performance.

For the cell in Fig. 2(A), the 95% confidence region along the dimension of spatial frequency was found to be slightly less than 1 octave (3.6–6.8 c/deg); the overall behavioral range of the macaque spans nearly 8 octaves (0.25–50.0 c/deg). Figure 2(B) shows the widths of the 95% confidence regions for the entire sample of cells; the average confidence region was 1.39 octaves (SD = 0.43). Interestingly, the confidence regions did not increase much at higher levels of confidence because the sides of the certainty function are quite steep; e.g., the average width of the 99% confidence region was only 0.2 octaves greater than the width of the 95% confidence region.

In the natural environment, more than one stimulus dimension is free to vary; therefore, it is of interest to consider identification performance for multiple dimensions. Figure 3(a) shows the certainty function for the cell in Fig. 1 along the dimensions of spatial frequency and contrast, given that a maximum response of 11 spikes was observed. As can be seen, the shape of the certainty function in the dimension of spatial frequency is approximately the same at all contrasts, indicating approximate probabilistic-independence of spatial frequency and contrast. Further, the most probable frequencies are confined to a relatively narrow region around the peak frequency, whereas the most probable contrasts are spread out over a considerable range. For this cell, the 95% confidence region ranges from 3.6 to 6.8 c/deg in spatial frequency and from 11% to 100% in contrast. The interesting result is that the confidence region for spatial frequency remained the same as in the one-dimensional case, even though the contrast was completely free to vary. Thus, when spatial frequency and contrast were completely uncertain, a single maximum response was sufficient to identify the stimulus as belonging to a narrow spatial frequency range. Furthermore, because maximum responses occurred at low contrasts (due to response saturation), spatial frequency identification was accurate over a wide range of contrasts.

Parallel results were obtained for other stimulus dimensions (which will be reported more fully in a subsequent paper). We found that the dimensions of position and contrast were also probabilistically independent; thus, when position and contrast were uncertain, a maximum response was sufficient to identify the stimulus as belonging to a narrow range of positions. Similarly, we found that direction and contrast were independent; thus, when direction of motion and contrast were uncertain, a maximum response was sufficient to identify the direction of motion (in those neurons with moderate to high direction selectivity). For example, a maximum response from the cell in Fig. 3 was sufficient to identify the direction of motion with almost 100% certainty (even though the ratio of the responses in the two directions was 0.32). Finally, when spatial frequency, direction of motion, and contrast were all uncertain (three dimensions of uncertainty), a maximum response was sufficient to identify the stimulus as belonging to a narrow range of spatial frequencies, moving in the preferred direction. It seems likely that probabilistic-independence will hold for other stimulus dimensions, and to the extent that it does, identification performance based upon a maximum response along any one dimension will be little affected by uncertainty along any other dimension.

Thus far, we have only considered identification performance when a maximum response has occurred. Figure 3(b) shows the certainty function for the cell in Fig. 1 (along the dimensions of spatial frequency and contrast) when the response was only 9% of the maximum. In this case, the identification performance is poor; the most probable frequencies and contrasts are not constrained to a small region. Although there is some probability that the stimulus is close to the center frequency, most of the volume under the certainty function is distributed away from the center frequency [unlike in Fig. 3(a)]. The 95% confidence region has a “horseshoe” shape in the contrast–spatial frequency plane, with a range of 1–11 c/deg (3.5 octaves) for spatial frequency, and 2–100% for contrast. In general, the smaller the response from a cortical cell, the larger the confidence region, and hence the poorer the identification performance.

Figure 4 plots the identification performance of the neuron (95% confidence regions) for 12 discrete response levels (0–11 spikes). Each layer defines the set of most probable spatial frequencies and contrasts for a particular response level; the summed probability of each set is 95% of the volume under the corresponding certainty function. For example, the sum over the set described by the one-spike layer is 95% of the volume under the certainty function shown in Fig. 3(b), and similarly, the sum

*To understand this result note that a near maximum response (11 spikes) was reached at low contrasts for near optimal spatial frequencies (see Fig. 1). Thus, small responses were rarely produced by near optimal frequencies except when the contrast was near zero.
over the set described by the 11-spike layer is 95% of the volume under the certainty function shown in Fig. 3(a). As noted above, at the lowest response level the 95% confidence region is horseshoe shaped and the set of most probable spatial frequencies spans more than 3 octaves. On the other hand, at the highest response level the 95% confidence region is more compact and the set of most probable spatial frequencies spans less than 1 octave.

