Neural Networks 22 (2009) 697–706

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2009 Special Issue

Error-probability noise benefits in threshold neural signal detection

Ashok Patel, Bart Kosko *

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

ARTICLE INFO

Article history:
Received 6 May 2009
Received in revised form 7 June 2009
Accepted 25 June 2009

Keywords:
Optimal noise intensity
Stochastic resonance
Threshold neurons
Signal detection
Error probability
Array detection

ABSTRACT

Five new theorems and a stochastic learning algorithm show that noise can benefit threshold neural signal detection by reducing the probability of detection error. The first theorem gives a necessary and sufficient condition for such a noise benefit when a threshold neuron performs discrete binary signal detection in the presence of additive scale-family noise. The theorem allows the user to find the optimal noise probability density for several closed-form noise types that include generalized Gaussian noise. The second theorem gives a noise-benefit condition for more general threshold signal detection when the signals have continuous probability densities. The third and fourth theorems reduce this noise benefit to a weighted-derivative comparison of signal probability densities at the detection threshold when the signal densities are continuously differentiable and when the noise is symmetric and comes from a scale family. The fifth theorem shows how collective noise benefits can occur in a parallel array of threshold neurons even when an individual threshold neuron does not itself produce a noise benefit. The stochastic gradient-ascent learning algorithm can find the optimal noise value for noise probability densities that do not have a closed form.

1. Neural noise benefits: Total and partial SR

Stochastic resonance (SR) occurs when a small amount of noise improves nonlinear signal processing (Amblard, Zozor, McDonnell, & Stocks, 2007; Chapeau-Blondeau & Rousseau, 2004; Gammaitoni, 1995; Kay, 2000; Kosko, 2006; Levy & Baxter, 2002; McDonnell, Stocks, Pearce, & Abbott, 2006, 2008; Moss, Ward, & Sannita, 2004; Patel & Kosko, 2009b; Rousseau & Chapeau-Blondeau, 2005b; Saha & Anand, 2003; Stocks, 2001). SR occurs in many types of subthreshold and suprathreshold neural signal detection (Bulsara, Jacobs, Zhou, Moss, & Kiss, 1991; Deco & Schürmann, 1998; Hänggi, 2002; Hoch, Wenning, & Obermayer, 2003; Li, Hou, & Xin, 2005; Mitaim & Kosko, 1998, 2004; Moss et al., 2004; Patel & Kosko, 2005, 2008; Sasaki et al., 2008; Stacey & Durand, 2000; Stocks, Appligham, & Morse, 2002; Stocks & Mannella, 2001; Wang & Wang, 1997; Wiesenfeld & Moss, 1995). Biological noise can arise from an internal source such as thermal noise (Faisal, Selen, & Wolpert, 2008; Manwani & Koch, 1999) or ion channel noise (Schneidman, Freedman, & Segev, 1998; White, Rubinstein, & Kay, 2000). Or it can arise from an external source such as synaptic transmission (Levy & Baxter, 2002; Markram & Tsodyks, 1996). We focus on noise-enhanced signal detection in threshold neurons where a user can control only the noise variance or dispersion (Läer et al., 2001; Pantazelou, Dames, Moss, Douglass, & Wilkens, 1995; Rao, Wolf, & Arkin, 2002). We measure detection performance with the probability of correct decision $P_{CD} = 1 - P_e$ when P_e is the probability of error (Patel & Kosko, 2009a).

We classify SR noise benefits as either total SR or partial SR. The SR effect is *total* if adding independent noise in the received signal reduces the error probability. Then the plot of detection probability versus noise intensity increases monotonically in some noise-intensity interval starting from zero. The SR effect is *partial* when the detection performance increases in some noise-intensity interval away from zero. Total SR ensures that adding small amounts of noise gives a better detection performance than not adding noise. Partial SR ensures only that there exists a noise intensity interval where the detection performance increases as the noise intensity increases. The same system can exhibit both total and partial SR. We derive conditions that screen for total or partial SR noise benefits in almost all suboptimal simple threshold detectors because the SR conditions apply to such a wide range of signal probability density functions (pdfs) and noise pdfs. Learnings laws can then search for the optimal noise intensity in systems that pass the screening conditions. Section 5 presents one such stochastic learning law.

We have already proven necessary and sufficient "*forbidden interval*" conditions on the noise mean or location for total SR in mutual-information-based threshold detection of discrete weak binary signals (Kosko & Mitaim, 2003, 2004): SR occurs if and only if the noise mean or location parameter μ obeys $\mu \notin (\theta - A, \theta + A)$ for threshold θ where $-A < A < \theta$ for bipolar subthreshold signal $\pm A$. More general forbidden interval theorems apply to many stochastic neuron models with Brownian or even Levy (jump)

* Corresponding author.
E-mail address: kosko@usc.edu (B. Kosko).

0893-6080/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2009.06.044

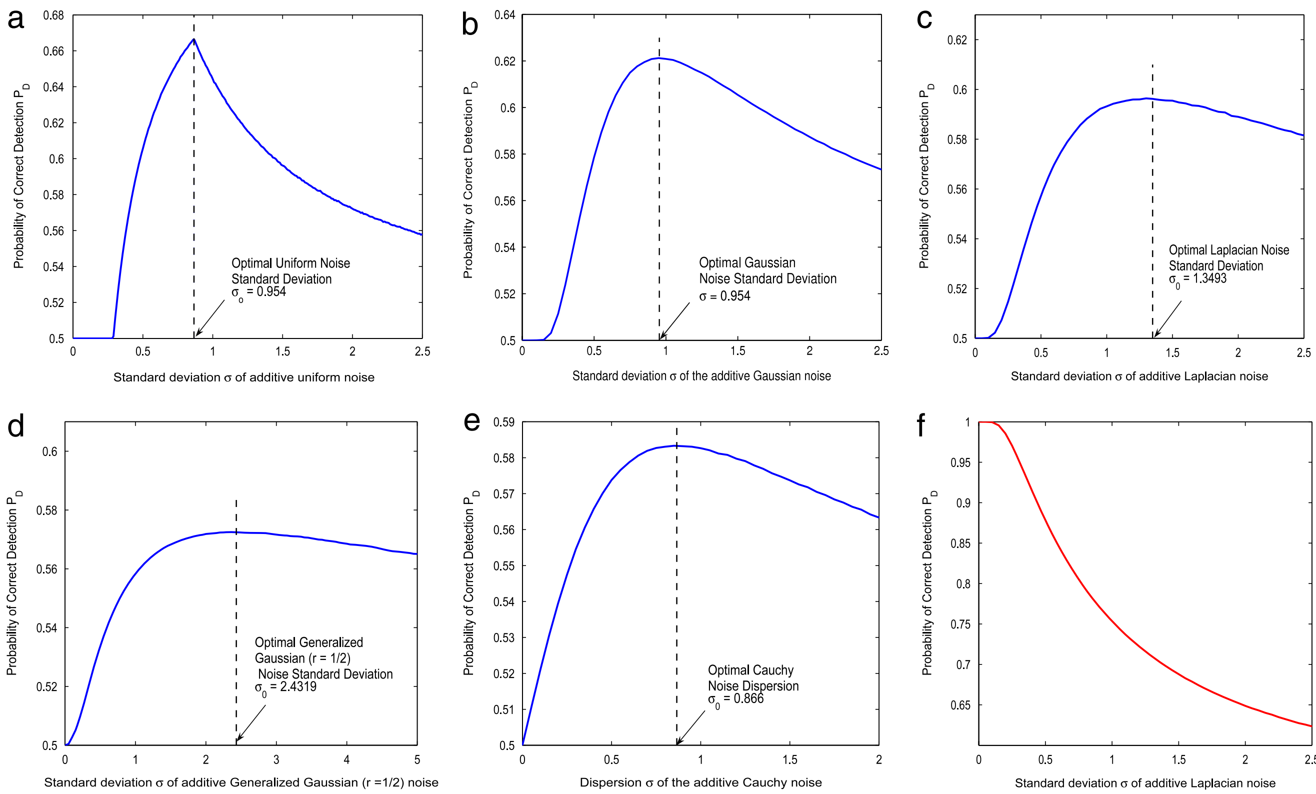

Fig. 1. Stochastic resonance (SR) noise benefits in binary neural signal detection for five out of six types of additive noise. The discrete signal X can take subthreshold values $s_0 = -1$ or $s_1 = 0$ with equal probability ($p_0 = P(s_0) = P(s_1) = p_1$) and $\theta = 1.5$ is the detection threshold. We decide $X = s_1$ if the observation $W = X + N > \theta$. Else $X = s_0$. There is a noise benefit in (a)–(e) because the zero-mean uniform, Gaussian, Laplacian, generalized Gaussian ($r = \frac{1}{2}$), and zero-location Cauchy noise satisfy condition (5) of Theorem 1. The SR effect is partial in (a) because condition (6) does not hold while the SR effect in (b)–(e) is total in each case because condition (6) holds. The dashed vertical lines show that the maximum SR effect occurs at the theoretically predicted optimal noise intensities. There is no noise benefit in (f) for Laplacian noise because its mean μ lies in the forbidden interval in accord with Corollary 1: $\mu = 1 \in (0.5, 1.5) = (\theta - s_1, \theta - s_0)$.

noise (Patel & Kosko, 2005, 2008). Corollary 1 gives a forbidden-interval necessary condition for SR in error-probability detection. But we did find necessary and sufficient conditions for both total and partial SR noise benefits in error-probability-based threshold signal detection when the noise has a scale-family distribution.

