Voltage Signal-to-Noise Ratio (SNR) Nonlinearity Resulting From Incoherent Summations

S. T. Lowe[1]

Simple scaling rules often are used to estimate signal-to-noise ratios (SNRs). For example, one commonly hears that voltage SNR varies as the square root of the number of data samples and linearly as the signal strength. For a variety of reasons, signals often are detected or measured by cross-correlating a small interval of sampled voltage data with a phase model and coherently integrating, followed by incoherently summing the resulting amplitudes over a number of such intervals. When this is done, the usual scaling rules do not always apply and can lead to decidedly incorrect conclusions (an example is given in Section V). This article derives analytic formulas for voltage SNR and some resulting scaling laws when incoherent amplitude sums are performed. In a common low-SNR situation, the correct rule states that voltage SNR varies as the square of the signal strength.

I. Background

Detecting a weak sinusoidal signal in a white noise environment usually involves cross-correlating sampled voltage data, $D(t)$, with a complex phase model, $\phi_m(t)$, and integrating over time:

$$\vec{\rho} = \int_0^T D(t)\, e^{-\mathrm{i}\,\phi_m(t)}\, dt \tag{1}$$

The resulting complex phasor, $\vec{\rho}$, will obtain its maximum amplitude when the phase model is accurate enough to approximately stop the cross-correlated signal's phase variations over the integration interval, T. The phase model usually is parameterized, and finding parameter values that produce a significantly large signal-to-noise ratio (SNR) implies the existence of a signal and approximately determines its phase characteristics.

The simplest and most common phase model is linear, corresponding to a monochromatic tone:

$$\phi_m(t) = \phi_0 + \nu t \tag{2}$$

[1] Tracking Systems and Applications Section.

Here, there are two model parameters, ν and ϕ_0, which represent the frequency and initial phase; however, the integrated phasor's amplitude is insensitive to ϕ_0, making this effectively a one-parameter model when the goal is signal detection.

II. Incoherent Sums

The phase model given by Eq. (2) seldom is accurate enough to apply over long integration intervals because ν is often time dependent. The model can be improved by creating additional parameters, but this requires handling higher dimensional parameter spaces, resulting in a much greater number of trial models. This both raises the threshold for false signal detections and increases the computational requirements. When large data sets and weak signals are involved, this option may not be feasible. A common approach is to apply Eq. (2) over short enough time intervals to ensure its accuracy and to sum the resulting integrated amplitudes over all intervals. This incoherent sum results in a lower overall SNR than does applying an accurate model over the entire interval and summing coherently, and it is this effect on the overall SNR that now is examined in detail.

Assume a sinusoidal signal is added to complex white Gaussian noise, where the real and imaginary components correspond to the in-phase and quadrature-phase noise components, respectively, and each has zero mean and variance σ^2. If the phase model exactly stops the signal's phase variations, so that the signal voltage is a constant, v, the resulting integrated complex voltage has amplitude V. The probability density function (PDF) of V is Ricean[2] [1], where the PDF, mean, and variance are given by

$$PDF(V) = \frac{V}{\sigma^2}\, e^{-(V^2+v^2)/2\sigma^2} I_0\left(\frac{v\,V}{\sigma^2}\right), \qquad 0 \le V < \infty \tag{3a}$$

$$\langle V\rangle = \sqrt{\frac{\pi}{2}}\,\sigma\, e^{-v^2/4\sigma^2}\left[\left(1+\frac{v^2}{2\sigma^2}\right) I_0\left(\frac{v^2}{4\sigma^2}\right) + \frac{v^2}{2\sigma^2} I_1\left(\frac{v^2}{4\sigma^2}\right)\right] \tag{3b}$$

$$\sigma_V^2 \equiv \langle V^2\rangle - \langle V\rangle^2 = v^2 + 2\sigma^2 - \langle V\rangle^2 \tag{3c}$$

and I_0 and I_1 are modified Bessel functions of the first kind. Assume a data set having t samples divided into k pieces, each having n samples (so $k\,n = t$), a sinusoidal signal with peak voltage R, and single-sample variance σ_1^2. The quantities $\langle V\rangle$ and σ_V^2 for each n-sample piece are calculated using Eq. (3) with $v = n\,R$ and $\sigma = \sqrt{n}\,\sigma_1$. This assumes the n-sample sum is completely coherent. Assuming errors on each n-sample piece are uncorrelated, these quantities are multiplied by k for the incoherent sum over all k pieces, leading to the overall amplitude, A, and amplitude variance, σ_A^2, for the entire t-sample data set:

$$\langle A\rangle = k\sqrt{\frac{\pi}{2}}\sqrt{n}\; e^{-(n\,S^2/4)}\left[\left(1+\frac{n\,S^2}{2}\right) I_0\left(\frac{n\,S^2}{4}\right) + \frac{n\,S^2}{2} I_1\left(\frac{n\,S^2}{4}\right)\right] \tag{4a}$$

$$\sigma_A^2 = k\,(n^2\,S^2 + 2n - \langle A\rangle_{k=1}^2) \tag{4b}$$

where $S \equiv R/\sigma_1$ can be thought of as the single-sample SNR. Removing the no-signal bias $\langle A\rangle_{S=0}$ and normalizing by $\sigma_{A|_{S=0}}$ so that the noise has zero mean and unit variance defines the voltage SNR:

[2] If you are not familiar with the Ricean distribution, it is easy to conceptualize. Imagine a two-dimensional Gaussian distribution centered v units from the origin and having both x and y variances equal to σ^2. Equation (3a) gives the resulting distribution of distances V from the origin.

$$\mathrm{SNR}_V \equiv \frac{\langle A \rangle - \langle A \rangle_{S=0}}{\sigma_{A|_{S=0}}} \tag{5a}$$

$$= \sqrt{k}\,\sqrt{\frac{\pi}{4-\pi}}\left\{e^{-(\epsilon^2/4)}\,\left[\left(1+\frac{\epsilon^2}{2}\right)\,I_0\left(\frac{\epsilon^2}{4}\right)+\frac{\epsilon^2}{2}\,I_1\left(\frac{\epsilon^2}{4}\right)\right]-1\right\} \tag{5b}$$

where $\epsilon \equiv \sqrt{n}\,S$ can be thought of as the n-sample SNR. Figure 1 shows SNR_V as a function of ϵ along with polynomial expansions for small and large ϵ. Combining these expansions at the crossover point, $\epsilon = 1.6755$, leads to an approximate expression for SNR_V that is good to better than 1 percent for all ϵ:

$$\mathrm{SNR}_V \simeq \sqrt{k}\,\sqrt{\frac{\pi}{4-\pi}}\left(\frac{\epsilon^2}{4}-\frac{\epsilon^4}{64}+\frac{\epsilon^6}{768}\right) \qquad \epsilon \le 1.6755 \tag{6a}$$

$$\simeq \sqrt{k}\,\sqrt{\frac{2}{4-\pi}}\left(\epsilon-\sqrt{\frac{\pi}{2}}+\frac{1}{2\epsilon}+\frac{1}{8\epsilon^3}+\frac{3}{16\epsilon^5}\right) \qquad \epsilon > 1.6755 \tag{6b}$$

III. Scaling Laws

First, note that the definition of SNR_V includes an arbitrary scale factor. Above, S was defined as the peak voltage of a sinusoidal signal, but another definition for S, for example the signal rms, would change the constant near the front of Eq. (5b). Thus, only relative scaling behavior is meaningful. The scaling laws for SNR_V, as defined above, fall into two regimes: $\epsilon \lesssim 1$ and $\epsilon \gtrsim 4$, where, again, ϵ is the SNR_V from a single n-sample piece of data.

The regime $\epsilon \lesssim 1$ is relevant for signal detection: The signal is too weak to detect in any one n-sample piece, but one hopes to detect it by incoherently summing k such pieces. Looking at the lowest order term in Eq. (6a),

$$\mathrm{SNR}_V \simeq \frac{1}{4}\sqrt{\frac{\pi}{4-\pi}}\,\sqrt{k}\,\epsilon^2 \qquad \epsilon \ll 1 \tag{7a}$$

$$= \frac{1}{4}\sqrt{\frac{\pi}{4-\pi}}\,\frac{t\,S^2}{\sqrt{k}} \tag{7b}$$

This shows that the overall SNR_V scales as S^2 in this regime, and not as S, which indicates that efforts to increase S are much more effective in increasing SNR_V here than in the strong-signal regime, where SNR_V scales linearly with S. Also, for a fixed number of pieces, k, SNR_V scales linearly with the amount of data, t, and not by the usual $\sqrt{t}$. Finally, for a given data set (fixed t and S), the observed SNR_V in this regime scales as $\sqrt{n}$. In other words, n should be maximized for signal detection.