**DISCUSSION**

In summary, we developed a method for measuring the identification performance of single neurons under conditions of high uncertainty, comparable to the stimulus uncertainty in the natural environment. Performance was quantified by measuring the region in stimulus space to which the stimulus could be classified with 95% accuracy (i.e., maximum-likelihood confidence regions). These regions decreased in size as response increased. Furthermore, when a near-maximum response occurred, the width of the confidence region along a given dimension was essentially unaffected by additional dimensions of uncertainty.

The results show that the contrast gain and exponent nonlinearities enhance identification performance over what would be possible with linear summation of excitation and inhibition alone (i.e., with a linear filter). Consider for example the performance of a linear neuron—a neuron which can be described as a linear filter*—under stimulus conditions where the mean luminance is fixed, but the contrast and other stimulus dimensions are free to vary. Such a linear neuron would exhibit good identification performance (a small 95% confidence region). We (e.g., Hamilton et al., 1989), and many others, have measured and described the receptive fields of visual neurons within the framework of linear systems analysis and have compared the behavior of the neurons with the behavior of a linear filter (for general reviews see Shapley & Lennie, 1985; De Valois & De Valois, 1988; Palmer, Jones & Stepnoski, 1991; Skottun, De Valois, Grosof, Movshon, Albrecht & Bonds, 1991).

**FIGURE 3.** Two-dimensional certainty functions of a typical cortical neuron. These plots show the probability of the possible stimuli (sine-wave gratings of arbitrary spatial frequency and contrast) given a particular response from a single cortical neuron in a 200 msec interval. These two-dimensional certainty functions were obtained from the descriptive functions shown in Fig. 1 using Bayes’ formula (see Appendix B). (a) The certainty function given a maximum response (11 spikes) is single humped. The most probable contrasts are spread out over a wide range because saturation occurs at a low contrast. The most probable spatial frequencies are confined to a narrow range because of the two nonlinearities: the expansive exponent and the contrast gain control. (The probabilities have been scaled to a peak value of 1.0; the absolute value of the peak was 0.0067.) (b) The certainty function given a weak response (one spike) is horseshoe shaped. The most probable contrasts and spatial frequencies are both spread out over a wide range. (The probabilities have been scaled to a peak value of 1.0; the absolute value of the peak was 0.0044.) Comparison of (b) with (a) shows that when a small response occurs, the neuron signals little about the stimulus spatial frequency; whereas when a large response occurs, the neuron signals that the stimulus belongs to a narrow spatial frequency range.

**FIGURE 4.** Stimulus identification performance for a typical neuron in the monkey visual cortex. Each layer shows the set of most probable spatial frequencies and contrasts (95% confidence regions), given a particular response level (a discrete number of action potentials occurring within a 200 msec fixation interval). For the lower response levels, the layers are horseshoe shaped. For the higher response levels the layers are no longer horseshoe shaped: the most probable spatial frequencies are tightly clustered around the center frequency. Spatial frequency axis, 0–20 c/deg; contrast axis, 0–100%. (Note that there are discrete layers because there can only be a whole number of action potentials in one trial.)

*We (e.g., Hamilton et al., 1989), and many others, have measured and described the receptive fields of visual neurons within the framework of linear systems analysis and have compared the behavior of the neurons with the behavior of a linear filter (for general reviews see Shapley & Lennie, 1985; De Valois & De Valois, 1988; Palmer, Jones & Stepnoski, 1991; Skottun, De Valois, Grosof, Movshon, Albrecht & Bonds, 1991).
confliction occurs. Unfortunately, near maximum responses can only occur in a linear neuron when the stimulus contrast is near 100%, and hence identification performance would be relatively poor over the typical range of contrasts in the environment. (This was verified by computing confidence regions for hypothetical linear neurons which had the same maximum response and response noise as typical cortical neurons.) The nonlinearities, on the other hand, allow near maximum responses to occur even at low contrasts, and hence enable good identification performance over a wide range of contrasts.

There is, however, a price paid for the enhanced identification performance created by the nonlinearities. The results suggest that the nonlinear mechanisms sacrifice identification performance along the dimension of contrast in order to enhance performance along other dimensions. Specifically, while the response saturation produced by the contrast gain control increases identification performance along many stimulus dimensions, it necessarily reduces identification performance along the dimension of contrast. Thus, in the process of enhancing identification performance, the nonlinearities create a fundamental dichotomy in the visual cortex between the encoding of contrast and the encoding of other stimulus dimensions.

