Stochastic resonance in noisy spiking retinal and sensory neuron models☆

Ashok Patel, Bart Kosko*

Department of Electrical Engineering, Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089-2564, USA

Abstract

Two new theorems show that small amounts of additive white noise can improve the bit count or mutual information of several popular models of spiking retinal neurons and spiking sensory neurons. The first theorem gives necessary and sufficient conditions for this noise benefit or stochastic resonance (SR) effect for subthreshold signals in a standard family of Poisson spiking models of retinal neurons. The result holds for all types of finite-variance noise and for all types of infinite-variance stable noise: SR occurs if and only if a sum of noise means or location parameters falls outside a 'forbidden interval' of values. The second theorem gives a similar forbidden-interval sufficient condition for the SR effect for several types of spiking sensory neurons that include the Fitzhugh-Nagumo neuron, the leaky integrate-and-fire neuron, and the reduced Type I neuron model if the additive noise is Gaussian white noise. Simulations show that neither the forbidden-interval condition nor Gaussianity is necessary for the SR effect.
© 2005 Elsevier Ltd. All rights reserved.

1. Stochastic resonance in spiking retinal and sensory neurons

Noise can help a nonlinear system as well as hurt it. The formal name for such a noise benefit is stochastic resonance (SR) (Bulsara and Zador, 1996; Deco and Schurmann, 1998; Gammaitoni, 1995; Godivier and Chapeau-Blondeau, 1998; Inchiosa et al., 2000; Jung, 1995; Kosko and Mitaim, 2001; Mitaim and Kosko, 1998; Wiesenfeld and Moss, 1995). Fig. 1 shows an SR noise benefit in a spiking retinal neuron. The neuron should emit more spikes when the brightness contrast level is low rather than high. The right amount of Gaussian noise helps the neuron discriminate between two levels of brightness contrast. The retinal neuron emits too few spikes if no noise corrupts the Bernoulli sequence of contrast levels. The neuron also emits too many spikes and emits many of them at the wrong time if too much noise corrupts the sequence.

The next section presents the first of two new SR theorems for spiking neurons. This first theorem gives necessary and sufficient conditions for an SR effect in standard models of spiking retinal neurons for almost all types of additive white noise. The last section presents a theorem that gives a sufficient condition for an SR noise benefit in standard models of spiking sensory neurons for Gaussian white noise. The converse also holds for the leaky integrate-and-fire neuron but need not hold for other spiking sensory neurons as simulations confirm. Lack of a converse broadens rather than lessens the potential scope of SR in spiking sensory neurons. Simulations also show that the SR effect can persist for other types of finite-variance and infinite-variance noise.

These new theorems extend our earlier results on SR in threshold neurons for subthreshold signals (Kosko and Mitaim, 2003; Kosko and Mitaim, 2004). These results characterize SR in terms of a technical condition on the noise mean or location parameter when mutual information measures system performance of a simple memoryless threshold neuron in the face of noisy Bernoulli input sequences: SR occurs if and only if the noise mean $E(n)$ does not lie in the 'forbidden interval' $(\theta - A, \theta + A)$ where $-A < A < \theta$ for threshold θ and signal amplitude $A > 0$. The sufficient or if-part of the theorem first appeared in (Kosko and Mitaim, 2003) while the converse only-if part first appeared in (Kosko and Mitaim, 2004). The result holds for all noise types that have finite variance and for all infinite-variance noise types from the broad family of stable distributions (Shao and Nikias, 1993). The proof technique assumes that the nonnegative mutual information is positive
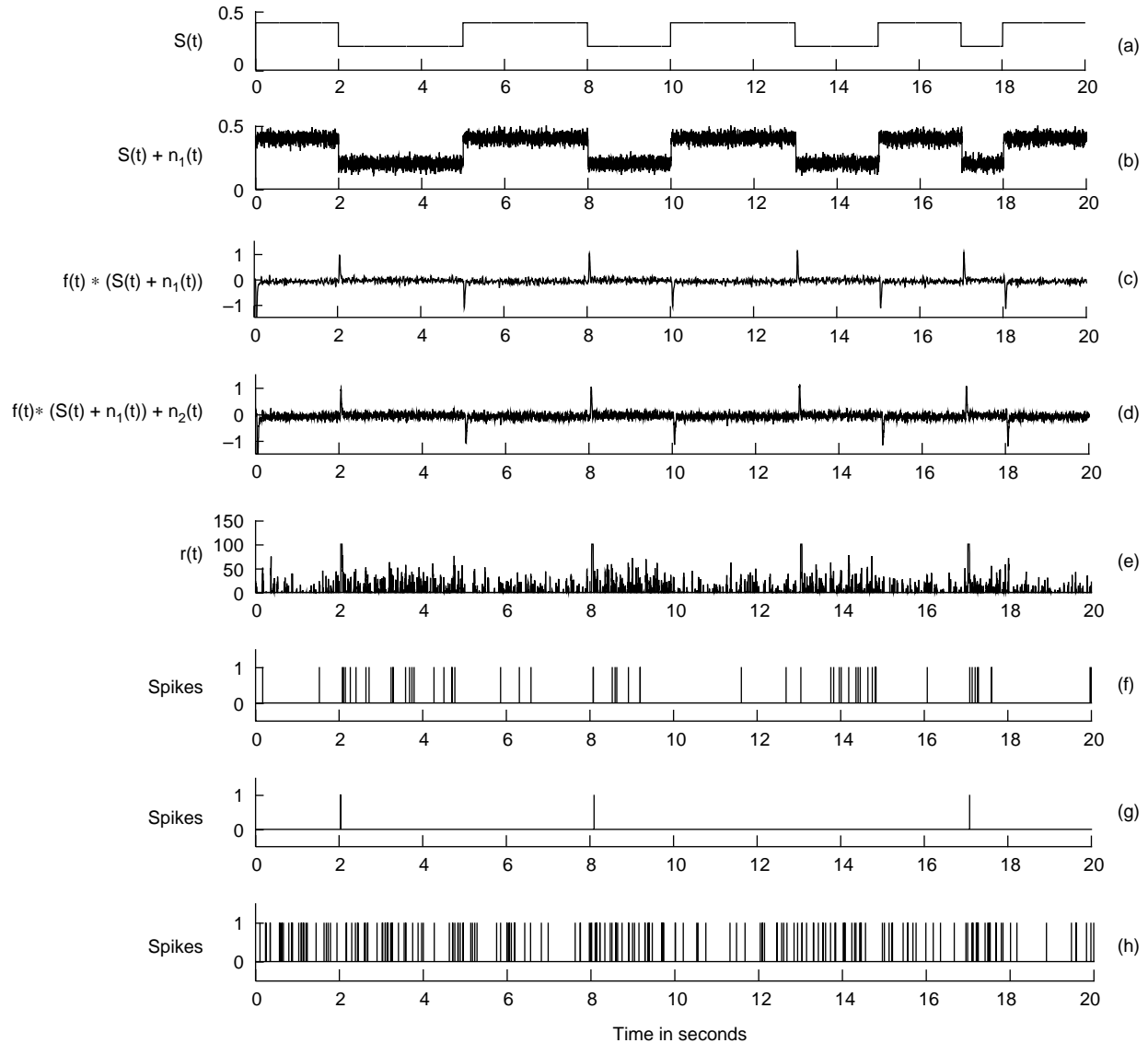
Fig. 1. Stochastic resonance in a spiking retinal neuron. The neuron should emit more spikes when the brightness contrast level is low rather than high. Noise improves the discrimination of subthreshold contrast stimuli in the retina model (1)–(3). (a) Bernoulli contrast signal S as a function of time t. (b) Contrast signal S plus Gaussian white noise n_1 with variance $\sigma_1^2 = 0.03^2$. (c) Signal in plot (b) filtered with f in (1). (d) Filtered noisy signal in (c) plus noise n_2 (synaptic and ion-channel noise) with variance $\sigma_2^2 = 0.06^2$. (e) Noisy spike rate $r(t)$. (f) SR effect: Output Poisson spikes that result from the noisy spike rate $r(t)$. (g) Output spikes in the absence of noise. (h) Output spikes in the presence of too much noise.

and then shows that it goes to zero as the noise variance or dispersion goes to zero-so the mutual information must increase as the noise dispersion increases from zero.

