Weakly Modulated Spike Trains: Significance, Precision, and Correction for Sample Size

CHOU P. HUNG, BENJAMIN M. RAMSDEN, AND ANNA WANG ROE
Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06520-8001

Hung, Chou P., Benjamin M. Ramsden, and Anna Wang Roe.
Weakly modulated spike trains: significance, precision, and correction for sample size. *J Neurophysiol* 87: 2542–2554, 2002; 10.1152/jn.00420.2001. Many single-unit electrophysiological studies of visual cortex have investigated strong evoked responses to simple stimuli such as oriented gratings. Experiments involving other types of stimuli, such as natural scenes, higher-order features, and surface brightness, produce single-unit responses that are more difficult to interpret. Experiments with brightness, in particular, evoke single-unit responses that are typically weakly modulated. When the brightness is generated by a visual illusion such as the Cornsweet illusion, statistical tests are often necessary to distinguish true responses from baseline fluctuations. Here, using data collected from cat Areas 17 and 18 in response to real and illusory brightness stimuli, we provide a method for detecting and quantifying weak but significant periodic responses. By randomizing spike trains (via bootstrap methods), we provide confidence levels for response significance, permitting the evaluation of both weak and strong responses. We show that because of a strong dependence on total spike number, response significance can only be appropriately determined with randomized spike trains of similar spike number. Such randomizations can be performed for both stimulus-elicited and spontaneously occurring spike trains. By developing a method for generating randomized modulated spike trains (phase-restricted randomization) from actual recordings, we calculate upper and lower confidence limits of *modulated* spike trains and describe how measurement precision varies as a function of total spike count. Finally, using this randomization method, we describe how a correction function can be determined to correct for measurement bias introduced at low spike counts. These methods may also be useful in the study of small but potentially significant responses in other systems.

INTRODUCTION

The ability to reliably quantify the magnitude and significance of weak single-unit electrophysiological responses is becoming increasingly important in neuroscience (Awiszus 1997; Moore et al. 1966; Vinje and Gallant 2000). Whereas robust responses are often recorded to strong sensory stimuli, responses to weak or illusory stimuli have been less studied. Responses to surface luminance and brightness, for example, are often weak (Hung et al. 2001, see Fig. 1; see also Rossi et al. 1996). The issue of weak signals also arises in other sensory and higher cortical areas, where minor differences in attention, memory, or planning are being recorded (Brumberg et al. 1996; Schulz et al. 2000).

To explore the mechanisms that underlie these classes of processing, one must be able to objectively judge whether a weak response is present or not. How does one distinguish an actual response from chance modulations in firing rate? The most direct way to examine this issue is to determine the significance of peaks or troughs relative to a baseline in the peristimulus time histogram (PSTH). This has been usually determined with regard to repeated but transient stimulus events (cf. Abeles 1982; Awiszus 1993a,b; Davey et al. 1986; Dörrscheidt 1981; Gerstein and Kiang 1960). Recent reports have also examined the role of sample size in assessing the significance of precise temporal patterns (Baker and Lemon 2000; Oram et al. 1999). However, very few published reports have tested the statistical significance of weak responses to sinusoidally modulated stimuli (Rossi et al. 1996). In this paper, we focus on the detection and quantification of weak sinusoidally modulated responses to brightness stimuli. We show how to account for differences in sample size when measuring the amplitude and significance of responses. We also show how to quantify response precision, the certainty of a response measurement, allowing us to determine whether two modulated responses are significantly different from each other. Although our methods are shown for sinusoidally modulated responses, they can also be applied to other types of periodic stimuli.

We have based our approach around two randomization methods. The first is a common test of response significance, which involves randomizing the order of spike arrival while preserving the interspike interval distribution (Cox and Lewis 1966; Manly 1997, Perkel et al. 1967; Rossi et al. 1996). This randomization method generates artificial spike trains that are unmodulated, even if the original spike train is a modulated brightness response. We refer to this method as *full randomization*. In this paper, we examine the rationale for using this method to determine response significance. We show that single-unit recordings from cat visual cortex support this measure of significance. However, we show that the number of spikes recorded affects response magnitude and measured significance. Furthermore, we show that the response significance should be determined from full randomization of the test response not spontaneous activity.

We also introduce a second randomization method that we call *phase-restricted (PR) randomization*. Unlike full randomization,

---

Address for reprint requests: A. W. Roe, Dept. of Neurobiology, POB 208001, Yale University School of Medicine, New Haven, CT 06520-8001 (E-mail: anna.roe@yale.edu).

0022-3077/02 $5.00 Copyright © 2002 The American Physiological Society

this method allows one to generate artificial spike trains with PSTHs that appear temporally similar to the original (seed) spike train (even if the original PSTH has sharp transient responses). Because artificial spike trains generated using this method preserve the modulation present in the original spike train, this method allows us to quantify the precision of modulated responses and thereby test whether two modulated responses are significantly different in strength. Using PR randomization, we show how to determine the minimum number of spikes needed to avoid mistaking a truly modulated response for a nonresponse. Finally, we show how this method can be used to determine and correct for biases that exist in PSTH responses. Our findings are not restricted to sinusoidally modulated stimuli and can be generalized to other types of periodic stimuli. They are especially relevant for weak PSTH responses and can be applied in other areas of the nervous system.

SINGLE-UNIT RESPONSES TO BRIGHTNESS MODULATION

Fifty-one single units were isolated in cat Areas 17 and 18 under pentothal anesthesia ($1–2$ mg $\cdot$ kg^{-1} $\cdot$ h^{-1}) and studied for electrophysiological response to real and illusory brightness modulation (Hung et al. 2001). Stimuli consisted of two rectangular surfaces separated by a stationary contrast border. Stimuli were positioned such that the recorded classical receptive field lay in the center of one of the surfaces, away from the contrast border (see Fig. 1A). In the "real" stimulus, the luminance of the two surfaces were sinusoidally modulated in counterphase at 0.5 Hz, such that increase in luminance of one surface was coupled with decrease in luminance of the other surface (8% contrast peak-to-peak; mean luminance, 32 Cd/m^2). In the illusory brightness stimulus ("Cornsweet" condition) (see Hung et al. 2001), only the immediate border contrast was modulated (16% contrast peak-to-peak; mean luminance, 32 Cd/m^2), but it produced a percept of distant surface brightness modulation very similar to that of the real stimulus (Burr
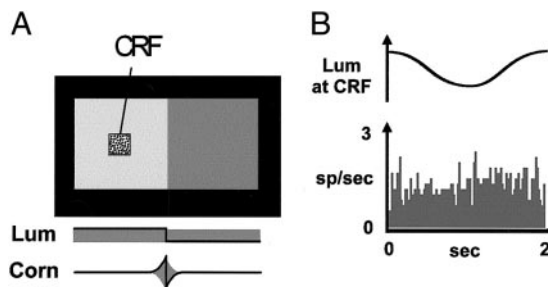

FIG. 1. *A*: schematic of luminance stimulus in which two uniform surfaces are separated by a contrast border that was positioned well away from recorded classical receptive field (CRF, small box). The luminances of the two surfaces are sinusoidally modulated in counterphase, such that increase in luminance of one surface is coupled with decrease in luminance of the other surface (0.5 Hz, 16 frames per modulation cycle, sign reversing around a mean luminance of 32 Cd/m^2, i.e., contrast incremented and then decremented in 16 luminance steps). In the illusory brightness stimulus ("Cornsweet" condition) (see Hung et al. 2001), only the immediate border contrast was modulated, but it produced a percept of distant surface brightness modulation very similar to that of the real stimulus. Luminance profiles of the luminance (Lum) and Cornsweet (Corn) stimuli are shown *below*. Regions shaded in gray indicate areas that are sinusoidally modulated in luminance. *B, top*: time course of the sinusoidal modulation of luminance around the mean. *Bottom*: an example perisimulus time histogram (PSTH) of a weak response in an Area 17 cell (192 bins, 11 ms/bin, 597 cycles).

1987; Cornsweet 1970). Response to the real luminance stimulus was compared with that to the illusory brightness stimulus. Responses were also compared with spontaneous activity (unmodulated gray screen; luminance, 32 Cd/m^2). In our stimuli, the strength of brightness modulation was weak (in the range where Cornsweet brightness percepts are most salient) (cf. Burr 1987) and often produced weakly modulated responses (e.g., Fig. 1B). We therefore felt the need to develop better methods to detect and quantify these weak responses and to objectively evaluate their significance.

RESPONSE TO WEAK LUMINANCE MODULATION IS BETTER DESCRIBED BY FIRING RATE MODULATION THAN BY AVERAGE FIRING RATE

It might be argued that average firing rate may be a sufficient index of neural response. However, cells in visual cortex exhibit a wide range of average firing rates during both spontaneous (i.e., "background") and stimulated conditions. We find that when the stimulus is a weak luminance modulation, average firing rate is not a good indicator of response. For some cells in our data set, average firing rate is higher during luminance modulation than during spontaneous activity, whereas for other cells it is less (Fig. 2). Indeed, across the population set, there is no significant difference between spontaneous and stimulated average firing rates ($n = 51$, Wilcoxon signed-rank test, $P = 0.5$).

Rather than exhibiting changes in *average* firing rate in response to luminance modulation, many cells are modulated in firing rate at the temporal frequency of the stimulus. Similar response modulations have been reported by other investigators for stimuli that include temporally varying luminance patches and moving gratings (Movshon et al. 1978a,b; Rossi et al. 1996; Skottun et al. 1991). To measure response modulation to our luminance stimuli, we fitted sinusoids by least-squares method to each PSTH and used the contrast ratio method (see Fig. 3A).

