# Information-Theoretic Stochastic Resonance in Noise-Floor Limited Systems: The Case for Adding Noise

M. E. Inchiosa*
*Space and Naval Warfare Systems Center, Code D364, San Diego, California 92152-5001*

J. W. C. Robinson†
*Defence Research Establishment, SE 172 90, Stockholm, Sweden*

A. R. Bulsara‡
*Space and Naval Warfare Systems Center, Code D364, San Diego, California 92152-5001*
(Received 3 April 2000)

We show that in systems whose output must compete with a noise source, stochastic resonance (maximization of output signal-noise separation as a nonmonotonic function of input noise strength) exists even when measured in terms of fundamental statistical measures and *optimal* detector performance. This is in contrast to the commonly considered scenario where, without the competing noise, the system (e.g., a driven, overdamped particle moving in a double well potential) is essentially invertible and optimal detector performance monotonically deteriorates with increasing input noise strength.

PACS numbers: 05.45.–a, 02.50.Ey, 05.40.Ca

Many archetypal stochastic nonlinear dynamical systems exhibit stochastic resonance (SR), whereby output signal power and certain output signal-noise separation indices, such as the signal-to-noise ratio, show a nonmonotonic behavior involving a local or global maximum as a function of input noise strength [1]. It has been speculated that the SR effect could be used to enhance information transfer in such systems.

The prototype SR system is an overdamped Brownian particle in a sinusoidally modulated bistable potential, a good first approximation to a number of noisy nonlinear dynamical systems. With deterministic and stochastic forcing, the response is *invertible* in the sense that the input, i.e., the total driving force, can be reconstructed from the output, i.e., the response. It follows immediately that *optimal* signal detection performance at the output of the system equals that at the input, because an optimal detector on the output could simply invert the system and apply optimal detection to the reconstructed input. Since adding Gaussian noise to the signal can only reduce the optimal detection performance at the input — and hence at the output — we would appear to have lost the SR effect by replacing the suboptimal detectors typically considered in SR studies (e.g., power at the signal frequency) with optimal detectors. In this Letter we show that the SR effect is present even within an optimal detection framework when one additional important physical consideration is retained in the model, viz., an *additional* noise source competing with the system's output.

Consider two signal processing blocks in series. The first block (the nonlinear "system") takes a noisy input $x(t)$, in which a target signal (e.g., a sinusoid) may be embedded, and outputs its response $y(t)$. The second block (the "detector") attempts to determine whether the target signal was present at the nonlinear system's input. A competing noise can be conceptualized in two ways. If adding this noise is viewed as the final operation performed in the first block, yielding a noise contaminated version $z$ of $y$, then we have replaced a possibly invertible system ($x \mapsto y$) with a noninvertible one ($x \mapsto z$). Alternatively, if adding competing noise is viewed as the first operation in the detector block, we have replaced a possibly optimal detector (acting on $y$) with a suboptimal — but more physical — detector (acting on $y$, but in a "randomized" manner). To see why including the competing or "measurement" noise makes the detector model more physical, suppose that for certain parameters the nonlinear system's mathematical model predicts a response with high signal-noise separation but arbitrarily low total energy. Without measurement noise, the optimal detection framework invokes the unphysical concept of processing the response with unwavering fidelity even as the total power in it goes to zero. In short, including a measurement noise floor means that two things, sufficient signal-noise separation *and* signal power, are required for good detector operation. This aspect of the basic SR setting seems to have been largely overlooked. An intimately related issue is whether or not one can improve the performance of a given *sub*optimal detector by adding noise. This is indeed possible and has been demonstrated in several SR settings [2] and, recently, in a low-complexity detector setting [3].

We treat the problem of maximization of signal-noise separation in noise-floor limited systems from an information-theoretic (or statistical inference) viewpoint. We show that, in otherwise ideal or optimal systems with measurement noise (manifesting itself, for example, as a noise floor at the optimal detector's input) of at least moderate magnitude, local or global maxima with respect to a naturally present or added (and controllable) input noise strength can occur, even in fundamental measures such as those mentioned above. This is demonstrated for both a simple static nonlinearity and a nonlinear dynamical system, with the former serving as a "cartoon" example in which the underlying mechanisms can be clearly uncovered and understood and the dynamical example showing the generality of the results and their relevance to SR.