This dichotomy between the encoding of contrast and other stimulus dimensions may have important consequences for understanding behavioral discrimination. For example, consider a population of striate neurons with a wide range of contrast sensitivities. As stimulus contrast is increased each neuron will eventually reach its saturated response level. When this occurs the neuron will no longer contribute to contrast discrimination. On the other hand, the neuron will continue to contribute to discrimination for the dimensions of spatial position, spatial frequency, direction of motion, and orientation. These facts may be related to the dissociation between discrimination along the dimension of contrast and discrimination along other stimulus dimensions, which has been demonstrated psychophysically. [For a recent review of the psychophysical literature see Bowne (1990).]

We have seen that a cortical neuron, responding near its maximum rate, strongly constrains the possible local features within the image, without the need for comparisons with other neurons. However, behavioral recognition of objects in the environment must surely involve combining the responses across a population of neurons. The results of this study suggest a possible strategy that subsequent brain mechanisms might use when combining V1 responses to perform object recognition. Consider the pattern of activity within the visual cortex when viewing a complex scene. Among the population of neurons, there will be a small subset that are responding at or near their maximum firing rate.* These neurons are the ones that will indicate most precisely the nature of the stimulus because of their small confidence regions; they will also have the greatest signal-to-noise ratio because the variance is proportional to the mean. Subsequent brain mechanisms might be able to make effective use of these two properties (the small confidence regions and the high signal-to-noise ratios) by employing a sequential scheme of image interpretation in which the most active neurons serve to restrict and guide the interpretation of the less active neurons. We note that the idea of beginning image interpretation on the basis of the least ambiguous information has appeared in other contexts. For example, Marr and Poggio (1979) suggested that the correspondence problem in binocular stereo vision might be solved by first matching the coarse features and then proceeding to finer features.

The results reported here, together with other recent evidence, suggest that cortical neurons can be described as C	728 filters (contrast gain/exponent filters) composed of four mechanisms: (a) a linear filter which establishes a neuron's stimulus selectivity through summation of excitation and inhibition; (b) a contrast gain control mechanism which makes selectivity invariant with contrast, in spite of response saturation; (c) an expansive nonlinearity which enhances selectivity; (d) a noise source which makes the variance of the response proportional to the mean, independent of the stimulus. This combination of mechanisms (which may be found in other regions of the cerebral cortex) provides an elegant solution to the problems of encoding and identifying stimulus attributes with neural elements that have limited dynamic response ranges and limited metabolic resources.

REFERENCES


*Most cortical neurons reach response saturation at contrasts well below 50% (e.g. Albrecht & Hamilton, 1982; Sclar et al., 1990). Furthermore, the average r.m.s. contrast in natural scenes has been estimated to be in the range of 40-50% (Geisler, unpublished observation). Thus, there should generally be a substantial subset of cortical cells for which the local stimulus is near optimal and for which the contrast is sufficiently high to produce a near maximum response.
Bayesian Analysis of Identification Performance

APPENDIX A

Description of Mean Response and Response Variability

The solid curves in Fig. 1(A, B) are the maximum-likelihood fit of the following equation:

\[
r(c, f) = \begin{cases} 
\frac{R_{max}}{c^2 + c_0} \exp \left( -\ln \left( \frac{f-f_0}{b_0} \right) \right) + R_0 & f \geq f_0 \\
\frac{R_{max}}{c^2 + c_0} \exp \left( -\ln \left( \frac{f-f_0}{b_0} \right) \right) + R_0 & f < f_0 
\end{cases}
\]

where \( c \) is contrast, \( f \) is spatial frequency, \( R_{max} \) is the maximum response, \( c_0 \) is the half-saturation contrast, \( n \) is an exponent, \( f_0 \) is the center frequency, \( b_0 \) is the half-height bandwidth for frequencies below \( f_0 \), \( b \) is the half-height bandwidth for frequencies below \( f_0 \), and \( R_0 \) is the spontaneous rate. (The maximum-likelihood estimates of the parameters for the cell in Fig. 1 are as follows: \( R_{max} = 10.73, c_0 = 9.21 \), \( n = 2.32, f_0 = 4.84, b_0 = 1.44, b = 2.29, R_0 = 0.03 \). This equation is a combination of a Naka-Rushton function

\[
r(c) = \frac{R_{max}}{c^2 + c_0}
\]

which has been used to describe mean response as a function of contrast (e.g. Albrecht & Hamilton, 1982; Sclar et al., 1990; DeAngelis et al., 1993), and a skewed Gabor function (i.e. a Gaussian function in the frequency domain, with different half-bandwidths above and below the center frequency).

