Supporting Information

Czanner et al. 10.1073/pnas.1505545112

What Does the SNR Estimate?

The SNR has been most studied for linear model systems $Y = X\beta + \varepsilon$ in which one has observations $y = (y_1, \ldots, y_n)'$ of a random vector $Y = (Y_1, \ldots, Y_n)'$, $X\beta$ is the signal, $X = (x_1, \ldots, x_p)$ is the $n \times p$ design matrix, x_k are fixed known vectors of covariates ($k = 1, \ldots, p$), β is a $p \times 1$ vector of unknown coefficients, and ε is an $n \times 1$ vector of independent, identically distributed Gaussian random errors with zero mean and variance σ_ε^2. The first column in X is a $n \times 1$ vector of 1s denoted as 1_n. The unconditional mean of the random vector Y can be defined as $EY = 1_n\beta_0$, where scalar parameter β_0 is typically unknown.

A standard way to define the SNR is as a ratio of variances as

$$\mathrm{SNR}_X = \frac{\sigma^2_{signal}}{\sigma^2_{noise}}, \qquad \textbf{[S1]}$$

where σ^2_{signal} is the variance of the signal representing the expected variability in the data induced by the signal, where

$$\sigma^2_{signal} = (X\beta - 1_n\beta_0)'(X\beta - 1_n\beta_0),$$

and $\sigma^2_{noise} = n\sigma^2_\varepsilon$ is the variance of noise. In other words, SNR is the true expected proportion of variance in the data due the signal divided by the variance due to the noise.

We can obtain an alternative interpretation of SNR (Eq. **S1**) if we view the two variances in terms of EPEs in the squared error sense. The variance of the noise can be viewed as the expected error of predicting Y when using covariates X (1). That is,

$$\sigma^2_{noise} = \mathrm{EPE}(Y, X\beta) = E[(Y - X\beta)'(Y - X\beta)]. \qquad \textbf{[S2]}$$

Analogously, the expected error of predicting Y when using $1_n\beta_0$ is

$$\mathrm{EPE}(Y, 1_n\beta_0) = E[(Y - 1_n\beta_0)'(Y - 1_n\beta_0)]. \qquad \textbf{[S3]}$$

Due to the Pythagorean property of EPE in linear Gaussian system (1–3), at the parameter values β_0 and β that minimize the EPE, the variance of the signal can be expressed as

$$\sigma^2_{signal} = \mathrm{EPE}(Y, 1_n\beta_0) - \mathrm{EPE}(Y, X\beta),$$

the EPE of predicting the values of Y with overall mean, $1_n\beta_0$ minus the EPE of predicting Y with the approximating model, $X\beta$ (4). Hence, σ^2_{signal} is the reduction in the EPE achieved by using the covariates X. This leads to an alternative definition of the SNR as

$$\mathrm{SNR}_X = \frac{\mathrm{EPE}(Y, 1_n\ \beta_0) - \mathrm{EPE}(Y, X\beta)}{\mathrm{EPE}(Y, X\beta)}, \qquad \textbf{[S4]}$$

which is the reduction EPE due to the signal, divided by the EPE due to noise, in the squared-error sense. For this reason, we will refer to Eqs. **S1** and **S4** as a variance-based or a squared-error-based SNR.

The variance-based SNR_X is the true expected SNR obtained if the parameters β and β_0 that give the minimum EPEs are known. In practice, however, the SNR_X is estimated by replacing the parameters β and β_0 by their least-squares estimates $\hat{\beta}$ and $\bar{y}$, respectively. This leads to the estimate of SNR_X (Eq. **S4**)

$$\widehat{\mathrm{SNR}}_X = \frac{SSResidual(y, 1_n\bar{y}) - SSResidual(y, X\hat{\beta})}{SSResidual(y, X\hat{\beta})} \qquad \textbf{[S5]}$$

where

$$SSResidual(y, 1_n\bar{y}) = (y - 1_n\bar{y})'(y - 1_n\bar{y})$$

$$SSResidual(y, X\hat{\beta}) = (y - X\hat{\beta})'(y - X\hat{\beta}).$$

In linear model analyses, $SSResidual(y, 1_n\bar{y})$ is the variance of the data around their estimated overall mean and $SSResidual(y, X\hat{\beta})$ is the estimated variability in the data around the estimated signal $X\hat{\beta}$, i.e., the variability that is not explained by the covariate X.

