Measuring the signal-to-noise ratio of a neuron


As Published
http://dx.doi.org/10.1073/pnas.1505545112

Publisher
National Academy of Sciences (U.S.)

Version
Final published version

Accessed
Sat Oct 13 05:19:53 EDT 2018

Citable Link
http://hdl.handle.net/1721.1/100795

Terms of Use
Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Detailed Terms

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Measuring the signal-to-noise ratio of a neuron

Gabriela Czanner,1, Sridive V. Sarma, Demba Ba, Uri T. Eden, Wei Wu, Emad Eskandar, Hubert H. Lim, Simona Temereanca, Wendy A. Suzuki, and Emery N. Brown

1Department of Biostatistics and Department of Eye and Vision Science, Faculty of Health and Life Sciences, University of Liverpool, Liverpool L69 3GA, United Kingdom; 2Department of Biomedical Engineering, Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD 21218; 3Department of Brain and Cognitive Sciences and Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139; 4Department of Mathematics and Statistics, Boston University, Boston, MA 02215; 5School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China; Departments of Neurosurgery and Anesthesiology, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; 6Department of Biomedical Engineering, Institute for Translational Neuroscience, University of Minnesota, Minneapolis, MN 55455; 7Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129; and 8Center for Neural Science, New York University, New York, NY 10003

Contributed by Emery N. Brown, March 27, 2015 (sent for review December 5, 2014; reviewed by Mingzhou Ding, Satish Iyengar, and Hualou Liang)

The signal-to-noise ratio (SNR), a commonly used measure of fidelity in physical systems, is defined as the ratio of the squared amplitude or variance of a signal relative to the variance of the noise. This definition is not appropriate for neural systems in which spiking activity is more accurately represented as point processes. We show that the SNR estimates a ratio of expected prediction errors and extend the standard definition to one appropriate for single neurons by representing neural spiking activity using point process generalized linear models (PP-GLM). We estimate the prediction errors using the residual deviations from the PP-GLM fits. Because the deviance is an approximate χ2 random variable, we compute a bias-corrected SNR estimate appropriate for single-neuron analysis and use the bootstrap to assess its uncertainty. In the analyses of four systems neuroscience experiments, we show that the SNRs are −10 dB to −3 dB for guinea pig auditory cortex neurons, −18 dB to −7 dB for rat thalamic neurons, −28 dB to −14 dB for monkey hippocampal neurons, and −29 dB to −20 dB for human subthalamic neurons. The new SNR definition makes explicit in the measure commonly used for physical systems the often-quoted observation that single neurons have low SNRs. The neuron’s spiking history is frequently a more informative covariate for predicting spiking propensity than the applied stimulus. Our new SNR definition extends to any GLM system in which the factors modulating the response can be expressed as separate components of a likelihood function.

Significance

Neurons represent both signal and noise in binary electrical discharges termed action potentials. Hence, the standard signal-to-noise ratio (SNR) definition of signal amplitude squared and divided by the noise variance does not apply. We show that the SNR estimates a ratio of expected prediction errors. Using point process generalized linear models, we extend the standard definition to one appropriate for single neurons. In analyses of four neural systems, we show that single neuron SNRs range from −29 dB to −3 dB and that spiking history is often a more informative predictor of spiking propensity than the signal or stimulus activating the neuron. By generalizing the standard SNR metric, we make explicit the well-known fact that individual neurons are highly noisy information transmitters.


Reviewers: M.D., University of Florida; S.I., University of Pittsburgh; and H.L., Drexel University.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

1To whom correspondence may be addressed. Email: czanner@liverpool.ac.uk or enb@neurostat.mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1505545112/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1505545112

PNAS | June 9, 2015 | vol. 112 | no. 23 | 7141–7146
(23). The SNR can be expressed in terms of the $R^2$ for linear and Poisson regression models. However, this relationship has not been exploited to construct an SNR estimate for neural systems or point process models. Finally, the SNR is a commonly computed statistic in science and engineering. Extending this concept to non-Gaussian systems would be greatly aided by a precise statement of the theoretical quantity that this statistic estimates (24, 25).

