

Measuring the signal-to-noise ratio of a neuron

Gabriela Czanner^{a,1}, Sridevi V. Sarma^b, Demba Ba^c, Uri T. Eden^d, Wei Wu^e, Emad Eskandar^f, Hubert H. Lim^g, Simona Temereanca^h, Wendy A. Suzukiⁱ, and Emery N. Brown^{c,j,k,1}

^aDepartment of Biostatistics and Department of Eye and Vision Science, Faculty of Health and Life Sciences, University of Liverpool, Liverpool L69 3GA, United Kingdom; ^bDepartment of Biomedical Engineering, Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD 21218;

^cDepartment of Brain and Cognitive Sciences and ^jInstitute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139; ^dDepartment of Mathematics and Statistics, Boston University, Boston, MA 02215; ^eSchool of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China; Departments of ^fNeurosurgery and ^kAnesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; ^gDepartment of Biomedical Engineering, Institute for Translational Neuroscience, University of Minnesota, Minneapolis, MN 55455; ^hMartinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129; and ⁱCenter 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 deviances from the PP-GLM fits. Because the deviance is an approximate χ^2 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.

SNR | signal-to-noise ratio | neuron | simulation | point processes

The signal-to-noise ratio (SNR), defined as the amplitude squared of a signal or the signal variance divided by the variance of the system noise, is a widely applied measure for quantifying system fidelity and for comparing performance among different systems (1–4). Commonly expressed in decibels as $10\log_{10}(\text{SNR})$, the higher the SNR, the stronger the signal or information in the signal relative to the noise or distortion. Use of the SNR is most appropriate for systems defined as deterministic or stochastic signals plus Gaussian noise (2, 4). For the latter, the SNR can be computed in the time or frequency domain.

Use of the SNR to characterize the fidelity of neural systems is appealing because information transmission by neurons is a noisy stochastic process. However, the standard concept of SNR cannot be applied in neuronal analyses because neurons transmit both signal and noise primarily in their action potentials, which are binary electrical discharges also known as spikes (5–8). Defining what is the signal and what is the noise in neural spiking activity is a challenge because the putative signals or stimuli for neurons differ appreciably among brain regions and experiments. For example, neurons in the visual cortex and in the auditory cortex respond respectively to features of light (9) and sound stimuli (10) while neurons in the somatosensory thalamus respond to tactile stimuli (11). In contrast, neurons in the rodent hippocampus respond robustly to the animal's position in its environment (11, 12), whereas monkey hippocampal neurons respond to the process of task learning (13). As part of responding to a putative stimulus,

a neuron's spiking activity is also modulated by biophysical factors such as its absolute and relative refractory periods, its bursting propensity, and local network and rhythm dynamics (14, 15). Hence, the definition of SNR must account for the extent to which a neuron's spiking responses are due to the applied stimulus or signal and to these intrinsic biophysical properties.

Formulations of the SNR for neural systems have been studied. Rieke et al. (16) adapted information theory measures to define Gaussian upper bounds on the SNR for individual neurons. Coefficients of variation and Fano factors based on spike counts (17–19) have been used as measures of SNR. Similarly, Gaussian approximations have been used to derive upper bounds on neuronal SNR (16). These approaches do not consider the point process nature of neural spiking activity. Moreover, these measures and the Gaussian approximations are less accurate for neurons with low spike rates or when information is contained in precise spike times.

Lyamzin et al. (20) developed an SNR measure for neural systems using time-dependent Bernoulli processes to model the neural spiking activity. Their SNR estimates, based on variance formulae, do not consider the biophysical properties of the neuron and are more appropriate for Gaussian systems (16, 21, 22). The Poisson regression model used widely in statistics to relate count observations to covariates provides a framework for studying the SNR for non-Gaussian systems because it provides an analog of the square of the multiple correlation coefficient (R^2) used to measure goodness of fit in linear regression analyses

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.

STATISTICS

Author contributions: G.C. and E.N.B. designed research; G.C., S.V.S., D.B., U.T.E., W.W., E.E., H.H.L., S.T., and W.A.S. performed research; and G.C. and E.N.B. wrote the paper. 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.