Defining the SNR for a Linear Gaussian Signal Plus Covariates Plus Noise System

If the system is driven by a signal and a nonsignal component and if the two components can be separated by an approximate linear additive model, then the SNR definition and estimate must be modified. We assume the covariate component of the linear model $Y = X\beta + \varepsilon$ can be partitioned as $X\beta = X_1\beta_1 + X_2\beta_2$, where the first component, $X_1\beta_1$, is a covariate not related to the signal, and the second component, $X_2\beta_2$, is the signal. There exist values of vectors β, β_1, and β_2 that give the minimum EPEs for describing Y in terms of minimizing $\mathrm{EPE}(Y, X\beta)$, $\mathrm{EPE}(Y, X_1\beta_1)$, and $\mathrm{EPE}(Y, X_2\beta_2)$, respectively. For this case, we can define SNR in which only a part of the variability in random vector Y is attributed to the signal to extend the SNR definition in Eq. **S1** by replacing $\mathrm{EPE}(Y, 1_n\beta_0)$ with $\mathrm{EPE}(Y, X_1\beta_1)$ to obtain

$$\mathrm{SNR}_{X_2} = \frac{\mathrm{EPE}(Y, X_1\beta_1) - \mathrm{EPE}(Y, X\beta)}{\mathrm{EPE}(Y, X\beta)} \qquad \textbf{[S6]}$$

where the first column of X_1 and of X is the vector 1_n. Eq. **S6** gives the expected SNR in Y about the signal, $X_2\beta_2$, while controlling for the effect of nonsignal component, $X_1\beta_1$. The numerator in Eq. **S6** is the reduction in the EPE due to the signal, $X_2\beta_2$, when controlling for $X_1\beta_1$, the systematic changes in random vector Y unrelated to the signal whereas the denominator is the EPE due to the noise. By analogy with Eq. **S5**, we can estimate the squared-error based SNR_{X_2} (Eq. **S6**) as

$$\widehat{\mathrm{SNR}}_{X_2} = \frac{SSResidual(y, X_1\hat{\beta}_1) - SSResidual(y, X\hat{\beta})}{SSResidual(y, X\hat{\beta})}, \qquad \textbf{[S7]}$$

where we replace $SSResidual(y, 1_n\ \bar{y})$ with $SSResidual(y, X_1\hat{\beta}_1)$ in Eq. **S5**.

Defining the SNR for GLM Systems

The SNR definition and estimate in Eqs. **S6** and **S7** extend to the GLM framework, the established statistical paradigm for conducting regression analyses when data from the exponential family are observed with covariates (5). We extend SNR to GLM systems in which the covariates may be partitioned into signal and nonsignal components by replacing the squared-error EPE in Eq. **S6** with the KL EPE of Y from the approximating model and by replacing the residual sums of squares in Eq. **S7** by the residual deviances (1, 3, 5). This leads to the following KL generalization of the true SNR in Y about the signal X_2, while taking into account the nonsignal effects X_1, for the system approximated by the GLM

$$\mathrm{SNR}_{X_2} = \frac{\mathrm{EPE}(Y, X_1\beta_1) - \mathrm{EPE}(Y, X\beta)}{\mathrm{EPE}(Y, X\beta)}, \qquad \textbf{[S8]}$$

and its deviance-based estimate

$$\hat{\mathrm{SNR}}_{X_2} = \frac{Dev(y, X_1\hat{\beta}_1) - Dev(y, X\hat{\beta})}{Dev(y, X\hat{\beta})} \qquad \textbf{[S9]}$$

$$\mathrm{EPE}(Y, X_1\beta_1) = E[-2\,log\,f(Y|X_1\beta_1)]$$

where the expectation is taken with respect to true generating probability distribution of random vector Y and the "2" in the definition makes the log-likelihood loss for the Gaussian distribution match squared-error loss. For this reason we refer to Eq. **S8** as a KL-based SNR and Eq. **S9** as its KL- or deviance-based SNR estimator.