We show that the SNR estimates a ratio of expected prediction errors (EPEs). Using point process generalized linear models (PP-GLM), we extend the standard definition to one appropriate for single neurons recorded in stimulus–response experiments. In analyses of four neural systems, we show that single-neuron SNRs range from ~29 dB to ~3 dB and that spiking history is often a more informative predictor of spiking propensity than the signal being represented. Our new SNR definition generalizes to any problem in which the modulatory components of a system’s output can be expressed as separate components of a GLM.

Theory
A standard way to define the SNR is as the ratio

$$\text{SNR} = \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{noise}}^2}$$  \hspace{1cm} [1]

where $\sigma_{\text{signal}}^2$ is structure in the data induced by the signal and $\sigma_{\text{noise}}^2$ is the variability due to the noise. To adapt this definition to the analysis of neural spike train recordings from a single neuron, we have: (i) define precisely what the SNR estimates; (ii) extend the definition and its estimate to account for covariates that, along with the applied stimulus or signal input, also affect the neural response; and (iii) extend the SNR definition and its estimate so that it applies to point process models of neural spiking activity.

By analyzing the linear Gaussian signal plus noise model (Supporting Information), we show that standard SNR computations (Eq. S5) provide an estimator of a ratio of EPEs (Eq. S4). For the linear Gaussian model with covariates, this ratio of EPEs is also well defined (Eq. S6) and can be estimated as a ratio of sum of squares of residuals (Eq. S7). The SNR definition further extends to the GLM with covariates (Eq. S8). To estimate the SNR for the GLM, we replace the sums of squares by the residual deviances, their extensions in the GLM framework Eqs. S9 and S10. The residual deviance is a constant multiple of the Kullback–Leibler (KL) divergence between the data and the model. Due to the Pythagorean property of the KL divergence of GLM models with canonical link functions (26–28) evaluated at the maximum likelihood estimates, the SNR estimator can be conveniently interpreted as the ratio of the explained KL divergence of the signal relative to the noise. We propose an approximate bias correction for the GLM SNR estimate with covariates (Eq. S11), which gives the estimator better performance in low signal-to-noise problems such as single-neuron recordings. The GLM framework formulated with point process models has been used to analyze neural spiking activity (5–7, 29). Therefore, we derive a point process GLM (PP-GLM) SNR estimate for single-neuron spiking activity recorded in stimulus–response experiments.

A Volterra Series Expansion of the Conditional Intensity Function of a Spiking Neuron. Volterra series are widely used to model biophysical systems (30), including neural spiking activity (16). We develop a Volterra series expansion of the log of the conditional intensity function to define the PP-GLM for single-neuron spiking activity (31). We then apply the GLM framework outlined in Supporting Information to derive the SNR estimate.

We assume that on an observation interval $(0, T]$, we record spikes at times $0 < t_1 < t_2 < \ldots < t_k < T$. If we model the spike events as a point process, then the conditional intensity function of the spike train is defined by

$$\lambda(t|H_t) = \beta_0 + \sum_{k=0}^{K} \beta_k \mathbf{X}_k(t) + \sum_{j=1}^{J} \beta H_j \nu_{t-j} \quad \text{[2]}$$

where $N(t)$ is the number of spikes in the interval $(0, t]$ for $t \in (0, T]$ and $H_t$ is the relevant history at $t$. It follows that for $\Delta$ small,

$$\text{Pr}(\text{spike in } (t, t+\Delta)|H_{t}) \approx \lambda(t|H_{t}) \Delta. \quad \text{[3]}$$

We assume that the neuron receives a stimulus or signal input and that its spiking activity depends on this input and its biophysical properties. The biophysical properties may include absolute and relative refractory periods, bursting propensity, and network dynamics. We assume that we can express $\log \lambda(t|H_{t})$ in a Volterra series expansion as a function of the signal and the biophysical properties (31). The first-order and second-order terms in the expansion are

$$\log \lambda(t|H_{t}) = \int_{0}^{t} s(t-u)\beta_3(u)du + \int_{0}^{t} \beta H(u) dN(t-u)$$