In our investigation, a central role will be played by a class of quantities closely related to fundamental limits in information theory and statistics, the Csizsar-Ali-Silvey $\phi$ divergences [4,5]. If $\lambda$ is a measure defined on some space $\mathcal{U}$ (e.g., Lebesgue measure on $\mathbb{R}$), and $p_0$, $p_1$ are two probability densities with respect to $\lambda$, the *$\phi$ divergence* $d_\phi(p_0, p_1)$ between $p_0$, $p_1$ is defined as $d_\phi(p_0, p_1) = \int_{\mathcal{U}} \phi(\frac{p_1}{p_0}) p_0 \, d\lambda$, where $\phi$ is a continuous convex function on $[0, \infty)$. A $\phi$ divergence is thus a functional of the *likelihood ratio* $p_1/p_0$, the central quantity in all the optimal solutions to the classical inference problems in statistics. In particular, if $H_i$ $(i = 0, 1)$ is the hypothesis that $p_i$ is the actual, currently "active" density on $\mathcal{U}$, then the optimal Bayesian test (i.e., detector) for deciding between $H_0$ and $H_1$ based on a sample $u \in \mathcal{U}$ is [6]

$$\text{decide } H_0 \text{ if } \frac{p_1(u)}{p_0(u)} \leq \gamma, \quad \text{else decide } H_1, \qquad (1)$$

where $\gamma = \alpha/(1 - \alpha)$ and $\alpha, 1 - \alpha \in (0, 1)$ are the *a priori* probabilities for $H_0$ and $H_1$, respectively, to be correct. For this decision problem, the Kolmogorov or error divergence $d_\varepsilon^{(\alpha)}(p_0, p_1)$, obtained for $\phi(v) = |(1 - \alpha)v - \alpha|$, is intimately related to the *minimal probability of error* $\tilde{P}_e^{(\alpha)}(p_0, p_1)$ [realized by (1)] through [7]

$$\tilde{P}_e^{(\alpha)}(p_0, p_1) = \frac{1}{2}[1 - d_\varepsilon^{(\alpha)}(p_0, p_1)]. \qquad (2)$$

A $\phi$ divergence can more generally be thought of as an index of *separation* between the probability densities $p_0$ and $p_1$. In particular, the value of $d_\varepsilon^{(\alpha)}(p_0, p_1)$ varies as $|1 - 2\alpha| \leq d_\varepsilon^{(\alpha)}(p_0, p_1) \leq 1$ with equality at the extreme cases for $p_1/p_0 = 1$ with probability one under $H_0$ ("equal" densities) and $p_1/p_0 = 0$ with probability one under $H_0$ (the densities "live" on disjoint parts of $\mathcal{U}$), respectively. The quantity $d_\varepsilon^{(\alpha)}(p_0, p_1)$ can therefore be regarded as a very basic signal-noise separation index in situations where the problem is to detect a signal's presence on the input or output of a stochastic nonlinear dynamical system. For other objectives, other $\phi$ divergences can be suitable as signal-noise separation indices [4].

The underlying mechanisms of noise-dependent optimization in noise-floor limited systems are easily understood by considering a simple example in the form of a static system with a continuous and invertible transfer characteristic (TC) $g$ of the piecewise-linear type, with $g(0) = 0$, $g'(x) = 1$ for $|x| \leq 3$ and $g'(x) = 10$ for $|x| > 3$. The input is represented by either one of two Gaussian probability densities $p_0, p_1$ having a common standard deviation $\sigma_i$ and means $\mu_0 = 0$ and $\mu_1 = 2$, respectively. The densities $p_0, p_1$ symbolize the noise-only ($H_0$) and the noise-plus-signal cases ($H_1$), respectively. The nonlinear TC transforms the densities $p_0$, $p_1$ into two corresponding pristine (P) output densities $q_0$, $q_1$ (Fig. 1).