The deviance is

$$Dev(y, X\hat{\beta}) = -2log\,\frac{L(y, X\,\hat{\beta})}{L(y,\,y)} \qquad \textbf{[S10]}$$

where $L(y, X\,\hat{\beta})$ is the likelihood evaluated at the maximum likelihood estimate $\hat{\beta}$ of the model parameter β. $L(y,\,y)$ is the saturated likelihood defined as the highest value of the likelihood (5).

By the Pythagorean property of the KL divergence estimate in a GLM with canonical link (1–3), the numerator in Eq. **S8** is the reduction in KL EPE due to the signal, $X_2\beta_2$, while controlling for the effect of the nonsignal component, $X_1\beta_1$. The KL-based SNR_{X_2} has squared error-based SNR_{X_2} as a special case in which the exponential family model has the Gaussian distribution. The numerator of the SNR estimate (Eq. **S9**) gives the reduction in deviance due to signal, $X_2\hat{\beta}_2$, while controlling for the nonsignal component, $X_1\hat{\beta}_1$. The estimates $\hat{\beta}$ and $\hat{\beta}_1$ are computed from two separate maximum-likelihood fits of the two models to data y (6).

We define a bias correction for the SNR estimator (Eq. **S9**), as this problem is especially prevalent in data with a weak signal (4, 7). By definition, the SNR estimate is always positive. Under regularity conditions, the asymptotic biases of the numerator and denominator in Eq. **S9** are respectively $\dim(\beta_1) - \dim(\beta)$ and $\dim(\beta)$, suggesting the approximate bias-corrected SNR estimate

$$\hat{\mathrm{SNR}}_{X_2} = \frac{Dev(y, X_1\hat{\beta}_1) - Dev(y, X\hat{\beta}) + \dim(\beta_1) - \dim(\beta)}{Dev(y, X\hat{\beta}) + \dim(\beta)}. \qquad \textbf{[S11]}$$

This SNR estimate remains biased because a ratio of unbiased estimators is not necessarily an unbiased estimator of the ratio. Our simulation studies in Fig. 4 (rows 4 and 5) suggest that the bias is small for neural spike trains (4).

Variance-Based and KL-Based SNR Are the Same in Linear Systems with Independent and Additive Gaussian Noise

We assume that $y_1, \ldots, y_n$ is a realization of independent random variables $Y_1, \ldots, Y_n$, from a linear regression model, with means $E[Y_i|X_i] = X_i\beta$, zero covariances, and a common random error variance, σ_ε^2. We also assume overall (unconditional) mean $E[Y_i] = \beta_0$. Furthermore, we assume a reduced model β_1, i.e., $\beta_1 \subset \beta$. An example of reduced model is a model with the overall mean, β_0, representing the background firing constant (see Eq. **6**), or a model with parameter vector β_1 for background firing constant and for nonsignal covariates. The full model is always the generating model or a good approximating model.

Then, under the above assumptions, the divergence between data $y_1, \ldots, y_n$ and the model $X\beta_1$ is

$$\mathrm{KL}(y_1..y_n, X\beta_1) = (y - X\beta_1)^T (y - X\beta_1), \qquad \textbf{[S12]}$$

and, assuming the vector value β_1 that minimizes EPE, the mean is equal to

$$\begin{aligned}\mathrm{EPE}_{KL}(Y_1..Y_n, X_1\beta_1) &= E[KL(Y_1..Y_n, X\beta_1)] = \sum E(Y_i - X_i\beta_1)^2 \\ &= \mathrm{EPE}_{SE}(Y_1..Y_n, X_1\beta_1)\end{aligned} \qquad \textbf{[S13]}$$

i.e., KL-based EPE reduces to squared-error-based EPE for the Gaussian linear system with independent noise. Furthermore,

$$\begin{aligned}\sum E(Y_i - X_i\beta_1)^2 &= \sum E(Y_i - X_i\beta_1)^2 + \sum (X_i\beta - X_i\beta_1)^2 \\ &= n\sigma_\varepsilon^2 + (X_i\beta - X_i\beta_1)^2\end{aligned} \qquad \textbf{[S14]}$$

where Y_i and X_i are ith component of Y and ith row of X, respectively. Hence, for a linear Gaussian system, we have $\mathrm{EPE}_{KL}(Y_1..Y_n, X_1\beta_1) = n\sigma_\varepsilon^2 + \sum (X_i\beta - X_i\beta_1)^2$, with a special case being $\beta_1 = \beta$ that gives $\mathrm{EPE}_{KL}(Y_1..Y_n, X\beta) = n\sigma_\varepsilon^2 = \sigma_{noise}^2$. If we substitute this into Eq. **S6**, we obtain