When $\epsilon \gtrsim 4$, the signal is detectable in a single n-sample piece, so the problems inherent in weak signals are not found here. The SNR is given approximately by

$$\mathrm{SNR}_V \simeq \sqrt{\frac{2}{4-\pi}}\,\sqrt{t}\,S \qquad \epsilon \gg 1 \tag{8}$$

The SNR_V loss, L, associated with breaking the data into k pieces is defined by the relation

$$\text{SNR}_V \equiv \text{SNR}_V \Big|_{k=1} (1 - L) \tag{9a}$$

Equation (6b) is used to determine L:

$$L \simeq \sqrt{\frac{\pi}{2}} \frac{\sqrt{k} - 1}{\sqrt{t}\, S} \qquad \epsilon \gg 1 \tag{9b}$$

This shows the strong-signal SNR_V loss associated with splitting a given data set into k coherent pieces scales as $\sqrt{k} - 1$ and inversely as the unsplit SNR.

IV. Numerical Verification

Equations (5) and (6) were verified by numerical simulation. A signal with $S = 0.05$ was added to $2^{20} = 1,048,576$ samples of normalized white noise, coherently integrated over all n-sample data pieces, and incoherently summed over all $2^{20}/n$ pieces, where n was all possible powers of 2. This was repeated 1000 times and the results averaged. Table 1 shows the resulting SNR_V along with the analytic values. Notice the big drop in SNR_V around $\epsilon = 4$, showing its two scaling regimes, $\epsilon \lesssim 1$ and $\epsilon \gtrsim 4$. Figure 2 shows the SNR_V simulation values, the analytic prediction from Eq. (5b), and the asymptotic approximations as a function of k. Figure 3 shows these same quantities as a function of ϵ.

Table 1. Simulation results.

n	k	ϵ	SNR_V	Eq. (4b)
1,048,576	1	51.20	75.528	76.254
524,288	2	36.20	73.891	75.476
262,144	4	25.60	75.213	74.385
131,072	8	18.10	72.753	72.860
65,536	16	12.80	68.685	70.738
32,768	32	9.05	65.877	67.808
16,384	64	6.40	63.157	63.807
8,192	128	4.53	57.921	58.441
4,096	256	3.20	50.850	51.474
2,048	512	2.26	42.629	43.033
1,024	1,024	1.60	34.400	34.022
512	2,048	1.13	26.347	25.702
256	4,096	0.80	18.691	18.846
128	8,192	0.57	13.433	13.582
64	16,384	0.40	9.851	9.698
32	32,768	0.28	7.070	6.892
16	65,536	0.20	4.799	4.885
8	131,072	0.14	3.418	3.459
4	262,144	0.10	2.409	2.447
2	524,288	0.07	1.682	1.731
1	1,048,576	0.05	1.250	1.224

Finally, to see SNR_V as a function of S, a second simulation was performed with $n = 2048$, $k = 16,384$, and S varying between 0.001 and 1. The results are plotted in Fig. 4 and clearly show the two scaling regimes. The long-dashed line shows the "usual" SNR_V scaling rule.

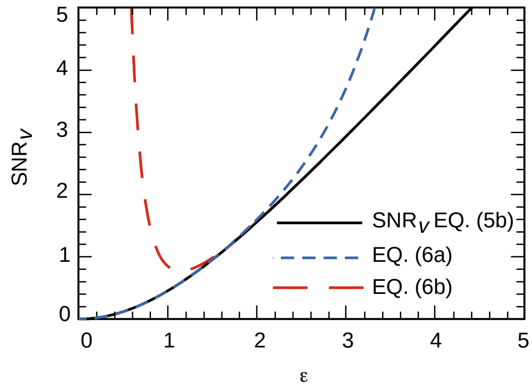

Fig. 1. The plot of $\mathbf{SNR}_V$, given by Eq. (5b), as a function of ε, for k = 1. Also shown are the $\mathbf{SNR}_V$'s polynomial expansions, Eqs. (6a) and (6b), also for k = 1.