Then, an additional independent zero-mean Gaussian measurement noise with probability density $h$ and standard deviation $\sigma_m$ is added onto the P output, producing the noise-contaminated (NC) output. The two corresponding probability densities of the NC output are $r_0 = q_0 * h$ and $r_1 = q_1 * h$ ($*$ denotes convolution), respectively (Fig. 2). From the graphs it appears that the separation between the P densities $q_0, q_1$ decreases monotonically with $\sigma_i$, and this is indeed true when measured in terms of $\phi$ divergences [since $g$ is invertible $d_\phi(p_0, p_1) = d_\phi(q_0, q_1)$ and $d_\phi(p_0, p_1)$ decreases monotonically; see, e.g., [7] ]. However, the separation between $r_0$ and $r_1$ *does not* appear to decrease monotonically because the right-hand "tail" of the density $r_1$ for $\sigma_i = 2$ is very long and "heavy" compared to that of $r_0$, whereas $r_0$ and $r_1$ are qualitatively very similar everywhere both for $\sigma_i = 1$ and $\sigma_i = 4$. Indeed, computing $\phi$ divergences verifies this observation.

In Fig. 3 the Kolmogorov divergence $d_\varepsilon^{(\alpha)} = d_\varepsilon^{(\alpha)}(r_0, r_1)$ is plotted as a function of input noise strength $\sigma_i$ and *a priori* probability $\alpha$ for $\sigma_m = 4$. The corresponding behavior of $\tilde{P}_e^{(\alpha)}(r_0, r_1)$ can then straightforwardly be obtained from (2) (qualitatively by "flipping" the surface in Fig. 3 upside down). For medium values of $\alpha$ a clear maximum of $d_\varepsilon^{(\alpha)}$ as a function of $\sigma_i$ emerges near the value $\sigma_i = 2.5$. The maximum occurs for a large range of values of $\sigma_m$, as illustrated in Fig. 4 where $d_\varepsilon^{(\alpha)}$ is plotted versus $\sigma_i$ and $\sigma_m$ for $\alpha = 0.5$. Intuitively, convolution with the measurement noise density $h$ "smears out" the densities $q_0, q_1$ and thus more actively involves

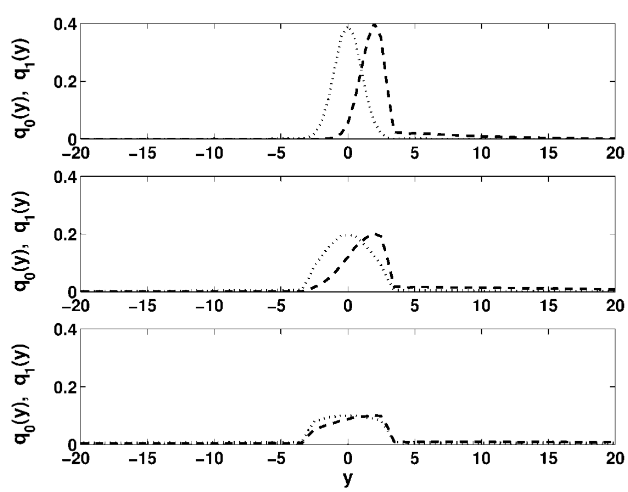


FIG. 1. Resulting probability densities $q_0$ (dotted curve) and $q_1$ (dashed curve) on the output of $g$ before addition of measurement noise, for $\sigma_i = 1$ (top), $\sigma_i = 2$ (middle), and $\sigma_i = 4$ (bottom). The separation between $q_0$ and $q_1$ decreases monotonically.

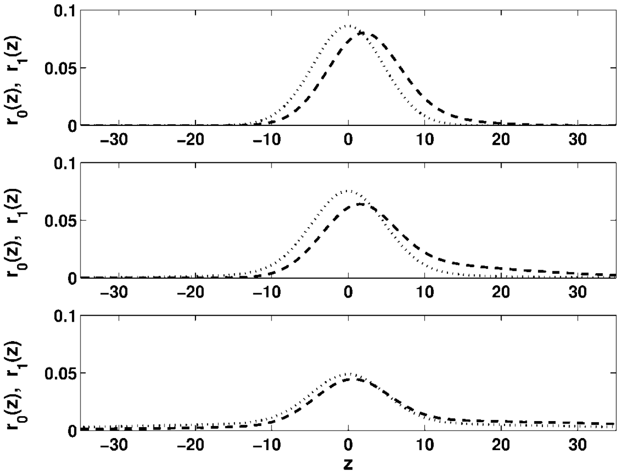


FIG. 2. Output densities $r_0$ (dotted curve) and $r_1$ (dashed curve) after addition of measurement noise with $\sigma_m = 4$. From top to bottom: $\sigma_i = 1, 2, 4$. The separation does not decrease monotonically: the long, heavy right-hand tail of $r_1$ makes the separation between $r_0$ and $r_1$ be the largest for $\sigma_i = 2$.