$$+ \int_{0}^{t} \int_{0}^{t} s(t-u)s(t-v)h_1(u,v)du dv$$

$$+ \int_{0}^{t} h_2(u,v)dN(t-u)dN(t-v)$$

$$+ \int_{0}^{t} h_3(u,v)s(t-u)dN(t-v) + \ldots. \quad \text{[4]}$$

where $s(t)$ is the signal at time $t$, $dN(t)$ is the increment in the counting process, $\beta_3(u)$ is the one-dimensional signal kernel, $\beta H(u)$ is the one-dimensional temporal or spike history kernel, $h_1(u,v)$ is the 2D signal kernel, $h_2(u,v)$ is the 2D temporal kernel, and $h_3(u,v)$ is the 2D signal–temporal kernel.

Eq. 4 shows that up to first order, the stimulus effect on the spiking activity and the effect of the biophysical properties of the neuron, defined in terms of the neuron’s spiking history, can be expressed as separate components of the conditional intensity function. Assuming that the second-order effects are not strong, then the approximate separation of these two components makes it possible to define the SNR for the signal, also taking account of the effect of the biophysical properties as an additional covariate and vice versa. We expand the log of the conditional intensity function in the Volterra series instead of the conditional intensity function itself in the Volterra series to ensure that the conditional intensity function is positive. In addition, using the log of the conditional intensity function simplifies the GLM formulation by using the canonical link function for the local Poisson model.

Likelihood Analysis Using a PP-GLM. We define the likelihood model for the spike train using the PP-GLM framework (5). We assume the stimulus–response experiment consists of $R$ independent trials, which we index as $r = 1, \ldots, R$. We discretize time within a trial by choosing $L$ large and defining the $L$ subintervals $\Delta = T/L$. We choose $L$ large so that each subinterval contains at most one spike. We index the subintervals $r = 1, \ldots, L$ and define $n_{r,t}$ to be 1 if, on trial $r$, there is a spike in the subinterval $((t-1)\Delta, t\Delta)$ and is 0 otherwise. We let $n_r = (n_{r,1}, \ldots, n_{r,t})$ be the set of spikes recorded on trial $r$ in $(0, T]$. Let $H_{t,r} = (n_{r,1}, \ldots, n_{r,t-1})$ be the relevant history of the spiking activity at time $\Delta t$. We define the discrete form of the Volterra expansion by using the first two terms of Eq. 4 to obtain

$$\log \lambda_r(t|H_{t,r}, \beta) \approx \beta_0 + \sum_{k=0}^{K} \beta_k \mathbf{X}_k(t) + \sum_{j=1}^{J} \beta H_j \nu_{t-j} \quad \text{[5]}$$
where \( \beta = (\beta_0, \beta_2, \beta_H) \), \( \beta_2 = (\beta_{S0}, ..., \beta_{SK}) \), and \( \beta_H = (\beta_{H1}, ..., \beta_{H2}) \), and hence the dependence on the stimulus goes back a period of \( K \Delta \), whereas the dependence on spiking history goes back a period of \( \Delta \). Exponentiating both sides of Eq. 5 yields

\[
\exp(\langle \Delta | H_\beta, \beta \rangle) \approx \exp\left\{ \beta_0 + \sum_{k=0}^{K} \sum_{i=0}^{k} \beta_{S{k}} s_{i} + \sum_{j=1}^{J} \beta_{H{j}} h_{i,j} \right\}. \tag{6}
\]

The first and third terms on the right side of Eq. 6 measure the intrinsic spiking propensity of the neuron, whereas the second term measures the effect of the stimulus or signal on the neuron’s spiking propensity.

The likelihood function for \( \beta \) given the recorded spike train is (5)

\[
L(n, \beta) = \exp\left\{ \sum_{r=1}^{\kappa} \left[ \sum_{i=1}^{L_r} \log \exp(\langle \Delta | H_\beta, H_{i,j} \rangle) \right] - \sum_{i=1}^{L} \lambda(\langle \Delta | \beta, H_{i,j} \rangle) \right\}. \tag{7}
\]

Likelihood formulations with between-trial dependence (32) are also possible but are not considered here.

