

## Threshold detection of wideband signals: A noise-induced maximum in the mutual information

Adi R. Bulsara\*

Naval Command, Control and Ocean Surveillance Center, RDT&E Division, Code 364, San Diego, California 92152-5000

Anthony Zador†

Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037

(Received 23 January 1996)

In some nonlinear dynamic systems, the addition of noise to a weak periodic signal can increase the detectability of the signal, a phenomenon belonging to a class of noise-induced cooperative behavior known as stochastic resonance (SR). There has been much recent speculation on the possible role of SR in signal processing by sensory neurons. However, most results have focused exclusively on increasing the output signal-to-noise ratio (SNR) of time-periodic signals, even though many real-world signals (e.g., those encountered in some neurophysiological and communications applications) are not of this form. Here we consider a generalization of SR, based on the Shannon mutual information between the transmitted and received signal. This generalization can be applied to cases (e.g., the information transmitted by the output spike train of an integrate-fire model neuron which we consider here), involving aperiodic input signals for which the output SNR might be ill-defined, uninformative, or irrelevant. Since the SR-like effect in the transmitted information disappears with the optimal choice of model parameters, we suggest that such an effect is likely to be particularly relevant to systems, e.g., neuronal populations, in which natural circuit constraints may render parameter optimization impractical. [S1063-651X(96)50109-1]

PACS number(s): 05.40.+j

Intuition suggests that when noise is added to a signal prior to or during transmission across a communication channel, the received signal will be more corrupted than if the uncorrupted signal had been transmitted. The amount of corruption is often quantified by the (suitably defined) signal-to-noise ratio (SNR) of the output: for a linear channel, the output SNR decreases monotonically with increasing noise intensity. For a large class of nonlinear channels, however, there is a seemingly paradoxical *increase* in the SNR, up to a maximum, with added noise; this effect has been widely studied under the occasionally misleading name of *stochastic resonance* (SR) [1–3]. Underlying SR is some form of thresholding in the communication channel; specifically, SR stems from a mismatch between the signal and the threshold.

The classical SR formulation in terms of the output SNR has two problems. First, SR is usually defined for systems with sinusoidal inputs; but in many cases it would be of interest to consider arbitrary input signals. A second and more basic problem is that for a nonlinear system driven by a broadband input signal, the output SNR may be either ill-defined or uninformative. Only when the input signal is sinusoidal is it straightforward to compute the output SNR, at the fundamental (or at any higher harmonic) of the input signal frequency.

Information theory [4] provides a natural framework for many problems in biological information processing [5]. Here we describe an information-theoretic generalization of SR that addresses the above-mentioned problems in the use of the output SNR as an information measure for systems

subject to aperiodic inputs [6]. Within this framework, the *mutual information* (MI)  $I[s(t), z(t)]$  between the input signal  $s(t)$  and the output signal  $z(t)$  replaces the output SNR. By analogy with the classical formulation, where SR is defined as a peak in the SNR vs noise characteristic, within the information-theoretic framework it is defined as a peak in the MI vs noise relation. SR has recently been observed in the cricket cercal system in the conventional manner, using the output SNR as a response measure (for a sinusoidal input signal), and via the above-mentioned “resonance” in a lower bound on the transinformation function (for a broadband stimulus) [7]. For the case of a fixed threshold (as in a level crossing detector), SR may be related to the well-known “dithering” effect [8]. We note that other researchers [9] have also suggested alternative measures of the response of nonlinear neuron models, to aperiodic inputs.

The basic approach can be illustrated with a very simple example. Consider a binary communication channel whose output is  $y = \theta(x + n)$ , where the signal  $x$  is a binary random variable that is 1 with probability  $P_x$  and  $-1$  with probability  $P_{\bar{x}} = 1 - P_x$ , and  $n$  is a Gaussian random variable with mean zero and variance  $\sigma_n^2$ . The output  $y$  is given by the threshold function  $\theta(\cdot)$  as  $y = 1$  if  $x + n > Q$ , and  $y = -1$  otherwise, where the constant  $Q$  denotes the threshold, and we define  $P_y$  and  $P_{\bar{y}} = 1 - P_y$  as the respective probabilities of these two occurrences. Then we can compute the mutual information  $I(X, Y)$  between the ensembles  $X$  and  $Y$  using the definition [4]

$$I(X, Y) = H(Y) - H(Y|X), \quad (1)$$

where  $H(Y) = -\langle \log_2 P_Y \rangle_y = -P_y \log_2 P_y - P_{\bar{y}} \log_2 P_{\bar{y}}$  is the entropy of  $Y$ , and