Theorem 1 gives the a simple necessary and sufficient SR condition for a noise benefit in threshold detection of discrete binary signals. This result appears in Section 3. The condition also determines whether the SR effect is total or partial if the noise density belongs to a *scale family*. Scale-family densities include many common densities such as the normal and uniform but not the Poisson. The condition implies that SR occurs in simple threshold detection of discrete binary signals only if the mean or location of additive location-scale family noise does not fall in an open forbidden interval. The uniform, Gaussian, Laplacian, and generalized Gaussian ($r = \frac{1}{2}$) noise in Fig. 1(a)–(e) produce a noise benefit because they satisfy condition (5) of Theorem 1. But the Laplacian noise in Fig. 1(f) violates this forbidden-interval condition of Corollary 1 and so there is no noise benefit. Section 4 shows that the SR condition of Theorem 1 also allows us to find the optimal noise dispersion that maximizes the detection probability for a given closed-form scale-family noise pdf. Section 5 shows that an adaptive gradient-ascent learning algorithm can find this optimal intensity from sample data even when the noise pdf does not have a closed form as with many thick-tailed noise pdfs.

Total SR can never occur in an optimal threshold system if we add only independent noise in the received signal. Kay and coworkers showed that the optimal independent additive SR noise is just a constant that minimizes the detection error probability of a given detection scheme (Kay, Michels, Chen, & Varshney, 2006). So total SR can never occur if the detection threshold location is optimal even when the overall detection scheme is suboptimal. But we show that partial SR can still occur in a single-threshold suboptimal system even if the detection threshold is optimal. Rousseau and Chapeau-Blondeau found earlier that what we call partial SR occurs in some special cases of optimal threshold detection (Rousseau & Chapeau-Blondeau, 2005b). Fig. 3 shows such a partial SR effect for the important but special case of an optimal threshold. Our result still holds in the general case of non-optimal thresholds. The suboptimality of the signal detection remains only a necessary condition for total SR noise benefits based on error probability.

Theorem 2 in Section 6 presents a related necessary and sufficient condition for a noise benefit in a more general case of threshold detectors when the signals have continuous pdfs and when the additive independent noise has a pdf from a scale family. Then Theorem 3 gives a necessary and sufficient condition for total SR with zero-mean discrete bipolar noise. Corollary 2 gives a necessary and sufficient condition for partial SR with zero-mean discrete bipolar noise when there is no total SR in Theorem 3. Theorems 3 and 4 each gives a necessary and sufficient condition for total SR when the additive noise is zero-mean discrete bipolar or when it comes from a finite-mean symmetric scale family. These two theorems compare weighted derivatives of continuously differentiable signal pdfs at the detection threshold to determine the total SR effect. Theorem 5 shows when noise produces a collective SR effect in parallel arrays of threshold neurons even when an individual threshold neuron does not produce an SR effect. The next section describes a general problem of threshold-based neural signal detection and defines the two SR effects based on error probability.

2. Binary signal detection based on error-probability

We now cast the problem of threshold-based neural signal detection as a statistical hypothesis test. So consider the binary hypothesis test where a neuron decides between $H_0 : f_X(x, H_0) = f_0(x)$ and $H_1 : f_X(x, H_1) = f_1(x)$ using a *single* noisy observation of $X + N$ and a detection threshold θ where noise N is independent of X. Here $X \in R$ is the original signal input to the threshold neuron and f_i is its pdf under the hypothesis H_i for $i = 0$ or 1. So we use the classical McCulloch–Pitts threshold neuron (McCulloch & Pitts, 1943) where the neuron's output Y has the form

$$Y = \begin{cases} 1 \text{ (accept } H_1) & \text{if } X + N > \theta \\ 0 \text{ (accept } H_0) & \text{else.} \end{cases} \quad (1)$$

This simple threshold neuron model has numerous applications (Auer, Burgsteiner, & Maass, 2008; Beiu, Member, Quintana, & Avedillo, 2003; Caticha, Palo Tejada, Lancet, & Domany, 2002; Freund & Schapire, 1999; Minsky & Papert, 1988).

The probability of correct decision $P_{CD}(\sigma) = 1 - P_e(\sigma)$ measures detection performance. Suppose that $p_0 = P(H_0)$ and $p_1 = P(H_1) = 1 - p_0$ are the prior probabilities of the respective hypotheses H_0 and H_1. Let $\alpha(\sigma)$ and $\beta(\sigma)$ be the respective Type-I and Type-II error probabilities when the intensity of the additive noise N is σ:

$$\alpha(\sigma) = P(\text{reject } H_0 | H_0 \text{ is true at noise intensity } \sigma) \quad (2)$$

$$\beta(\sigma) = P(\text{accept } H_0 | H_1 \text{ is true at noise intensity } \sigma). \quad (3)$$

Then define the probability of error as the usual probability-weighted sum of decision errors (Proakis & Salehi, 2008)

$$P_e(\sigma) = p_0\alpha(\sigma) + p_1\beta(\sigma). \quad (4)$$

We assume that the additive noise N is a scale-family noise with pdf $f_N(\sigma, n)$ where σ is the noise intensity (standard deviation or dispersion): $f_N(\sigma, n) = \frac{1}{\sigma} f(\frac{n}{\sigma})$ where f is the standard pdf for the family (Casella & Berger, 2001). Then the noise cumulative distribution function (CDF) is $F_N(\sigma, n) = F(\frac{n}{\sigma})$ where F is the standard CDF for the family.

We next define SR effects in neural signal detection based on error probability. A binary signal detection or hypothesis testing system exhibits the SR effect in the noise intensity interval (a, b) for $0 \le a < b < \infty$ iff $P_{CD}(\sigma_1) < P_{CD}(\sigma_2)$ for any two noise intensities σ_1 and σ_2 such that $a \le \sigma_1 < \sigma_2 \le b$. The SR effect is *total* if $a = 0$ and *partial* if $a \ne 0$. We say that the SR effect occurs at the noise intensity σ iff the SR effect occurs in some noise intensity interval (a, b) and $\sigma \in (a, b)$.

3. Noise benefits in threshold detection of discrete binary signals

We first consider the binary signal detection problem where the signal X is a binary discrete random variable with the two values s_0 and s_1 so that $s_0 < s_1$ and that $P(X = s_0) = p_0$ and $P(X = s_1) = p_1$. Then Theorem 1 gives a necessary and sufficient condition for an SR effect in the threshold neuron model (1) for discrete binary signal detection if the additive noise comes from an absolutely continuous scale-family distribution.

Theorem 1. *Suppose that the additive continuous noise N has scale-family pdf $f_N(\sigma, n)$ and that the threshold neuron model is (1). Suppose that signal X is a binary discrete random variable with the two values s_0 and s_1 so that $P(X = s_0) = p_0$ and $P(X = s_1) = p_1$. Then the SR noise benefit occurs in a given noise intensity interval (a, b) if and only if*

$$p_0(\theta - s_0)f_N(\sigma, \theta - s_0) < p_1(\theta - s_1)f_N(\sigma, \theta - s_1) \quad (5)$$

for almost every noise intensity $\sigma \in (a, b)$. The SR effect is total if

$$\lim_{\sigma \downarrow 0} p_0(\theta - s_0)f_N(\sigma, \theta - s_0) < \lim_{\sigma \downarrow 0} p_1(\theta - s_1)f_N(\sigma, \theta - s_1). \quad (6)$$

Proof. The signal X is a binary discrete random variable with the two values s_0 and s_1. So the Type-I and Type-II error probabilities (2) and (3) become

$$\alpha(\sigma) = 1 - F_N(\sigma, \theta - s_0) \quad (7)$$

$$\beta(\sigma) = F_N(\sigma, \theta - s_1) \quad (8)$$

where F_N is the absolutely continuous CDF of the additive noise random variable N. Then the error probability $P_e(\sigma) = p_0\alpha(\sigma) + p_1\beta(\sigma)$ is an absolutely continuous function of σ in any closed interval $[c, d] \subset \mathbf{R}^+$ where $c > 0$. Then the above definition of SR effects and the fundamental theorem of calculus (Folland, 1999) imply that the SR effect occurs in the noise intensity interval (a, b) if and only if $\frac{dP_e(\sigma)}{d\sigma} < 0$ for almost all $\sigma \in (a, b)$. So the SR effect occurs in the noise intensity interval (a, b) if and only if

$$0 < -p_1 \frac{\partial F_N(\sigma, \theta - s_1)}{\partial \sigma} - p_0 \frac{\partial [1 - F_N(\sigma, \theta - s_0)]}{\partial \sigma} \quad (9)$$

for almost all $\sigma \in (a, b)$. Rewrite (9) as

$$0 < -p_1 \frac{\partial F(\frac{\theta - s_1}{\sigma})}{\partial \sigma} - p_0 \frac{\partial [1 - F(\frac{\theta - s_0}{\sigma})]}{\partial \sigma} \quad (10)$$

where F is the standard scale-family CDF of the additive noise N. Then (10) gives

$$0 < p_1 \frac{(\theta - s_1)}{\sigma} f\left(\frac{\theta - s_1}{\sigma}\right) - p_0 \frac{(\theta - s_0)}{\sigma} f\left(\frac{\theta - s_0}{\sigma}\right)$$
$$= p_1(\theta - s_1)f_N(\sigma, \theta - s_1) - p_0(\theta - s_0)f_N(\sigma, \theta - s_0) \quad (11)$$

because the additive noise N has scale-family pdf $f_N(\sigma, n) = \frac{1}{\sigma} f(\frac{n}{\sigma})$ and because the noise scale σ is always positive. Inequality (5) now follows from (11). The definition of a limit implies that condition (5) holds for all $\sigma \in (0, b)$ for some $b > 0$ if (6) holds. So (6) is a sufficient condition for the total SR effect in the simple threshold detection of discrete binary random signals. □

Theorem 1 lets users screen for total or partial SR noise benefits in discrete binary signal detection for a wide range of noise pdfs. This screening test can prevent a fruitless search for nonexistent noise benefits in many signal-noise contexts. The inequality (5) leads to the simple stochastic learning law in Section 5 that can find the optimal noise dispersion when a noise benefit exists. The learning algorithm does not require a closed-form noise pdf.