The maximum likelihood estimate of \( \beta \) can be computed by maximizing Eq. 7 or, equivalently, by minimizing the residual deviance defined as

\[
\text{Dev}(n, \hat{\beta}) = -2 \left[ \log L(n, \hat{\beta}) - \log L(n, \beta) \right], \tag{8}
\]

where \( n = (n_1, ..., n_\kappa) \) and \( L(n, \beta) \) is the saturated model or the highest possible value of the maximized log likelihood (26). Maximizing \( \log L(n, \beta) \) to compute the maximum likelihood estimate of \( \beta \) is equivalent to minimizing the deviance, because \( L(n, \beta) \) is a constant. The deviance is the generalization to the GLM of the sum of squares from the linear Gaussian model (33).

As in the standard GLM framework, these computations are carried out efficiently using iteratively reweighted least squares. In our PP-GLM likelihood analyses, we use Akaike’s Information Criterion (AIC) to help choose the order of the discrete kernels \( \beta_2 \) and \( \beta_H \) (34). We use the time-rescaling theorem and Kolmogorov–Smirnov (KS) plots (35) along with analyses of the Gaussian transformed interspike intervals to assess model goodness of fit (36). We perform the AIC and time-rescaling goodness-of-fit analyses using cross-validation to fit the model to half of the trials in the experiments (training data set) and then evaluating AIC, the KS plots on the second half of the trials (test data set). The model selection and goodness-of-fit assessments are crucial parts of the SNR analyses. They allow us to evaluate whether our key assumption is valid, that is, that the conditional intensity function can be represented as a finite-order Volterra series whose second-order terms can be neglected. Significant lack of fit could suggest that this assumption did not hold and would thereby weaken, if not invalidate, any subsequent inferences and analyses.

**SNR Estimates for a Single Neuron.** Applying Eq. S1, we have that for a single neuron, the SNR estimate for the signal given the spike history (biophysical properties) with the approximate bias corrections is

\[
\text{SNR}_S = \frac{\text{Dev}(n, \hat{\beta}_0, \hat{\beta}_H) - \text{Dev}(n, \hat{\beta}) - \dim(\hat{\beta}_0) - \dim(\hat{\beta}_H) + \dim(\hat{\beta})}{\text{Dev}(n, \hat{\beta}) + \dim(\hat{\beta})}. \tag{9}
\]

and that for a single neuron, the SNR estimates of the spiking propensity given the signal is

\[
\text{SNR}_H = \frac{\text{Dev}(n, \hat{\beta}_S) - \text{Dev}(n, \hat{\beta}) - \dim(\hat{\beta}_S) + \dim(\hat{\beta})}{\text{Dev}(n, \hat{\beta}) + \dim(\hat{\beta})}. \tag{10}
\]

where \( \dim(\hat{\beta}) \) is the dimension or the number of parameters in \( \hat{\beta} \). Application of the stimulus activates the biophysical properties of the neuron. Therefore, to measure the effect of the stimulus, we fit the GLM with and without the stimulus and use the difference between the deviances to estimate the SNR (Eq. 9). Similarly, to measure the effect of the spiking history, we fit the GLM with and without the spike history and use the difference between the deviances to estimate the SNR (Eq. 10).

Expressed in decibels, the SNR estimates become

\[
\text{SNR}_S^{dB} = 10 \log_{10}(\text{SNR}_S) \tag{11}
\]

\[
\text{SNR}_H^{dB} = 10 \log_{10}(\text{SNR}_H). \tag{12}
\]

**Applications**

**Stimulus–Response Neurophysiological Experiments.** To illustrate our method, we analyzed neural spiking activity data from stimulus–response experiments in four neural systems. The stimulus applied in each experiment is a standard one for the neural system being studied. The animal protocols executed in experiments 1–3 were approved by the Institutional Animal Care and Use Committees at the University of Michigan for the guinea pig studies, the University of Pittsburgh for the rat studies, and New York University for the monkey studies. The human studies in experiment 4 were approved by the Human Research Committee at Massachusetts General Hospital.