We now extend the above memoryless SR theorem to the more complex models of retinal and sensory neurons that produce spike trains. We prove that a general retinal model with two noise sources and a piecewise-linear sigmoidal function exhibits SR if and only if the sum of the two noise means does not lie in the forbidden interval $(\theta_1 - v_1, \theta_2 - v_2)$ that depends on the threshold values θ_1 and θ_1 and on the subthreshold signal values v_1 and v_2. The only-if part holds in the sense that the system performs better without noise than with it when the interval condition fails. We then show that the SR effect holds for a general family of nonlinear

sensory neural models if the additive noise is Gaussian white noise. These models include the popular FitzHugh-Nagumo (FHN) model (Chialvo et al., 1997; Collins et al., 1995) and the integrate-and-fire model (Collins et al., 1996; Gerstner and Kistler, 2002), and the reduced Type I neuron model (Lindner et al., 2003).

2. Stochastic resonance in spiking retinal models

Theorem 1 below characterizes SR in spiking retinal models. It states that standard spiking retinal models benefit from additive white noise if and only if a joint noise mean or location parameter does not fall in a forbidden interval of

threshold-based values. Theorem 1 holds for all finite-variance noise and for all impulsive or infinite-variance stable noise (Grigoriu, 1995; Kosko and Mitaim, 2001; Shao and Nikias, 1993). The performance measure is the input-output Shannon mutual-information (Cover and Thomas, 1991) bit count $I(S, R) = H(R) - H(R|S)$ for input signal random variable S and output response random variable R. Fig. 1 shows a simulation instance of Theorem 1 for Gaussian white noise that corrupts a random Bernoulli sequence of brightness contrast levels in a Poisson-spiking retinal neuron.

The retina model of Theorem 1 is a noisy version of a common Wiener-type cascade model (Chander and Chichilnisky, 2001; Kim and Rieke, 2001; Korenberg and Hunter, 1986; Rieke et al., 1996; Sakai et al., 1995):

$$r(t) = r_0 h \left[\int_{-\infty}^{\infty} f(z)\{S(t-z) + n_1(t)\}dz + n_2(t) \right] \quad (1)$$

where S is the input stimulus defined below, r is the instantaneous Poisson spike rate that gives the exponential interspike-interval density function as $p(t) = r(t)\exp\left[-\int_0^t r(\tau)d\tau\right]$, f is a band-pass linear filter function, and h is a memoryless monotone-nondecreasing function. Here n_1 denotes the combined stimulus and photoreceptor noise (AlaLaurila et al., 2004; Lamb, 1987; Rieke et al., 1996) and n_2 denotes the combined ion-channel noise (Schneidman et al., 1998; Van Rossum et al., 2003) and the synaptic noise (Freed, 2000; Levy and Baxter, 2002; Manwani and Koch, 1999).

The input stimulus S is Michelson's visual contrast signal (Burkhardt et al., 1984): $S = (L_c - L_s)/(L_c + L_s)$. L_c is the amount of light that falls on the center of the ganglion cell's receptive field. L_s is the light that falls on its surround region.

The sigmoid-shaped memoryless function h approximates the spike threshold and saturation level. We define h as a piecewise-linear approximation of a sigmoidal nonlinearity (Yu et al., 2005):

$$h(x) = \begin{cases} \theta_2 - \theta_1 & \text{if } x > \theta_2 \\ x - \theta_1 & \text{if } \theta_1 \leq x \leq \theta_2 \\ 0 & \text{if } x < \theta_1 \end{cases} \quad (2)$$

and so

$$r(w(t)) = \begin{cases} r_0(\theta_2 - \theta_1) & \text{if } w(t) > \theta_2 \\ r_0(w(t) - \theta_1) & \text{if } \theta_1 \leq w(t) \leq \theta_2 . \\ 0 & \text{if } w(t) < \theta_1 \end{cases} \quad (3)$$

The Shannon mutual information $I(S, R)$ between the input contrast signal S and the output average spiking rate r measures the neuron's bit count and allows us to detect the noise enhancement or SR effect.

The subthreshold contrast signal $S(t) \in \{A, B\}$ is a random Bernoulli sequence with $P[S(t) = A] = p$ and

$P[S(t) = B] = 1 - p$. The time duration of each signal value A and B in $S(t)$ is much larger than the time constant of the linear filter $f(t)$. We define $v(t)$ as the filtered output of the contrast signal $S(t)$ without noise $n_1(t)$ and such that

$$v(t)|_{S(t)=A} = v_1 \quad (4)$$

and

$$v(t)|_{S(t)=B} = v_2 \quad (5)$$

in steady-state, where $v_1 > v_2$ and $\max(v_1, v_2) < \theta_1 < \theta_2$. So the input signal $S(t)$ is subthreshold. We measure the average spike rate for each symbol only when the corresponding value of $v(t)$ is in steady-state. Then the filtered noise z is $z(t) = f(t)*n_1$ where '*' denotes convolution.

Theorem 1 below gives necessary and sufficient conditions for an SR noise effect in the retina neuron model (1)–(3) for either noise source n_1 or n_2. The theorem shows that some increase in such noise *must* increase the neuron's mutual information $I(S, R)$-and thus must increase the neuron's ability to discriminate subthreshold contrast signals-if the noise mean or location parameter obeys a simple interval constraint. This SR effect holds for all finite-variance probability density functions. The result is robust because it further holds for all infinite-variance stable noise densities such as impulsive Cauchy or Levy noise (Grigoriu, 1995; Kosko and Mitaim, 2001; Kosko and Mitaim, 2003; Kosko and Mitaim, 2004) and the uncountably many other stable densities that obey a generalized central limit theorem (Shao and Nikias, 1993). The proof follows the technique of (Kosko and Mitaim, 2003; Kosko and Mitaim, 2004).

Theorem 1. *Suppose that the noise sources n_1 and n_2 in the retina model (1)–(3) are white and have finite-variance (or finite-dispersion in the stable case) probability density functions $p_1(n)$ and $p_2(n)$ with corresponding variances (dispersions) σ_1^2 and σ_2^2 (γ_1 and γ_2). Suppose that the input signal S is subthreshold ($v_2 < v_1 < \theta_1 < \theta_2$) and that there is some statistical dependence between the input contrast random variable S and the output random variable R so that $I(S, R) > 0$. Then the retina model (1)–(3) exhibits the nonmonotone SR effect in the sense that $I(S, R) \to 0$ as σ_1^2 and σ_2^2 (or γ_1 and γ_2) decrease to zero if and only if the mean sum $E(n_1) \times \int f(\tau)d\tau + E(n_2)$ (or the location parameter sum in the stable case) does not lie in the interval $(\theta_1 - v_1, \theta_2 - v_2)$. The only-if part holds in the sense that the system performs better without noise than with it when the interval condition fails.*

Proof. Assume $0 < P_S(s) < 1$ to avoid triviality when $P_S(s) = 0$ or 1.