$$\begin{aligned}\mathrm{SNR}_{X_1} &= \frac{\mathrm{EPE}_{KL}(Y, X_1\beta_1) - \mathrm{EPE}_{KL}(Y, X\beta)}{\mathrm{EPE}_{KL}(Y, X\beta)} \\ &= \frac{n\sigma_\varepsilon^2 + \sum (X_i\beta - X_i\beta_1)^2 \; - n\sigma_\varepsilon^2}{n\sigma_\varepsilon^2} \\ &= \frac{(X\beta - 1_n\beta_0)^T (X\beta - 1_n\beta_0)}{\sigma_{noise}^2}.\end{aligned}$$

That is,

$$\mathrm{SNR}_{KL,X_1} = \mathrm{SNR}_{SE,X_1} \qquad \textbf{[S15]}$$

in systems that are linear with additive, independent, and Gaussian noise. Lastly, for completeness, we note here that the scale parameter of a linear Gaussian system is $\phi = \sigma_\varepsilon^2$.

Variance-Based and KL-Based SNR Are Not the Same for Independent Binomial Observations

We assume that data $y_1, \ldots, y_L$ are recorded at 1-ms resolution and that they are realizations of independent random variables $Y_1, \ldots, Y_L$, from a Bernoulli distribution with parameters K and p_l, $l = 1, \ldots, L$ i.e., their means are $K \times p_l$ and the variances are $K \times p_l \times (1 - p_l)$. Then the overall expected probability of an event (such as a spike) is $p = L^{-1} \sum K \times p_l$ and the total variance is $\sum Var(y_l) = \sum K \times p_l \times (1 - p_l)$, and hence the squared-error-based SNR (Eqs. **S1** and **S4**) can be shown to be

$$\mathrm{SNR} = \frac{\sum (K \times p_l - K \times p)^2}{\sum K \times p_l \times (1 - p_l)}. \qquad \textbf{[S16]}$$

The formula in Eq. **S16** can be used to calculate SNR for spike trains when spikes trains are independent across trials, and when times of spikes are independent within each trial, such as when there is no spike history dependence. Then, one can summarize the data into a 1-ms peristimulus time histogram, which can be seen as a realization of independent Binomial random variables. Then the SNR numerator in Eq. **S16** is the variance of signal, and the denominator contains the sum of variances of Binomial random variables across L bins. This idea was used in ref. 8, and it was extended to incorporate the spike history, which was estimated in a sequential manner rather than in one single analysis. Nevertheless, our simulations in Fig. 5 indicate that expected variance-based SNR (Eq. **S16**) is smaller than KL-based expected SNR (Eq. **S8**).

Variance-Based SNR and the Coefficient-of-Determination

In linear models with Gaussian noise, the coefficient of determination, R^2, is a commonly used measure of the fit of the model to the data. The coefficient of determination ranges from 0 to 1, with 1 indicating perfect fit. Specifically,

$$R^2 = \frac{(1_n\bar{y} - X\hat{\beta})^T(1_n\bar{y} - X\hat{\beta})}{(y - 1_n\bar{y})^T(y - 1_n\bar{y})} = \frac{SSResidual(y, 1_n\bar{y}) - SSResidual(y, X\hat{\beta})}{SSResidual(y, 1_n\bar{y})}, \quad \textbf{[S17]}$$

i.e., the numerators of SNR estimator (see Eq. **S5**) and of R^2 are the same, and it is the sum-of-squares explained by the model (i.e., the signal), and it is often referred to as *SSModel* or *SSRegression* in statistical software output. The denominators of R^2 and SNR are different. The denominator in R^2 is the sum of squares around the grand mean, $SSResidual(y, 1_n\bar{y})$, representing the total variability in the data and hence often referred to as *SSTotal*. On the other hand, the variability of the data around the estimated linear function is summarized in the term $SSResidual(y, X\hat{\beta})$, which is often referred to in the statistical software as *SSResidual*. In summary, the R^2 can be written as

$$R^2 = \frac{SSModel}{SSTotal} = \frac{SSModel}{SSModel + SSResidual}, \quad \textbf{[S18]}$$