Inequality (5) differs from similar inequalities in standard detection theory for likelihood ratio tests. It specifically resembles but differs from the maximum a posteriori (MAP) likelihood ratio test in detection theory (Proakis & Salehi, 2008):

$$\begin{cases} \text{Reject } H_0 & \text{if } p_0 f_N(\sigma, z - s_0) < p_1 f_N(\sigma, z - s_1) \\ \text{Else accept } H_0. \end{cases} \quad (12)$$

The MAP rule (12) minimizes the detection-error probability in optimal signal detection. But it requires the noisy observation z of the received signal $Z = X + N$ whereas inequality (5) does not. Inequality (5) also contains the differences θ-s_0 and θ-s_1. So Theorem 1 gives a general way to detect SR noise benefits in *suboptimal* detection.

Theorem 1 implies a forbidden-interval necessary condition if the noise pdf comes from a location-scale family.

Corollary 1. *Suppose that the additive noise N has location-scale family pdf $f_N(\sigma, n)$. Then the SR noise benefit effect occurs only if the noise mean or location μ obeys the forbidden-interval condition $\mu \notin (\theta - s_1, \theta - s_0)$.*

Proof. The equivalent signal detection problem is H_0: $s_0+\mu$ versus H_1: $s_1+\mu$ in the additive scale-family noise $\tilde{N} = N-\mu$ if we absorb μ in the signal X. Then inequality (5) becomes

$$p_0(\theta - s_0 - \mu)f_{\tilde{N}}(\sigma, \theta - s_0 - \mu) < p_1(\theta - s_1 - \mu)f_{\tilde{N}}(\sigma, \theta - s_1 - \mu). \quad (13)$$

So the SR noise benefit does not occur if $\mu \in (\theta-s_1, \theta-s_0)$ because then the right-hand side of inequality (13) would be negative and the left-hand side would be positive. □

Corollary 1 differs from similar forbidden-interval conditions based on mutual information. The corollary shows that the interval condition $\mu \notin (\theta - s_1, \theta - s_0)$ is only necessary for a noise benefit based on error probability. It is not sufficient. But the same interval condition is both necessary and sufficient for a noise benefit based on mutual information (Kosko & Mitaim, 2003, 2004).

Fig. 1(a)–(e) show simulation instances of Theorem 1 for zero-mean uniform, Gaussian, Laplacian, generalized Gaussian ($r = \frac{1}{2}$), and for zero-location Cauchy noise when $s_0 = 0, s_1 = 1$, $p_0 = p_1$, and $\theta = 1.5$. There is a noise benefit in (a)–(e) because the zero-mean uniform, Gaussian, Laplacian, generalized Gaussian ($r = \frac{1}{2}$), and zero-location Cauchy noise satisfy condition (5) of Theorem 1. Uniform noise gives the largest SR effect but the detection performance degrades quickly as the noise intensity increases beyond its optimal value. Laplacian, generalized Gaussian ($r = \frac{1}{2}$), and Cauchy noise give a more sustained SR effect but with less peak detection performance. The SR effect is partial in Fig. 1(a) because condition (6) does not hold while the SR effects in Fig. 1(b)–(e) are total because (6) holds for Gaussian, Laplacian, generalized Gaussian ($r = \frac{1}{2}$), and Cauchy noise.

Fig. 1(f) shows a simulation instance of Corollary 1 for $s_0 = 0$, $s_1 = 1$, and $\theta = 1.5$ with additive Laplacian noise N with mean $\mu = 1$: $N \sim L(1, \sigma^2)$. Noise does not benefit the detection performance because the noise mean $\mu = 1$ lies in the forbidden interval $(\theta - s_1, \theta - s_0) = (0.5, 1.5)$.

4. Closed-form optimal SR noise

Theorem 1 permits the exact calculation of the optimal SR noise in some special but important cases of closed-form noise pdfs. The noise pdf here comes from a scale family and has zero mean or location. The binary signals are "weak" in the sense that they are subthreshold: $s_0 < s_1 < \theta$. Then Theorem 1 implies that the optimal noise intensity (standard deviation or dispersion) σ_o obeys

$$p_0(\theta - s_0)f_N(\sigma, \theta - s_0) = p_1(\theta - s_1)f_N(\sigma, \theta - s_1) \quad (14)$$

if the noise density is unimodal. Eq. (14) may be nonlinear in terms of σ and so may require a root-finding algorithm to compute the optimal noise intensity σ_o. But we can still directly compute the optimal noise values for several common closed-form noise pdfs that include generalized Gaussian pdfs.

The generalized Gaussian distribution (Nadarajah, 2005) is a two-parameter family of symmetric continuous pdfs. The scale-family pdf $f(\sigma, n)$ of a generalized Gaussian noise has the form

$$f(\sigma, n) = \frac{1}{\sigma}f_{gg}\left(\frac{n}{\sigma}\right) = \frac{1}{\sigma}\frac{r}{2}\frac{[\Gamma(3/r)]^{1/2}}{[\Gamma(1/r)]^{3/2}}e^{-B|\frac{n}{\sigma}|^r} \quad (15)$$

where f_{gg} is the standard pdf of the family, r is a positive shape parameter, Γ is the gamma function, $B = \left[\frac{\Gamma(3/r)}{\Gamma(1/r)}\right]^{\frac{r}{2}}$, and σ is the scale parameter (standard deviation). This family of pdfs includes all normal ($r = 2$) and Laplace distributions ($r = 1$). It includes in the limit ($r \to \infty$) all continuous uniform distributions on bounded real intervals. This family can also model symmetric platykurtic densities whose tails are heavier than normal ($r < 2$) or symmetric leptokurtic densities whose tails are lighter than normal ($r > 2$). Applications include noise modeling in image, speech, and multimedia processing (Bazi, Bruzzone, & Melgani, 2007; Gazor & Zhang, 2003; Krupiński & Purczyński, 2006). Putting (15) in (14) gives the optimal intensity of generalized Gaussian noise as

$$\sigma_o = \left[\frac{B(|\theta_0 - s_0|^r - |\theta_1 - s_1|^r)}{ln(\theta_0 - s_0) - \ln(\theta_1 - s_1) + \ln(p_0) - \ln(p_1)}\right]^{\frac{1}{r}}. \quad (16)$$

We can now state the closed-form optimal noise dispersions for uniform, Gaussian, Laplacian, generalized Gaussian ($r = \frac{1}{2}$), and Cauchy noise.

- *Uniform noise*: Let N be uniform noise in the interval $[-v, v]$ so that $f_N(\sigma, n) = \frac{1}{2v}$ if $n \in [-v, v]$ and $f_N(\sigma, n) = 0$ else. Then the noise standard deviation $\sigma = v/\sqrt{3}$. So inequality (5) holds if and only if either $\theta - s_1 < v < \theta - s_0$ or $\theta - s_0 < v$ but $\frac{p_1}{p_0} > \frac{\theta - s_0}{\theta - s_1}$. So then the SR effect occurs if and only if $\sigma \in \left(\frac{(\theta - s_1)}{\sqrt{3}}, \frac{(\theta - s_0)}{\sqrt{3}}\right)$ when $p_0 = p_1$. Fig. 1(a) shows that the SR effect is partial and the unimodal detection performance is maximal at the optimal noise standard deviation $\sigma_o = \frac{(\theta - s_0)}{\sqrt{3}} = 0.866$ when $s_0 = 0, s_1 = 1$, and $\theta = 1.5$.
- *Gaussian noise* ($r = 2$): Eq. (16) implies that the unimodal detection performance for Gaussian noise is maximal at the noise standard deviation

$$\sigma_o = \left[\frac{\left[(\theta - s_0)^2 - (\theta - s_1)^2\right]}{2\left[\ln(\theta - s_0) - \ln(\theta - s_1) + \ln(p_0) - \ln(p_1)\right]}\right]^{\frac{1}{2}}. \quad (17)$$

  Fig. 1(b) shows a Monte–Carlo simulation plot (10^6 runs) of detection performance when $s_0 = 0, s_1 = 1, p_0 = p_1$, and $\theta = 1.5$. The dashed vertical line shows that the maximal detection performance occurs at the predicted optimal noise intensity $\sigma_o = 0.954$.
- *Laplacian noise* ($r = 1$): Eq. (16) gives the optimal standard deviation σ_o of the Laplacian noise as

$$\sigma_o = \frac{2(s_1 - s_0)^2}{[\ln(\theta - s_0) - \ln(\theta - s_1) + \ln(p_0) - \ln(p_1)]^2}. \quad (18)$$