The data were fitted with sinusoids because the stimulus was sinusoidally modulated. Historically, sinusoidal fitting has been the most common method of fitting this type of data (Movshon et al. 1978a,b), although a sliding box filter has also been used (Rossi et al. 1996). Deviants from sinusoidal responses included responses with narrow or transient peaks. We developed fits for some of these but found the sinusoidal fit to be the most clear and reasonable. For both luminance and spontaneous conditions, sinusoids were fitted at the temporal frequency of the luminance stimulus (F1). We also fitted sinusoids at twice (F2) and three times the temporal frequency (F3). Over half the cells (28/51) showed a predominantly F1 response (i.e., F1 contrast ratio > F2, F3; see following text), one-third (17/51) showed a predominantly F2 response, and 12% (6/51) were dominated by F3. F1 and F2 response types correspond to cells that prefer the ON part of the cycle, the OFF part of the cycle, or both. Thus 88% of our luminance responses can be classified as F1 or F2, supporting the use of sinusoidal fitting for our data set. Except where specified, the remainder of this paper will deal with measurements based on the F1 sinusoid, which is the easiest to interpret in terms of a brightness response.

We defined the response contrast ratio as (max $-$ min)/(max $+$ min) of the fitted sinusoid. When modulation response
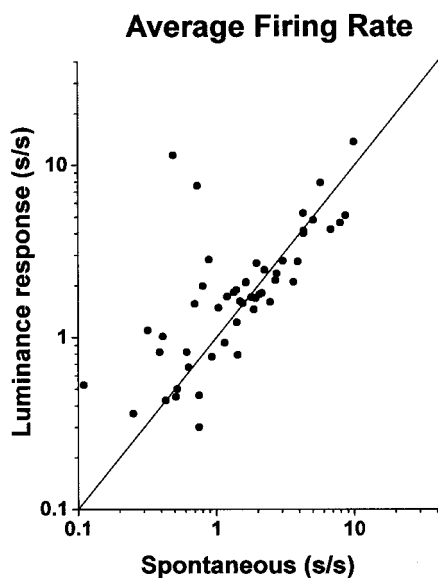
Average Firing Rate

FIG. 2. As a population, average firing rate was not significantly greater during luminance modulation vs. spontaneous activity ($n = 51$; Wilcoxon signed-rank test, $P = 0.5$). Average firing rates varied widely among individual cells in Areas 17 and 18, from 0.1 spikes/s to >10 spikes/s (note log scale). —, no difference in average firing rate between spontaneous and luminance modulation conditions.

is thus defined, most cells show greater response under luminance modulation than during spontaneous activity (Fig. 3B, ●). This demonstrates that, in this case, contrast ratio is better than average firing rate for measuring response magnitude.

Although response modulation is better than average firing rate for revealing significant responses, there remain significant responses that by this measure alone fall within the spontaneous range. For this data set, the 95th percentile of the distribution of spontaneous contrast ratios was ~0.17 (the 95th percentile calculation was based on the 3rd largest contrast ratio of 51 neurons). Thus for response contrast ratios <0.17 (see shaded area), the measured luminance responses could not be distinguished from random fluctuations in spike activity by these methods.

In our study of illusory brightness, we found a subpopulation of cells in Area 18 that appeared weakly modulated by our Cornsweet stimulus (cf. Hung et al. 2001). However, the responses to the Cornsweet stimulus, shown as ○ in Fig. 3B, tended to be <0.17. Although these responses fell below the cutoff (<0.17), most responses were nevertheless greater than the spontaneous contrast ratios obtained from the same cell (○ above —). This suggested that basing the cutoff on the 95th percentile of the spontaneous population may be too conservative. We surmised that the contrast ratio method might be improved by using bootstrap methods and accounting for differences in sample size. To determine which responses in this range were truly significant, we devote the next few sections to examining how sample size might affect recorded response amplitudes and significance.

MEASURED CONTRAST RATIO OF SPONTANEOUS RECORDING DEPENDS ON NUMBER OF SPIKES RECORDED

We began by considering whether the contrast ratio might vary with differences in spike count during *spontaneous re-*

cording. Intuitively, spontaneous spike trains with a large number of spikes should tend toward a "flat" PSTH (i.e., contrast ratio tends toward 0). A hypothetical spike train with only one spike, however, would yield a contrast ratio >1 (because the trough of the fitted sinusoid is negative; e.g., for a PSTH with 192 bins, a contrast ratio of 2.3 is obtained). This suggests that contrast ratios based on fewer spikes may be overestimating the presence of actual modulation.

To investigate this further, we re-evaluated our spontaneous spike data with an emphasis on underlying spike counts. Figure 4A replots the spontaneous contrast ratios, sorted by total number of spikes recorded (contrast ratios were computed from sinusoids at the same periodicity as the brightness stimuli). As hypothesized, the higher spontaneous contrast ratios tended to arise from recordings with fewer spikes. Statistical analysis
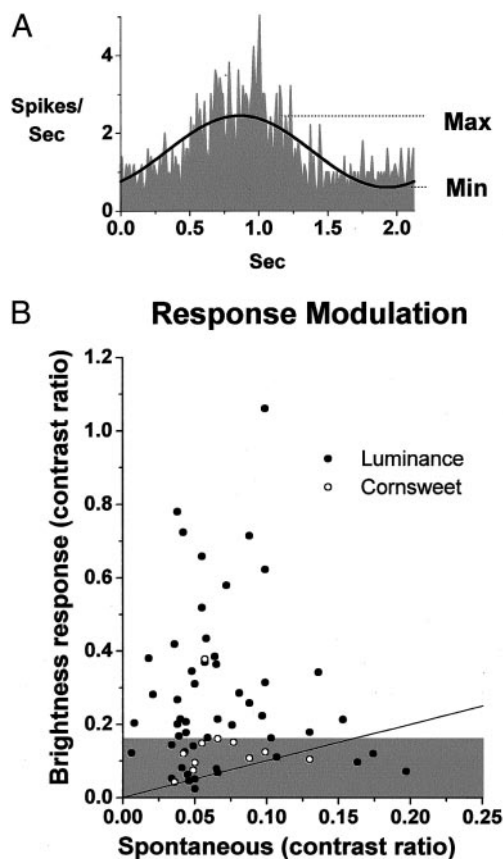

FIG. 3. Modulation of firing rate by the luminance cycle. A: the strength of firing rate modulation was measured by the contrast ratio method. A PSTH (192 bins, 11 ms/bin) of the response was generated. Response profiles were fitted by a sinusoid, and the peak and trough of the sinusoid were designated max and min, respectively. Response contrast ratio was measured as (max − min)/(max + min). A 0 contrast ratio indicates a flat histogram, and a contrast ratio of 1 indicates a trough at 0 spikes/s. On rare occasions, the best fitting sinusoid can have a trough <0 spikes/s, resulting in a contrast ratio greater than one. For the PSTH shown, the contrast ratio is 0.56. B: comparison of measured contrast ratio under spontaneous activity vs. luminance modulation (●, same 51 cells shown in Fig. 1) or illusory brightness modulation (○, Cornsweet stimulus, from a subpopulation of brightness cells in Area 18, $n = 11$). —, (slope = 1) no difference in contrast ratio between spontaneous activity and brightness modulation conditions. Most cells exhibited larger sinusoidal modulation in firing rate in response to brightness modulation compared with spontaneous activity (cells above —). ▨, responses below the 95th percentile of the spontaneous distribution (0.17). Most responses to the Cornsweet stimulus (○) were <0.17, making them indistinguishable from spontaneous by this method.
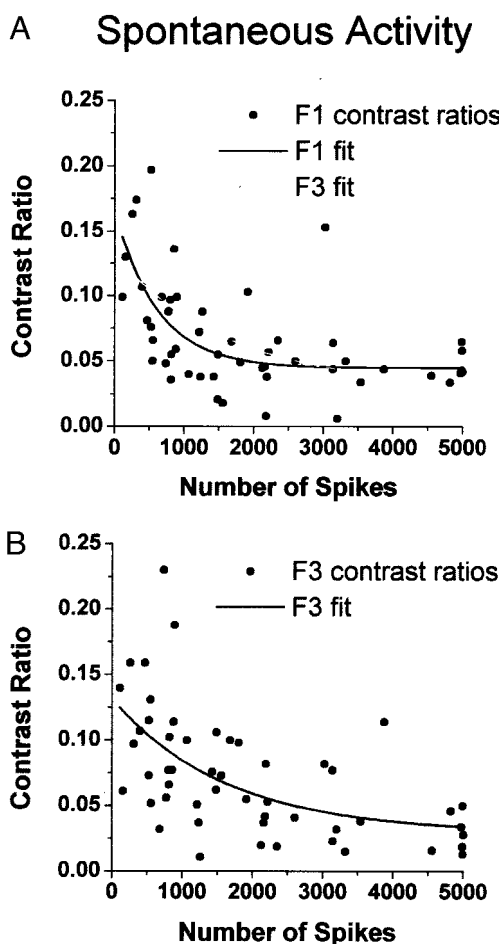
A Spontaneous Activity

FIG. 4. Measured contrast ratio of spontaneous activity of 51 cells depends on the number of spikes recorded. *A*: results calculated at the same temporal frequency (~0.5 Hz) as the luminance modulation condition. —, exponential fit to the distribution. Spike trains with more spikes tend to have lower contrast ratios. *B*: results calculated at 3 times the temporal frequency. Exponential fit to the distribution is shown by — and, in *A*, by - - -. Again, higher spike counts tend to coincide with lower contrast ratios.

revealed that the magnitude of measured spontaneous contrast ratio was significantly correlated with the number of spikes recorded (Fig. 4*A*, $R = -0.45$, $P = 0.001$, $n = 51$). The effect of sample size was similar when contrast ratios were measured from sinusoids at half and one-third the periodicity (F2 and F3 sinusoids, respectively). Results at one-third periodicity are shown in Fig. 4*B*. Our data thus suggest that the number of spikes recorded significantly affects the magnitudes of measured spontaneous contrast ratios.