(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

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

where σ_{signal}^2 is structure in the data induced by the signal and σ_{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: to (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 biological 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 $0 < u_1 < u_2 < \dots < u_J < T$. If we model the spike events as a point process, then the conditional intensity function of the spike train is defined by (5)

$$\lim_{\Delta \rightarrow 0} \frac{\Pr(N(t + \Delta) - N(t)|H_t)}{\Delta} = \lambda(t|H_t), \quad [2]$$

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 Δ small,

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

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

$$\begin{aligned} \log \lambda(t|H_t) = & \int_0^t s(t-u)\beta_S(u)du + \int_0^t \beta_H(u)dN(t-u) \\ & + \int_0^t \int_0^t s(t-u)s(t-v)h_1(u,v)dudv \\ & + \int_0^t \int_0^t h_2(u,v)dN(t-u)dN(t-v) \\ & + \int_0^t \int_0^t h_3(u,v)s(t-u)dN(t-v) + \dots, \end{aligned} \quad [4]$$

where $s(t)$ is the signal at time t , $dN(t)$ is the increment in the counting process, $\beta_S(u)$ is the one-dimensional signal kernel, $\beta_H(t)$ 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 $r = 1, \dots, R$. We discretize time within a trial by choosing L large and defining the L subintervals $\Delta = T^{-1}L$. We choose L large so that each subinterval contains at most one spike. We index the subintervals $\ell = 1, \dots, L$ and define $n_{r,\ell}$ to be 1 if, on trial r , there is a spike in the subinterval $((\ell-1)\Delta, \ell\Delta)$ and it is 0 otherwise. We let $n_r = (n_{r,1}, \dots, n_{r,L})$ be the set of spikes recorded on trial r in $(0, T]$. Let $H_{r,\ell} = \{n_{r,\ell-j}, \dots, n_{r,\ell-1}\}$ be the relevant history of the spiking activity at time $\ell\Delta$. We define the discrete form of the Volterra expansion by using the first two terms of Eq. 4 to obtain

$$\log \lambda_r(\ell\Delta | H_{r,\ell}, \beta) \approx \beta_0 + \sum_{k=0}^K \beta_{S,k} s_{\ell-k} + \sum_{j=1}^J \beta_{H,j} n_{r,\ell-j}, \quad [5]$$

where $\beta = (\beta_0, \beta_S, \beta_H)'$, $\beta_S = (\beta_{S,0}, \dots, \beta_{S,K})'$, and $\beta_H = (\beta_{H,1}, \dots, \beta_{H,J})'$, 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 $J\Delta$. Exponentiating both sides of Eq. 5 yields

$$\lambda_r(\ell\Delta | H_\ell, \beta) \approx \exp \left\{ \beta_0 + \sum_{k=0}^K \beta_{S,k} s_{\ell-k} + \sum_{j=1}^J \beta_{H,j} n_{r,\ell-j} \right\}. \quad [6]$$

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 β given the recorded spike train is (5)

$$L(n, \beta) = \exp \left\{ \sum_{r=1}^R \left[\sum_{\ell=1}^L n_{r,\ell} \log \lambda(\ell\Delta | \beta, H_{r,\ell}) \Delta - \sum_{\ell=1}^L \lambda(\ell\Delta | \beta, H_{r,\ell}) \Delta \right] \right\}. \quad [7]$$

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

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

$$Dev(n, \hat{\beta}) = -2[\log L(n, \hat{\beta}) - \log L(n, n)], \quad [8]$$

where $n = (n_1, \dots, n_R)$ and $L(n, n)$ 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 β is equivalent to minimizing the deviance, because $L(n, n)$ 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 β_H and β_S (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 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. S11, 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

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

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

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

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 \hat{SNR}_S (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 \hat{SNR}_H (Eq. 10).

Expressed in decibels, the SNR estimates become

$$\hat{SNR}_S^{dB} = 10 \log_{10}(\hat{SNR}_S) \quad [11]$$

$$\hat{SNR}_H^{dB} = 10 \log_{10}(\hat{SNR}_H). \quad [12]$$

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.