  The dashed vertical line in Fig. 1(c) shows that the maximum detection performance occurs at the predicted optimal noise scale $\sigma_o = 1.3493$ for $s_0 = 0, s_1 = 1$, and $\theta = 1.5$.
- *Generalized Gaussian* ($r = \frac{1}{2}$) noise: Eq. (16) gives the optimal standard deviation σ_o of generalized Gaussian noise with $r = \frac{1}{2}$ as

$$\sigma_o = \left[\frac{(120)^{\frac{1}{4}}\left[(\theta - s_0)^{\frac{1}{2}} - (\theta - s_1)^{\frac{1}{2}}\right]}{\ln(\theta - s_0) - \ln(\theta - s_1) + \ln(p_0) - \ln(p_1)}\right]^2. \quad (19)$$

  Fig. 1(c) shows that the maximal detection performance occurs at the predicted optimal noise scale $\sigma_o = 2.4319$ when $s_0 = 0, s_1 = 1$, and $\theta = 1.5$ Fig. 1(b)–(e) also show that the peak SR effect decreases and the related optimal noise intensity increases as the shape parameter r of the generalized Gaussian pdf decreases. Simulations showed that the SR effect decayed slowly with increasing noise intensity far beyond σ_o as r decreased.
- *Cauchy noise*: A zero-location infinite-variance Cauchy noise with scale parameter or dispersion σ has pdf $f_N(\sigma, n) = \frac{\sigma}{\pi(\sigma^2+n^2)}$. Then (16) implies that

$$\sigma_o = \sqrt{(\theta - s_0)(\theta - s_1)} \quad (20)$$

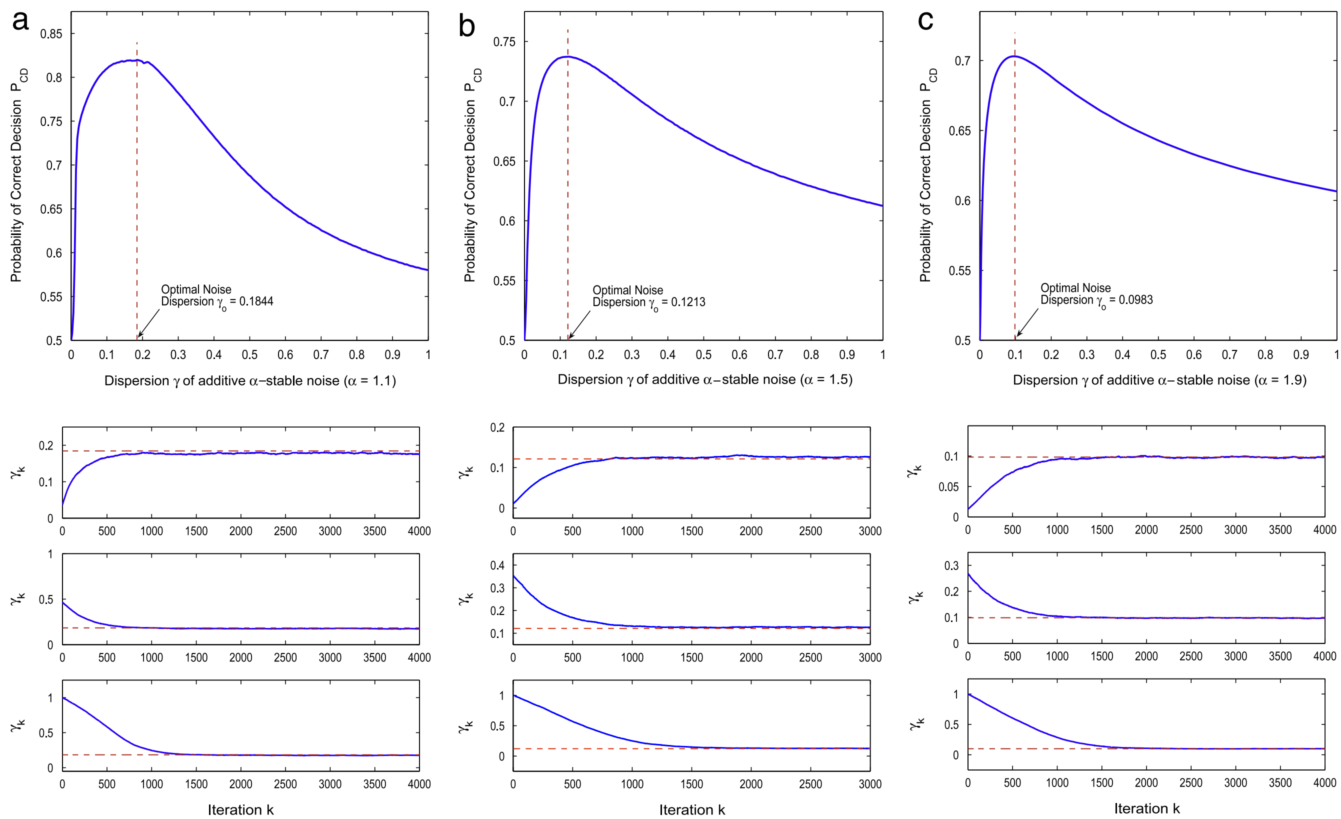

Fig. 2. Adaptive SR. A gradient-ascent learning algorithm learns the optimal noise dispersions for infinite-variance α-stable noise. The bipolar signal X can take either the value $s_0 = -0.4$ or $s_1 = 0.4$ with equal probability ($p_0 = P(s_0) = P(s_1) = p_1$) when $\theta = 0.5$ is the detection threshold. We decide $X = s_1$ if the noisy observation $X + N > \theta$ and otherwise $X = s_0$. The additive noise is α-stable with tail-thickness parameter α: (a) $\alpha = 1.1$, (b) $\alpha = 1.5$, and (c) $\alpha = 1.9$. The graphs at the top show the nonmonotonic signatures of SR where peak detection occurs at the optimal dispersion γ_o. The sample paths in the bottom plots show the convergence of the noise dispersion γ_k from the initial noise scale to the optimum γ_o. The initial noise dispersions are 0.05, 0.5, and 1 with constant learning rate $c_k = 0.1$.

is the optimal Cauchy noise dispersion. The dashed vertical line in Fig. 1(d) shows that the maximal detection performance occurs at the predicted optimal noise scale $\sigma_o = 0.866$ for $s_0 = 0, s_1 = 1$, and $\theta = 1.5$.

5. Adaptive noise optimization

Noise adaptation can find the optimal noise variance or dispersion when a closed-form noise pdf is not available (Kosko & Mitaim, 2004; Mitaim & Kosko, 1998, 2004). This applies to almost all zero-location symmetric α-stable ($S\alpha S$) random variables because their bell-curve pdfs have no known closed form. These thick-tailed bell curves often model impulsive noise in many environments (Nikias & Shao, 1995). An $S\alpha S$ random variable does have a characteristic function φ with a known exponential form (Grigoriu, 1995; Nikias & Shao, 1995):

$$\varphi(\omega) = \exp(j\delta\omega - \gamma|\omega|^{\alpha}) \tag{21}$$

where finite δ is the location parameter, $\gamma = \sigma^{\alpha} > 0$ is the dispersion that controls the width of the bell curve, and $\alpha \in (0, 2]$ controls the bell curve's tail thickness. The $S\alpha S$ bell curve's tail gets thicker as α falls. The only known closed-form $S\alpha S$ pdfs are the thick-tailed Cauchy with $\alpha = 1$ and the thin-tailed Gaussian with $\alpha = 2$. The Gaussian pdf alone among $S\alpha S$ pdfs has a finite variance and finite higher-order moments. The rth lower-order moments of an α-stable pdf with $\alpha < 2$ exist if and only if $r < \alpha$. The location parameter δ serves as the proxy for the mean for $1 < \alpha \le 2$ and as only the median for $0 < \alpha \le 1$.

An adaptive SR learning law updates the dispersion parameter based on how the parameter increases the system performance measure (Mitaim & Kosko, 1998). So using the probability of correct detection P_{CD} gives the following gradient-ascent algorithm:

$$\gamma_{k+1} = \gamma_k + c_k \frac{\partial P_{CD}}{\partial \gamma}. \tag{22}$$

Here c_k is an appropriate learning coefficient at iteration k. The chain rule of calculus implies that $\frac{\partial P_{CD}}{\partial \gamma} = \frac{\partial P_{CD}}{\partial \sigma}\frac{\partial \sigma}{\partial \gamma} = \frac{\partial P_{CD}}{\partial \sigma}\alpha\gamma^{\frac{1-\alpha}{\alpha}}$. Here $\frac{\partial P_{CD}}{\partial \sigma} = p_1(\theta - s_1)f_N(\theta - s_1) - p_0(\theta - s_0)f_N(\theta - s_0)$ because $P_{CD}(\sigma) = 1 - P_e(\sigma)$ and $\frac{\partial P_e(\sigma)}{\partial \sigma} = p_0(\theta - s_0)f_N(\theta - s_0) - p_1(\theta - s_1)f_N(\theta - s_1)$ from (11). Then the learning law (22) becomes

$$\gamma_{k+1} = \gamma_k + c_k \left[p_{1_k}(\theta - s_1)f_{N_k}(\theta - s_1) - p_{0_k}(\theta - s_0)f_{N_k}(\theta - s_0)\right]\alpha\gamma_k^{\frac{1-\alpha}{\alpha}}. \tag{23}$$

Fig. 2 shows how (23) can find the optimal noise dispersion for zero-location symmetric α-stable ($S\alpha S$) random variables that do not have closed-form densities. We need to estimate the signal probabilities p_{i_k} and noise density f_{N_k} at each iteration k. So we generated 500 signal-noise random samples $\{s_l, n_l\}$ for $l = 1, \ldots, 500$ at each k and then used them to estimate the signal probabilities and noise density with their respective histograms. Fig. 2 shows the SR profiles and noise-dispersion learning paths for different α-stable noise types. We used a constant learning rate $c_k = 0.1$ and started the noise level from several initial conditions with different noise seeds. All the learning paths quickly converged to the optimal noise dispersion γ_0.