the high-gain flanks of the TC, enhancing the difference in the tail behavior of $q_0, q_1$ and overcoming their similarity near the origin. Several other $\phi$ divergences (not shown) consistently reflect this separation enhancement of the NC output densities.

The peaks shown in Figs. 3 and 4 are modest; however, the phenomenon becomes much more pronounced in a dynamical system exemplified by the overdamped Duffing oscillator:

$$\dot{y}(t) = f(y) + A_0 \sin(\omega_0 t + \theta) + \sqrt{2D_i}\, \xi_i(t), \quad (3)$$

where $D_i > 0$, $\xi_i(t)$ is white Gaussian noise with mean zero and autocorrelation $\langle \xi_i(t)\xi_i(t + \tau)\rangle = \delta(\tau)$, $\theta$ is a constant randomly drawn from the uniform distribution on $[0, 2\pi)$, and $f(y) = y - y^3$. Here, if we regard $x(t) = A_0 \sin(\omega_0 t + \theta) + \sqrt{2D_i}\, \xi_i(t)$ as the input and $y(t)$ as the P output, all $\phi$ divergences based on the information in trajectories over $[0, T]$ are preserved between input and output since the system is invertible [8]; hence, the minimal probability of error in detection ($A_0 = 0$ vs $A_0 = \text{const} > 0$) is the same on the input as on the P output. However, if a measurement noise is added to the output, yielding an NC output $z(t)$ according to, e.g.,

$$z(t) = y(t) + \sqrt{2D_m}\, \xi_m(t), \quad (4)$$

where $D_m > 0$ and $\xi_m(t)$ is another zero-mean white Gaussian noise source with autocorrelation $\langle \xi_m(t)\xi_m(t + \tau)\rangle = \delta(\tau)$, the situation changes completely and SR within the optimal detection framework can occur. It is in general difficult to compute the NC output divergences and corresponding minimal probability of error in detection $\tilde{P}_e^{(\alpha)}$ for observations based on a trajectory, so we will consider bounds for these quantities instead.

In Figs. 5 and 6 we have computed, respectively, an upper bound $P_e^U$ for $\tilde{P}_e^{(\alpha)}$ and a lower bound $d_\varepsilon^L$ for the Kolmogorov divergence $d_\varepsilon^{(\alpha)}$ for the NC output of the system (3), (4) with $A_0 = 0.086$, $\omega_0 = 0.1$, and $T = 2010.62$ (which corresponds to 32 periods of the sinusoid). The bound $P_e^U$ was obtained by using Monte Carlo simulation to calculate, for each value of $D_i$, $D_m$, and $\alpha$, the error probability $P_e^U$ of the optimal detector of a harmonic signal of frequency $\omega_0$ and unknown phase (the *incoherent detection* problem [6]) in the *linearized* response of the system output. The error probabilities of this suboptimal detector clearly satisfy $P_e^U \geq \tilde{P}_e^{(\alpha)}$. We define $d_\varepsilon^L \equiv (1 - 2P_e^U)$, which is easily seen to give a lower bound for $d_\varepsilon^{(\alpha)}$ since by (2) and the definition of $d_\varepsilon^L$ we have $\frac{(1-d_\varepsilon^L)}{2} = P_e^U \geq \tilde{P}_e^{(\alpha)} = \frac{(1-d_\varepsilon^{(\alpha)})}{2}$, and it follows that $d_\varepsilon^{(\alpha)} \geq d_\varepsilon^L$.