\*Electronic address: bulsara@nosc.mil

†Electronic address: zador@salk.edu

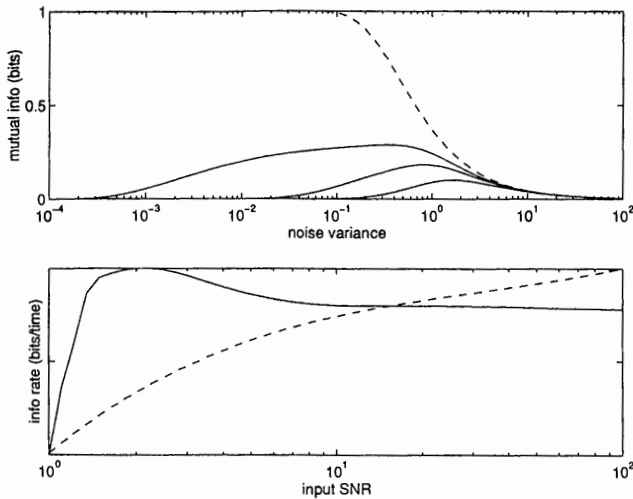


FIG. 1. (Top) Simple example of noise-induced increase in information transmission through a binary channel. Zero-mean Gaussian noise is added to a binary input  $x$  and the sum is thresholded at  $Q$  to produce a binary output  $y$ . The mutual information between  $x$  and  $y$  is plotted as a function of the noise variance  $\sigma_n^2$  for  $Q=0$  (dashed curve),  $Q=1.25$  (top curve),  $Q=2.5$  (middle curve) and  $Q=5$  (bottom curve). The dashed curve represents the maximum information obtainable for each noise level. (Bottom) Noise-induced increase in information through a spiking neuron. The mutual information (3) between the output “spike train” and the input is plotted as a function of the input SNR, defined as  $1 + \sigma_s^2/\sigma_n^2$ . For both curves,  $\tau=20$ ,  $Q=20$ , the signal variance was fixed at  $\sigma_s = (0.4)^2$  and the noise variance  $\sigma_n^2$  was varied to change the SNR. For the solid curve,  $\mu=0.8$ , while for the dashed curve  $\mu=0.9$ . Note that the information has been normalized separately for each curve: the normalization factor for the dashed curve was five times that for the solid curve, and in absolute units always exceeds it. The information rate was estimated by Monte Carlo methods.

$$\begin{aligned}
 H(Y|X) = & -\langle \log_2 P_{Y|X} \rangle_x = P_x (-P_{y|x} \log_2 P_{y|x} \\
 & - P_{\bar{y}|x} \log_2 P_{\bar{y}|x}) + P_{\bar{x}} (-P_{y|\bar{x}} \log_2 P_{y|\bar{x}} \\
 & - P_{\bar{y}|\bar{x}} \log_2 P_{\bar{y}|\bar{x}}) \quad (2)
 \end{aligned}$$

is the conditional entropy of  $Y$  given  $X$ . Equation (1) is a definition, while the expressions for  $H(Y)$  and  $H(Y|X)$  apply only if the output distribution  $Y$  is binary. For this simple example, the relevant quantities are straightforward to compute. For example, the output is 1 whenever either the signal  $x$  is 1 and the noise  $n$  is smaller than  $1-Q$ , or  $x$  is  $-1$  and  $n$  is larger than  $-1-Q$ . Thus the probability  $P_y$  of the event  $y=1$  is the sum of these two conjunctions,