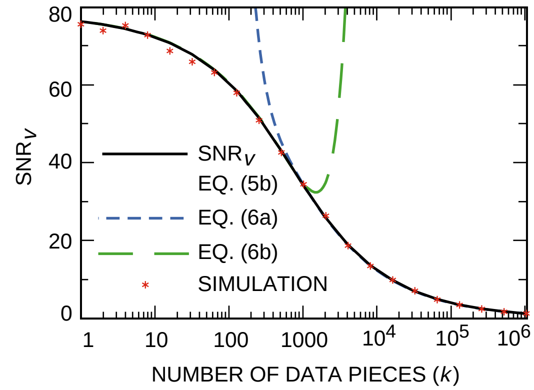

Fig. 2. The plot of $\mathbf{SNR}_V$ as a function of k for the simulation, along with the analytic prediction given by Eq. (5b). Also shown are the asymptotic expressions, Eqs. (6a) and (6b).

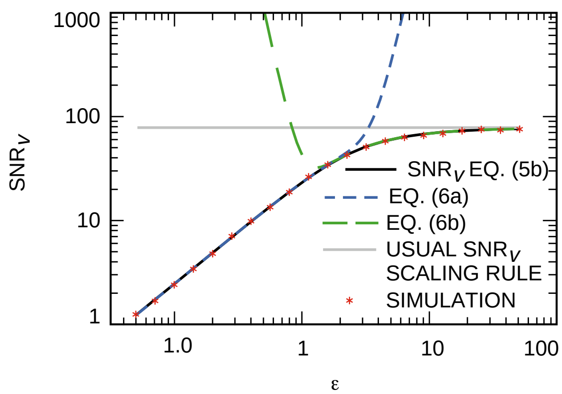

Fig. 3. The plot of $\mathbf{SNR}_V$ as a function of ε for the simulation, along with the analytic prediction given by Eq. (5b). Note that ε = $\sqrt{n}S$ and S is fixed at 0.05, showing $\mathbf{SNR}_V$ as a function of 0.05$\sqrt{n}$. Also shown are the asymptotic expressions, Eqs. (6a) and (6b).

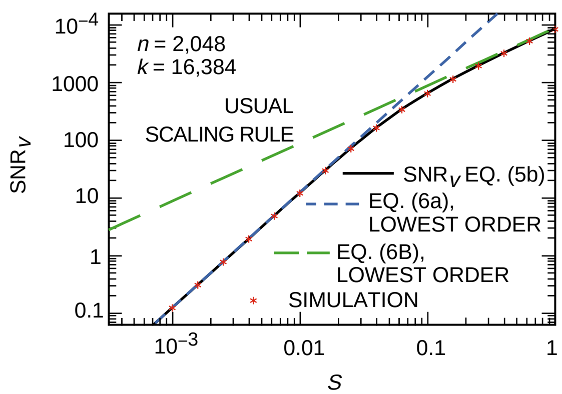

Fig. 4. The plot of $\mathbf{SNR}_V$ as a function of S for the simulation, along with the analytic prediction given by Eq. (5b). Also shown are the lowest-order asymptotic expressions, Eqs. (6a) and (6b) (first terms). The Eq. (6b) line corresponds to the usual $\mathbf{SNR}_V$ scaling rule.

V. Conclusions

It has been shown that SNR_V derived from incoherent voltage sums cannot always be scaled using the "usual" scaling rules. An example makes this clear. Assume a time series of 33.554×10^6 sampled voltages is analyzed by continuously dividing the data into 2048-sample subsets and performing a fast Fourier transform (FFT) of each subset. Also assume the coherence time of any signal is approximately equal to the time interval of a 2048-sample subset. The amplitudes of the resulting 16,384 FFTs are summed, and a signal having an SNR_V of 750 is seen. A second signal, known to have a voltage 50 times smaller than the observed signal could be present. Can this second signal be detected? The usual SNR scaling would indicate the second signal will have $\mathrm{SNR}_V \approx 750/50 \approx 15$ and can be observed. However, the correct answer is that it cannot be observed in this manner. Figure 4 shows the curve for $n = 2048$ and $k = 16,384$, indicating this signal has $S = 0.1$. Moving down by a factor of 50 to $S = 0.002$ shows the expected SNR_V is about 0.5, making the signal unobservable. This example shows that Eq. (5) should be used when scaling SNR_Vs that are derived from incoherent summations. Note that the signal could be detected if longer subsets were used and the signal were coherent over this interval.

Acknowledgment

Many thanks go to George Purcell for his very careful reading of this article.

Reference

[1] A. Thompson, J. Moran, and G. Swenson, *Interferometry and Synthesis in Radio Astronomy*, New York: John Wiley and Sons, 1987.