SIGNIFICANCE AND THE NULL HYPOTHESIS

To test the null hypothesis, that a cell's luminance response is not significantly different from its spontaneous activity, one must be able to define what level of contrast ratio is indistinguishable from spontaneous activity. Figure 4 suggests that this threshold level depends on the number of spikes recorded. Given the effect of spike count on measured contrast ratio, how does one effectively test the null hypothesis? One method would be to use only spike trains with "many" spikes. A predetermined minimum spike count could be feasibly derived (e.g., for Fig. 4 data, a spike count threshold of 2,000). Al-

though it is preferable to record many spikes for a good measurement, the low firing rates of some cortical cells in some experimental protocols means that sometimes one must be satisfied with only a few hundred spikes from a 10-min recording. In this section, we address analytical methods that allow for the testing of the null hypothesis when total spike count is low.

Spike train randomization captures the range of spontaneous modulation

One way to test the null hypothesis, and determine the confidence limits by which to judge whether a response is truly modulated, is to base them on the spontaneous activity of each cell. The approach is to generate artificial spike trains from actual spontaneous spike trains by randomizing the order of interspike intervals (Fig. 5*A*). Such an approach, usually referred to as the "bootstrap" or "randomization" analysis, has been well documented for a broad range of data (Baker and Lemon 2000; Manly 1997; Oram et al. 1999). By using bootstrap methods to generate artificial spike trains, one can determine confidence limits for chance modulation (Cox and Lewis 1966; Perkel et al. 1967; Rossi et al. 1996).

The validity of the bootstrap method rests on the assumption that spikes occur as renewal processes. However, second- and higher-order interactions between successive interspike intervals are known to occur (e.g., bursts of firing, or short intervals alternating with long intervals) (cf. Baker and Lemon 2000; Oram et al. 1999). We therefore wondered whether this method truly captured the range of spontaneous fluctuations in the PSTH. For example, is the calculated 95% confidence level truly above 95% of recorded spontaneous activity? We based our analysis of spontaneous data on these methods as follows. We took each recorded spontaneous spike train and fully randomized the sequence of interspike intervals (ISIs) to create 1,000 new spike trains with the same interspike interval distribution (henceforth called "full randomization"). The contrast ratios of the randomized spontaneous spike trains were derived and sorted, and then the cell's actual luminance-modulated response contrast ratio was compared with this sorted population to determine its *confidence level*. We then did this for all 51 cells studied. One would expect that out of 100 spontaneous recordings, N of them will fall below the $N\%$ bootstrapped confidence level, e.g., 95 of them will fall below the 95% confidence level (i.e., a plot of actual vs. expected rank should yield a line of slope 1). Figure 5*B* shows the confidence levels of recorded spontaneous activity for our population of cells, arranged by confidence level ($n = 51$). As might be expected, the recordings yielded bootstrapped confidence levels that evenly spanned the entire range (i.e., the two distributions line up along the 1:1 diagonal and are not significantly different; $P = 0.8$, Kolmogorov-Smirnov test). Had the recordings resulted in predominantly low or high confidence levels, it would indicate that the randomization method was over- or underestimating, respectively, the range of modulation levels in spontaneous activity. The linearity of the plot (a quantile-quantile plot) (cf. Hamilton 1992) supports the validity of this randomization method in determining confidence levels.
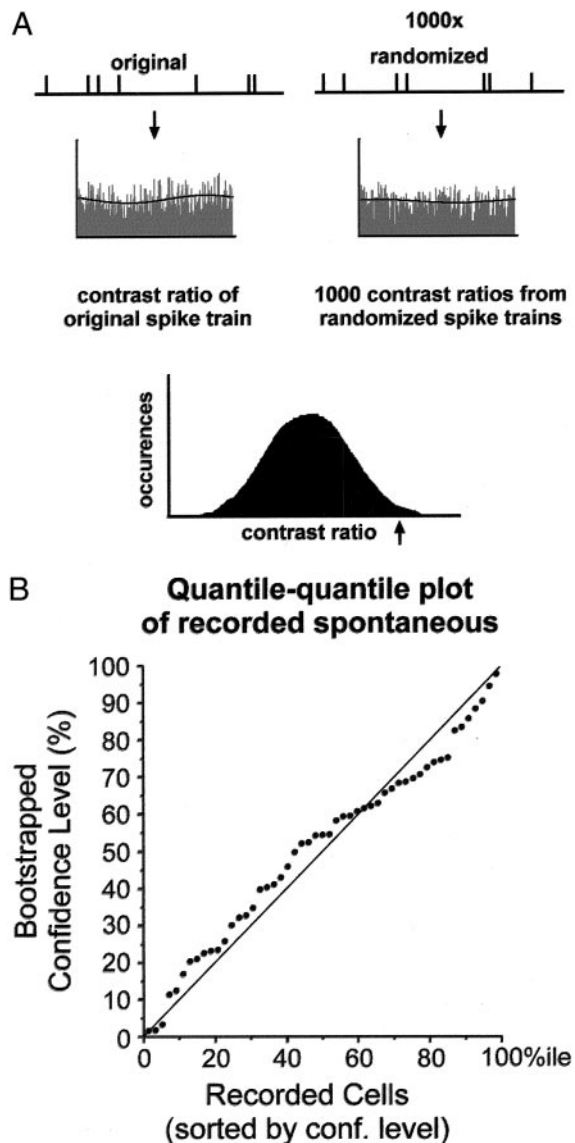
FIG. 5. *A*: the full randomization method of determining response significance. One thousand random spike trains, *right*, are generated from an actual spike train, *left*, by randomizing the sequence of interspike intervals. Contrast ratios of the 1,000 randomized spike trains are sorted, *bottom*, and the contrast ratio of the original spike train is compared against this distribution to arrive at the bootstrapped significance level. ↑, ~95th percentile of the distribution. *B*: quantile-quantile plot of bootstrapped confidence levels of spontaneous activity recorded from 51 cells (spike collections ranged from 96 to 1,421 "cycles" of ~2 s each). The abscissa and ordinate indicate the actual and expected confidence level rankings, respectively. For each recorded spike train, its bootstrapped confidence level (the value along the ordinate) was calculated as follows: the sequence of interspike intervals was randomized ("full-randomization") to generate 1,000 artificial spike trains. Contrast ratios of these randomized spike trains were sorted and compared against the measured contrast ratio of each spontaneous recording to arrive at a confidence level for each recording. For example, if a measured contrast ratio were ranked 351 smallest of 1,000, then the bootstrapped confidence level would be 35.1%. The bootstrapped confidence levels of the 51 responses are then sorted and arranged to lie on the abscissa between 0 and 100 (percentile ranking; values are $\{[(1/51, 2/51 \cdots 51/51) - 1/102]*100\}$. Results lie along the diagonal and span the entire range ($P = 0.8$, Kolmogorov's test), suggesting that the randomization method (which assumes spikes occur as renewal processes) is appropriate for determining the confidence level.

Bootstrap significance levels are also dependent on number of spikes recorded

To further test whether the bootstrapped spike trains have the same behavior as real spike trains, we examined the dependency of bootstrapped spike trains on total spike count. Figure 6*A* shows how randomization-derived confidence levels change with spike count. Here we plot the results of full randomization of spontaneous activity for each of the cells ($n = 51$). Large open circles indicate 50% confidence level. Smaller upper and lower open circles indicate 95% and 5% confidence levels, respectively. For comparison, the measured spontaneous contrast ratios from all cells (filled circles, same as Fig. 4) are also shown. Thus for each real measurement, indicated by a filled circle, there are three bootstrapped measurements at 5, 50, and 95%, indicated by open circles. Comparison of the solid curve (exponential fit of filled circles) with the thicker dotted curve (exponential fit of large open circles, 50% confidence levels) reveals the similarity between measured and bootstrapped values and their similar dependence on spike number.

The dependence on number of spikes exists not only for the 50% confidence level but also for other confidence levels (95 and 5% confidence levels shown by upper and lower open circles, fitted by dotted curves). When sample size is taken into account, the 95% confidence level ranges from 0.3 at 100 spikes to ~0.08 at 5,000 spikes, a dramatic improvement over the previous 0.17 cutoff. In sum, these findings suggest that to appropriately test the null hypothesis (i.e., that a response is not present), the number of spikes recorded must be taken into account. The full randomization bootstrap is one means to create objective sample size-corrected confidence levels.

Significance level depends on number of spikes in the test response not number of spikes during spontaneous activity