A. If-part

We show that S and R are asymptotically independent: $I(\sigma_1, \sigma_2)$ as $\sigma_1 \to 0$ and $\sigma_2 \to 0$. This is equivalent to $I(\sigma) \to 0$ as $\sigma \to 0$ where σ is the variance of the total noise $n = z + n_2$. Independence of n_1 and n_2 implies that z and n_2 are

independent and hence $\sigma^2 = \mathrm{Var}(z) + \sigma_2^2$, where $\mathrm{Var}(z) = \sigma_1^2 \int f^2(t)\mathrm{d}t$. Recall that $I(S, R) = 0$ if and only if S and R are statistically independent (Cover and Thomas, 1991). So we need to show only that $f_{SR}(s, r) = P_S(s)f_R(r)$ or $f_{R|S}(r|s) = f_R(r)$ as $\sigma \to 0$ for signal symbols $s \in \{A, B\}$ and $r \in [0, r_0(\theta_2 - \theta_1)]$ where f_{SR} is a joint probability density function and $f_{S|R}$ is a conditional density function. This is equivalent to $F_{R|S} = F_R$ as $\sigma \to 0$ where $F_{R|S}$ is the conditional distribution function (Durrett, 1996). The well-known theorem on total probability and the two-symbol alphabet set $\{A, B\}$ give

$$F_R(r) = \sum_s F_{R|S}(r|s)P_S(s)$$
$$= F_{R|S}(r|A)P_S(A) + F_{R|S}(r|B)P_S(B)$$
$$= F_{R|S}(r|A)P_S(A) + F_{R|S}(r|B)(1 - P_S(A))$$
$$= (F_{R|S}(r|A) - F_{R|S}(r|B))P_S(A) + F_{R|S}(r|B)$$

So we need to show that $F_{R|S}(r|A) - F_{R|S}(r|B) \to 0$ as $\sigma \to 0$ for all r in the closed interval $[0, r_0(\theta_2 - \theta_1)]$. This condition implies that $F_R(r) = F_{R|S}(r|B)$ and $F_R(r) = F_{R|S}(r|A)$. Note that $F_{R|S}(r|A) = F_{R|S}(r|B) = 1$ for $r = r_0(\theta_2 - \theta_1)$ because $r_0(\theta_2 - \theta_1)$ is the maximum firing rate. So we need to show only that $F_{R|S}(r|A) - F_{R|S}(r|B) \to 0$ as $\sigma \to 0$ for all r in the half-open interval $[0, r_0(\theta_2 - \theta_1))$.

Consider $s = A$: Then (3) implies that

$$F_{R|S}(r|A) = \mathrm{Pr}\{r_0 h(v + n) \le r\}|_{s=A}$$
$$= \mathrm{Pr}\{r_0 h(v_1 + n) \le r\} \text{ by (4)}$$
$$= \mathrm{Pr}\{h(v_1 + n) \le r/r_0\} \text{ because } r_0 > 0$$
$$= \mathrm{Pr}\{v_1 + n \le \sup[h^{-1}(r/r_0)]\}$$
$$\qquad \text{because } h \text{ is monotonic nondecreasing}$$
$$= \mathrm{Pr}\{n \le \sup[h^{-1}(r/r_0)] - v_1\}$$
$$= \int_{-\infty}^{\sup[h^{-1}(r/r_0)]-v_1} p(n)\mathrm{d}n$$

where $p(n)$ is the probability density function of the total noise $z + n_2$. A symmetric argument shows that

$$F_{R|S}(r|B) = \int_{-\infty}^{\sup[h^{-1}(r/r_0)]-v_2} p(n)\mathrm{d}n$$

So we need to show that

$$\int_{-\infty}^{\sup[h^{-1}(r/r_0)]-v_2} p(n)\mathrm{d}n - \int_{-\infty}^{\sup[h^{-1}(r/r_0)]-v_1} p(n)\mathrm{d}n$$
$$= \int_{\sup[h^{-1}(r/r_0)]-v_2}^{\sup[h^{-1}(r/r_0)]-v_1} p(n)\mathrm{d}n \to 0 \text{ as } \sigma \to 0$$

But Eq. (2) implies that $\theta_1 \le \sup[h^{-1}(r/r_0)] < \theta_2$. So

$$\int_{\sup[h^{-1}(r/r_0)]-v_1}^{\sup[h^{-1}(r/r_0)]-v_2} p(n) \, \mathrm{d}n \le \int_{\theta_1-v_1}^{\theta_2-v_2} p(n)\mathrm{d}n$$

and so it is enough to show that

$$\int_{\theta_1-v_1}^{\theta_2-v_2} p(n)\mathrm{d}n \to 0 \text{ as } \sigma \to 0$$

We first consider the case of *finite* variance noise. Let the mean of the total noise $n = z + n_2$ be $m = E(z) + E(n_2)$. Suppose that $m < \theta_1 - v_1$ since $m \notin (\theta_1 - v_1, \theta_2 + v_2)$ where for convenience only we ignore the measure-zero case of $m = \theta_1 - v_1$. Pick $\varepsilon = (1/2)(\theta_1 - v_1 - m)$. So $\theta_1 - v_1 - \varepsilon = \theta_1 - v_1 - \varepsilon + m - m = m + (\theta_1 - v_1 - m) - \varepsilon = m + 2\varepsilon - \varepsilon = m + \varepsilon$. Then

$$F_{R|S}(r|A) - F_{R|S}(r|B)$$
$$= \int_{\theta_1-v_1}^{\theta_2-v_2} p(n)\mathrm{d}n$$
$$\le \int_{\theta_1-v_1}^{\infty} p(n)\mathrm{d}n$$
$$\le \int_{\theta_1-v_1-\varepsilon}^{\infty} p(n)\mathrm{d}n = \mathrm{Pr}\{n \ge m + \varepsilon\}$$
$$= \mathrm{Pr}\{n - m \ge \varepsilon\} \le \mathrm{Pr}\{|n - m| \ge \varepsilon\}$$
$$\le \frac{\sigma^2}{\varepsilon^2} \text{ by Chebychev's inequality}$$
$$\to 0 \text{ as } \sigma_1 \to 0 \text{ and } \sigma_2 \to 0$$

because $\sigma^2 = \sigma_1^2 \int f^2(t)\mathrm{d}t + \sigma_2^2$. A symmetric argument shows that for $m > \theta_2 - v_2$

$$F_{R|S}(r|A) - F_{R|S}(r|B) \le \frac{\sigma^2}{\varepsilon^2} \to 0$$

as $\sigma_1 \to 0$ and $\sigma_2 \to 0$.

We next consider the case of *infinite* variance noise. Note that if n_1 and n_2 are alpha-stable noise then $z = n_1 * f$ and $z + n_2$ are also alpha-stable noise (Grigoriu, 1995). Let m be the location parameter of the total alpha-stable noise $n = z + n_2$. The characteristic function $\phi(\omega)$ of alpha-stable noise density $p(n)$ reduces to a simple exponential in the zero dispersion limit (Kosko and Mitaim, 2003):

$$\lim_{\gamma \to 0} \phi(\omega) = \exp\{im\omega\}$$

for *all* α's, skewness β's, and location m's because

$$\phi(\omega) = \exp\left\{im\,\omega - \gamma|\omega|^\alpha \left(1 + i\beta\,\mathrm{sign}(\omega)\tan\frac{\alpha\pi}{2}\right)\right\}$$

for $\alpha \ne 1$

and

$$\phi(\omega) = \exp\{im\omega - \gamma|\omega|(1 - 2i\,\beta\,\ln|\omega|\text{sign}(\omega)/\pi)\}$$

for $\alpha = 1$

where

$$\text{sign}(\omega) = \begin{cases} 1 & \text{if } \omega > 0 \\ 0 & \text{if } \omega = 0 \\ -1 & \text{if } \omega < 0 \end{cases}$$

with $i = \sqrt{-1}$, $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, and $\gamma > 0$. So Fourier transformation gives the corresponding density function in the limiting case ($\gamma \to 0$) as a translated delta function

$$\lim_{\gamma \to 0} p(n) = \delta(n - m).$$

Then

$$F_{R|S}(r|A) - F_{R|S}(r|B) = \int_{\theta_1 - v_1}^{\theta_2 - v_2} p(n)\mathrm{d}n$$

$$\to \int_{\theta_1 - v_1}^{\theta_2 - v_2} \delta(n - m)\mathrm{d}n = 0$$

because $m \notin (\theta_1 - v_1, \theta_2 - v_2)$.

B. Only-if part

Suppose that $m \in (\theta_1 - v_1, \theta_2 - v_2)$ where m is the mean or location parameter of the total noise $n = z + n_2$. Then exactly one of the following four cases holds:

Case (1): $v_2 + m \leq \theta_1 < v_1 + m \leq \theta_2$
Case (2): $\theta_1 < v_2 + m < v_1 + m \leq \theta_2$
Case (3): $\theta_1 < v_2 + m < \theta_2 < v_1 + m$
Case (4): $v_2 + m \leq \theta_1 < \theta_2 < v_1 + m$

Suppose that Case (1) or Case (4) holds. Then define a new random variable $Y = g(R)$ such that

$$y = g(r) = \begin{cases} 0 & \text{if } r = 0 \\ 1 & \text{if } r > 0 \end{cases}$$

Suppose next that Case (2) holds. Then define

$$y = g(r) = \begin{cases} 0 & \text{if } r \leq r_0(v_2 + m + a) \\ 1 & \text{if } r > r_0(v_2 + m + a) \end{cases}$$

where $a = (v_1 - v_2)/2$.