The relationship between the mean and variance of cortical neuron responses has been described by a power function, with two parameters

\[
\sigma^2 = Kr
\]

where the value of \( K \) ranges from approx. 1.2 to 1.5 and the value of \( r \) ranges from approx. 1.0 to 1.2 (Tolhurst et al., 1983; Vogels, Spileers & Orban, 1989; Snowden, Trune & Anderson, 1992; Softky & Koch, 1993). However, we found that a one-parameter proportionality rule was adequate

\[
\sigma^2 = Kr'
\]

Specifically, the two functions above were fitted to the variability data measured from each neuron, using maximum-likelihood methods. We then compared the adequacy of the two fits for each neuron, using an F-test, and found that the power function did not provide a significantly better fit. (One consequence of this result is that the identification performances reported here are essentially unaffected by the choice of equation.) Both the power function and the more parsimonious constant proportionality rule were able to account for more than 91% of the total variability in the SD data, for the entire sample of cells. The solid curves in Fig. 1 show the simultaneous fit of equations (A1)-(A3).

APPENDIX B

Determination of Identification Performance

In order to determine identification performance using Bayes' formula [equation (1)], we must first determine the probability of a response
given a stimulus, \( p(R | x) \). This probability density function was obtained from the means and variances under the assumption that the responses were approximately normally distributed:

\[
p(R | x) = \frac{1}{\sqrt{2\pi Kr(x)}} \exp \left( -0.5 \frac{(R - r(x))^2}{Kr(x)} \right).
\]

The probability density function was then substituted into Bayes’ formula to obtain the certainty function (also known as the a posteriori probability density function):

\[
p(x | R) = \frac{\exp \left( -0.5 \frac{(R - r(x))^2}{Kr(x)} \right)}{\int \frac{1}{\sqrt{r(x)}} \exp \left( -0.5 \frac{(R - r(x))^2}{Kr(x)} \right) dx}.
\]

In these equations, \( x \) is a point in a stimulus space composed of one or more dimensions (e.g. a contrast and a spatial frequency), and \( r(x) \) is a function which describes the mean response over the stimulus space [e.g. equation (2)]. The constant \( K \) is the proportionality factor that describes the relationship between the mean and variance of the neuron’s response, and thus \( Kr(x) \) is the variance. Both the descriptive function, \( r(x) \), and the constant \( K \) were fitted to the mean and variance data using a maximum-likelihood procedure.

The above equation was derived assuming that the response probability density, \( p(R | x) \), is approximately normal, with the variance proportional to the mean. In fact, at low mean responses, the measured probability densities tend to be skewed to the right (like the Poisson density). To assess whether such deviations from normality would affect the accuracy of equation (B2), we computed certainty functions for skewed distributions (like the Poisson) and for the normal distribution; we found them to be nearly indistinguishable (the differences in the certainty functions were generally <5%).

Once the certainty function, \( p(x | R) \), has been determined, the maximum-likelihood confidence region can be obtained by summing stimulus probabilities until the criterion confidence level, \( \beta \), is reached.

The summation begins with the most probable stimulus, \( \hat{x} \) and then proceeds to include the next most probable stimulus etc. Formally, this process can be expressed by the following equation:

\[
\beta = \int_{[x \in CR_{ss}]} p(x | R) \, dx
\]

where \( \beta \) is some level of stimulus probability. For example, a 95% confidence region is derived by starting with the largest value of \( \beta \) (\( \beta = p(\hat{x} | R) \)) and then decreasing \( \beta \) in small steps until the sum (the integral in this equation) equals 0.95. We refer to this final value of \( \beta \) as \( \beta_5 \). The 95% confidence region, \( CR_{ss} \), is the set of stimuli whose probability density exceeds \( \beta_5 \), i.e.

\[
CR_{ss} = \{ x : p(x | R) > \beta_5 \}.
\]

The smaller this set the better the identification performance. The 95% confidence region is a useful quantitative index of identification performance.

Another way to quantify identification performance would be to first define particular regions of stimulus space as categories, and then determine the accuracy with which the neuron could identify the stimulus as belonging to these categories. Maximum-likelihood identification accuracy, \( P(C) \), is obtained by summing the response probability density function [equation (B1)] over the different stimulus categories and over the possible responses:

\[
P(C) = \sum_{i} \max \left\{ \int_{[x \in S_i]} p(R | x)p(x) \, dx \right\}
\]

where \( S_i \) is the set of stimuli making up the \( i \)th category, and \( p(x) \) is the a priori probability density function. These calculations are also straightforward, but we chose here to report 95% confidence regions because they are easier to interpret and because they do not require defining arbitrary stimulus categories.

The procedures described above should be applicable to other cortical neurons and other stimulus domains; they can also be generalized to measure identification performance for ensembles of neurons.