![Raster plots of neural spiking activity](image)

Fig. 1. Raster plots of neural spiking activity. (A) Forty trials of spiking activity recorded from a neuron in the primary auditory cortex of an anesthetized guinea pig in response to a 200 μs biphasic electrical pulse applied in the inferior colliculus at time 0. (B) Fifty trials of spiking activity from a rat thalamic neuron recorded in response to a 50 mm/s whisker deflection repeated eight times per second. (C) Twenty-five trials of spiking activity from a monkey hippocampal neuron recorded while executing a location scene association task. (D) Forty trials of spiking activity recorded from a subthalamic nucleus neuron in a Parkinson’s disease patient before and after a hand movement in each of four directions (dir.): up (dir. U), right (dir. R), down (dir. D), and left (dir. L).
In experiment 1 (Fig. 1A), neural spike trains were recorded from 12 neurons in the primary auditory cortex of anesthetized guinea pigs in response to a 200 μs-pulse biphase electrical pulse at 44.7-μA applied in the inferior colliculus (10). Note that the neural recordings were generally multi-unit responses recorded on 12 sites but we refer to them as neurons in this paper. The stimulus was applied at time 0, and spiking activity was recorded from 10 ms before the stimulus to 50 ms after the stimulus during 40 trials. In experiment 2, neural spiking activity was recorded in 12 neurons from the ventral posteromedial (VPM) nucleus of the thalamus (VPm thalamus) in rats in response to whisker stimulation (Fig. 1B) (11). The stimulus was deflection of the whisker at a velocity of 50 mm/s at a repetition rate of eight deflections per second. Each deflection was 1 mm in amplitude and began from the whiskers' neutral position as the trough of a single sine wave and ended smoothly at the same neutral position. Neural spiking activity was recorded for 3,000 ms across 51 trials.

In experiment 3 (Fig. 1C), neural spiking activity was recorded in 13 neurons in the hippocampus of a monkey executing a location scene association task (13). During the experiment, two to four novel scenes were presented along with two to four well-learned scenes in an interleaved random order. Each scene was presented for between 25 and 60 trials. In experiment 4, the data were recorded from 10 neurons in the subthalamic nucleus of human Parkinson's disease patients (Fig. 1D) executing a directed movement task (15). The four movement directions were up, down, left, and right. The neural spike trains were recorded in 10 trials per direction beginning 200 ms before the movement cue and continuing to 200 ms after the cue.

The PP-GLM was fit to the spike trains of each neuron using likelihood analyses as described above. Examples of the model goodness of fit for a neuron from each system are shown in Supporting Information. Examples of the model estimates of the stimulus and history effects for a neuron from each system are shown in Fig. 2.

SNR of Single Neurons. We found that the SNR\[^{db}\] estimates (Eq. 11) of the stimulus controlling for the effect of the biophysical model properties were (median [minimum, maximum]): −6 dB [−10 dB, −3 dB] for guinea pig auditory cortex neurons; −9 dB [−18 dB, −7 dB] for rat thalamic neurons; −20 dB [−28 dB, −14 dB] for the monkey hippocampus; and −23 dB [−29 dB, −20 dB] for human subthalamic neurons (Fig. 3, black bars). The higher SNRs (from Eq. 11) in experiments 1 and 2 (Fig. 3A and B) are consistent with the fact that the stimuli are explicit, i.e., an electrical current and mechanical displacement of the whisker, respectively, and that the recording sites are only two synapses away from the stimulus. It is also understandable that SNRs are smaller for the hippocampus and thalamic systems in which the stimuli are implicit, i.e., behavioral tasks (Fig. 3C and D).

We found that SNR\[^{ff}\] estimates (from Eq. 12) of the biophysical properties controlling for the stimulus effect were: 2 dB [−9 dB, 7 dB] for guinea pig auditory cortex; −13 dB [−22 dB, −8 dB] for rat thalamic neurons; −15 dB [−24 dB, −11 dB] for the monkey hippocampal neurons; and −12 dB [−16 dB, −5 dB] for human subthalamic neurons (Fig. 3, gray bars). They were greater than SNR\[^{db}\] for the guinea pig auditory cortex (Fig. 3A), the monkey hippocampus (Fig. 3C), and the human subthalamic experiments (Fig. 3D), suggesting that the intrinsic spiking propensities of neurons are often greater than the spiking propensity induced by applying a putatively relevant stimulus.