Suppose last that Case (3) holds. Then define

$$y = g(r) = \begin{cases} 0 & \text{if } r < r_0(\theta_2 - \theta_1) \\ 1 & \text{if } r = r_0(\theta_2 - \theta_1) \end{cases}$$

We show that $I(S, Y) \to H(S)$ as $\sigma \to 0$. Recall that $H(S) \geq I(S, R)$ because $I(S, R) = H(S) - H(S|R)$ and S is a discrete random variable, and that $I(S, R) \leq I(S, Y = g(R))$ by data processing inequality (Cover and Thomas, 1991). Then $I(S, R)$ converges to its maximum value $H(S)$ as $\sigma \to 0$ and hence the SR effect does not exist in the sense that the system performs better without noise than with it when the interval condition fails.

We first give the proof for Case (1) and Case (4). Note that $v_2 + m < \theta_1$ implies $m < \theta_1 - v_2$ where for convenience only we ignore the measure-zero case of $m = \theta_1 - v_2$. Suppose that m is the mean of the *finite* variance total noise $z + n_2$. Pick $\varepsilon = (1/2)d(m, \theta_1 - v_2) > 0$. Then $\theta_1 - v_2 - \varepsilon = m + \varepsilon$. Write

$$\begin{aligned}
P_{Y|S}(0|B) &= \Pr\{r_0(n + v) = 0\}|_{S=B} \\
&= \Pr\{n + v_2 \leq \theta_1\} \text{ by (3) and (5)} \\
&= \Pr\{n \leq \theta_1 - v_2\} \\
&\geq \Pr\{n \leq \theta_1 - v_2 - \varepsilon\} = \Pr\{n \leq m + \varepsilon\} \\
&= 1 - \Pr\{n - m > \varepsilon\} \\
&\geq 1 - \Pr\{|n - m| > \varepsilon\} \\
&\geq 1 - \frac{\sigma^2}{\varepsilon^2} \text{ by Chebychev's inequality} \\
&\to 1 \text{ as } \sigma^2 \to 0
\end{aligned}$$

So $P_{Y|S}(0|B) = 1$.

Similarly for $P_{Y|S}(1|A)$: Note that $\theta_1 < v_1 + m \Rightarrow \theta_1 - v_1 < m$. Now pick $\varepsilon = (1/2)d(\theta_1 - v_1, m) > 0$. Then $\theta_1 - v_2 + \varepsilon = m - \varepsilon$. Write

$$\begin{aligned}
P_{Y|S}(1|A) &= \Pr\{r_0(n + v) > 0\}|_{S=A} \\
&= \Pr\{n + v_1 \geq \theta_1\} \text{ by (3) and (4)} \\
&= \Pr\{n \geq \theta_1 - v_1\} \geq \Pr\{n \geq \theta_1 - v_1 + \varepsilon\} \\
&= \Pr\{n \geq m - \varepsilon\} = 1 - \Pr\{n - m < -\varepsilon\} \\
&\geq 1 - \Pr\{|n - m| > \varepsilon\} \\
&\geq 1 - \frac{\sigma^2}{\varepsilon^2} \text{ by Chebychev's inequality} \\
&\to 1 \text{ as } \sigma^2 \to 0
\end{aligned}$$

So $P_{Y|S}(1|A) = 1$.

Suppose next that m is the location parameter of the total alpha-stable noise $z + n_2$. Then

$$\begin{aligned}
P_{Y|S}(0|B) &= \Pr\{n \leq \theta_1 - v_2\} \\
&= \int_{-\infty}^{\theta_1 - v_2} p(n)\mathrm{d}n \\
&\to \int_{-\infty}^{\theta_1 - v_2} \delta(n - m)\mathrm{d}n = 1
\end{aligned}$$

as $\gamma \to 0$ because $m < \theta_1 - v_2$.

Similarly

$$P_{Y|S}(1|A) = \Pr\{n \geq \theta_1 - v_1\}$$

$$= \int_{\theta_1 - v_1}^{\infty} p(n)\mathrm{d}n$$

$$\rightarrow \int_{\theta_1 - v_1}^{\infty} \delta(n-m)\mathrm{d}n = 1$$

as $\gamma \rightarrow 0$ because $m > \theta_1 - v_1$.

The two conditional probabilities for both the finite-variance and infinite variance cases likewise imply that $P_{Y|S}(0|A) = P_{Y|S}(1|B) = 0$ as $\sigma \rightarrow 0$ or $\gamma \rightarrow 0$. These four probabilities further imply that

$$H(Y|S) = \sum_s \sum_y P_{SY}(s, y)\log_2 P_{Y|S}(y|s)$$

$$= \sum_s P_s(s) \sum_y P_{Y|S}(y|s)\log_2 P_{Y|S}(y|s) \rightarrow 0$$

where we have used the fact (L'Hospital) that $0 \log_2 0 = 0$. The unconditional entropy $H(Y)$ becomes

$$H(Y) = -\sum_y P_Y(y)\log_2 P_Y(y)$$

$$\rightarrow -\sum_s P_S(s)\log_2 P_S(s) = H(S)$$

because

$$P_Y(y) = \sum_s P_{Y|S}(y|s)P_S(s)$$

$$= P_{Y|S}(y|A)P_S(A) + P_{Y|S}(y|B)P_S(B)$$

$$= P_{Y|S}(y|A)P_S(A) + P_{Y|S}(y|B)(1 - P_S(A))$$

$$= (P_{Y|S}(y|A) - P_{Y|S}(y|B))P_S(A) + P_{Y|S}(y|B)$$

$$= (P_{Y|S}(y|B) - P_{Y|S}(y|A))P_S(B) + P_{Y|S}(y|A)$$

$$\rightarrow \begin{cases} P_S(A) & \text{if } y = 1 \\ P_S(B) & \text{if } y = 0 \end{cases}$$

Thus $H(Y|S) \rightarrow 0$ and $H(Y) \rightarrow H(S)$ as $\sigma \rightarrow 0$ or $\gamma \rightarrow 0$. Then $I(S, Y) \rightarrow H(S)$ as $\sigma \rightarrow 0$ or $\gamma \rightarrow 0$ because $I(S, Y) = H(Y) - H(Y|S)$. $H(S)$ is the maximum of $I(S, Y)$ because $I(S, Y) = H(S) - H(S|Y)$ and $H(S|Y) \geq 0$ (Cover and Thomas, 1991). So $I(S, R)$ converges to its maximum value $H(S)$ as $\sigma \rightarrow 0$ and hence the system performs better without noise than with it for Case (1) and Case (4).

We next prove the claim for Case (2). We show only that $P_{Y|S}(0|B) = P_{Y|S}(1|A) = 1$ as $\sigma \rightarrow 0$ because the rest of the proof proceeds as in Case (1).

$$P_{Y|S}(0|B) = \Pr\{r_0(n + v_2) \leq r_0(v_2 + m + a)\}$$

$$= \Pr\{n + v_2 \leq v_2 + m + a\}$$

$$= \Pr\{n \leq m + a\}$$

$$\geq \Pr\{n \leq m + \varepsilon\} \quad \text{for } \varepsilon = \frac{a}{2}$$

$$= 1 - \Pr\{n - m > \varepsilon\} \geq 1 - \Pr\{|n - m| > \varepsilon\}$$

$$\geq 1 - \frac{\sigma^2}{\varepsilon^2} \text{ by Chebychev's inequality}$$

$$\rightarrow 1 \quad \text{as } \sigma^2 \rightarrow 0$$

So $P_{Y|S}(0|B) = 1$.