6. Noise benefits in threshold detection of signals with continuous probability densities

Consider a more general binary hypothesis test $H_0 : F_0$ versus $H_1 : F_1$ where F_0 and F_1 are the absolutely continuous signal CDFs under the respective hypotheses H_0 and H_1. Assume again the threshold neuron model of (1). Then Theorem 2 gives a necessary and sufficient condition for an SR effect in general threshold signal detection. The proof extends the proof of Theorem 1 and uses the theory of generalized functions.

Theorem 2. *Suppose that the signal CDFs F_0 and F_1 are absolutely continuous and that the additive noise N has scale-family pdf $f_N(\sigma, n)$. Then the SR noise benefit occurs in a given noise intensity interval (a, b) if and only if*

$$p_0 \int_R nf_0(\theta - n)f_N(\sigma, n)\mathrm{d}n < p_1 \int_R nf_1(\theta - n)f_N(\sigma, n)\mathrm{d}n \tag{24}$$

for almost all $\sigma \in (a, b)$. The above condition also holds if the noise is discrete when appropriate sums replace the integrals.

Proof. Write the respective Type-I and Type-II error probabilities (2) and (3) as

$$\alpha(\sigma) = \int_R [1 - F_0(\theta - n)]f_N(\sigma, n)\mathrm{d}n \tag{25}$$

$$\beta(\sigma) = \int_R F_1(\theta - n)f_N(\sigma, n)\mathrm{d}n \tag{26}$$

where appropriate sums replace the integrals if the noise is discrete. Then the error probability $P_e(\sigma) = p_0\alpha(\sigma) + p_1\beta(\sigma)$ is an absolutely continuous function of σ in any closed interval $[c, d] \subset \mathbf{R}^+$ where $c > 0$. Then the above definition of SR effects and the fundamental theorem of calculus (Folland, 1999) imply that the SR effect occurs in the noise intensity interval (a, b) if and only if $\frac{\mathrm{d}P_e(\sigma)}{\mathrm{d}\sigma} < 0$ for almost all $\sigma \in (a, b)$. So the SR effect occurs in the noise intensity interval (a, b) if and only if

$$\begin{aligned} 0 &< -p_1 \frac{\partial}{\partial\sigma} \int_R F_0(\theta - n)\frac{1}{\sigma} f\left(\frac{n}{\sigma}\right)\mathrm{d}n \\ &\quad - p_0 \frac{\partial}{\partial\sigma} \int_R [1 - F_1(\theta - n)]\frac{1}{\sigma} f\left(\frac{n}{\sigma}\right)\mathrm{d}n \end{aligned} \tag{27}$$

$$\begin{aligned} &= -p_1 \frac{\partial}{\partial\sigma} \int_R F_0(\theta - \sigma\tilde{n})f(\tilde{n})\mathrm{d}\tilde{n} \\ &\quad - p_0 \frac{\partial}{\partial\sigma} \int_R [1 - F_1(\theta - \sigma\tilde{n})]f(\tilde{n})\mathrm{d}\tilde{n} \end{aligned} \tag{28}$$

for almost all $\sigma \in (a, b)$. The last equality follows from the change of variable from $\frac{n}{\sigma}$ to $\tilde{n}$.

We next use the theory of distributions or generalized functions to interchange the order of integration and differentiation. The error probabilities $\alpha(\sigma)$ and $\beta(\sigma)$ are locally integrable (Zemanian, 1987) in the space $\mathbf{R}^+$ of σ because they are bounded. Then $P_e(\sigma)$ is a generalized function of σ (Zemanian, 1987) and hence its distributional derivative always exists. The terms $F_0(\theta - \sigma\tilde{n})f(\tilde{n})$ and $F_1(\theta - \sigma\tilde{n})f(\tilde{n})$ in (28) are also generalized functions of σ for all $\tilde{n} \in \mathbf{R}$ because they too are bounded. Then we can interchange the order of integration and distributional derivative in (28) (Jones, 1982). So the SR effect occurs in the noise intensity interval (a, b) if and only if

$$\begin{aligned} 0 &< -p_1 \int_R \frac{\partial F_0(\theta - \sigma\tilde{n})}{\partial\sigma} f(\tilde{n})\mathrm{d}\tilde{n} \\ &\quad - p_0 \int_R \frac{\partial[1 - F_1(\theta - \sigma\tilde{n})]}{\partial\sigma} f(\tilde{n})\mathrm{d}\tilde{n} \end{aligned} \tag{29}$$

$$= p_1 \int_R \tilde{n}f_0(\theta - \sigma\tilde{n})f(\tilde{n})\mathrm{d}\tilde{n} - p_0 \int_R \tilde{n}f_1(\theta - \sigma\tilde{n})f(\tilde{n})\mathrm{d}\tilde{n} \tag{30}$$

for almost all $\sigma \in (a, b)$. Then inequality (24) follows if we substitute back $\sigma\tilde{n} = n$ and $f_N(\sigma, n) = \frac{1}{\sigma}f(\frac{n}{\sigma})$ in (30). □

Corollary 2 gives a sufficient condition for the SR effect using a zero-mean bipolar discrete noise N.

Corollary 2. *Suppose that the signal pdfs f_0 and f_1 are continuous and that there exist positive numbers r_1 and r_2 such that*

$$p_0[f_0(\theta - r_2) - f_0(\theta + r_1)] < p_1[f_1(\theta - r_2) - f_1(\theta + r_1)] \tag{31}$$

holds. Suppose also that N is additive scale-family noise with standard family pdf $P(N' = -\sqrt{\frac{r_1}{r_2}}) = r_2/(r_1 + r_2)$ and $P(N' = \sqrt{\frac{r_2}{r_1}}) = r_1/(r_1 + r_2)$ so that $N = \sigma N'$ is zero-mean bipolar noise with variance σ^2. Then an SR effect occurs at the noise standard deviation $\sqrt{r_1 r_2}$.

Proof. $N = \sigma N'$ has pdf $P(N = -\sigma\sqrt{\frac{r_1}{r_2}}) = r_2/(r_1 + r_2)$ and $P(N = \sigma\sqrt{\frac{r_2}{r_1}}) = r_1/(r_1 + r_2)$. Then Theorem 2 implies that the SR effect occurs at the noise standard deviation $\sqrt{r_1 r_2}$ if and only if

$$\begin{aligned} &p_0\left[f_0\left(\theta - \sigma\sqrt{\frac{r_2}{r_1}}\right) - f_0\left(\theta + \sigma\sqrt{\frac{r_1}{r_2}}\right)\right] \\ &\quad < p_1\left[f_1\left(\theta - \sigma\sqrt{\frac{r_2}{r_1}}\right) - f_1\left(\theta + \sigma\sqrt{\frac{r_1}{r_2}}\right)\right] \end{aligned} \tag{32}$$

for almost all σ in some open interval that contains $\sqrt{r_1 r_2}$. The inequality (32) holds for $\sigma = \sqrt{r_1 r_2}$ from the hypothesis (31). Then (32) holds for all σ in some open interval that contains $\sqrt{r_1 r_2}$ because the signal pdfs f_0 and f_1 are continuous. □

The next theorem gives a necessary and sufficient condition for the total SR effect when the signal pdfs are continuously differentiable and when the additive discrete noise N is zero-mean bipolar.

Theorem 3. *Suppose that the signal pdfs f_0 and f_1 are continuously differentiable at θ. Suppose also that the additive discrete noise N is zero-mean bipolar. Then the total SR effect occurs if $p_0 f_0'(\theta) > p_1 f_1'(\theta)$ and only if $p_0 f_0'(\theta) \geq p_1 f_1'(\theta)$.*

Proof. Note that $p_0 f_0'(\theta) > p_1 f_1'(\theta)$ if and only if

$$\begin{aligned} &p_0 \lim_{\sigma\downarrow 0}\left[\frac{f_0\left(\theta + \sigma\sqrt{\frac{r_1}{r_2}}\right) - f_0\left(\theta - \sigma\sqrt{\frac{r_2}{r_1}}\right)}{\sigma\left(\sqrt{\frac{r_2}{r_1}} + \sqrt{\frac{r_1}{r_2}}\right)}\right] \\ &\quad > p_1 \lim_{\sigma\downarrow 0}\left[\frac{f_1\left(\theta + \sigma\sqrt{\frac{r_1}{r_2}}\right) - f_1\left(\theta - \sigma\sqrt{\frac{r_2}{r_1}}\right)}{\sigma\left(\sqrt{\frac{r_2}{r_1}} + \sqrt{\frac{r_1}{r_2}}\right)}\right]. \end{aligned}$$

This implies inequality (32) for $\sigma \in (0, b)$ for some $b > 0$. □

Fig. 3 shows a simulation instance of Corollary 2 and Theorem 3 when $\theta = 0.7379$ is the optimal detection threshold in the absence of noise. The signal pdf is equally likely to be either a trimodal Gaussian mixture $f_0 = \frac{1}{3}\frac{e^{-(n+3)^2/2}}{\sqrt{2\pi}} + \frac{1}{3}\frac{e^{-n^2/2}}{\sqrt{2\pi}} + \frac{1}{3}\frac{e^{-(n-3)^2/2}}{\sqrt{2\pi}}$ or a bimodal Gaussian mixture $f_1 = 0.4\frac{e^{-(n+2)^2/2}}{\sqrt{2\pi}} + 0.6\frac{e^{-(n-2)^2/2}}{\sqrt{2\pi}}$. These multimodal mixture densities arise in some neural systems (Min & Appenteng, 1996; Wimmer, Hildebrandt, Hennig, & Obermayer, 2008). The optimal (minimum error-probability) detection in this case requires four thresholds to partition the signal space into acceptance and rejection regions. But the neurons have only one