$$P_y = \frac{1}{2} \left[ P_x \operatorname{erfc} \left( \frac{1-Q}{\sigma_n \sqrt{2}} \right) + P_{\bar{x}} \operatorname{erfc} \left( \frac{-1-Q}{\sigma_n \sqrt{2}} \right) \right],$$

where the erfc arises from the cumulative density of the Gaussian noise. Using similar arguments, we see that

$$P_{y|x} = \frac{1}{2} \operatorname{erfc} \left( \frac{1-Q}{\sigma_n \sqrt{2}} \right), \quad P_{y|\bar{x}} = \frac{1}{2} \operatorname{erfc} \left( \frac{-1-Q}{\sigma_n \sqrt{2}} \right)$$

$$P_{\bar{y}|x} = 1 - P_{y|x}, \quad \text{and} \quad P_{\bar{y}|\bar{x}} = 1 - P_{y|\bar{x}}.$$

Figure 1 (top) shows the mutual information  $I(X, Y)$  as a

function of the input noise variance  $\sigma_n^2$  for four values of the threshold,  $Q=0, 1.25, 2.5, 5$ . For  $Q=0$ , in fact for all  $|Q| < 1$ , the transmitted information is unity up to a critical noise level and then falls off steeply; thus in the regime  $|Q| < 1$ , the channel works best without added noise. However for  $|Q| > 1$ , i.e., when  $Q$  exceeds the dynamic range  $[-1, 1]$  of the input  $x$ , in the absence of noise the output  $y$  is independent of the input and  $I(X, Y)=0$ . As noise is added to induce threshold crossings, the transmitted information increases to a maximum, after which added noise degrades the signal. Thus for every threshold  $Q$  in this regime, there is an optimal noise level that maximizes the transmitted information. Note, however, that if  $Q$  could be varied, then the maximum information would be a monotonically decreasing function of  $\sigma_n^2$ ; indeed, it would be precisely the sigmoidal envelope given by the optimal value  $Q=0$  (dashed line). Thus the SR-like effect can only be realized if the threshold  $Q$  is not a free parameter.

We can now apply these observations to a question of substantial interest in neurophysiology. A typical neuron receives a continual barrage of synaptic inputs from other neurons; with the exception of primary sensory transduction neurons driven directly by external stimuli, this is the *only* input that the neuron receives. The signal generated by the synaptic barrage is, in general, aperiodic. Since a necessary (but not sufficient) condition for useful computation is that the output of a neuron preserve information about its input, it is reasonable to ask how the mutual information between a neuron's input and its output spike train depends on the input noise, when the input signal is nonsinusoidal.

We consider a standard model of neuronal dynamics, the “leaky” integrate-and-fire model [10]. The input  $y(t)$  to the neuron is the sum of a signal  $s(t)$  and some noise  $n(t)$ ,  $y(t) = n(t) + s(t)$ . The neuron dynamics are given by  $\dot{v}(t) = [v(t)/\tau] + y(t) + \mu$ , with initial condition  $v(t_0) = 0$  for  $v(t) < Q$ , where  $\tau$  is the membrane time constant,  $\mu$  is the steady state input,  $Q$  is the firing threshold, and  $v(t)$  denotes the membrane voltage. When  $v(t) = Q$ , the neuron emits a spike and resets, deterministically, to  $v(t_0)$ . The output  $x(t)$  is completely characterized by a sequence  $t_1, t_2, \dots, t_i, \dots$  of spikes, called the *spike train*, i.e., the times at which the threshold was crossed. The sequence can equivalently be represented by the list of interspike intervals (ISIs)  $T_i = t_{i+1} - t_i$ . This model is completely Markovian: no history prior to the last reset is preserved. Note that if both  $n(t)$  and  $s(t)$  are Gaussian and white, then  $y(t)$  is also Gaussian and white, and the distinction between “signal” and “noise” is purely a matter of definition: no measure based solely on output statistics—such as the output SNR—could provide a complete insight into the information being transmitted.