A Simulation Study of Single-Neuron SNR Estimation. To analyze the performance of our SNR estimation paradigm, we studied simulated spiking responses of monkey hippocampal neurons with specified stimulus and history dynamics. We assumed four known SNRs of −8.3 dB, −17.4 dB, −28.7 dB, and −∞ dB corresponding, respectively, to stimulus effects on spike rate ranges of 500, 60, 10, and 0 spikes per second (Fig. 4, row 1). For each of the stimulus SNRs, we assumed spike history dependence (Fig. 4, row 2) to be similar to that of the neuron in Fig. 1C. For each of four stimulus effects, we simulated 300 experiments, each consisting of 25 trials (Fig. 4, row 3). To each of the 300 simulated data sets at each SNR level, we applied our SNR estimation paradigm: model
and the variance-based estimates; and (row 4) histogram of the 300 squared dB, where the same spike history component was used in each simulation. For each SNR level, 300 25-trial simulations were performed. Shown are (row 1) the true signal; (row 2) the true spike history component; (row 3) a raster plot of a representative simulated experiment; (row 4) histogram of the 300 SNR\(^{\text{KL}}\), the SNR estimates due to the stimulus correcting for the spiking history; and (row 5) histogram of the 300 SNR\(^{d}\), the SNR estimates due to the intrinsic biophysics of the neuron correcting for the stimulus. The vertical lines in rows 4 and 5 are the true SNR.

A Simulation Study of SNR Estimation for Single Neurons with No History Effect. We repeated the simulation study with no spike history dependence for the true SNR values of \(-1.5\) dB, \(-16.9\) dB, \(-27.9\) dB, and \(-\infty\) dB, with 25 trials per experiment and 300 realizations per experiment (Fig. 5). Removing the history dependence makes the simulated data within and between trials independent realizations from an inhomogeneous Poisson process. The spike counts across trials within a 1-ms bin obey a binomial model with \(n = 25\) and the probability of a spike defined by the values of the true conditional intensity function times 1 ms. Hence, it is possible to compute analytically the SNR and the bias in the estimates. We used our paradigm to compute SNR\(^{d}\). For comparison, we also computed the variance-based SNR proposed by Lyamzin et al. (20) Both SNR\(^{d}\) and the variance-based estimates were computed from the parameters obtained from the same GLM fits (see Eq. S16). For each simulation in Fig. 5, the true SNR value based on our paradigm is shown (vertical lines).

The histograms of SNR\(^{d}\) (Fig. 5, row 3) are spread symmetrically about the true expected SNR. The variance-based SNR estimate overestimates the true SNR in Fig. 5A and underestimates the true SNR in Fig. 5B and C. These simulations illustrate that the variance-based SNR is a less refined measure of uncertainty, as it is based on only the first two moments of the spiking data, whereas our estimate is based on the likelihood that uses information from all of the moments. At best, the variance-based SNR estimate can provide a lower bound for the information content in the non-Gaussian systems (16). Variance-based SNR estimators can be improved by using information from higher-order moments (37), which is, effectively, what our likelihood-based SNR estimators do.

Discussion

Measuring the SNR of Single Neurons. Characterizing the reliability with which neurons represent and transmit information is an important question in computational neuroscience. Using the PP-GLM framework, we have developed a paradigm for estimating the SNR of single neurons recorded in stimulus response experiments. To formulate the GLM, we expanded the log of the conditional intensity function in a Volterra series (Eq. 4) to represent, simultaneously, background spiking activity, the stimulus or signal effect, and the intrinsic dynamics of the neuron. In the application of the methods to four neural systems, we found that the SNRs of neuronal responses (Eq. 11) to putative stimuli—signals—ranged from \(-29\) dB to \(-3\) dB (Fig. 1). In addition, we showed that the SNR of the intrinsic dynamics of the neuron (Eq. 12) was frequently higher than the SNR of the stimulus (Eq. 11). These results are consistent with the well-known observation that, in general, neurons respond weakly to putative stimuli (16, 20). Our approach derives a definition of the SNR appropriate for neural spiking activity modeled as a point process. Therefore, it offers important improvements over previous work in which the SNR estimates have been defined as upper bounds derived from Gaussian approximations or using Fano factors and coefficients of variation applied to spike counts. Our SNR estimates are straightforward to compute using the PP-GLM framework (5) and public domain software that is readily available (38). Therefore, they can be computed as part of standard PP-GLM analyses.