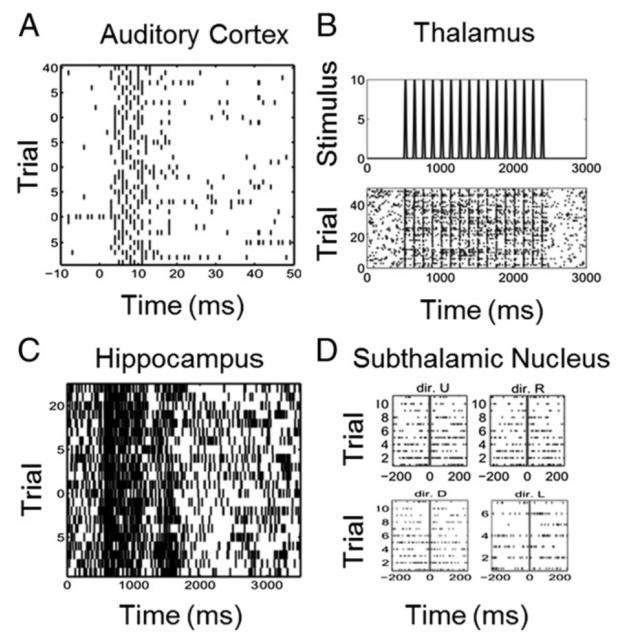


Fig. 1. Raster plots of neural spiking activity. (A) Forty trials of spiking activity recorded from a neuron in the primary auditory cortex of an anesthetized guinea pig in response to a 200 μ s/phase biphasic electrical pulse applied in the inferior colliculus at time 0. (B) Fifty trials of spiking 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 μ s/phase biphasic electrical pulse at 44.7- μ A 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 is 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 \hat{SNR}_S^{dB} estimates (Eq. 11) of the stimulus controlling for the effect of the biophysical model

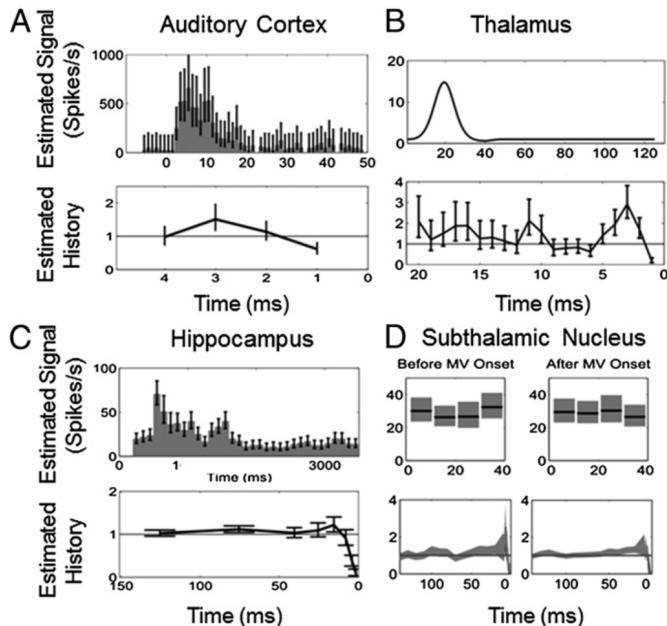


Fig. 2. Stimulus and history component estimates from the PP-GLM analyses of the spiking activity in Fig. 1. (A) Guinea pig primary auditory cortex neuron. (B) Rat thalamic neuron. (C) Monkey hippocampal neuron. (D) Human subthalamic nucleus neuron. The stimulus component (*Upper*) is the estimated stimulus-induced effect on the spike rate in A, C, and D and the impulse response function of the stimulus in B. The history components (*Lower*) show the modulation constant of the spike firing rate.