We are interested in  $I[S(t), Z(t)]$ , the mutual information between the signal and the spike train, because it tells us how much information (in bits/time) the spike train conveys about the signal. In general this is a difficult quantity to compute, because it involves the joint probability distribution of all possible signals and resultant spike trains—a high-dimensional distribution. However, if the ISIs are independent, then we can express the information in terms of  $P(T)$ , the probability distribution of a single ISI. Since by assumption the neuronal dynamics are forgetful, correlations in the

ISI distribution can only arise from correlations in the signal itself; if we consider only input signals  $S(t)$  drawn from a white Gaussian distribution, and uncorrelated with the noise  $N(t)$ , then correlations cannot arise from the input and ISIs are guaranteed to be independent. In this case  $I(S(t), Z(t))$  can be expressed as the product of  $I(S(t), T)$ , the mutual information per spike between the signal and a single ISI, and the average spike rate  $\bar{R} = 1/\langle T \rangle$  [which is calculated directly from the distribution  $P(T)$ ],

$$I(S(t), Z(t)) = \bar{R} I(S(t), T) = \bar{R} [H(T) - H(T|S(t))], \quad (3)$$

where we have used the expression for the mutual information given by Eq. (1). The right-hand side is straightforward to evaluate, since it depends only on the ISI distribution  $P(T)$  and the conditional distribution  $P(T|s(t))$ , rather than the distribution of all spike trains  $P(Z(t))$ .

In general, both  $P(T)$  and  $P(T|s(t))$  depend on the statistics of the signal and noise ensembles. Thus, the entropy of the ISI distribution is  $H(T) = -\int_0^\infty P(T) \log_2 P(T) dT$ , where we utilize a new realization of the signal and noise after every threshold crossing, ultimately computing the density function  $P(T)$ . The conditional entropy is similarly defined in terms of the ISI distribution conditional on the signal

$$H(T|S(t)) = - \left\langle \int_0^\infty P(T|s(t)) \log_2 P(T|s(t)) dT \right\rangle_{s(t)}. \quad (4)$$

The evaluation of the conditional entropy can be thought of as involving two steps. First, some particular realization  $s(t)$  of the signal is generated. A different realization of the noise is introduced following each threshold crossing and the conditional density  $P(T|s(t))$ , and hence the entropy  $H(T|s(t))$ , computed for that realization of the signal. This quantifies the input-output ‘‘channel’’ and may be regarded as a measure of the reliability of the response to the particular input  $s(t)$ ; in general, the reliability decreases as the noise variance increases. This procedure is then repeated for different realizations of the signal  $s(t)$  drawn from the ensemble, and the average over the signal ensemble is obtained; this is the reliability averaged over the ensemble of inputs. For any particular choice of signal and noise statistics, the above expression can be estimated through Monte Carlo methods.

We now consider the effect on the information rate of added noise, where the noise level at the input is characterized by the SNR which is by definition,  $\sigma_y^2/\sigma_n^2 = 1 + \sigma_s^2/\sigma_n^2$ . Increasing the noise  $\sigma_n^2$  while holding the signal variance  $\sigma_s^2$  fixed decreases the input SNR but increases  $\sigma_y^2$ . This exerts two competing effects on the mutual information. First, the increase in  $\sigma_y^2$  increases the spike rate  $\bar{R}$ . Second, the decrease in the input SNR decreases the information per spike  $I(S(t), T)$ . The relative contribution of these two effects to the information rate  $\bar{R} I(S(t), T)$  depends on the model parameters. When the first effect dominates at low noise and the second effect dominates at high noise, an SR-like effect is observed.

Figure 1 (bottom) (*solid curve*) shows an example of an SR-like relation between the input SNR and the transmitted information (shown in normalized units of bits/time). We

have considered the regime  $\mu\tau < Q$  in which, for the weak signal strengths considered here, the threshold crossings are largely noise-induced and  $\bar{R} \ll 1/\tau$  for  $\sigma_n^2 = 0$ . This corresponds (in the zero-noise limit) to a near-exponential ISI distribution with maximal information transfer [11]. In this so-called ‘‘Poisson regime,’’ the assumptions leading to (4) are expected to be rigorous and the spike rate is a very sensitive function of  $\sigma_y^2$ , so that a small amount of added noise greatly increases the spike rate  $\bar{R}$  but exerts only a moderate amount on the information per spike. Thus, at low noise (high SNR) the information rate *increases* with added noise, until the mutual information passes through a maximum. This maximum is analogous to the maximum in the output SNR in the classical formulation of SR, and is seen to occur at an input SNR  $\approx 2$ . We note that, for the solid curve of Fig. 1 (bottom), the product  $\bar{R}\tau$  decreases from 1.0 to 0.01 with increasing input SNR; at the maximum,  $\bar{R}\tau = 0.1$ .