Given that confidence levels vary with spike count, the significance of a test response should ideally be evaluated with a spontaneous spike train with matching spike count. However, in actual recordings the number of spikes collected under the test condition is often very different from that collected under the spontaneous condition. Figure 6*B* illustrates luminance-modulated (triangles) and spontaneous (dots) response contrast ratios for nine cells in our sample. These are arranged along the abscissa by number of spikes recorded. For clarity, only luminance responses <0.40 are shown. Gray lines are shown to indicate the luminance-spontaneous pairings. Dotted curves drawn indicate 5, 50, and 95% confidence levels (of values shown in Fig. 6*A*). As can be seen, the number of spikes recorded for the same cell under luminance versus spontaneous conditions can vary greatly. One example (straight black line) spans almost the entire range, from 256 spikes (279 spontaneous "cycles") to 4,981 spikes (205 luminance cycles, 0.5 Hz). For this cell, the contrast ratio of the luminance response (0.10) is below the spontaneous contrast ratio (0.17) recorded from the same cell. In addition, the luminance response is well below the 95% confidence limit at 256 spikes (0.27, indicated by arrow), based on the distribution of contrast ratios for its randomized *spontaneous* spike trains. However, it would be incorrect to conclude that this cell is unresponsive to luminance, as its luminance response is clearly above the 95%
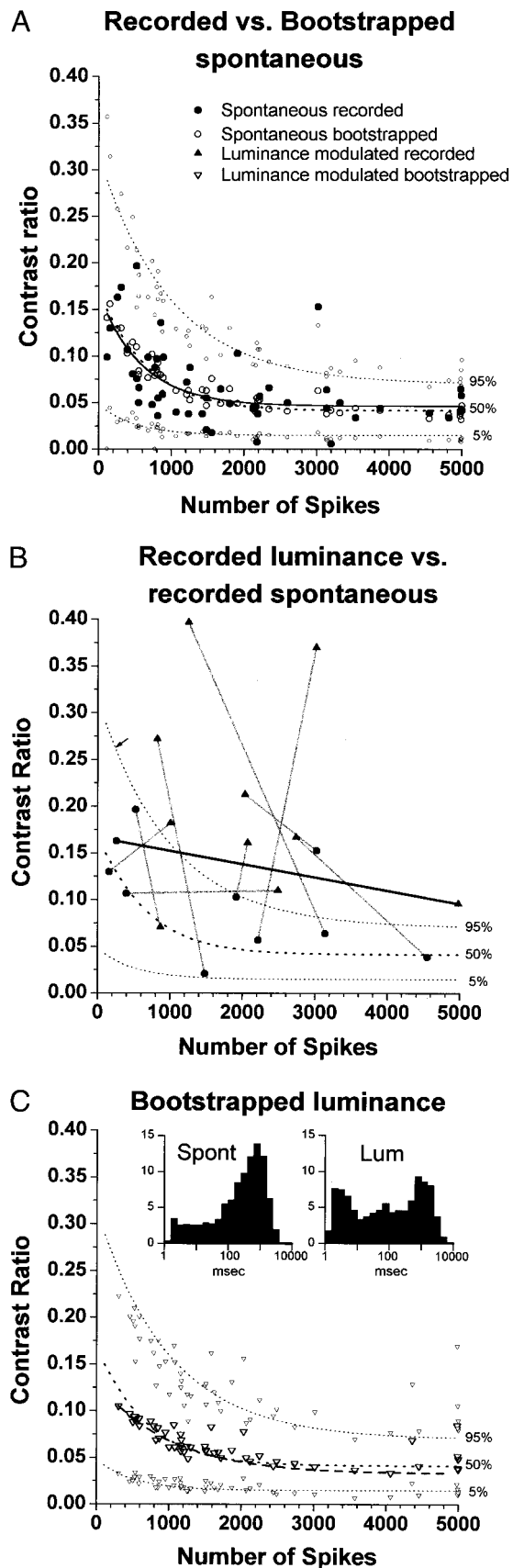
confidence limit (0.08) calculated at 4,981 spikes (the number of spikes in the *luminance-modulated* spike train). This example shows that, when determining significance level, significance should be based on the same number of spikes as the test response spike train rather than the number of spikes in the spontaneous recording. Such a spike count-matched spontaneous spike train can be generated with bootstrap randomization of the test response spike train.

Randomized luminance spike trains provide similar confidence limits compared with randomized spontaneous spike trains

Because spike number is critical in determining significance, we would like to use the test response spike train itself as a seed for generating randomized spike trains. However, as the two spike trains were recorded under very different conditions (luminance modulated vs. spontaneous), there may be differences in ISI distribution that could affect these bootstrap results. From our observations, whereas spontaneous ISI distributions typically have only one peak, luminance-modulated ISI distributions can have two peaks due to bursts in the spike train (example shown in Fig. 6C, insets). We wondered whether this difference in ISI distribution would affect the bootstrapped confidence levels of luminance modulated spike trains. Thus before using the luminance response spike train for bootstrap randomization, we must first confirm that the ISI distributions of the stimulated and spontaneous spike trains result in similar confidence levels.

Figure 6C compares the distributions of confidence levels derived from the full randomization of spontaneous versus luminance-modulated spike trains. For the randomized *luminance-modulated* spike trains, confidence levels are indicated by large hollow triangles (50%, fitted by dashed curve) and

FIG. 6. *A*: the 95% confidence level of spontaneous activity is not a fixed level of contrast ratio but instead depends on the number of spikes recorded. Contrast ratios of spontaneous activity (filled circles, $n = 51$) are shown arranged by number of spikes. For each cell, 50% and 5%/95% bootstrapped confidence levels of spontaneous activity (large open circles, and lower and upper small open circles, respectively) were also calculated by full randomization. Fits to the recorded contrast ratio and to the 5/50/95% confidence levels are shown by solid and dotted curves, respectively. The triangles in the figure legend refer to points in *B* and *C*. *B*: response significance should not be based on the 95% confidence level of spontaneous activity because of differences in number of spikes recorded. Most luminance-modulated responses (filled triangles) have a different number of spikes compared with spontaneous activity (filled circles) recorded from the same cells (for 10 cells, pairing of recordings from the same cell indicated by gray or black lines). For clarity, only contrast ratios <0.40 are shown. For comparison, 5/50/95% confidence levels of spontaneous activity are shown. For the pair indicated by black line, the luminance response contrast ratio is lower than that of the paired spontaneous activity and below the 95% confidence level of the paired spontaneous recording (indicated by arrow at 0.27 contrast ratio). However, the luminance response is significant (greater than the 95% confidence level) after accounting for the number of spikes in the luminance response. *C*: luminance-modulated and spontaneous spike trains yield similar bootstrapped confidence levels despite differences in interspike interval (ISI) distribution. Comparison of 5/50/95% confidence levels of full-randomized luminance-modulated spike trains (triangles, $n = 51$) vs. spontaneous spike trains (dotted lines, same as *B*). Fit to the 50% confidence levels of the randomized luminance-modulated spike train is indicated by dashed (luminance) curve. *Insets*: histograms of ISI distributions recorded from an Area 17 cell during spontaneous activity (*left*) vs. luminance modulation (*right*). Ordinate shows percentage of ISIs represented by each bin. Note logarithmic abscissa, which shifts the peak of the histogram.

smaller upper (95%) and lower (5%) triangles. For comparison, the confidence levels of randomized *spontaneous* spike trains are indicated by lower, middle, and upper dotted curves (5/50/95%, same as shown in Fig. 6, *A* and *B*). As can be seen, full randomizations of both luminance (triangles) and spontaneous (dotted lines) spike trains yield similar, overlapping distributions of confidence levels.

Of 51 cells, only 3 (at 4,000–5,000 spikes) appear to deviate from the 95% confidence level. The histograms shown in the *insets* are not from among the outliers (their 95% confidence levels lie very close to the fitted 95% confidence level curve). We have examined the ISI histograms of luminance responses of the longer spike trains (4,000–5,000 spikes). Eight of eight have clearly discernable double peaks. The three that appear to be outliers (95% confidence level, >0.10) are different from the others in that the bulk of their ISIs (>50%) are short (<10 ms) with very weak peaks near 1,000–2,000 ms. For the largest outlier, >90% of the ISIs are <5 ms. The remaining five have comparatively stronger and broader peaks near 1,000–2,000 ms. We surmise that the outliers result from extreme burstiness.

Thus with the exception of a few very bursty cells, there is no difference between the 95% confidence limits of spontaneous versus luminance-modulated spike trains despite their different ISI distributions. Together, these findings suggest that one can randomize the test response, rather than spontaneous activity, to determine response significance.

PR RANDOMIZATION FOR MODULATED SPIKE TRAINS

In the previous section, we discussed the full-randomization method. Although full randomization is suitable for determining the response significance of a single cell (compared with spontaneous), it is unsuitable for determining whether two stimulated responses are significantly different from each other. This is because full randomization can generate only flat PSTHs (i.e., any modulation present in the original spike train is lost; see Fig. 6*C,* full-randomized luminance spike trains are as flat as randomized spontaneous). Here, we introduce a new randomization method in which permutations are restricted based on phase proximity. This method preserves the modulation found in the original spike train, generating artificial spike trains whose PSTHs appear similar to the original. We call this method phase-restricted randomization, or PR randomization. Because PR-randomized spike trains remain modulated, it allows us to measure the effect of sample size on *modulated* spike trains in ways that are impossible with other methods such as full randomization. In the next section, we will use it to measure how sample size affects precision (i.e., the certainty) of measured responses. Finally, this method will allow us to generate a function to correct for sample size-related bias in the measurement of weak responses.

The PR randomization method is based on the idea that spikes occur as renewal processes. There are conditions under which renewal processes can be an inappropriate model (e.g., nonstationarity, bursting, precise synchronization) (cf. Abeles 1991; Gerstein et al. 1989). However, it is an adequate model for our purposes and is supported by the linearity of the quantile-quantile plot of the full-randomized spontaneous distribution (Fig. 5*B,* indicating the good match between expected and actual confidence level distributions) and by previous studies (Baker and Lemon 2000; Cox and Lewis 1966; Perkel et al. 1967). The PR randomization method further assumes that modulated spike trains occur as a result of a time-dependent ISI distribution but makes no assumption as to the shape (e.g., Poisson) (cf. Abeles 1982; Dörrscheidt 1981) or time-dependent nature (e.g., sinusoidal) of the ISI distribution.

The PR randomization method is depicted in Fig. 7*A*, and results of several randomizations are shown in Fig. 7, *B*, *C*, *E,* and *F*. At the top of each panel is a PSTH, and below it the corresponding peristimulus time ISI plot (PSTISI). The abscissa indicates time within the stimulus cycle. The ordinate of the PSTISI indicates the time interval between each point (spike) on the plot and the subsequent spike. To generate the randomized spike train, one begins by taking an actual spike train recorded in response to luminance modulation (Fig. 7*A*). The start times of ISIs of this spike train are sorted relative to the stimulus trigger and plotted along the abscissa in Fig. 7*A, bottom*. We have used this ISI data to generate a series of new randomized spike trains as follows. First, a spike is chosen at random from the spike train (consequently, time was not evenly covered in the randomization). From the randomly chosen spike at t_1 (see Fig. 7*A*, *bottom*), one waits the corresponding ISI (in this case, the ISI for the spike at t_1 is 500 ms) until t_2. The duration of the subsequent ISI is randomly selected from a randomization window of 10 ISIs (encircled in gray, 5 ISIs before and 5 after t_2; note that all ISIs were sorted relative to the start of each *cycle*), and this ISI continues from the first, beginning at t_2. We wish to emphasize that this randomization is *not adjacent in time*, as the 10 ISIs are not sequential and may be separated from each other by many cycles. The four arrows in the figure illustrate 4 of the 10 possible paths the random sequence may take at t_2. They are meant to show that some paths have longer ISIs, and others have shorter ISIs. The tips of the arrows suggest where t_3 might occur. This process of random jumps is repeated until the desired number of spikes has been generated.