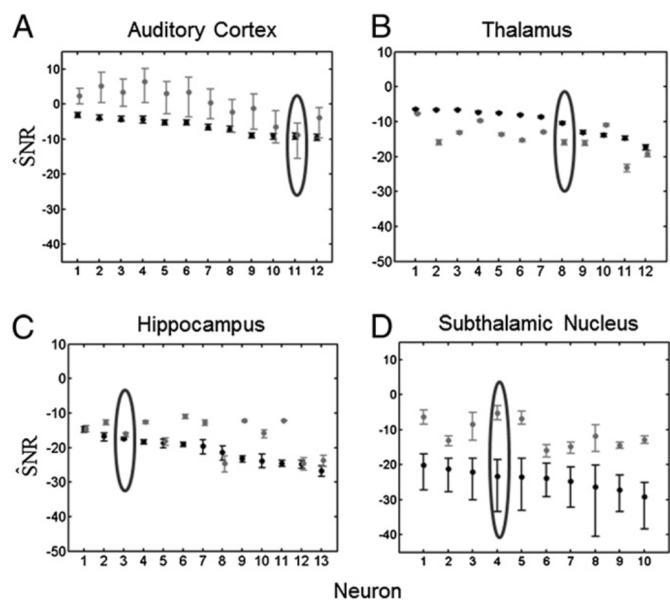


Fig. 3. KL-based SNR for (A) 12 guinea pig auditory cortex neurons, (B) 12 rat thalamus neurons, (C) 13 monkey hippocampal neurons, and (D) 10 subthalamic nucleus neurons from a Parkinson's disease patient. The black dots are \hat{SNR}_S^{dB} , the SNR estimates due to the stimulus correcting for the spiking history. The black bars are the 95% bootstrap confidence intervals for SNR_S^{dB} . The gray dots are \hat{SNR}_H^{dB} , the SNR estimates due to the intrinsic biophysics of the neuron correcting for the stimulus. The gray bars are the 95% bootstrap confidence intervals for SNR_H^{dB} . The encircled points are the SNR and 95% confidence intervals for the neural spike train raster plots in Fig. 1.

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 \hat{SNR}_H^{dB} 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 \hat{SNR}_S^{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 $-\infty$ dB corresponding, respectively, to stimulus effects on spike rates 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

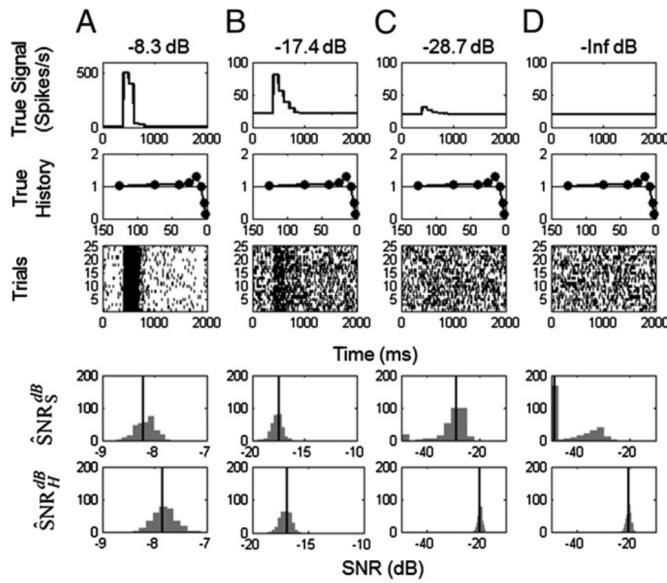


Fig. 4. KL-based SNR of simulated neurons with stimulus and history components. The stimulus components were set at four different SNRs: (A) -8.3 dB , (B) -17.4 dB , (C) -28.7 dB , and (D) $-\infty\text{ dB}$, where the same spike history component was used in each simulation. For each SNR level, 300 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 $\hat{\text{SNR}}_S^{\text{dB}}$, the SNR estimates due to the stimulus correcting for the spiking history; and (row 5) histogram of the 300 $\hat{\text{SNR}}_H^{\text{dB}}$, 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 SNRs.

fitting, model order selection, goodness-of-fit assessment, and estimation of $\hat{\text{SNR}}_S^{\text{dB}}$ (Fig. 4, row 4) and $\hat{\text{SNR}}_H^{\text{dB}}$ (Fig. 4, row 5).

The bias-corrected SNR estimates show symmetric spread around their true SNRs, suggesting that the approximate bias correction performed as postulated (Fig. 4, rows 4 and 5). The exception is the case in which the true SNR was $-\infty$ and our paradigm estimates $\hat{\text{SNR}}_S^{\text{dB}}$ as large negative numbers (Fig. 4D, row 4). The $\hat{\text{SNR}}_S^{\text{dB}}$ are of similar magnitude as the SNR estimates in actual neurons (see $\text{SNR} = -18.1\text{ dB}$ in the third neuron in Fig. 3C versus -17.4 dB in the simulated neuron (Fig. 4B).