Fortunately, for certain parameter values these bounds are tight. Since $A_0$ is small and we have used the detector

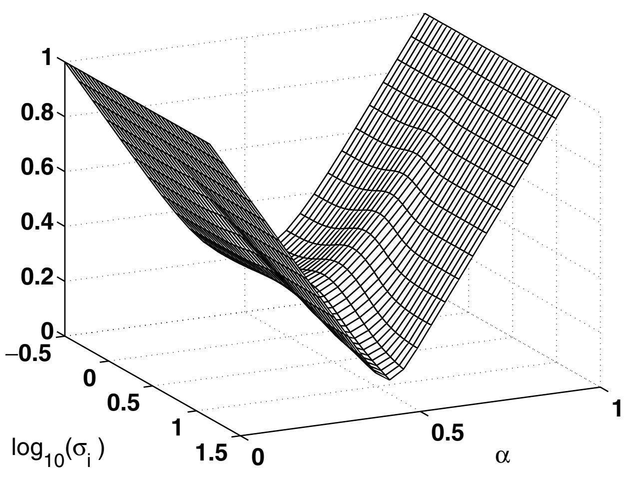


FIG. 3. Kolmogorov divergence $d_\varepsilon^{(\alpha)}$ for the output of the static nonlinearity with added measurement noise as a function of $\sigma_i$ and $\alpha$, for $\sigma_m = 4$.

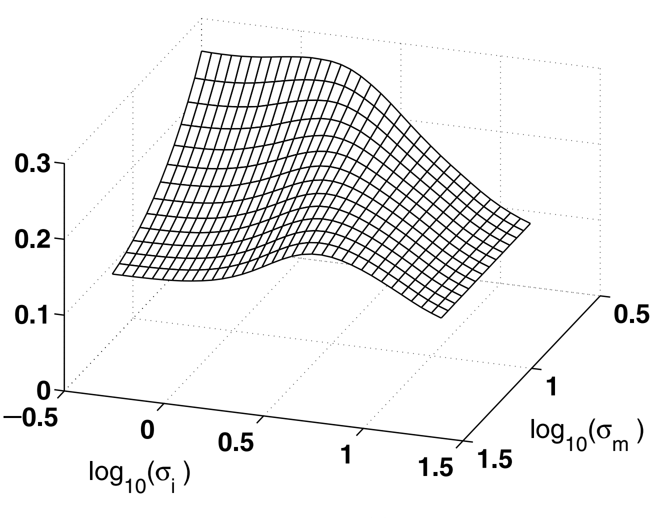


FIG. 4. Kolmogorov divergence $d_\varepsilon^{(\alpha)}$ for the output with measurement noise added as a function of $\sigma_i$ and $\sigma_m$, for $\alpha = 0.5$.

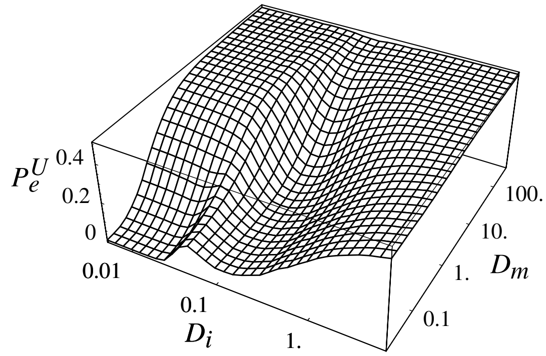


FIG. 5. The upper bound $P_e^U$ for $\tilde{P}_e^{(\alpha)}$ based on the measurement-noise contaminated output $z(t)$ in (4) of the double well (3) as a function of $D_i$ and $D_m$, for $\alpha = 0.5$.

given by linear theory, for small $D_i$ (linear response) the performance of the optimal detector for the linear case is in fact essentially optimal for the nonlinear system as well, and the values $\tilde{P}_e^{(\alpha)}$ and $d_\varepsilon^{(\alpha)}$ are essentially achieved. Also, in the high input noise limit $D_i \to \infty$, $d_\varepsilon^{(\alpha)}$ goes to its minimum possible value of $|1 - 2\alpha|$ and, for all detectors, the probability of error tends to $\tilde{P}_e^{(\alpha)} = (1 - |1 - 2\alpha|)/2$. Hence, even though the plot of $P_e^U$ provides only a bound, there must, at least for each fixed value of $D_m$ in the order of 1–10, be a pronounced minimum in $\tilde{P}_e^{(\alpha)}$ as a function of $D_i$ at some optimal $\tilde{D}_i \in (0, \infty)$. A similar argument applies to $d_\varepsilon^L$.