The SR-like maximum disappears outside the Poisson firing regime, i.e., when the condition  $\bar{R} \ll 1/\tau$  is not met: the effect is observed only when the noise-induced increase in the firing rate compensates for the concomitant noise-induced decrease in the information per spike. The *dashed curve* in Fig. 1 (bottom) shows the effect of increasing  $\bar{R}$  by increasing  $\mu$ . In this case, the signal is effectively supra-threshold (although we still have  $\mu\tau < Q$ ) and the information rate is a monotonically increasing function of the input SNR. Note that the information rate is shown in normalized units; the actual peak information rate is five times lower for the *solid curve* in which information rate was increased through noise than for the *dashed curve* in which the information rate was increased by varying the threshold. The same effect is obtained by increasing  $\sigma_s^2$  or decreasing  $Q$ . However, the difference in scales between the curves of Fig. 1 (bottom) illustrates that when trying to enhance the detectability of a weak signal in a noise background, the best results are obtained by lowering the detector threshold, rather than by adding more noise although, for nonlinear detectors, the latter procedure has been shown to enhance the ‘‘detection probability’’ under the SR scenario [12]. We note that a SR-like effect has recently been quantified, via the Fisher information, in a similar model driven by a dc signal in white Gaussian noise [13] as well as via a dynamical entropy [14] in a sinusoidally driven Schmitt trigger.

In related work, DeWeese and Bialek used an information-theoretic approach to study a different model of neuronal dynamics [6]. They considered the linear-filtered threshold crossing model and showed that if, as in our simple example [Fig. 1 (top) *dashed curve*] above, the model parameters (e.g., the threshold) were set to optimize information rate, the rate was a monotonic function of the SNR of the input, and no SR effect was observed. They therefore suggested that adding noise could be viewed as a strategy for overcoming an ‘‘incorrect’’ setting of model parameters.

Although, in the integrate-and-fire model that we have considered, the SR similarly disappears with the appropriate tuning of model parameters, SR-like effects may nevertheless be relevant to the nervous system when constraints make it difficult or inappropriate to maximize the transmitted information. Consider the mammalian cortex, where a *population* of neurons within a cortical column represents some

sensory feature. Each neuron within the column transforms similar inputs from thousands of other neurons into a single output spike train; that is, the input entropy to each neuron exceeds the maximum possible output entropy.

A caricature of this situation illustrates how constraints might make SR-like effects relevant. Consider a small modification of the simple binary model analyzed above, where we now allow the input  $x$  to have higher entropy (e.g., by drawing it from a continuous distribution) than the output  $y$ ; here  $y_j$  is analogous to the output from the  $j$ th neuron in the column, and  $x = \sum_i u_i$  is the analog sum of the other input elements  $u_i$  (outside the column) connected to  $y_j$ . Under these conditions, the signal  $x$  can be represented using an ensemble of  $N$  binary elements, each receiving the same  $x$ ; the question is how  $I$ , the information about the input  $x$  represented by the population  $y$ , depends on  $N$ . Under the most efficient population coding strategy, wherein each element represents a binary "digit," the information  $I$  is directly proportional to the population size,  $I \propto N$ , but this optimum requires a sophisticated encoding-decoding scheme that might be awkward to implement using simple summation elements wherein we limit ourselves to decoding  $z = \sum_j y_j$ . A much simpler strategy involves simply adding noise to each input, so that  $I \propto \log N^{1/2} = \frac{1}{2} \log N$ ; this more ro-

bust approach satisfies only a constant factor of 2 in the information. Thus the role of SR may not be to achieve the globally optimal mutual information, which in this case would require an implausibly sophisticated decoder, but rather to make efficient use of the simplest architecture.