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 $-\infty\text{ dB}$, 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 $\hat{\text{SNR}}_S^{\text{dB}}$. For comparison, we also computed the variance-based SNR proposed by Lyamzin et al. (20) Both $\hat{\text{SNR}}_S^{\text{dB}}$ 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 $\hat{\text{SNR}}_S^{\text{dB}}$ (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. 5 B 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

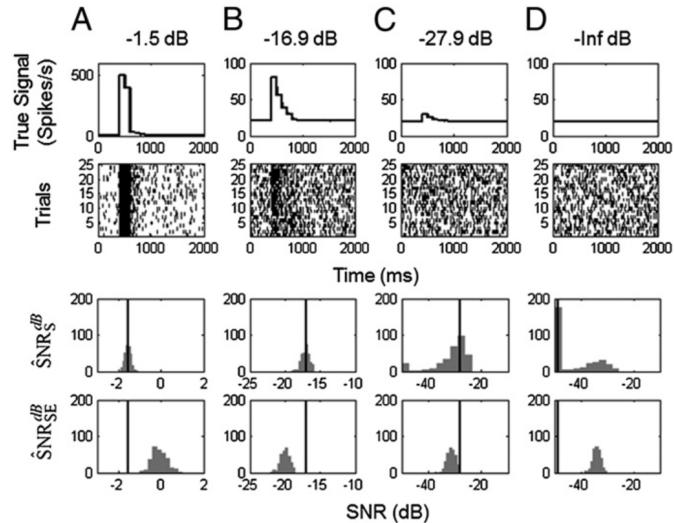


Fig. 5. A comparison of SNR estimation in simulated neurons. The stimulus components were set at four different SNRs: (A) -1.5 dB , (B) -16.9 dB , and (C) -27.9 dB with no history component. For each SNR level, 300 25-trial simulations were performed. Shown are (row 1) the true signal; (row 2) a raster plot of a representative simulated experiment; (row 3) histogram of the 300 KL-based SNR estimates, $\hat{\text{SNR}}_S^{\text{dB}}$; and (row 4) histogram of the 300 squared error-based SNR estimates, $\hat{\text{SNR}}_{\text{SE}}^{\text{dB}}$ (20). The vertical lines in rows 3 and 4 are the true SNRs.

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 statistic for assessing the strength of the association between data Y and covariates X_2 . The generalized SNR estimator can be seen as generalized effect size. This observation is especially important because it can be further developed for planning neurophysiological 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 Research Engineering Partnership Award R01-DA015644 (to E.N.B. and W.A.S.), Pioneer Award DP1 OD003646 (to E.N.B.), and Transformative Research Award GM 104948 (to E.N.B.); and NIH grants that supported the guinea pig experiments P41 EB2030 and T32 DC00011 (to H.H.L.).