The simulation study (Fig. 5) showed that our SNR methods provide a more accurate SNR estimate than recently reported variance-based SNR estimate derived from a local Bernoulli model (20). In making the comparison between the two SNR estimates, we derived the exact prediction error ratios analytically, and we used the same GLM fit to the simulated data to construct the SNR estimates. As a consequence, the differences are only due to differences in the definitions of the SNR. The more accurate
performance of our SNR estimate is attributable to the fact that it is based on the likelihood, whereas the variance-based SNR estimate uses only the first two sample moments of the data. This improvement is no surprise, as it is well known that likelihood-based estimates offer the best information summary in a sample given an accurate or approximately statistical model (34). We showed that for each of the four neural systems, the PP-GLM accurately described the spike train data in terms of goodness-of-fit assessments.

A General Paradigm for SNR Estimation. Our SNR estimation paradigm generalizes the approach commonly used to analyze SNRs in linear Gaussian systems. We derived the generalization by showing that the commonly computed SNR statistic estimates a ratio of EPEs (Supporting Information); the expected prediction of the error of the signal representing the data corrected for the nonsignal covariates relative to the EPE of the system noise. With this insight, we used the work of ref. 26 to extend the SNR definition to systems that can be modeled using the GLM framework in which the signal and relevant covariates can be expressed as separate components of the likelihood function. The linear Gaussian model is a special case of a GLM. In the GLM paradigm, the sum of squares from the standard linear Gaussian model is replaced by the residual deviance (Eq. S10). The residual deviance may be viewed as an estimated KL divergence between data and model (26). To improve the accuracy of our SNR estimator, particularly given the low SNRs of single neurons, we devised an approximate bias correction, which adjusts separately the numerator and the denominator (Eqs. 9 and 10). The bias-corrected estimator performed well in the limited simulation study we reported (Figs. 4 and 5). In future work, we will replace the separate bias corrections for the numerator and denominator with a single bias correction for the ratio, and extend our paradigm to characterize the SNR of neuronal ensembles and those of other non-Gaussian systems.

In Supporting Information, we describe the relationship between our SNR estimate and several commonly used quantities in statistics, namely the \( R^2 \), coefficient of determination, the \( F \) statistic, the likelihood ratio (LR) test statistic and \( f^2 \), Cohen's effect size. Our SNR analysis offers an interpretation of the \( F \) statistic that is not, to our knowledge, commonly stated. The \( F \) statistic may be viewed as a scaled estimate of the SNR for the linear Gaussian model, where the scale factor is the ratio of the degrees of freedom (Eq. S21). The numerator of our GLM SNR estimate (Eq. S9) is a LR test statistics for assessing the strength of the association between data \( Y \) and covariates \( X \). The generalized SNR estimator can be seen as generalized effect size. This observation is especially important because it can be further developed for planning neuropsychological experiments, and thus may offer a way to enhance experimental reproducibility in systems neuroscience research (39).

In summary, our analysis provides a straightforward way of assessing the SNR of single neurons. By generalizing the standard SNR metric, we make explicit the well-known fact that individual neurons are noisy transmitters of information.

ACKNOWLEDGMENTS. This research was supported in part by the Clinical Eye Research Centre, St. Paul’s Eye Unit, Royal Liverpool and Broadgreen University Hospitals National Health Service Trust, United Kingdom (G.C.); the US National Institutes of Health (NIH) Biomedical Engineering Research Partnership Award R01-DK015644 (to E.N.B. and W.A.S.); Pioneer Award DP1 OD004546 (to E.N.B.); and Transformative Research Award Grant 104948 (to E.N.B.); and NIH grants that supported the guinea pig experiments P41 EB02030 and T32 DC00011 to H.H.L.