Most of the SR literature to date [1-3] has focused on time-sinusoidal signals in Gaussian noise backgrounds. While the output SNR may be an adequate measure for characterizing the response to such signals, it is not, in general, the most appropriate measure when considering more complicated signals of the type considered in this work. The "conventional" SR effect may also be (for a sinusoidal input signal) connected [15,16] with a synchronization between the characteristic deterministic (the input signal period) and stochastic (the Kramers rate for zero input signal) time scales associated with the system. Clearly, the above-described phenomena may be regarded as a generalization of the "conventional" SR effect for more complex signals.

A.Z. acknowledges support from the Howard Hughes Medical Institute to Charles F. Stevens, and A.R.B. was supported by the Office of Naval Research (Physics Division) and by NATO CRG No. 931464. We also acknowledge valuable discussions with W. Bialek (NEC) and M. DeWeese (Salk Inst.).

- 
- [1] See, e.g., *Proceedings of the NATO Advanced Research Workshop on Stochastic Resonance in Physics and Biology*, edited by F. Moss, A. Bulsara, and M. Shlesinger, a special issue of *J. Stat. Phys.* **70** (1993).
- [2] *Proceedings of the International Workshop on Fluctuations in Physics and Biology: Stochastic Resonance, Signal Processing, and Related Phenomena, Elba, Italy 1994*, edited by R. Mannella and P. V. E. McClintock, special issue of *Nuovo Cimento* **17D** (1995).
- [3] K. Wiesenfeld and F. Moss, *Nature* **373**, 33 (1995); A. Bulsara and L. Gammitoni, *Phys. Today* **49**(3), 39 (1996). A complete bibliography may also be found at <http://www.pg.infn.it/sr/>.
- [4] C. Shannon, *Proc. IRE* **37**, 10 (1949).
- [5] R. de Ruyter van Steveninck and W. Bialek, *Proc. R. Soc. London Ser. B* **234**, 379 (1988); W. Bialek, M. de Weese, F. Rieke, and D. Warland, *Physica A* **200**, 581 (1993).
- [6] M. deWeese, Ph.D. thesis, Princeton University, 1995 (unpublished); M. deWeese and W. Bialek, in Ref. [2].
- [7] J. Levin and J. Müller, *Nature* **380**, 165 (1996).
- [8] L. Gammitoni, *Phys. Rev. E* **52**, 4691 (1995); *Phys. Lett. A* **208**, 315 (1995).
- [9] J. Collins, C. Chow, and T. Imhoff, *Nature* **376**, 236 (1995); *Phys. Rev. E* **52**, R3321 (1995); C. Heneghan, J. Collins, C. Chow, T. Imhoff, S. Lowen, and M. Teich, this issue, *Phys. Rev. E* **54**, R2228 (1996); L. Kiss (unpublished).
- [10] H. Tuckwell, *Introduction to Theoretical Neurobiology* (Cambridge University Press, Cambridge, 1988); J. Hopfield, *Phys. Today* **47**(2), 40 (1994).
- [11] C. F. Stevens and A. Zador, in *Advances in Neural Information Processing Systems 8*, edited by M. Mozer, D. S. Touretzky, and M. Hasselmo (MIT Press, Cambridge, MA, in press).
- [12] M. Inghiosa and A. Bulsara, *Phys. Rev. E* **53**, R2021 (1996).
- [13] M. Stemmler, *Network* (to be published).
- [14] A. Neiman, B. Shulgin, V. Anischenko, W. Ebeling, L. Schimansky-Geier, and J. Freund, *Phys. Rev. Lett.* **23**, 4299 (1996).
- [15] L. Gammitoni, F. Marchesoni, and S. Santucci, *Phys. Rev. Lett.* **74**, 1052 (1995).
- [16] A. Bulsara, S. Lowen, and D. Rees, *Phys. Rev. E* **49**, 4989 (1994); A. Bulsara, T. Elston, C. Doering, S. Lowen, and K. Lindenberg, *ibid.* **53**, 3958 (1996).