Similarly

$$P_{Y|S}(1|A) = \Pr\{r_0(n + v_2) > r_0(v_2 + m + a)\}$$

$$= \Pr\{n + v_1 > v_2 + m + a\}$$

$$\text{because } a = \frac{v_1 - v_2}{2}$$

$$= \Pr\{n + v_1 > v_1 + m - a\}$$

$$= \Pr\{n > m - a\} \geq \Pr\{n > m - \varepsilon\}$$

$$\text{by picking } \varepsilon = \frac{a}{2}$$

$$= 1 - \Pr\{n - m < -\varepsilon\} \geq 1 - \Pr\{|n - m| > \varepsilon\}$$

$$\geq 1 - \frac{\sigma^2}{\varepsilon^2} \text{ by Chebychev's inequality}$$

$$\rightarrow 1 \text{ as } \sigma^2 \rightarrow 0$$

So $P_{Y|S}(1|A) = 1$.

The proof for Case (3) proceeds as in Case (1). Q.E.D.

Simulation results confirm this mathematical result that noise in retinal signal processing can help retinal neurons detect subthreshold contrast signals. Figs. 2 and 3 show detailed simulation instances of the predicted SR effect in Theorem 1. Fig. 2 shows a 3-D plot of the Shannon mutual information versus the standard deviations of Gaussian white noise sources n_1 and n_2 in (1). Fig. 3 shows their respective cross-section plots for the values $\sigma_1 = 0.01$ and $\sigma_2 = 0.02$. We computed the bit count $I(S, R)$ using a discrete density of R based on the number of spikes in 1-second intervals for each input symbol. Each plot shows the nonmonotonic signature of SR.

3. Stochastic resonance in spiking sensory neuron models

Theorem 2 below describes the SR noise benefit in a wide range of spiking sensory neuron models. It states its own 'forbidden-interval' sufficient condition for SR in the special but ubiquitous case of additive Gaussian white noise. Proposition 1 shows that the converse also holds for the leaky integrate-and-fire neuron. Fig. 6 further shows that
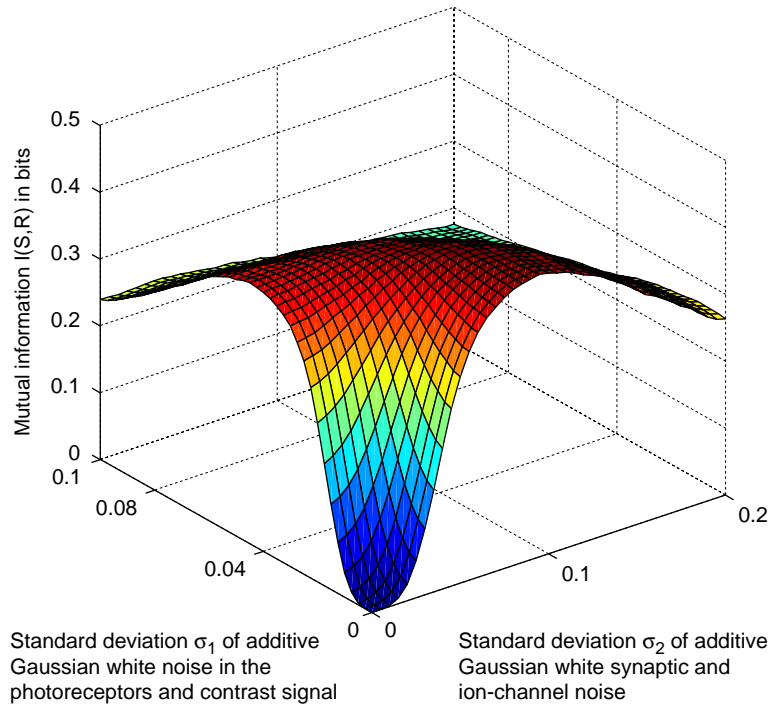
Fig. 2. Stochastic resonance (SR) in the spiking retina model with additive Gaussian white noise. The noisy retina model has the spiking Poisson form (1)–(3) with thresholds $\theta_1 = 0$ and $\theta_2 = 0.3$. The maximum firing rate is 100 spikes/s. The Bernoulli contrast signal takes the value of 0.2 with success probability $p = 1/2$ and takes the value of 0.4 otherwise. The graph shows the retina model's smoothed input-output mutual information surface as a function of the noise standard deviations σ_1 and σ_2.
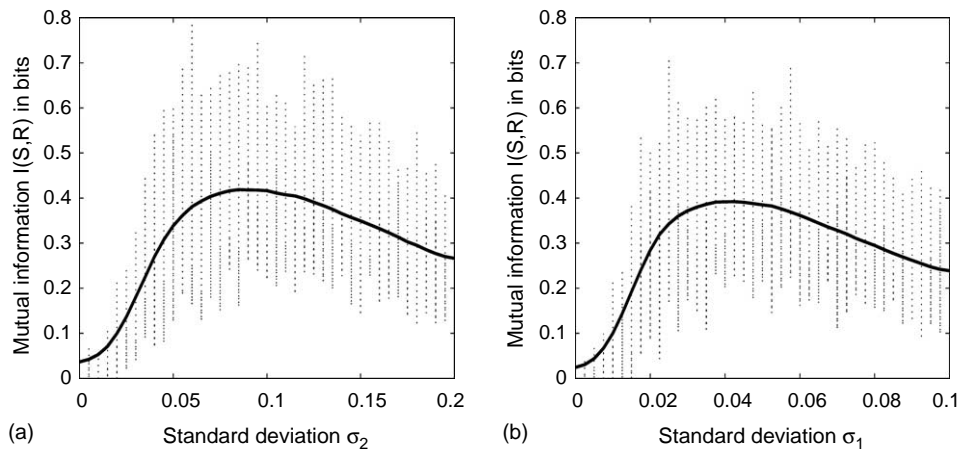

Fig. 3. Stochastic resonance in the model (1)–(3) with additive Gaussian white noise. Plots (a) and (b) show the respective cross-sections of the mutual-information surface of Fig. 2 for $\sigma_1 = 0.01$ and $\sigma_2 = 0.02$. Each simulation trial produced 10,000 input-output samples $\{s(t), r(t)\}$ that estimated the Poisson spiking rate $r(t)$ to obtain the mutual information. Thick lines show the average mutual information. Vertical lines show the total min–max deviations of the mutual information in 1000 trials.

the SR result persists in this case even for infinite-variance stable noise. Proposition 2 shows that the converse need not hold for the Fitzhugh-Nagamo neuron—SR can still occur inside the forbidden interval.

Theorem 2 specifically shows that these and other spiking neuron models enjoy an SR noise benefit if the noise mean $E(n)$ falls to the left of a bound and if their average firing rates depend on the Kramers rate solution (Kramers, 1940) of the Fokker-Planck diffusion equation.

Theorem 2 applies to popular spiking sensory neuron models such as the FitzHugh-Nagumo (FHN) model (Chialvo et al., 1997; Collins et al., 1995), the leaky integrate-and-fire model (Collins et al., 1996; Gerstner and Kistler, 2002), and the reduced Type I neuron model (Lindner et al., 2003). Fig. 7 shows that SR can still occur in the FHN neuron model even if $E(n)$ falls to the right of this bound. So the interval condition in Theorem 2 is not necessary.

The FHN neuron model has the form

$$\varepsilon\dot{v} = -v\left(v^2 - \frac{1}{4}\right) - w + A_T - d + n, \tag{6}$$

$$\dot{w} = v - w \tag{7}$$

where v is the membrane voltage (fast) variable, w is a recovery (slow) variable, $A_T = -5/(12\sqrt{3})$ is a threshold voltage, S is the input signal, $d = B - S$, B is the constant signal-to-threshold distance, and n is independent Gaussian white noise. The input signal is subthreshold when $d > 0$ and so then $S < B$.