1. Chen Y, Beaulieu N (2007) Maximum likelihood estimation of SNR using digitally modulated signals. *IEEE Trans Wirel Comm* 6(1):210–219.
2. Kay S (1993) *Fundamentals of Statistical Signal Processing* (Prentice-Hall, Upper Saddle River, NJ).
3. Welvaert M, Rosseel Y (2013) On the definition of signal-to-noise ratio and contrast-to-noise ratio for FMRI data. *PLoS ONE* 8(11):e77089.
4. Mendel J (1995) *Lessons in Estimation Theory for Signal Processing, Communications, and Control* (Prentice-Hall, Upper Saddle River, NJ), 2nd Ed.
5. Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol* 93(2):1074–1089.
6. Brillinger DR (1988) Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol Cybern* 59(3):189–200.
7. Brown EN, Barbieri R, Eden UT, Frank LM (2003) Likelihood methods for neural data analysis. *Computational Neuroscience: A Comprehensive Approach*, ed Feng J (CRC, London), pp 253–286.
8. Brown EN (2005) Theory of point processes for neural systems. *Methods and Models in Neurophysics*, eds Chow CC, Gutkin B, Hansel D, Meunier C, Dalibard J (Elsevier, Paris), pp 691–726.
9. MacEvoy SP, Hanks TD, Paradiso MA (2008) Macaque V1 activity during natural vision: Effects of natural scenes and saccades. *J Neurophysiol* 99(2):460–472.
10. Lim HH, Anderson DJ (2006) Auditory cortical responses to electrical stimulation of the inferior colliculus: Implications for an auditory midbrain implant. *J Neurophysiol* 96(3):975–988.
11. Temereanca S, Brown EN, Simons DJ (2008) Rapid changes in thalamic firing synchrony during repetitive whisker stimulation. *J Neurosci* 28(44):11153–11164.
12. Wilson MA, McNaughton BL (1993) Dynamics of the hippocampal ensemble code for space. *Science* 261(5124):1055–1058.
13. Wirth S, et al. (2003) Single neurons in the monkey hippocampus and learning of new associations. *Science* 300(5625):1578–1581.
14. Dayan P, Abbott LF (2001) *Theoretical Neuroscience* (Oxford Univ Press, London).
15. Sarma SV, et al. (2010) Using point process models to compare neural spiking activity in the subthalamic nucleus of Parkinson's patients and a healthy primate. *IEEE Trans Biomed Eng* 57(6):1297–1305.
16. Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W (1997) *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA).
17. Optican LM, Richmond BJ (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J Neurophysiol* 57(1):162–178.
18. Softky WR, Koch C (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J Neurosci* 13(1):334–350.
19. Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J Neurosci* 18(10):3870–3896.
20. Lyamzin DR, Macke JH, Lesica NA (2010) Modeling population spike trains with specified time-varying spike rates, trial-to-trial variability, and pairwise signal and noise correlations. *Front Comput Neurosci* 4(144):144.
21. Soofi ES (2000) Principal information theoretic approaches. *Am Stat* 95(452):1349–1353.
22. Erdogmus D, Larsson EG, Yan R, Prince JC, Fitzsimmons JR (2004) Measuring the signal-to-noise ratio in magnetic resonance imaging: A caveat. *Signal Process* 84(6):1035–1040.
23. Mittlböck M, Waldhör T (2000) Adjustments for R^2 -measures for Poisson regression models. *Comput Stat Data Anal* 34(4):461–472.
24. Kent JT (1983) Information gain and a general measure of correlation. *Biometrika* 70(1):163–173.
25. Alonso A, et al. (2004) Prentice's approach and the meta-analytic paradigm: A reflection on the role of statistic in the evaluation of surrogate endpoints. *Biometrics* 60(3):724–728.
26. Hastie T (1987) A closer look at the deviance. *Am Stat* 41(1):16–20.
27. Simon G (1973) Additivity of information in exponential family probability laws. *J Am Stat Assoc* 68(342):478–482.
28. Cameron AC, Windmeijer FAG (1997) An R -squared measure of goodness of fit for some common nonlinear regression models. *J Econom* 77(2):329–342.
29. Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. *Network* 15(4):243–262.
30. Marmarelis VZ (2004) *Nonlinear Dynamic Modeling of Physiological Systems* (John Wiley, Hoboken, NJ).
31. Plourde E, Delgutte B, Brown EN (2011) A point process model for auditory neurons considering both their intrinsic dynamics and the spectrotemporal properties of an extrinsic signal. *IEEE Trans Biomed Eng* 58(6):1507–1510.
32. Czanner G, Sarma MA, Eden UT, Brown EN (2008) A signal-to-noise ratio estimator for generalized linear model systems. *Lect Notes Eng Comput Sci* 2171(1):1063–1069.
33. McCullagh P, Nelder JA (1989) *Generalized Linear Models* (Chapman and Hall, New York).
34. Pawitan Y (2013) *In All Likelihood. Statistical Modelling and Inference Using Likelihood* (Oxford Univ Press, London).
35. Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM (2002) The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput* 14(2):325–346.
36. Czanner G, et al. (2008) Analysis of between-trial and within-trial neural spiking dynamics. *J Neurophysiol* 99(5):2672–2693.
37. Sekhar SC, Sreenivas TV (2006) Signal-to-noise ratio estimation using higher-order moments. *Signal Process* 86(4):716–732.
38. Cajigas I, Malik WQ, Brown EN (2012) nSTAT: Open-source neural spike train analysis toolbox for Matlab. *J Neurosci Methods* 211(2):245–264.
39. Collins FS, Tabak LA (2014) Policy: NIH plans to enhance reproducibility. *Nature* 505(7485):612–613.