We used a window size of 10 for all PR randomizations in this paper. We have also tried other window sizes (e.g., 2, 25, and 100). Although wider window sizes may provide better randomization, they produce noticeable blurring and flattening of the PSTH. Narrower window sizes better correspond to the width of transient responses in the PSTH (we are assuming that the neuron, within the time window, is receiving roughly the same pattern of inputs regardless of cycle). However, we remained concerned about the effective level of randomization at narrow window sizes. Although a narrow window size of 2 provides, for a 100-spike spike train, $2 \wedge 100$ (= $10 \wedge 34$) possible permutations, the effective number of permutations remains limited by the shape of the ISI distribution (e.g., the bursty outliers shown in Fig. 6*C*). Our choice of a randomization window size of 10 corresponds to a time window of ~3 ms (ranging from 1.66 to 6.64 ms, depending on firing rate) for the 6,988-spike train example shown in Fig. 7*A*. For comparison, the PSTH bins are ~11 ms wide.

Figure 7, *B* and *C,* shows two randomized trains of 1,600 spikes each generated from the train in Fig. 7*A*. Note that both the PSTH (*top*) and the ISI distribution (*bottom*) appear similar to the original. The close match in both PSTH and ISI distribution would not be possible with other randomization methods such as a sinusoidally-modulated Poisson distribution. This

Luminance response
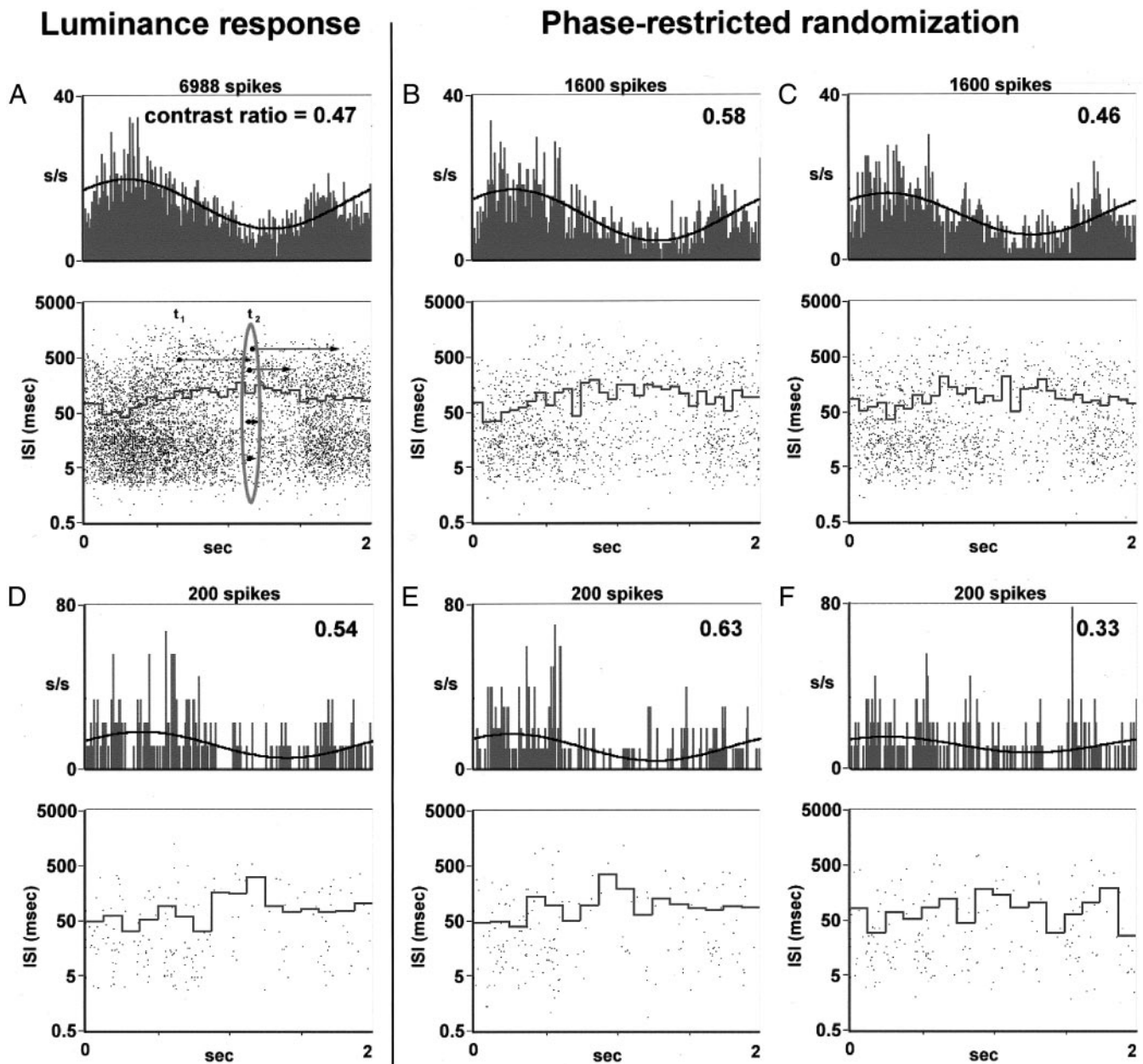
Phase-restricted randomization

FIG. 7. The phase-restricted (PR)-randomization method for generating artificial spike trains for modulated responses. *A, top*: the PSTH of a cell's response to luminance modulation. Black line indicates sinusoidal fit to the histogram. *Bottom*: the peristimulus time ISI plot (PSTISI) of the recording. Stepped gray line indicates mean ISI of sequential 67-ms bins. Artificial spike trains, shown in *B*, *C*, *E*, and *F*, were generated from this spike train under the assumption that spikes are renewal processes and that the shape of the underlying probability distribution varies predictably during the stimulus cycle. Note that no assumption was made of the actual shape of the probability distribution (e.g., Poisson). The set of ISIs was sorted relative to the time from the stimulus trigger. Following a predetermined ISI from t_1 to t_2, the magnitude of the subsequent ISI is randomly chosen from the set of 10 ISIs nearest in time to t_2 (encircled in gray). The subsequent ISI begins at t_2, and the process is repeated until the desired number of spikes has been generated. The 4 arrows illustrate 4 of the 10 possible paths the random sequence may take at t_2. They are meant to show that some paths have longer ISIs, and others have shorter ISIs. The tips of the arrows suggest where t_3 might occur. For this spike train, a randomization window of 10 spikes corresponds to ~3 ms (6,988 spikes, 301 cycles, 2.134 s/cycle). The value 10 was arbitrarily chosen and provides a good balance between the need for randomization and the desire to preserve precise temporal relationships in the PSTH. *B* and *C*: examples of 2 1,600-spike trains artificially generated from *A*. *D*: recorded response shown in *A*, truncated to 200 spikes. Mean ISI (stepped gray line, *bottom*) is binned at 133 ms each. *E* and *F*: truncated artificial spike trains of *B* and *C*, 200 spikes each. Compare with recorded spike train shown in *D*.

demonstrates that artificial spike trains generated by PR randomization exhibit similar temporal structure to recorded spike trains and have comparable contrast ratios.