Thus, in a very simple setting we have shown an SR effect in fundamental statistical measures for two different types of nonlinear systems (static and dynamic) with additive noise superimposed on the output to represent the effects of a measurement noise-floor limitation. The two systems share a common feature: nonlinear interaction between the signal and noise, which, loosely speaking, provides a noise-controlled nonlinear gain effect. The enhanced effect observed in the dynamical system can perhaps be attributed to a larger degree of nonlinearity away from the origin (in function space) and the fact that the infinite dimensional space that the probability densities live in allows for more "room" for the noise only and signal-plus-noise densities to differ. The underlying premise of this Letter is, however, that even the *optimal* detection performance (based on the measurement-noise contaminated output) can be enhanced via a generalized version of the SR effect, using controlled added input noise. In the absence of the noise floor, of course, one expects the information-theoretic distances to be monotonically decreasing functions of the input noise since the systems under consideration are invertible in this case. Since most practical detection or quantification systems have measurement and input noise, our results suggest wide applicability of SR as a tool for improving detection performance by adding noise in these systems.

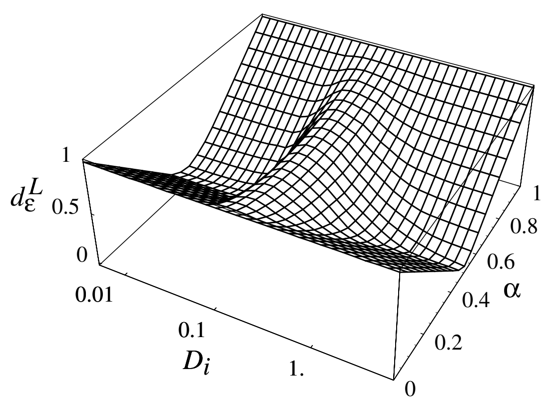


FIG. 6. The lower bound $d_\varepsilon^L$ for $d_\varepsilon^{(\alpha)}$ based on the measurement-noise contaminated output $z(t)$ of the double well as a function of $D_i$ and $\alpha$, for $D_m = 3.266$.

A. R. B. and M. E. I. were supported by the Office of Naval Research (Grant No. N00014-00-WX-20481). J. W. C. R. acknowledges support from FOA Project No. E6022.

*Email address: inchiosa@spawar.navy.mil
†Email address: john@sto.foa.se
‡Email address: bulsara@spawar.navy.mil

[1] For good overviews, see K. Wiesenfeld and F. Moss, Nature (London) **373**, 33 (1995); A. Bulsara and L. Gammaitoni, Phys. Today **49**, No. 3, 39 (1996); L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, Rev. Mod. Phys. **70**, 223 (1998).
[2] M. E. Inchiosa and A. R. Bulsara, Phys. Rev. E **53**, R2021 (1996); **58**, 115 (1998); M. E. Inchiosa, A. R. Bulsara, J. F. Lindner, B. K. Meadows, and W. L. Ditto, in *Chaotic, Fractal, and Nonlinear Signal Processing,* edited by R. A. Katz, AIP Conf. Proc. No. 375 (AIP, New York, 1996).
[3] S. Kay, IEEE Signal Process. Lett. **7**, 8 (2000).
[4] J. Rung and J. W. C. Robinson, in *STOCHAOS: Stochastic and Chaotic Dynamics in the Lakes,* edited by D. S. Broomhead, E. A. Luchinskaya, P. V. E. McClintock, and T. Mullin (AIP, Melville, NY, 2000).
[5] F. Liese and I. Vajda, *Convex Statistical Distances* (Teubner, Leipzig, 1987).
[6] H. V. Poor, *An Introduction to Signal Detection and Estimation* (Springer-Verlag, New York, 1994).
[7] J. W. C. Robinson, D. E. Asraf, A. R. Bulsara, and M. E. Inchiosa, Phys. Rev. Lett. **81**, 2850 (1998).
[8] To guarantee invertibility, and rigorous definition of the SDE (3), the potential must be modified so that $|f(y)|$ grows no faster than linearly asymptotically as $|y| \to \infty$.