and we have that

$$\hat{\text{S}}\text{NR} = \frac{SSModel}{SSResidual}.$$

It follows that

$$1/R^2 = \frac{SSModel + SSResidual}{SSModel} = 1 + \frac{SSResidual}{SSModel} = 1 + 1/\hat{\text{S}}\text{NR}$$

and that

$$\hat{\text{S}}\text{NR} = \begin{cases} \dfrac{R^2}{1-R^2} \ \mathit{if}\ R^2 \neq 1 \\ \mathit{Inf}\ \mathit{if}\ R^2 = 1 \\ 0\ \mathit{if}\ R^2 = 0 \end{cases} \quad \textbf{[S19]}$$

$$\hat{\text{S}}\text{NR}^{dB} = \begin{cases} 10log_{10}\left(\dfrac{R^2}{1-R^2}\right) \mathit{if}\ R^2 \neq 1 \\ \mathit{Inf}\ \mathit{if}\ R^2 = 1 \\ -\mathit{Inf}\ \mathit{if}\ R^2 = 0 \\ 0\ \mathit{if}\ R^2 = 0.5 \end{cases}. \quad \textbf{[S20]}$$

Hence, by Eq. **S20**, we have that squared-error-based $\hat{SNR}$ is an increasing function of R^2 (Fig. S1). Furthermore, both quantities R^2 and SNR decrease with increasing level of noise (Fig. S2).

A well-known problem with R^2 is that it always increases, even if unimportant covariates are added to the model. Hence an adjusted R^2 was proposed (6, 9) that adjusts for the number of explanatory terms in a model. Unlike R^2, the adjusted R^2 increases only if the new term improves the model more than would be expected by chance. The adjusted R^2 can be negative—just like bias-adjusted SNR—and will always be less than or equal to the R^2. While R^2 is a measure of fit, the adjusted R^2 is used for comparison of nested models and for feature (i.e., variable) selection in model building and machine learning. By analogy, the adjusted SNR can also be used for feature selection in biological systems to quantify the amount of information in features.

There are many generalizations of R^2 for GLM models (called pseudo-R^2). Some generalizations are based on likelihoods (9, 10). Their bias-adjusted versions for independent data are known and implemented in statistical software (e.g., statistical software R). These bias-adjusted pseudo-R^2 measures can be directly used to obtain the bias-adjusted SNR via Eq. **S20**. However, even if an unbiased R^2 estimate is used in Eq. **S20** under the assumption that the data are independent, then the SNR estimate can still be biased because the ratio of unbiased estimates is not necessarily an unbiased estimate.

Variance-Based SNR and *F*-Test Statistic

In linear regression models with independent Gaussian errors, the F test is a commonly used test to evaluate the importance of a set of covariates, X, in explaining the variability of dependent variable, Y. The F-test statistic has the form

$$F = \frac{SSModel/df(Model)}{SSResidual/df(Residual)}$$

where $df(Model) = k - 1$, $df(Residual) = n - k - 1$ are degrees of freedom of the model and residuals, and k is the number of covariates (i.e., the number of columns of X). Hence, using Eq. **S5**,

$$F = \hat{\text{S}}\text{NR} \times \frac{df(Residual)}{df(Model)} \quad \textbf{[S21]}$$

i.e., the bias-unadjusted SNR estimate Eq. **S5** is a multiple of the F statistic.

If there is no signal, then $(\sigma^2_{signal}/\sigma^2_{noise}) = 0$, i.e., SNR = 0 (in Eq. **S1**). In this case, none of the covariates in matrix X is related to Y. In other words, the true generating model is a model with a constant only. In this case, the F statistic has a central Fisher distribution with degrees of freedom $df(Model)$ and $df(Residual)$. It is easy to see that the mean of the F statistic (if $df(Residual) > 2$) is

$$E(F) = \frac{df(Residual)}{df(Residual) - 2}$$

and hence, when there is no signal, it follows from Eq. **S21** and properties of the central F distribution that the mean of the variance-based $\hat{\text{S}}$NR is

$$E(\hat{\text{S}}\text{NR}) = \frac{df(Residual)}{df(Residual) - 2} \times \frac{df(Model)}{df(Residual)} = \frac{df(Model)}{df(Residual) - 2},$$

while the true SNR = 0; hence the bias of $\hat{\text{S}}$NR is $df(Model)/[df(Residual) - 2]$, which converges to zero when the ratio of data size to number of parameters becomes large.