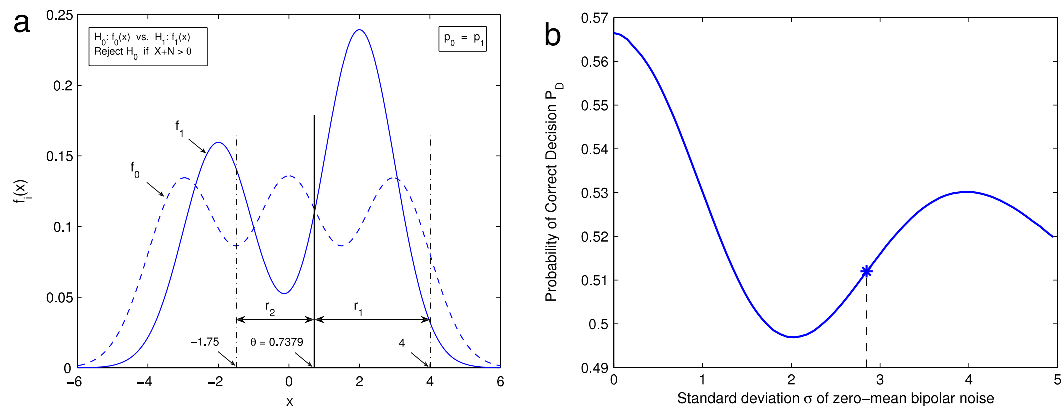

Fig. 3. Partial stochastic resonance (SR) in threshold detection of a binary signal X with continuous pdf in the presence of zero-mean additive discrete bipolar scale-family noise N. (a) The signal pdf is equally likely to be either H_0: $f_0(n) = \frac{1}{3}\frac{e^{-(n+3)^2/2}}{\sqrt{2\pi}} + \frac{1}{3}\frac{e^{-n^2/2}}{\sqrt{2\pi}} + \frac{1}{3}\frac{e^{-(n-3)^2/2}}{\sqrt{2\pi}}$ or H_1: $f_1(n) = 0.4\frac{e^{-(n+2)^2/2}}{\sqrt{2\pi}} + 0.6\frac{e^{-(n-2)^2/2}}{\sqrt{2\pi}}$. The thick vertical line shows the optimal detection threshold $\theta = 0.7379$ when there is no noise. We decide H_1 if the noisy observation $X + N > \theta$. Else we decide H_0. $N = \sigma N'$ is additive scale-family noise with standard family pdf $P(N' = -\sqrt{\frac{r_1}{r_2}}) = r_2/(r_1 + r_2)$ and $P(N' = \sqrt{\frac{r_2}{r_1}}) = r_1/(r_1 + r_2)$. Thus N is zero-mean bipolar noise with variance σ^2. (b) f_0 and f_1 satisfy condition (31) of Corollary 2 for $r_1 = 4 - \theta = 3.2621$ and $r_2 = \theta + 1.75 = 2.4879$. Hence the additive noise N shows a partial SR effect at $\sigma_N = \sqrt{r_1 r_2} = 2.8488$ (marked by the dashed vertical line and ∗). But the total SR effect does not occur because f_0 and f_1 do not satisfy the condition $p_0 f_0'(\theta) > p_1 f_1'(\theta)$ of Theorem 3.

threshold for signal detection. The optimal location is at 0.7379 for such a single detection threshold θ.

Fig. 3(a) shows that f_0 and f_1 satisfy condition (31) of Corollary 2 for $r_1 = 4 - \theta = 3.2621$ and $r_2 = \theta + 1.75 = 2.4879$. Then the SR effect occurs at $\sigma = \sqrt{r_1 r_2} = 2.8488$ because we choose the discrete noise N as in Corollary 2. The signal pdfs f_0 and f_1 do not satisfy the the condition $p_0 f_0'(\theta) > p_1 f_1'(\theta)$ of Theorem 3 at $\theta = 0.7379$. Hence Fig. 3(b) shows that the total SR effect does not occur but a partial SR effect does occur at $\sigma = 2.8488$.

Theorem 4 extends the necessary and sufficient total-SR-effect condition of Theorem 3 to any finite-mean symmetric scale-family additive noise N. This noise can be $S\alpha S$ with infinite variance so long as $\alpha > 1$.

Theorem 4. *Suppose that additive noise N has a finite mean and has a symmetric scale-family pdf. Suppose signal pdfs f_0 and f_1 are bounded and continuously differentiable at θ. Then the total SR effect occurs if*

$$p_0 f_0'(\theta) > p_1 f_1'(\theta) \tag{33}$$

and only if $p_0 f_0'(\theta) \geq p_1 f_1'(\theta)$.

Proof. Inequality (24) of Theorem 2 implies that the total SR effect occurs if and only if

$$p_0 \int_0^\infty n\left(f_0(\theta + n) - f_0(\theta - n)\right) \frac{1}{\sigma} f\left(\frac{n}{\sigma}\right) \mathrm{d}n$$
$$> p_1 \int_0^\infty n\left(f_1(\theta + n) - f_1(\theta - n)\right) \frac{1}{\sigma} f\left(\frac{n}{\sigma}\right) \mathrm{d}n \tag{34}$$

for almost all $\sigma \in (0, b)$ for some $b > 0$ because $f_N(\sigma, n) = \frac{1}{\sigma} f(\frac{n}{\sigma})$ is a symmetric noise pdf. Putting $n = \sigma\tilde{n}$ in (34) gives

$$p_0 \int_0^\infty \sigma\tilde{n} f_0(\theta + \sigma\tilde{n}) - f_0(\theta - \sigma\tilde{n}) f(\tilde{n}) \mathrm{d}\tilde{n}$$
$$> p_1 \int_0^\infty \sigma\tilde{n} f_1(\theta + \sigma\tilde{n}) - f_1(\theta - \sigma\tilde{n}) f(\tilde{n}) \mathrm{d}\tilde{n}. \tag{35}$$

for almost all $\sigma \in (0, b)$. Then for any positive constant L: the definition of a limit implies that (35) holds if

$$\lim_{\sigma\to 0} p_0 \int_0^L \sigma\tilde{n} f_0(\theta + \sigma\tilde{n}) - f_0(\theta - \sigma\tilde{n}) f(\tilde{n}) \mathrm{d}\tilde{n}$$
$$+ \lim_{\sigma\to 0} p_0 \int_L^\infty \sigma\tilde{n} f_0(\theta + \sigma\tilde{n}) - f_0(\theta - \sigma\tilde{n}) f(\tilde{n}) \mathrm{d}\tilde{n}$$
$$> \lim_{\sigma\to 0} p_1 \int_0^L \sigma\tilde{n} f_1(\theta + \sigma\tilde{n}) - f_1(\theta - \sigma\tilde{n}) f(\tilde{n}) \mathrm{d}\tilde{n}$$
$$+ \lim_{\sigma\to 0} p_1 \int_L^\infty \sigma\tilde{n} f_1(\theta + \sigma\tilde{n}) - f_1(\theta - \sigma\tilde{n}) f(\tilde{n}) \mathrm{d}\tilde{n}. \tag{36}$$

where $|\sigma\tilde{n} f_i(\theta + \sigma\tilde{n}) - f_i(\theta - \sigma\tilde{n}) f(\tilde{n})| \leq |\sigma\tilde{n} Q f(\tilde{n})|$ for some number Q because the pdfs f_i are bounded. So Lebesgue's dominated convergence theorem (Folland, 1999) implies that the limit of the second term on both sides of (36) is zero because the additive noise has a finite mean. Then inequality (36) holds if

$$\lim_{\sigma\to 0} p_0 \int_0^L \tilde{n} \frac{f_0(\theta + \sigma\tilde{n}) - f_0(\theta - \sigma\tilde{n})}{\sigma} f(\tilde{n}) \mathrm{d}\tilde{n}$$
$$> \lim_{\sigma\to 0} p_1 \int_0^L \tilde{n} \frac{f_1(\theta + \sigma\tilde{n}) - f_1(\theta - \sigma\tilde{n})}{\sigma} f(\tilde{n}) \mathrm{d}\tilde{n}. \tag{37}$$

The mean-value theorem (Folland, 1999) implies that for any $\epsilon > 0$

$$\left| \frac{f_i(\theta + \sigma\tilde{n}) - f_i(\theta - \sigma\tilde{n})}{2\sigma\tilde{n}} \right| \leq \sup_{|\theta - u| \leq \epsilon} f_i'(u) \tag{38}$$

for all $|\sigma n| \leq \epsilon$. The right-hand side of (38) is bounded for a sufficiently small ϵ because the pdf derivatives f_0' and f_1' are continuous at θ. So Lebesgue's dominated convergence theorem applies to (37) and thus the limit passes under the integral:

$$p_0 \int_0^L 2\tilde{n}^2 \lim_{\sigma\to 0} \frac{f_0(\theta + \sigma\tilde{n}) - f_0(\theta - \sigma\tilde{n})}{2\sigma\tilde{n}} f(\tilde{n}) \mathrm{d}\tilde{n}$$
$$> p_1 \int_0^L 2\tilde{n}^2 \lim_{\sigma\to 0} \frac{f_1(\theta + \sigma\tilde{n}) - f_1(\theta - \sigma\tilde{n})}{2\sigma\tilde{n}} f(\tilde{n}) \mathrm{d}\tilde{n}. \tag{39}$$