Kramers rate formula gives the average firing rate of the FHN neuron model with subthreshold input signals ($S(t) \ll B$) (Collins et al., 1996)

$$E(r(t)) = \frac{B}{2\pi\sqrt{3}\varepsilon}\exp\left[\frac{-2\sqrt{\varepsilon}[B^3 - 3B^2 S(t)]}{3\sigma^2}\right]. \tag{8}$$

The average spike rate model poorly estimated the averge firing rates of the FHN model in simulations. So we instead fitted the equation

$$E(r(t)) = a\exp\left[\frac{-bB^3 + cB^2 S(t)}{\sigma^2}\right] \tag{9}$$

to the simulation data. Nonlinear least-squares gave the parameters a, b, and c in (9). Fig. 4 shows that the fitted model (9) closely estimates the average spike rates of the FHN neuron model because the coefficient of determination was $r^2 = 0.9976$.
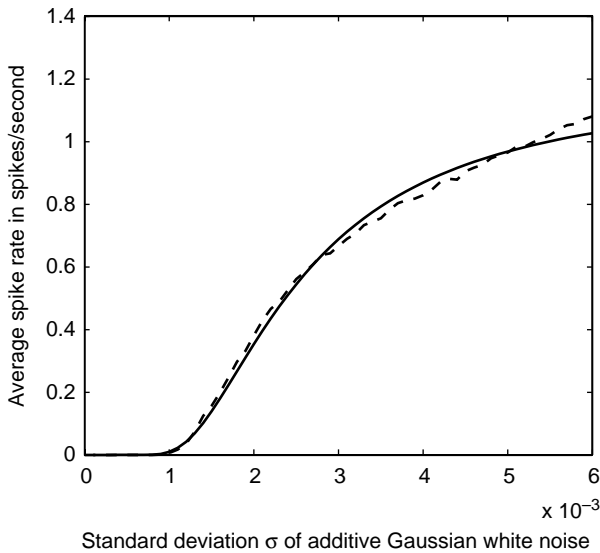

Fig. 4. Approximation of the average firing rate. The estimated firing rate (solid line) closely approximates the average firing rate (dashed line) for the FHN neuron model in (9). The model parameters are $A_T = -5/(12\sqrt{3})$, $B = 0.07$, and $S = 0.01$. Nonlinear least-squares fitted the parameters in (9) as $a = 1.1718$, $b = 0.0187$, and $c = 0.0680$ with coefficient of determination $r^2 = 0.9976$.

The leaky integrate-and-fire neuron model has the form (Collins et al., 1996)

$$\dot{v} = -av + a - \delta + S + n \tag{10}$$

where v is the membrane voltage, a and δ are constants, δ/a is the barrier height of the potential, S is an input signal, and n is independent Gaussian white noise. The input signal S is subthreshold when $S < \delta$. The neuron emits a spike when the membrane voltage v crosses the threshold value of 1 from below to above. The membrane voltage v resets to $1 - \delta/a$ just after the neuron emits a spike. Then the ensemble-averaged spike rate $E(r(t))$ for subthreshold inputs ($S^2 \ll \delta$) has the form (Collins et al., 1996)

$$E(r(t)) = \frac{\delta}{\sqrt{\sigma^2\pi}}\exp\left[\frac{-\delta^2 + 2\delta S(t)}{\sigma^2 a}\right] \tag{11}$$

where σ^2 is the variance of n.

Theorem 2 applies to the reduced Type I neuron model in (12) below. The reduction procedure in (Gutkin and Ermentrout, 1998; Hoppensteadt and Izhikevich, 1997) gives a simple one-dimensional normal form (Lindner et al., 2003) of the multi-dimensional dynamics of Type I neuron models:

$$\dot{v} = \beta + v^2 + \sigma n \tag{12}$$

where v is the membrane potential, β is the value of input signal, and σ is the standard deviation of Gaussian white noise n. The firing rate of the reduced model (12) for subthreshold or excitable regime ($\beta < 0$) and weak noise ($\sigma^2 \ll 2|\beta|^{3/2}$) is (Lindner et al., 2003)

$$E(r(t)) = \frac{\sqrt{|\beta|}}{\pi}\exp\left[\frac{-8|\beta|^{3/2}}{3\sigma^2}\right]. \tag{13}$$

We can combine (9), (11), and (13) into the general form

$$E(r(t)) = g(B, S(t), \sigma)\exp\left[\frac{h(B, S(t))}{k\sigma^2}\right] \tag{14}$$

where $E(r(t))$ is the average firing rate and k is a constant. The functions $g(B, S, \sigma)$ and $h(B, S)$ depend on the potential barrier B, the subthreshold input signal S, and on the variance σ^2 of the additive Gaussian white noise n so that $E(r(t)) \rightarrow 0$ as $\sigma \rightarrow 0$. We note that the formula for the average Poisson spike rate in excitable cells due to the voltage-gated ion channels dynamics has a form similar to (14) (Bezrukov and Vodyanoy, 1998).

We can now state Theorem 2. This theorem gives a sufficient condition for SR to occur in spiking sensory neuron models if their average output spike rates have the general form (14). The proof again follows the proof in (Kosko and Mitaim, 2003; Kosko and Mitaim, 2004).

Theorem 2. *Suppose that the average spike rate of a sensory neuron model has the form* (14) *and that $E(n)$ is the mean of the model's additive Gaussian white noise n. Suppose that input signal $S(t) \in \{s_1, s_2\}$ is subthreshold:*

$S(t) < B$. Suppose that there is some statistical dependence between the input signal random variable S and the output average spike-rate random variable R so that $I(S, R) > 0$. Then the spiking sensory neuron exhibits the nonmonotone SR effect in the sense that $I(S, R) \to 0$ as the noise intensity $\sigma \to 0$ if $E(n) < B - s_2$.

Proof. Again we need to show only that

$$F_{R|S}(r|s_1) - F_{R|S}(r|s_2) \to 0$$
$$\text{iff} \quad F_{R|S}(r|s_1) - 1 + 1 - F_{R|S}(r|s_2) \to 0$$
$$\text{iff} \quad \Pr\{R \geq r|S = s_2\} - \Pr\{R \geq r|S = s_1\} \to 0$$
$$\text{as} \quad \sigma \to 0 \quad \text{for all } r \text{ in } (0, r_{max}) \tag{15}$$

We can write

$$\Pr\{R \geq r|S = s_1\} \leq \frac{E(R|S = s_1)}{r}$$

by Markov's inequality for all r and similarly

$$\Pr\{R \geq r|S = s_2\} \leq \frac{E(R|S = s_2)}{r}$$

If the expression of $E(r(t))$ has the form (14) then we need only show that $E(R|S = s_1) \to 0$ and $E(R|S = s_2) \to 0$ as $\sigma \to 0$. We can absorb $E(n)$ into the input signal $S(t)$ because the noise n is additive in the model of spiking sensory neuron. Then the new input signal is $S'(t) = S(t) + E(n)$ and $S'(t)$ is subthreshold ($S'(t) < B$) because $E(n) < B - s_2$ where $s_2 = \max\{s_1, s_2\}$. Thus $E(r(t))$ has the form of (14). This proves (15) and hence the Theorem 2.   Q.E.D.

Fig. 5 shows a simulation instance of the SR effect in Theorem 2 for the special but important case of the FHN neuron model. The mutual-information plot shows the predicted nonmonotonic signature of SR. The leaky integrate-and-fire neuron model produces similar nonmonotonic SR plots. Fig. 6 goes beyond the scope of Theorem 2 and shows a simulation instance of the SR effect in the leaky integrate-and-fire neuron model with implusive infinite-variance α-stable white noise.

Proposition 1. *The converse of* Theorem 2 *holds for the leaky integrate-and-fire neuron model* (10) *in the sense that the system performs better without noise than with it when the interval condition $E(n) < B - s_2$ fails.*

Proof. Suppose that $E(n) > B - s_2$. Then exactly one of the following two is true:

    Case (1): $s'_1 = s_1 + E(n)$ is subthreshold and $s'_2 = s_2 + E(n)$ is superthreshold.
    Case (2): Both s'_1 and s'_2 are superthreshold.