In the general case, when the true variance-based SNR $\neq 0$, then the associated F statistic (Eq. **S21**) has a noncentral Fisher distribution with degrees of freedom $df(Model)$ and $df(Residual)$ and with a noncentrality parameter equal to $\sigma^2_{signal}/\sigma^2_{\varepsilon} = n \times$ SNR. In such a case, it can be shown that

$$E[\hat{\mathrm{SNR}}] = E\left[F\frac{df(Model)}{n - df(Model) - 1}\right]$$
$$= \frac{df(Model)}{df(Residual) - 2} + \mathrm{SNR}\frac{n}{df(Residual) - 2}$$

and the confidence intervals for SNR can be constructed using quantiles of noncentral Fisher distribution (11).

The equivalent theory for the bias correction and confidence intervals of SNR is not available in GLM models with history dependence. Therefore, here we offered a simple bias correction Eq. **S11**) that removes some bias, and we showed that it can work well in simulations. However, it can be proved that our bias correction is asymptotically equivalent to the bias correction above for independent data from linear Gaussian model.

SNR and LR Test

The concept of SNR is also related to the concept of the LR test (5). Specifically, the scaled numerator of the generalized SNR estimate (Eq. **S9**) is an LR test statistic for testing the association between covariates X and variable Y in GLMs. Under independence of the observations, the LR test statistics have asymptotically χ^2 distributions with degrees of freedom equal to the number of estimated parameters associated with the covariates. Hence, low levels of LR lead to the conclusion that there is not enough evidence for the association, which corresponds to low values of SNR estimate (Eq. **S9**).

Variance-Based SNR and Effect Size for Linear Regression

Another related measure is effect size. Cohen's effect size for linear regression models (6, 12), defined as $f^2 = R^2/(1 - R^2)$, is the same as the squared-error-based SNR in Eq. **S5**. Cohen's f^2 is not typically reported in studies, but it is often used for sample size calculations in linear regression. For linear regressions, effect sizes of 0.02, 0.15, and 0.35 are considered small, medium, and large, respectively. These three effect sizes correspond to an R^2 of 0.02, 0.13, and 0.26 SNR of −17 dB, −8.2 dB, and −4.6 dB, which are consistent with the SNR values that we reported for some of the neurons.

1. Hastie T (1987) A closer look at the deviance. *Am Stat* 41(1):16–20.
2. Simon G (1973) Additivity of information in exponential family probability laws. *J Am Stat Assoc* 68(342):478–482.
3. 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.
4. Czanner G, Sarma SV, 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.
5. McCullagh P, Nelder JA (1989) *Generalized Linear Models* (Chapman and Hall, New York).
6. Cohen J (1992) A power primer. *Psychol Bull* 112(1):155–159.
7. Chen Y, Beaulieu N (2007) Maximum likelihood estimation of SNR using digitally modulated signals. *IEEE Trans Wirel Comm* 6(1):210–219.
8. 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.
9. Agresti A, Caffo B (2002) Measures of relative model fit. *Comput Stat Data Anal* 39:127–136.
10. Nagelkerke NJD (1991) A note on a general definition of the coefficient of determination. *Biometrika* 78(3):691–692.
11. Kelley K (2007) Confidence intervals for standardized effect sizes: Theory, application, and implementation. *J Stat Softw* 20(8):1–24.
12. Fritz CO, Morris PE, Richler JJ (2012) Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol Gen* 141(1):2–18.

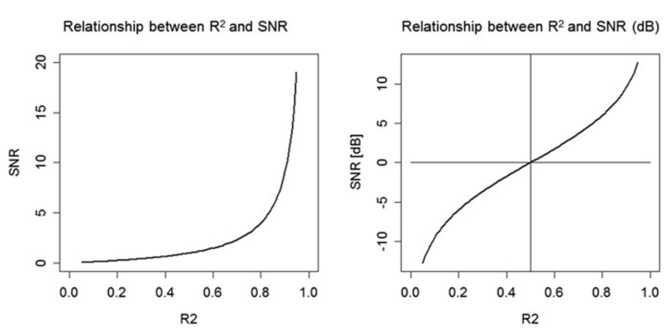

Fig. S1. Relationship between SNR and R^2. Both plots are created for R^2 values between 0.05 and 0.95. (*Left*) Eq. **S19** and (*Right*) Eq. **S20**.