We have also observed that, because of smaller sample size, shorter spike trains exhibit greater variation in their contrast ratios (see also next section, Fig. 8). Figure 7D shows a typical example of a 200-spike train truncated (not randomized) from the luminance response shown in Fig. 7A. Its contrast ratio is
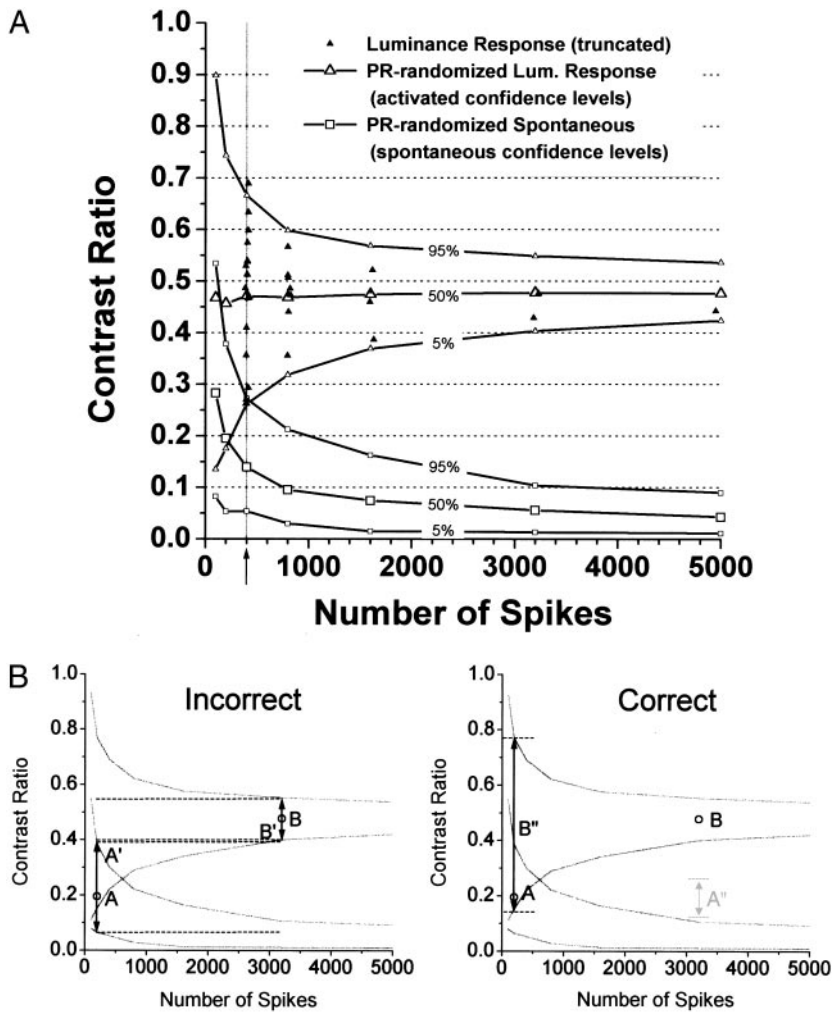
FIG. 8. *A*: precision of measured luminance response as a function of spike count. Same cell as in Fig. 7. Contrast ratios of truncated luminance spike trains (closed circle) demonstrate that measured contrast ratio is less precise in shorter spike trains. Recorded responses lie within the 5–95% confidence interval of PR-randomized spike trains (open triangle, randomization window of 10) generated from the same luminance response. PR-randomized spike trains of the spontaneous recording from the same cell (open square) yielded confidence levels consistent with full-randomized spontaneous responses (Fig. 6*A*). Confidence levels were determined from 1,000 randomizations. Significant overlap can be seen between the PR-randomized luminance and spontaneous distributions for spike trains with <400 spikes (indicated by arrow and gray vertical line). *B*: confidence levels determined by PR randomization can be used to test whether 2 contrast ratios are significantly different. Circles A and B indicate the 2 contrast ratios to be compared. Spike train A has 200 spikes and a contrast ratio of 0.20. Spike train B has 3200 spikes and a contrast ratio of 0.48. Gray curves suggest possible confidence bands of A and B. *Left*: incorrect method for testing significance. A′ and B′ indicate confidence intervals around A and B. Comparisons of A against B′, and B against A′, incorrectly find that A and B lie outside the confidence intervals and are significantly different. *Right*: correct method for testing significance. A″ and B″ denote confidence intervals calculated from spike trains A and B at spike counts of B and A, respectively. Because PR randomization fails to extrapolate to higher spike counts, A″ (shaded in gray) cannot be used to compare against B. Comparison of B″ against A, however, finds that A lies within the confidence band of B. Thus A and B are not significantly different ($P > 0.05$, 2-tailed test, 2.5–97.5% confidence band).

0.54. Other 200-spike trains truncated from the same luminance response have contrast ratios ranging from 0.2 to 0.8 (see filled triangles in Fig. 8*A*). Figure 7, *E* and *F,* shows randomized trains of 200 spikes each, truncated from the spike trains shown in *B* and *C*. Their contrast ratios are 0.63 and 0.33, well within the 0.2–0.8 range obtained from truncations of the actual recording.

Thus PR randomization produces artificial spike trains that preserve the original spike train's modulation structure and desired number of spikes. Although the example we show is a sinusoidally modulated spike train, this method works equally well for spike trains with transient responses. Previous studies have also included methods for generating appropriate surrogate data (Oram et al. 1999; Baker and Lemon 2000). They differ from this method in that they simulate the spike train by matching a hypothetical instantaneous firing rate, whereas this method deals only with the PSTISI plot. All of these methods capture temporal modulations in firing rate, although both they and PR randomization are limited in the maximum number of spikes they can generate (they should not be used to extrapolate beyond the number of spikes in the original spike train). In the next section, we use PR randomization to determine the how the precision of measurement varies as a function of spike count. Finally, we show how to determine the bias in measured contrast ratio as a function of spike count, thereby permitting the comparison of modulation responses from different cells and different recordings.

PRECISION OF MEASURED CONTRAST RATIOS OF LUMINANCE RESPONSES

We have shown how to generate artificial modulated spike trains using PR randomization, and we now apply this method to examine the certainty of measured luminance response modulation (i.e., response precision). Our method provides an estimate of modulation strength contained in the neural response. When two contrast ratio estimates are compared, each has an uncertainty associated with it. Given these uncertainties, how can we be sure that a measured contrast ratio of 0.50, for example, is significantly more modulated than a contrast ratio of 0.40? The answer, again, depends on the number of spikes recorded. Recordings with more spikes tend to provide a greater degree of precision, meaning that repeated presentations of the same stimulus yield smaller differences in measured contrast ratios. As with the randomized spontaneous spike trains, this precision can also be quantified and indicated by confidence levels. To differentiate confidence levels of spontaneous activity from that of luminance modulated spike trains, we will use the terms "spontaneous confidence levels" and "activated confidence levels," as indicated in Fig. 8.

To derive the activated confidence levels for the luminance response, we generated for the spike train of Fig. 7A (using the PR method) 1,000 modulated spike trains for each spike count (100, 200, 400, 800, 1,600, 3,200, and 5,000 spikes; 7 × 1,000 spike trains). The 7,000 PR-randomized spike trains were not subsets of each other. This produced an activated confidence band (indicated by open triangles) as shown in Fig. 8A. Five, 50, and 95% confidence levels (lower, middle, and upper open triangles, respectively) are shown. Comparison of this measure of precision with measured contrast ratios of truncated portions of the original spike train (filled triangles) suggests that this method is well suited for activated spike trains.

To examine whether this method also produces reasonable spontaneous confidence levels, we also generated 1,000 PR-randomized spike trains derived from spontaneous activity recorded from the same cell. As shown by open squares in Fig. 8A, the 5, 50, and 95% spontaneous confidence levels are comparable to those in Fig. 6A (generated by full randomization), supporting the validity of this method for generating spontaneous confidence levels as well. That our choice of window size (10 spikes) is sufficiently large, and thus our spike trains sufficiently random, is indicated by the similarity between these two distributions.

This application of PR randomization allows one to evaluate the response difference between two spike trains. For example, two spike trains' contrast ratios are significantly different from each other at the $P = 0.05$ level (a 2-tailed comparison) if each contrast ratio lies outside the 2.5–97.5% confidence *band* of the other contrast ratio. The 5 or 95% confidence level should be used for one-tailed comparisons at $P = 0.05$.

We stress that the confidence band, not the confidence interval, should be used. By confidence band, we mean that the calculation of confidence interval should take into account the number of spikes as the variance is greater at low spike counts. The correct test in these cases is to test the contrast ratio of one spike train against the confidence interval of other *at the same spike count*. Figure 8B shows examples of inappropriate and appropriate significance testing between two contrast ratios. In these examples, circles A and B indicate the two contrast ratios to be compared (A: contrast ratio 0.20, 200 spikes; B: contrast ratio 0.48, 3,200 spikes). Gray curves suggest possible confidence bands of A and B. In the example at left, the confidence interval at A, denoted by A′, is used to test the significance of B, and B′ is used to test the significance of A. Because A is not within B′, and B is not within A′, this test suggests that the contrast ratio measurements are significantly different. The problem with this example, however, is that the test has not been corrected for sample size. In the example at *right*, A″ and B″ denote confidence intervals calculated from spike trains A and B at spike counts of B and A, respectively. Thus this test takes into account the fact that the confidence band of B is wider at fewer spikes. As a result, it correctly finds that A is within B″, and the difference between A and B is deemed to be nonsignificant at the $P = 0.05$ level (2-tailed test, 2.5–97.5% confidence band).

In the example at *right*, it would be inappropriate to compare A″ against B. This is because PR randomization cannot extrapolate from a 200-spike spike train what the contrast ratio might be at 3,200 spikes. In virtually all contrast ratio comparison tests, one spike train will be longer than the other, and the test must be performed by comparing the confidence band of the *longer* spike train against the contrast ratio of the *shorter* spike train, as shown in Fig. 8B, *right*. In rare cases where both spike trains are of the same length, differences in variance between the two contrast ratios can result in one contrast ratio lying *within* the confidence band of the other, but the other contrast ratio lying *outside* the confidence band of the former. An example of this can be seen in the 50% confidence levels of Fig. 8A at 200 spikes. We believe this arises from asymmetry in the probability distribution (because the contrast ratio is not normally distributed). In these cases, we suggest that if either value falls within the confidence band of the other, then the two values are not significantly different from each other.

This method can also be used to determine the minimum number of spikes needed to avoid falsely rejecting a true activated response. In Fig. 8A, substantial overlap can be seen between the luminance-modulated and spontaneous distributions for spike counts <400. This overlap suggests that, for low spike counts, errors can arise from the use of the 95% spontaneous confidence limit. For spike trains with <400 spikes, the lower precision of contrast ratio measurements increases the likelihood of a true luminance response being misclassified as a nonresponse (incorrect acceptance of the null hypothesis, i.e., a type II error) (cf. Kvanli et al. 1989). For this cell, a minimum of 400 spikes is needed to avoid such errors. A cell with a weaker luminance response would require even more spikes for the response to be clearly distinguishable from spontaneous. For example, a cell with a luminance contrast ratio of 0.35 might need a minimum of 800 spikes. In our experience, the spontaneous confidence levels do not differ much from cell to cell. Thus the minimum spike number is largely determined by the strength of the luminance response.

Sample size correction function for measured contrast ratios

Contrast ratios are a popular tool for quantifying both sinusoidal and transient responses. As we have demonstrated for recordings of spontaneous activity (Fig. 4), differences in sample size can dramatically affect the measured contrast ratio. This bias is present not only in spontaneous activity but in luminance responses as well. When recorded responses are weakly modulated, a correction function (correcting the measured contrast ratio for the number of spikes recorded) may be required to dissociate the test response population from the spontaneous population (i.e., using population statistics such as the Wilcoxon signed-rank test). Here, we show how to generate such a correction function.