Then L'Hospital's rule gives

$$p_0 \int_0^L 2\tilde{n}^2 f_0'(\theta) f(\tilde{n}) \mathrm{d}\tilde{n} > p_1 \int_0^L 2\tilde{n}^2 f_1'(\theta) f(\tilde{n}) \mathrm{d}\tilde{n}. \tag{40}$$

or

$$p_0 f_0'(\theta) \int_0^L 2\tilde{n}^2 f(\tilde{n}) \mathrm{d}\tilde{n} > p_1 f_1'(\theta) \int_0^L 2\tilde{n}^2 f(\tilde{n}) \mathrm{d}\tilde{n}. \tag{41}$$

Thus $p_0 f_0'(\theta) > p_1 f_1'(\theta)$ because the integrals in (41) are positive and finite. □

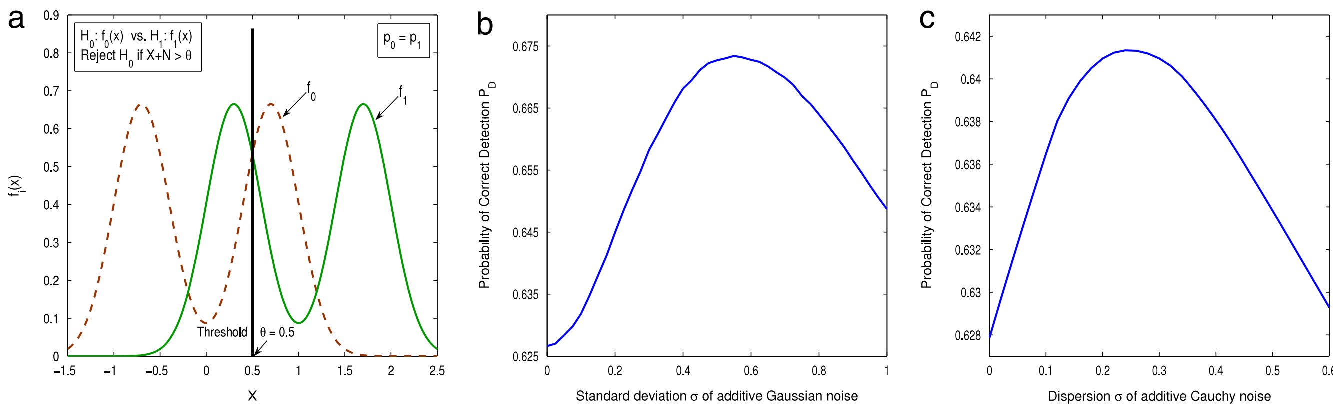

Fig. 4. Total stochastic resonance (SR) in threshold detection of a binary signal X with continuous pdf in the presence of zero-mean additive symmetric scale family noise. (a) The bimodal signal pdf is equally likely to be either H_0: $f_0(n) = \frac{1}{2}\frac{e^{-(n+0.7)^2/2(0.3)^2}}{\sqrt{2\pi}0.3} + \frac{1}{2}\frac{e^{-(n-0.7)^2/2(0.3)^2}}{\sqrt{2\pi}0.3}$ or H_1: $f_1(n) = \frac{1}{2}\frac{e^{-(n-0.3)^2/2(0.3)^2}}{\sqrt{2\pi}0.3} + \frac{1}{2}\frac{e^{-(n-1.7)^2/2(0.3)^2}}{\sqrt{2\pi}0.3}$. The thick vertical line indicates the detection threshold $\theta = 0.5$. We decide H_1 if the noisy observation $X + N > \theta$. Else decide H_0. These signal pdfs satisfy the condition $p_0 f_0'(\theta) > p_1 f_1'(\theta)$ of Theorem 4 at $\theta = 0.5$. The detection performance curves in (b) and (c) show the respective predicted total SR effect for zero-mean Gaussian noise and conjectured SR effect for zero-location Cauchy noise.

Fig. 4(a) plots the following two bimodal Gaussian-mixture signal pdfs

$$f_0(n) = \frac{1}{2}\frac{e^{-(n+0.7)^2/2(0.3)^2}}{\sqrt{2\pi}0.3} + \frac{1}{2}\frac{e^{-(n-0.7)^2/2(0.3)^2}}{\sqrt{2\pi}0.3} \quad (42)$$

$$f_1(n) = \frac{1}{2}\frac{e^{-(n-0.3)^2/2(0.3)^2}}{\sqrt{2\pi}0.3} + \frac{1}{2}\frac{e^{-(n-1.7)^2/2(0.3)^2}}{\sqrt{2\pi}0.3} \quad (43)$$

with detection threshold $\theta = 0.5$. This figure also shows that these signal pdfs satisfy the inequality $p_0 f_0'(\theta) > p_1 f_1'(\theta)$ of Theorem 4. Fig. 4(b) shows the predicted total SR effect for zero-mean uniform noise. The total SR effect in Fig. 4(c) exceeds the scope of Theorem 4 because the mean of the Cauchy noise does not exist. This suggests the conjecture that $p_0 f_0'(\theta) > p_1 f_1'(\theta)$ is a sufficient condition for the total SR effect even when the symmetric scale-family noise has no expected value.

7. Noise benefits in parallel arrays of threshold neurons

We show last how a simple calculus argument ensures a noise benefit for maximum likelihood detection with a large number of parallel arrays or networks of threshold neurons. This SR noise benefit need not occur for a single neuron and so is a genuine *collective* noise benefit.

This total SR result for arrays applies to maximum likelihood detection of two alternative signals. So it resembles but differs from the array SR result in Patel and Kosko (2009c) that applies instead to the Neyman–Pearson detection of a constant signal in infinite-variance symmetric alpha-stable channel noise with a single array of noisy quantizers. We note that Stocks et al. (2002) first showed that adding noise in an array of parallel-connected threshold elements improves the mutual information between the array's input and output. Then Rousseau and Chapeau-Blondeau (Rousseau, Anand, & Chapeau-Blondeau, 2006; Rousseau & Chapeau-Blondeau, 2005a) used such a threshold array for signal detection. They first showed an SR noise benefit for Neyman–Pearson detection and for Bayesian detection. Researchers have also shown mutual-information noise benefits in arrays of threshold neurons (Hoch et al., 2003; McDonnell et al., 2008; Stocks, 2001; Stocks & Mannella, 2001).

Suppose first that the signal pdfs f_i are equally likely ($p_0 = p_1$) and symmetric around m_i so that

$$f_i(m_i + x) = f_i(m_i - x) \quad \text{for all } x \quad (44)$$

where $m_0 < m_1$. Then $\theta = \frac{m_0+m_1}{2}$ is a reasonable detection threshold for individual neurons because of this symmetry. Suppose there are K parallel arrays such that each parallel array has M noisy threshold neurons. Suppose that the kth parallel array receives an independent sample X_k of the input signal X. The output of the mth threshold neuron in the kth array is

$$Y_{m,k}(\sigma_N) = sign(X_k + N_m - \theta) \quad (45)$$

where σ_N is the scale parameter of an additive signal independent symmetric noise N_m in the mth threshold neuron. Assume that the noise random variables N_m are i.i.d. for each neuron and for each array. Suppose last that the output W_k of the kth array sums all M threshold neuron outputs $Y_{m,k}$:

$$W_k(\sigma_N) = \sum_{m=1}^{M} Y_{m,k}(\sigma_N). \quad (46)$$

A summing-threshold neuron combines the outputs W_k from each parallel array into Λ and then uses $sign(\Lambda)$ for maximum-likelihood detection:

$$\Lambda(\sigma_N) = \sum_{k=1}^{K} W_k(\sigma_N) \underset{H_0}{\overset{H_1}{\gtrless}} 0 \quad (47)$$

since the pdf of Λ is symmetric around zero because the signal and noise pdfs are symmetric.

Now define $\mu_i(\sigma_N)$ and $\sigma_i^2(\sigma_N)$ as the respective mean and variance of Λ under the hypothesis H_i when σ_N is the neuron's noise intensity. Then $\mu_0(\sigma_N) = -\mu_1(\sigma_N)$ and $\sigma_0^2(\sigma_N) = \sigma_1^2(\sigma_N)$ for all σ_N also because all signal and noise pdfs are symmetric. The pdf of Λ is approximately Gaussian for either hypothesis because the central limit theorem applies to the sum (47) if the sample size K is large since the summands are i.i.d. (Casella & Berger, 2001). Then Theorem 5 gives a necessary and sufficient condition for an SR effect (total or partial) in the parallel-array detector (45)–(47).