Suppose that the input signal s'_i is superthreshold. Then the interspike interval T_i in the absence of additive noise n is (Gerstner and Kistler, 2002)
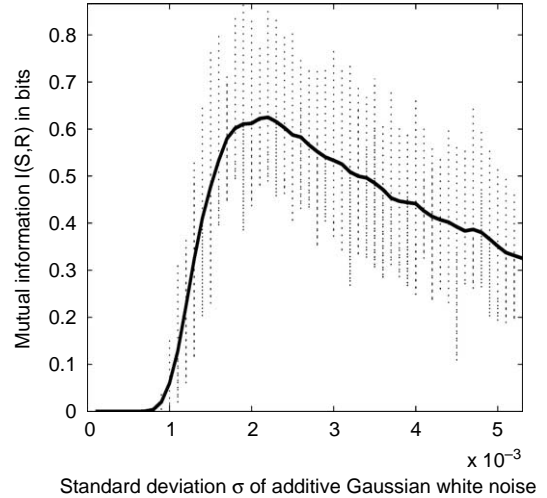

Fig. 5. Stochastic resonance in the FHN spiking neuron model—a simulation instance of Theorem 2. The model parameters are $A_T = -5/12(\sqrt{3})$, $B = 0.07$, and $S = \pm0.005$. The solid curve shows the average mutual information. The dashed vertical lines show the total min–max deviations of mutual information in 100 simulation trials.

$$T_i = \tau_m \ln \frac{v_i^\infty - v_r}{v_i^\infty - T_h}$$

where v_i^∞ and v_r are the respective values of the membrane potential at steady-state and at the reset, τ_m is a time-constant of the membrane potential, and T_h is a threshold for spike generation. The interspike interval has a Gaussian distribution in the presence of Gaussian white noise n in (10) (Gerstner and Kistler, 2002). The probability density of
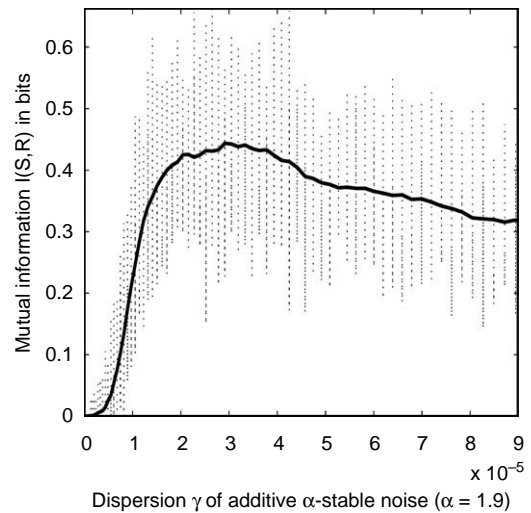

Fig. 6. Stochastic resonance in the integrate-and-fire spiking neuron model with subthreshold input signals and infinite-variance α-stable noise ($\alpha = 1.9$). The model parameters are $a = 0.5$, $\delta = 0.01$, $s_1 = 0.0025$, and $s_2 = 0.005$. The solid curve shows the smoothed average mutual information. The dashed vertical lines show the total min–max deviations of mutual information in 100 simulation trials.

interspike interval τ_i is

$$f(\tau_i) = \frac{v'_i}{\sqrt{\pi}\sigma} \exp\left[-\frac{v'^2_i(\tau_i - T_i)^2}{\sigma^2}\right]$$

where $E(\tau_i) = T_i$, $v'_i = (dv_i(t)/dt)$ evaluated at $t = T_i$, and σ is the standard deviation of the additive white noise. Then

$$\Pr\{|\tau_i - T_i| > \varepsilon\} \leq \frac{\sigma^2}{2v'\varepsilon} \quad \text{for all } \varepsilon > 0$$
$$\to 0 \quad \text{as } \sigma^2 \to 0$$

Thus if s'_i is superthreshold then $\tau_i \to T_i$ in probability as $\sigma \to 0$ by the definition of convergence in probability. Then the corresponding output spike rate $r|_{S'=s'_i} = r_i \to (1/T_i)$ in probability because $r_i = (1/\tau_i)$. So

$$\Pr\left\{\left|r_i - \frac{1}{T_i}\right| > \varepsilon\right\} \to 0 \quad \text{for all } \varepsilon > 0 \quad \text{as } \sigma^2 \to 0 \qquad (16)$$

Suppose that Case (1) holds. Then define

$$y = g(r) = \begin{cases} 0 & \text{if } r \leq \dfrac{1/T_2}{2} \\[2ex] 1 & \text{if } r > \dfrac{1/T_2}{2} \end{cases}.$$

Suppose that Case (2) holds. Then define

$$y = g(r) = \begin{cases} 0 & \text{if } r \leq \dfrac{1}{T_1} + a \\[2ex] 1 & \text{if } r > \dfrac{1}{T_2} - a \end{cases}$$

where $a = ((1/T_2) - (1/T_1))/2$. Note that $a > 0$ because $(1/T_2) > (1/T_1)$.

We need to show only that $P_{Y|S'}(0|s'_1) = P_{Y|S'}(1|s'_2) = 1$ as $\sigma \to 0$ because the rest of the proof is similar to the only-if part of the proof of Theorem 1.

Suppose that Case (1) holds. Then

$$P_{Y|S}(0|s'_1) = \Pr\left\{r \leq \frac{1/T_2}{2}\middle| S' = s'_1\right\}$$
$$= 1 - \Pr\left\{r > \frac{1/T_2}{2}\middle| S' = s'_1\right\}$$
$$\geq 1 - \frac{E(R|S' = s'_1)}{\frac{1/T_2}{2}} \quad \text{by Markov's inequality}$$
$$\to 1 \quad \text{as } \sigma \to 0$$

because s'_1 is subthreshold and $E(r(t)) \to 0$ for (14).

$$P_{Y|S'}(1|s'_2) = \Pr\left\{r > \frac{1/T_2}{2}\middle| S' = s'_2\right\}$$
$$= 1 - \Pr\left\{r < \frac{1/T_2}{2}\middle| S' = s'_2\right\}$$
$$\geq 1 - \Pr\left\{|r_2 - \frac{1}{T_2}| > \frac{1/T_2}{2}\right\}$$
$$\text{because } r|_{S'=s'_i} = r_i$$
$$\to 1 \quad \text{by (16)}.$$

Suppose now that Case (2) holds. Then

$$P_{Y|S}(0|s'_1) = \Pr\left\{r \leq \frac{1}{T_1} + a\middle| S' = s'_1\right\}$$
$$= \Pr\left\{r_1 \leq \frac{1}{T_1} + a\right\}$$
$$\text{because } r|_{S'=s'_i} = r_i$$
$$= 1 - \Pr\left\{r_1 > \frac{1}{T_1} + a\right\}$$
$$\geq 1 - \Pr\left\{|r_1 - \frac{1}{T_1}| > a\right\}$$
$$\to 1 \quad \text{by (16)}$$

and

$$P_{Y|S}(1|s'_2) = \Pr\left\{r > \frac{1}{T_2} - a\middle| S' = s'_2\right\}$$
$$= \Pr\left\{r_2 > \frac{1}{T_2} - a\right\}$$
$$= 1 - \Pr\left\{r_2 < \frac{1}{T_2} - a\right\}$$
$$\geq 1 - \Pr\left\{|r_2 - \frac{1}{T_2}| > a\right\}$$
$$\to 1 \quad \text{by (16)} \qquad \text{Q.E.D.}$$

Proposition 2. *The converse of* Theorem 2 *does not hold for the FHN neuron model* (6)–(7).

Fig. 7 confirms Proposition 2 because it shows that SR can still occur when the noise mean $E(n)$ falls to the right of $B - \max\{s_1, s_2\}$.