We have considered two forms of sample-size correction functions. The first form is based on a conversion from raw contrast ratio to a significance scale (Fig. 5). The second form corrects the contrast ratio to the most likely strength of the underlying modulation (i.e., the most likely contrast ratio one would measure if all recordings had many spikes). We tested both forms with the Wilcoxon signed-rank test and found them to be comparable in power when testing for the significance of a population response (both tended to improve P values by roughly an order of magnitude). However, we found the second form to be more useful overall. The first form (significance) fails to accurately reflect the relative strength between stronger responses. This distortion arises at higher contrast ratios, where significance values are related primarily to the number of spikes recorded and only indirectly related to the underlying

strength of modulation (picture the slope of the higher confidence levels in Fig. 6A). In contrast, we believe the second form fairly represents the magnitudes of both weak and strong responses and their relationship to sample size.

The difference between these two methods is evident if one examines the following hypothetical scenario. Suppose that under two conditions, A and B, all recorded measurements yield a contrast ratio of ~0.5. Further suppose that the only difference between results from the two conditions is that recordings from A have 1,000 spikes, but recordings from B have 5,000 spikes (perhaps due to a difference in firing rate). The significance scale method would find that condition B has significantly more modulation than condition A, whereas the correction function method would find, correctly, that both conditions have the same magnitude of modulation response. For this reason, we have used the correction function method in our analyses (see Figs. 6 and 7 in Hung et al. 2001) and will show how to derive it here.

We begin by finding the 50% activated confidence levels of a selection of modulated responses. We choose the 50% level because the PR randomized response distributions are unimodal, and the 50% level is both the median and peak of the distribution. The 50% activated confidence level, therefore, indicates the most likely contrast ratio for a given strength and spike number. Figure 9, left, shows 50% activated confidence levels of 12 cells (open triangles, thin lines are interpolated). These values were derived from 8,400 (12 cells × 7 spike counts × 100 randomizations each) PR randomized spike trains based on twelve seed spike trains of >5,000 spikes each. These randomly generated spike trains did not overlap in any way (were not subsets of each other).

Because each 50% activated confidence level was generated from a single cell, the strength of the response for each curve should be constant, independent of the number of spikes (i.e., independent of the length of data collection). Thus each of these curves should have a single response measure after correction for the number of spikes. To create a response measure that can accurately describe the response strength of all cells, a correction function (see APPENDIX) was fitted to the 50% confidence levels using least squares method. The term modulation index (MI) was used to refer to the contrast ratio corrected for spike count. We then generated a family of iso-MI curves based on this function (Fig. 9, right). Thus each line at right represents an iso-MI level. For practical purposes, it was adequate to adjust iso-MI levels to approximate the contrast ratios at 5,000 spikes. For example, a 600-spike train with a contrast ratio of 0.09 (indicated by arrow) would be corrected to have a modulation index of 0.06 (i.e., fall along 0.06 iso-MI curve). Responses from all cells in our data set are then corrected in this fashion (cf. Hung et al. 2001, Figs. 7 and 8).

Figure 10 demonstrates the usefulness of this correction function. Two histograms are shown comparing response magnitudes recorded during Cornsweet stimulation versus spontaneous activity. The first histogram (Fig. 10A) shows raw contrast ratios, and the second (Fig. 10B) shows contrast ratios after using the correction function of Fig. 9 (MIs). Greater dissociation between the two conditions can be seen in the second histogram. In practice, the function lowers the contrast ratios of spontaneous activity but has little effect on activated responses.

This method of sample size correction is not limited to sinusoidal contrast ratio measurements. We suggest that for any data obtained from periodic stimuli, this method of PR-randomization, followed by fitting a function to the 50% con-
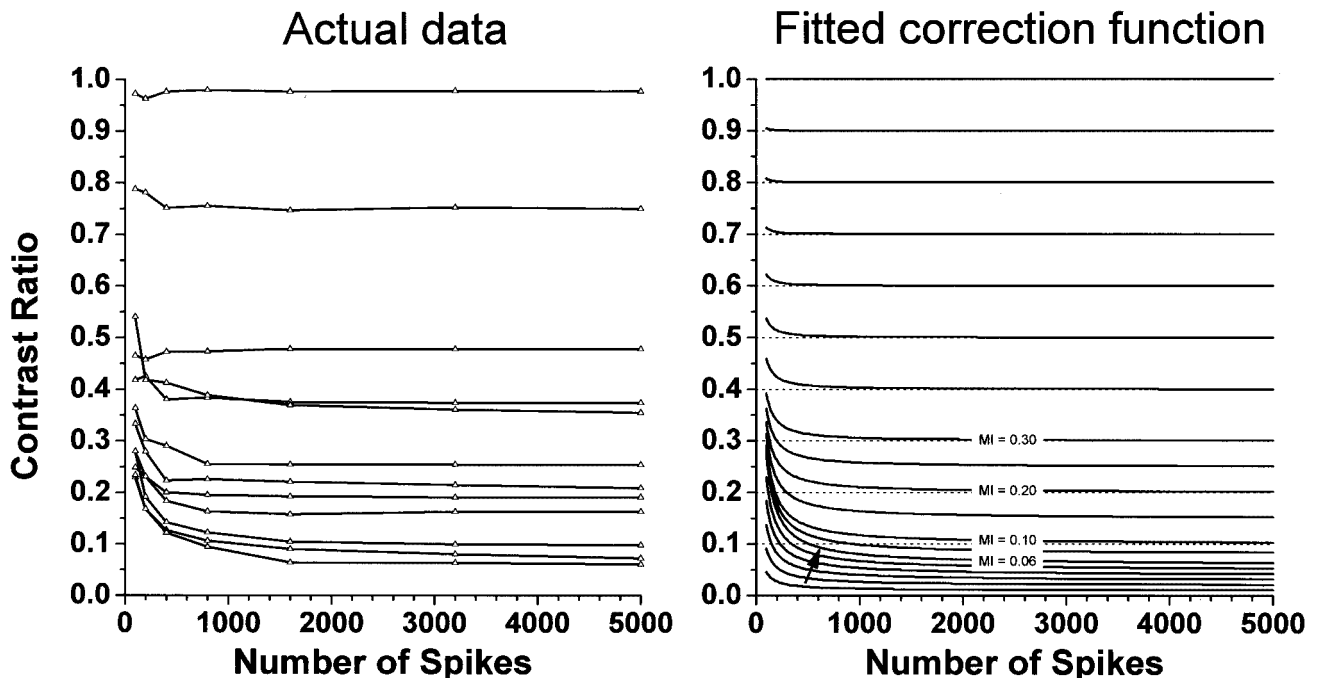

FIG. 9.   Construction of spike-correction function. *Left*: each line indicates the 50% confidence level (the 50th percentile of 1,000 PR-randomized spike trains) generated from an individual cell's luminance response. Note that higher contrast ratios are less affected by low spike count. *Right*: spike count correction function. Function was fitted to the 50% confidence levels shown by lines at *left*. Curved lines in the correction function indicate iso-MI levels (levels resulting in the same modulation index).
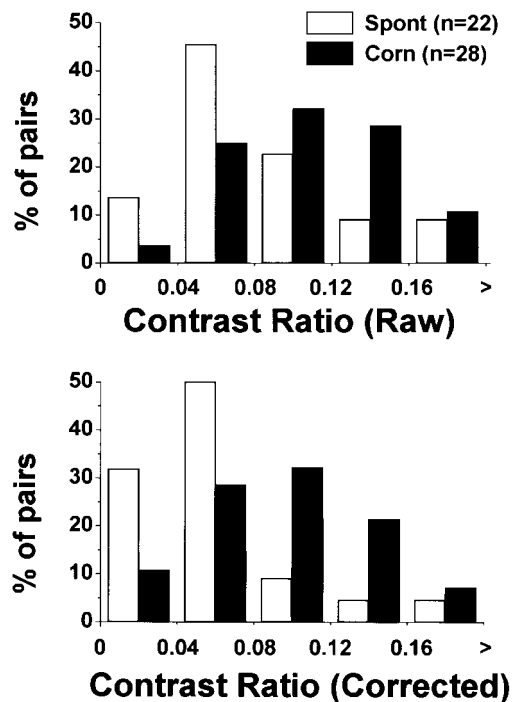
FIG. 10. Histograms of Cornsweet and spontaneous contrast ratios before and after sample-size correction. *A*: histograms of raw contrast ratios recorded under Cornsweet (■, *n* = 28) and spontaneous (□, *n* = 22) conditions. Not all cells were tested under both conditions. Cells were recorded from Areas 17 and 18 and were chosen because their primary sinusoidal response was at the Cornsweet stimulus frequency (vs. 2 or 3 times the stimulus frequency). *B*: same histogram after correction for number of spikes. The distributions are more clearly separated, and the level of significance is improved from $P = 0.023$ to $P = 0.0015$ (Wilcoxon signed-rank test, $n = 20$).

fidence levels of response strengths from several cells, can be used to generate a compensated MI for each recorded response.

DISCUSSION

In this paper, we have addressed the quantification and determination of response significance of weakly modulated spike trains. We have shown how to apply two randomization methods in the analysis of weak responses. The first method, full randomization, is commonly used in testing for response significance. We have shown a quantile-quantile plot indicating that this method correctly predicts the range of responses of spike trains recorded during spontaneous activity from Areas 17 and 18 of anesthetized cats. We have also shown that this test for response significance is best determined by randomizing the test response spike train itself, rather than randomizing the spike train of the recorded spontaneous activity. The second method, PR randomization, is novel and allows for the generation of modulated randomized responses. We have shown how this second method can be used to visualize the precision of measurement in modulated responses, allowing one to compare whether two modulated responses are significantly different in strength from each other. It is also useful for determining the minimum number of spikes needed to distinguish true responses from spontaneous activity. Finally, we have shown how to use PR randomization to create a sample-size correction function for the measurement of weak responses.