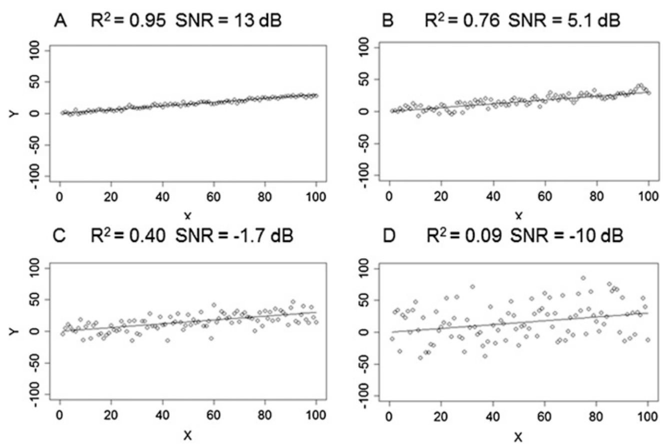

Fig. S2. Simulation analysis of the relationship between SNR and R^2. One hundred observations were simulated from the linear model $Y = 0.3X + \varepsilon$, where ε, the errors, are independent Gaussian with zero mean and SDs of 2, 5, 10, and 30. These models give (*A*) SD = 2, $R^2 = 0.95$, SNR = 13 dB; (*B*) SD = 5, $R^2 = 0.76$, SNR = 5.1 dB; (*C*) SD = 10, $R^2 = 0.40$, SNR = −1.7 dB; and (*D*) SD = 30, $R^2 = 0.09$, SNR = −10 dB.

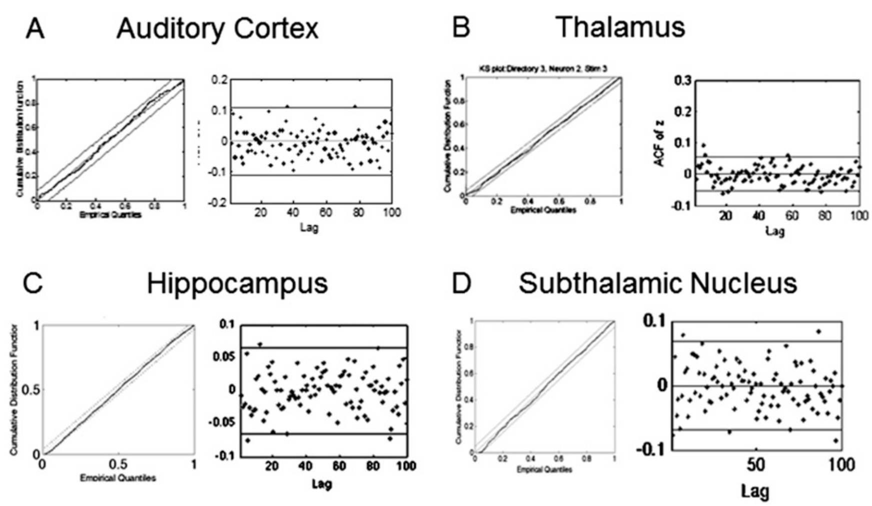

Fig. S3. Examples of goodness-of-fit analysis of GLM for a single neuron from the (*A*) primary auditory cortex of an anesthetized guinea pig, (*B*) rat thalamus, (*C*) monkey hippocampus, and (*D*) human subthalamic nucleus neuron. (*Left*) The KS plot of the time-rescaled interspike intervals The parallel 45° lines are the 95% confidence interval. The KS plot (dark curve) lies within the 95% confidence intervals, suggesting agreement between the GLM and the data. (*Right*) The partial autocorrelation function of the interspike intervals transformed into Gaussian random variables. The horizontal parallel lines are the 95% confidence. The Gaussian transformed interspike intervals falling within the 95% confidence intervals suggests lack of correlations up to lag 100. Lack of correlation is consistent with the transformed times being independent and further supports the goodness of fit of the GLM.