4. Conclusion

Theorems 1 and 2 and their progeny present technical interval conditions that screen for whether an SR noise benefit occurs in spiking neurons. Several of the most popular models of biological models provably benefit from adding the right amount of white noise subject to these interval conditions. Spiking retinal models benefit from almost all types of noise because Theorem 1 holds both for all finite-variance noise and for the large class
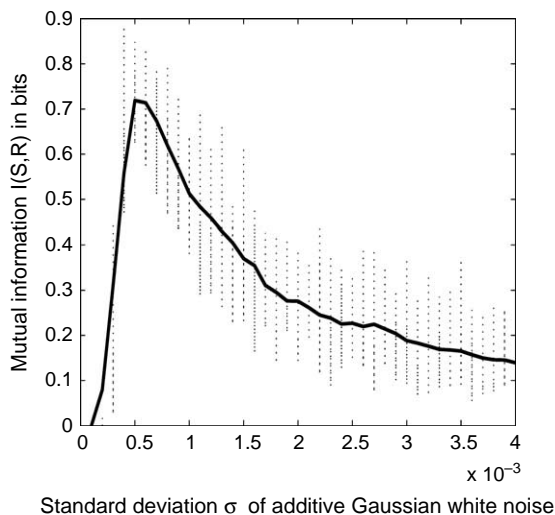
Fig. 7. The interval condition in Theorem 2 is not necessary. Stochastic resonance in the FHN spiking neuron model with superthreshold input signals and additive Gaussian white noise. The model parameters are $A_T = -5/(12\sqrt{3})$, $B = 0.07$, $s_1 = 0.56$, and $s_2 = 0.565$. $E(n) = 0 > B - s_2 = -0.495$ implies that $E(n)$ does not satisfy the interval condition of Theorem 2. The solid curve shows the smoothed average mutual information. The dashed vertical lines show the total min–max deviations of mutual information in 100 simulation trials.

of infinite-variance stable noise. Theorem 2 applies only to additive Gaussian white noise although simulations confirm a comparable noise benefit for infinite-variance stable noise (see Fig. 6) and other types of finite-variance noise. An open research question is whether other noise types produce a noise benefit in sensory neurons or in the more complex cortical neurons that take spikes as input as well as emit spikes as output. These theorems suggest a potential engineering noise benefit from controlled noise injection in artificial retinas, neural prosthetics, low-light imaging and night vision, and infrared imaging and object detection.

Acknowledgements

References

Ala-Laurila, P., Donner, K., & Koskelainen, A. (2004). Thermal activation and photoactivation of visual pigments. *Biophysical Journal*, *86*, 3653–3662.

Berzukov, S. M., & Vodyanoy, I. (1998). Stochastic resonance in thermally activated reactions: Applications to biological ion channel. *Chaos*, *8*(3), 557–566.

Bulsara, A. R., & Zador, A. (1996). Threshold detection of wideband signals: A noise induced maximum in the mutual information. *Physical Review E*, *54*(3), R2185–R2188.

Burkhardt, D. A., Gottesman, J., Kersten, D., & Legge, G. E. (1984). Symmetry and constancy in the perception of negative and positive luminance contrast. *Jornal of the Optical Society of America A*, *1*, 309–316.

Chander, D., & Chichilnisky, E. J. (2001). Adaptation to temporal contrast in primate and Salamander retina. *Journal of Neuroscience*, *21*(24), 9904–9916.

Chialvo, D. R., Longtin, A., & Muller-Gerkin, J. (1997). Stochastic resonance in models of neuronal ensembles revisited. *Physical Review E*, *55*(2), 1798–1808.

Collins, J. J., Chow, C. C., Capela, A. C., & Imhoff, T. T. (1996). Aperiodic stochastic resonance. *Physical Review E*, *54*(5), 5575–5584.

Collins, J. J., Chow, C. C., & Imhoff, T. T. (1995). Aperiodic stochastic resonance in excitable systems. *Physical Review E*, *52*(4), R3321–R3324.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley pp. 12–49.

Deco, G., & Schurmann, B. (1998). Stochastic resonance in the mutual information between input and output spike trains of noisy central neurons. *Physica D*, *117*, 276–282.

Durrett, R. (1996). *Probability: Theory and examples*. Duxbury Press pp. 23–32.

Freed, M. A. (2000). Rate of quantal excitation to a retinal Ganglion cell evoked by sensory input. *Journal of Neurophysiology*, *83*, 2956–2966.

Gammaitoni, L. (1995). Stochastic resonance and the dithering effect in threshold physical systems. *Physical Review E*, *52*(5), 4691–4698.

Gerstner, W., & Kistler, W. M. (2002). *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge University Press pp. 93–200.

Godivier, X., & Chapeau-Blondeau, F. (1998). Stochastic resonance in the information capacity of a nonlinear dynamic system. *International Journal of Bifurcation and Chaos*, *8*(3), 581–589.

Grigoriu, M. (1995). *Applied non-gaussian processes*. Englewood Cliffs, NJ: Prentice Hall pp. 99–119 (see also 314–329).

Gutkin, B. S., & Ermentrout, G. B. (1998). Dynamics of membrane excitability determine interspike interval variability: A link between spike generation mechanism and cortical spike train statistics. *Neural Computation*, *10*, 1047–1065.

Hoppensteadt, F. C., & Izhikevich, E. M. (1997). *Weakly connected neural networks*. New York: Springer pp. 111–142.

Inchiosa, M. E., Robinson, J. W. C., & Bulsara, A. R. (2000). Information-theoretic stochastic resonance in noise-floor limited systems: The case for adding noise. *Physical Review Letters*, *85*, 3369–3372.

Jung, P. (1995). Stochastic resonance and optimal design of threshold detectors. *Physics Letters A*, *207*, 93–104.

Kim, K. J., & Rieke, F. (2001). Temporal contrast adaptation in the input and output signals Signals of Salamander retinal ganglion cell. *Journal of Neuroscience*, *21*, 287–299.

Korenberg, M. J., & Hunter, I. W. (1986). The identification of nonlinear biological systems: LNL cascade models. *Biological Cybernetics*, *55*, 125–134.

Kosko, B., & Mitaim, S. (2001). Robust stochastic resonance: Signal detection and adaptation in impulsive noise. *Physical Review E*, *64*(5), 051110/1–051110/11.

Kosko, B., & Mitaim, S. (2003). Stochastic resonance in noisy threshold neurons. *Neural Networks*, *14*(6–7), 755–761.

Kosko, B., & Mitaim, S. (2004). Robust stochastic resonance for simple threshold neurons. *Physical Review E*, *70*, 031911/1–031911/10.

Kramers, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reaction. *Physica*, *7*, 284–304.

Lamb, T. D. (1987). Sources of noise in photoreceptor transduction. *Jornal of the Optical Society of America A*, *4*, 2295–2300.

Levy, W. B., & Baxter, R. A. (2002). Energy-efficient neural computation via quantal synaptic failures. *Journal of Neuroscience*, *22*(11), 4746–4755.

Lindner, B., Longtin, A., & Bulsara, A. R. (2003). Analytic expressions for rate and CV of a type I neuron driven by White Gaussian noise. *Neural Computation*, *15*, 1761–1788.

Manwani, A., & Koch, C. (1999). Detecting and estimating signals in noisy cable structures. I. Neuronal noise sources. *Neural Computation*, *11*, 1797–1829.

Mitaim, S., & Kosko, B. (1998). Adaptive stochastic resonance. *Proceedings of the IEEE*, *86*(11), 2152–2183.

Patel, A., & Kosko, B. (2005). *Noise benefits in spiking retinal and sensory neuron models Proceedings of the international joint conference on neural networks, July 31–August 4*. Montreal, Canada.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1996). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press pp. 38–48 (see also 267–277).

Sakai, H. M., Wang, J.-L., & Naka, K.-I. (1995). Contrast gain control in the lower vertebrate retinas. *Journal of General Physiology*, *105*, 815–835.

Schneidman, E., Freedman, B., & Segev, I. (1998). Ion channel stochasticity may be critical in determining the reliability and precision of spike timing. *Neural Computation*, *10*, 1679–1703.

Shao, M., & Nikias, C. L. (1993). Signal processing with fractional lower order moments: Stable processes and their applications. *Proceedings of the IEEE*, *81*, 984–1010.

Van Rossum, M. C. W., O'Brien, B. J., & Smith, R. G. (2003). Effects of noise on the spike timing precision of retinal ganglion cells. *Journal of Neurophysiolgy*, *89*, 2406–2419.

Wiesenfeld, K., & Moss, F. (1995). Stochastic resonance and the benefits of noise: From ice ages to Crayfish and SQUIDs. *Nature*, *373*, 33–36.

Yu, Y., Potetz, B., & Lee, T. S. (2005). The role of spiking nonlinearity in contrast gain control and information transmission. *Vision Research*, *45*(5), 583–592.