Key to our determination of precision and spike-count correction is the PR-randomization method. Unlike full randomization, this method is able to generate modulated spike trains. This method has an advantage over other randomization methods for generating modulated spike trains —namely, its reliance on relatively few assumptions about the actual mechanisms underlying spike generation. Its principal assumption, that spikes occur as renewal processes, is supported by our data and by recent analyses of the occurrence of repeating spike patterns (Baker and Lemon 2000). However, this view is not universally supported, and there are situations in which this assumption would not be appropriate (Abeles 1991; Abeles et al. 1993; Engel et al. 1992; Eskandar et al. 1992a,b; Gerstein et al. 1989; Grammont and Riehle 1999; Riehle et al. 1997; Singer and Gray 1995; Von der Malsburg 1995). Another assumption, that spike probability varies with the phase of the stimulus cycle, is reasonable and supported by PSTISI plots generated from many cells.

Both full and PR randomization assume stationarity of the spike train over minutes of recording. If these methods were applied to nonstationary recordings, response magnitudes measured over longer epochs would not reflect the variability at shorter epochs. This can lead to either over- or underestimation of confidence intervals within epochs if one does not correct for sample size in these epochs. We have looked for effects of nonstationarity in our measurements by testing the contrast ratio and spike counts of successive epochs in our longer records. We found that, other than gradual increases or decreases in firing rate, our records were largely stable. For example, the luminance response in Figs. 7 and 8*A* varied in firing rate with a SD 23% of the mean. Contrast ratio did not progressively increase or decrease over time (although variance was greater in epochs with fewer spikes, as expected). Even in the spontaneous recordings, epochs with fewer spikes did not necessarily predict higher contrast ratios. Thus we believe nonstationarity (on a minutes time scale) does not factor greatly into our results.

In this paper, PR randomization has been used to examine the measurement of contrast ratios of PSTHs. It may also be extended to quantify the precision of joint peristimulus time histograms (JPSTHs), cross-correlograms, spike patterns, etc. The primary disadvantage of this method is its dependence on actual recorded data. One must be sure that the original spike train is free from contamination (i.e., single unit) and has sufficient number and density of spikes (for example, 5,000 spikes for cycles 2 s in duration, yielding a density of 2.5 spikes/ms). Although it is possible to generate more artificial spikes than in the original spike train, doing so is an extrapolation beyond the actual data. Also, one must choose a randomization window size large enough to yield sufficient randomization, yet small enough to accurately reflect the PSTH. For a spike density of 2.5 spikes/ms, a randomization window size of 10 spikes corresponds to 4 ms, a reasonable temporal resolution for the measurement of contrast ratios.

The analysis methods we have shown are applicable for analyzing both weak and robust responses. These methods can be generalized across systems. Although this paper focuses on responses to sinusoidally modulated stimuli, the same methods may also be applied to correct for sample size in responses to other periodic stimuli.

APPENDIX

Sample size correction function

The form of the function was determined by trial and error, and the exact fit was determined by relaxation methods by reducing the total squared error between the fitted function and all the points at Fig. 9, *left*. To fit the function, a putative modulation index was assigned for each line at left, based on each line's contrast ratio at 5,000 spikes. During fitting, we made slight adjustments to the twelve modulation indices as necessary. The values at lower spike counts were weighted less in the total squared error (the weight at 100 spikes was one-quarter the weight at 5,000 spikes). A custom-made computer program individually adjusted each variable in the function (slight adjustments, such as ± 0.00001) and tested its effect on the total squared error. Adjustment sizes were controlled by the user as needed. For practical purposes, only the inverse function was determined

$$\text{Raw} = \left\{ (1 - \text{MI})^{-\text{spikes}} + \text{MI} + \frac{A \times (B - \text{MI})}{\text{spikes}^{MIC+1}} \right\}$$
$$\times \frac{1 + \tanh[(\text{MI} - E) \times D]}{2} + \frac{\text{MI}}{E} \times \left\{ (1 - E)^{-\text{spikes}} + E + \frac{A \times (B - E)}{\text{spikes}^{EC+1}} \right\}$$
$$\times \frac{1 + \tanh[(E - \text{MI}) \times D]}{2}$$

where: Raw = raw contrast ratio; MI = modulation index; spikes = number of spikes; $A = 0.52305$; $B = 46.1622$; $C = 1.30107$; $D = 6198$; $E = 0.06$.

Note the tanh functions, which effectively separate the function into two regions at MI = 0.06. Above 0.06, the function was determined by trial and error. Below 0.06, the function was graded down to zero. We based this decision on the shape of the lower confidence levels (see Fig. 6).

REFERENCES

ABELES M. Quantification, smoothing, and confidence limits for single units' histograms. *J Neurosci Methods* 5: 317–325, 1982.
ABELES M. *Corticonics.* Cambridge, UK: Cambridge Univ. Press, 1991.
ABELES M, BERGMAN H, MARGALIT E, AND VAADIA E. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J Neurophysiol* 70: 1629–1638, 1993.
AWISZUS F. Quantification and statistical verification of neuronal stimulus responses from noisy spike train data. *Biol Cybern* 68: 267–274, 1993a.
AWISZUS F. Sensitivity of different stimulus-timing strategies for the detection of small excitations in noisy spike train data. *Biol Cybern* 68: 553–558, 1993b.
AWISZUS F. Spike train analysis. *J Neurosci Methods* 74: 155–166, 1997.
BAKER SN AND LEMON RN. Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *J Neurophysiol* 84: 1770–1780, 2000.
BRUMBERG JC, PINTO DJ, AND SIMONS DJ. Spatial gradients and inhibitory summation in the rat whisker barrel system. *J Neurophysiol* 76: 130–140, 1996.
BURR DC. Implications of the Craik-O'Brien illusion for brightness perception. *Vision Res* 27: 1903–1913, 1987.
CORNSWEET TN. *Visual Perception.* New York: Academic, 1970.
COX DR AND LEWIS PAW. *The Statistical Analysis of Series of Events.* New York: Wiley, 1966.
DAVEY NJ, ELLAWAY PH, AND STEIN RB. Statistical limits for detecting change in the cumulative sum derivative of the peristimulus time histogram. *J Neurosci Methods* 17: 153–166, 1986.
DÖRRSCHEIDT GH. The statistical significance of the peristimulus time histogram (PSTH). *Brain Res* 220: 397–401, 1981.
ENGEL AK, KONIG P, KREITER AK, SCHILLEN TB, AND SINGER W. Temporal coding in the visual cortex: new vistas on integration in the nervous system. *Trends Neurosci* 15: 218–226, 1992.
ESKANDAR EN, OPTICAN LM, AND RICHMOND BJ. Role of inferior temporal neurons in visual memory. II. Multiplying temporal waveforms related to vision and memory. *J Neurophysiol* 68: 1296–1306, 1992a.
ESKANDAR EN, RICHMOND BJ, AND OPTICAN LM. Role of inferior temporal neurons in visual memory. I. Temporal encoding of information about visual images, recalled images, and behavioral context. *J Neurophysiol* 68: 1277–1295, 1992b.
GERSTEIN GL, BEDENBAUGH P, AND AERTSEN AMHJ. Neuronal assemblies. *IEEE Trans Biomed Eng* 36: 4–14, 1989.
GERSTEIN GL AND KIANG NY-S. An approach to the quantitative analysis of electrophysiological data from single neurons. *Biophys J* 1: 15–28, 1960.
GRAMMONT F AND RIEHLE A. Precise spike synchronization in monkey motor cortex involved in preparation for movement. *Exp Brain Res* 128: 118–122, 1999.
HAMILTON LC. *Regression with Graphics: A Second Course in Applied Statistics.* Belmont, CA: Duxbury, 1992, p. 13–15.
HUNG CP, RAMSDEN BM, CHEN LM, AND ROE AW. Building surfaces from borders in Areas 17 and 18 of the cat. *Vision Res* 41: 1389–1407, 2001.
KVANLI AH, GUYNES CS, AND PAVUR RJ. *Introduction to Business Statistics* (2nd ed). St. Paul, MN: West Publishing, 1989.
MANLY BFJ. *Randomization. Bootstrap and Monte Carlo Simulation Procedures in Biology.* London, UK: Chapman and Hall, 1997.
MOORE GP, PERKEL DH, AND SEGUNDO JP. Statistical analysis and functional interpretation of neuronal spike data. *Annu Rev Physiol* 28: 493–522, 1966.
MOVSHON JA, THOMPSON ID, AND TOLHURST DJ. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J Physiol (Lond)* 283: 55–77, 1978a.
MOVSHON JA, THOMPSON ID, AND TOLHURST DJ. Receptive field organization of complex cells in the cat's striate cortex. *J Physiol (Lond)* 283: 79–99, 1978b.
ORAM MW, WIENER MC, LESTIENNE R, AND RICHMOND BJ. Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *J Neurophysiol* 81: 3021–3033, 1999.
PERKEL DH, GERSTEIN GL, AND MOORE GP. Neuronal spike trains and stochastic point processes. I. The single spike train. *Biophys J* 7: 391–418, 1967.
RIEHLE A, GRUN S, DIESMANN M, AND AERTSEN A. Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* 278: 1950–1953, 1997.
ROSSI AF, RITTENHOUSE CD, AND PARADISO MA. The representation of brightness in primary visual cortex. *Science* 273: 1104–1107, 1996.
SCHULZ DE, SOSNIK R, EGO V, HAIDARLIU S, AND AHISSAR E. A neuronal analogue of state-dependent learning. *Nature* 403: 549–552, 2000.
SINGER W AND GRAY CM. Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci* 18: 555–586, 1995.
SKOTTUN BC, DE VALOIS RL, GROSOF DH, MOVSHON JA, ALBRECHT DG, AND BONDS AB. Classifying simple and complex cells on the basis of response modulation. *Vision Res* 31: 1079–1086, 1991.
VINJE WE AND GALLANT JL. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287: 1273–1276, 2000.
VON DER MALSBURG C. Binding in models of perception and brain function. *Curr Opin Neurobiol* 5: 520–526, 1995.