Theorem 5. *Suppose that* $\Lambda(\sigma_N)|H_0 \sim N(\mu_0(\sigma_N), \sigma_0^2(\sigma_N))$ *and* $\Lambda(\sigma_N)|H_1 \sim N(\mu_1(\sigma_N), \sigma_1^2(\sigma_N))$ *where* $\mu_0(\sigma_N) = -\mu_1(\sigma_N)$ *and* $\sigma_0^2(\sigma_N) = \sigma_1^2(\sigma_N)$ *for the threshold-neuron array model* (45)–(47). *Then*

$$\sigma_1(\sigma_N)\mu_1'(\sigma_N) > \mu_1(\sigma_N)\sigma_1'(\sigma_N) \quad (48)$$

is necessary and sufficient for an SR effect at the noise intensity σ_N *in the parallel-array maximum-likelihood detector.*

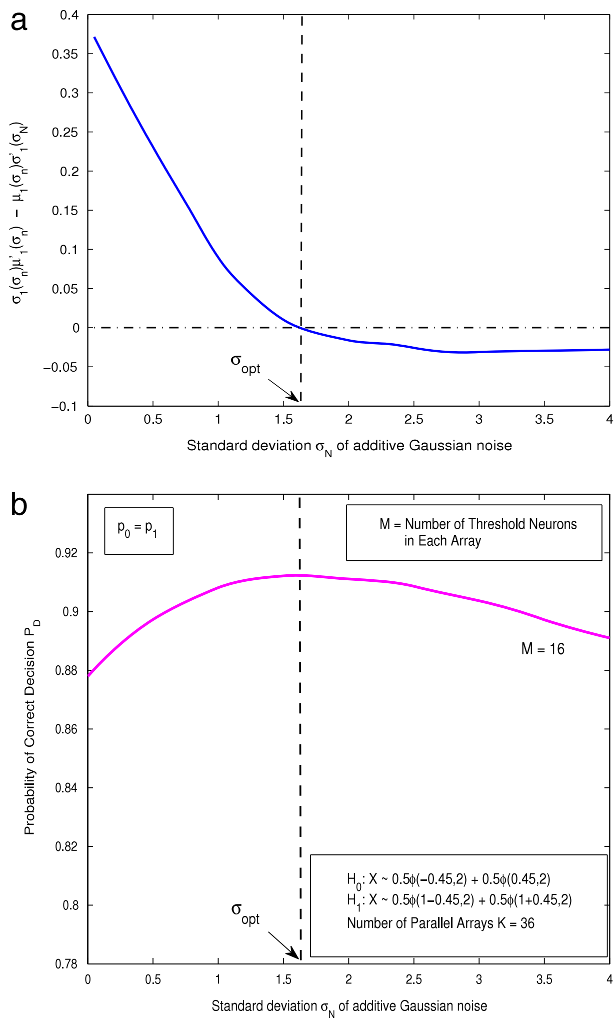

Fig. 5. SR noise benefits in neural array signal detection. (a) The smoothed plot of $\sigma_1(\sigma_N)\mu_1'(\sigma_N) - \mu_1(\sigma_N)\sigma_1'(\sigma_N)$ versus the standard deviation σ_N of additive Gaussian noise. The zero-crossing occurs at the noise standard deviation $\sigma_N = \sigma_{opt}$. (b) The solid line and square markers show the respective plots of the detection probabilities P_D. Adding small amounts of Gaussian noise N to each threshold neuron improves the array detection probability P_D. This SR effect occurs until inequality (48) holds. So σ_{opt} maximizes the detection probability.

Proof. The detection probability obeys

$$P_D = \frac{1}{2}P(\Lambda(\sigma_N) < 0|H_0) + \frac{1}{2}P(\Lambda(\sigma_N) > 0|H_1) \tag{49}$$

$$= \Phi\left(\frac{\mu_1(\sigma_N)}{\sigma_0(\sigma_N)}\right) \tag{50}$$

because $\mu_0(\sigma_N) = -\mu_1(\sigma_N)$ and $\sigma_0^2(\sigma_N) = \sigma_1^2(\sigma_N)$ as discussed above. Here Φ is the CDF of the standard normal random variable. Then the chain and quotient rules of differential calculus give

$$\frac{\mathrm{d}P_D}{\mathrm{d}\sigma_N} = \tilde{\phi}\left(\frac{\mu_1(\sigma_N)}{\sigma_1(\sigma_N)}\right)\frac{\sigma_1(\sigma_N)\mu_1'(\sigma_N) - \mu_1(\sigma_N)\sigma_1{}'(\sigma_N)}{\sigma_1^2(\sigma_N)}. \tag{51}$$

So $\sigma_1(\sigma_N)\mu_1'(\sigma_N) > \mu_1(\sigma_N)\sigma_1'(\sigma_N)$ is necessary and sufficient for the SR effect ($\frac{\mathrm{d}P_D}{\mathrm{d}\sigma_N} > 0$) at the noise intensity σ_N because $\tilde{\phi}$ is the pdf of the standard normal random variable. □

Fig. 5 shows a simulation instance of the SR condition in Theorem 5 for the parallel-array maximum-likelihood detection of Gaussian mixture signals in the hypothesis test of

$$H_0 : f_0(x) = \frac{1}{2}\phi(-0.45, 2, x) + \frac{1}{2}\phi(0.45, 2, x) \tag{52}$$

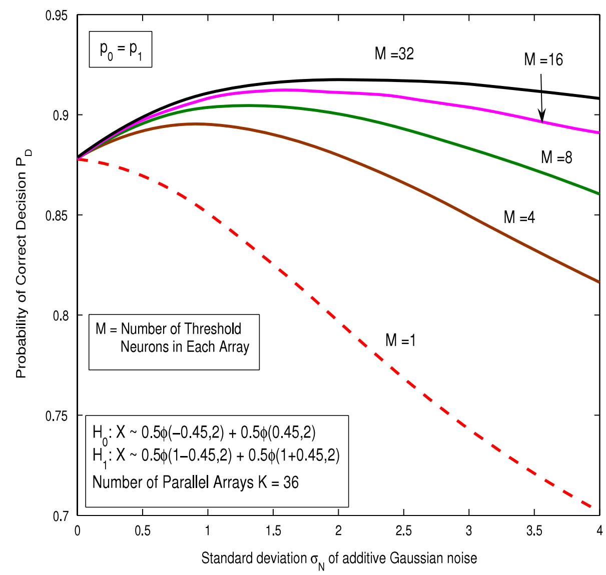

Fig. 6. Collective noise benefits in parallel-array maximum-likelihood signal detection for different values of M (the number of threshold neurons in each of $K = 36$ arrays). The solid lines show that the detection probability P_D improves initially as the noise intensity σ_N increases. The solid lines also show that the SR effect increases as the number M of threshold neurons increases. The dashed line shows that the SR effect does not occur if $M = 1$.

$$\text{versus } H_1 : f_1(x) = \frac{1}{2}\phi(1 - 0.45, 2, x) + \frac{1}{2}\phi(1 + 0.45, 2, x) \tag{53}$$

where $\phi(v, d, x)$ is the pdf of the Gaussian random variable with mean v and variance d^2. We used $K = 36$ parallel arrays and so took $K = 36$ i.i.d. random samples of the input signal X. Each parallel array has $M = 16$ threshold neurons and each receives the same input signal sample. The threshold θ is $\frac{m_0+m_1}{2} = 0.5$ because these signal pdfs f_i are symmetric around $m_0 = 0$ and $m_1 = 1$. The noise N in the neurons are i.i.d. Gaussian random variables. Fig. 5(a) plots smoothed $\sigma_1(\sigma_N)\mu_1'(\sigma_N) - \mu_1(\sigma_N)\sigma_1'(\sigma_N)$ versus the standard deviation σ_N of the additive Gaussian noise. Adding small amounts of noise N before thresholding improves the array's overall detection probability P_D in Fig. 5(b). This total SR effect occurs until the inequality (48) holds in Fig. 5(a).

Fig. 6 shows that the collective noise benefit increases as the number M of threshold neurons increases in each of $K = 36$ arrays. The dashed line shows that the SR effect does not occur for an individual threshold neuron ($M = 1$) because inequality (33) of Theorem 4 does not hold for the specific signal pdfs f_i in (52)–(53). But Fig. 6 shows that the SR effect does occur if we use more than one threshold neuron (if $M > 1$) in a large number ($K > 30$) of parallel arrays. And Fig. 4 shows that the two bimodal Gaussian-mixture signal pdfs f_0 and f_1 in (42)–(43) satisfy inequality (33) of Theorem 4. So one threshold neuron can produce the SR effect.

8. Conclusion

The above five theorems extend and depart from earlier forbidden-interval results that characterize mutual-information SR effects in some types of neural signal detection. These new necessary and sufficient conditions for error-probability SR effects can act as a type of screening procedure that predicts whether a noise benefit will occur in a given system. The error-probability SR learning law can then search for the best noise settings for a predicted noise benefit when a closed-form solution is not available. Other learning laws may arise if the user lacks sufficient or accurate training data. More general SR theorems may also hold when the additive noise's probability density comes from an asymmetric scale family or from still more general families of noise densities.

References

Amblard, P.-O., Zozor, S., McDonnell, M. D., & Stocks, N. G. (2007). *Pooling networks for a discrimination task: Noise-enhanced detection*: *Vol. 6602* (p. 66020S). SPIE.

Auer, P., Burgsteiner, H., & Maass, W. (2008). A learning rule for very simple universal approximators consisting of a single layer of perceptrons. *Neural Networks*, *21*(5), 786–795.

Bazi, Y., Bruzzone, L., & Melgani, F. (2007). Image thresholding based on the em algorithm and the generalized Gaussian distribution. *Pattern Recognition*, *40*(2), 619–634.

Beiu, V., Member, S., Quintana, J. M., & Avedillo, M. J. (2003). Vlsi implementations of threshold logic — a comprehensive survey. *IEEE Transactions on Neural Networks*, *14*, 1217–1243.

Bulsara, A., Jacobs, E., Zhou, T., Moss, F., & Kiss, L. (1991). Stochastic resonance in a single neuron model: Theory and analog simulation. *Journal of Theoretical Biology*, *152*, 531–555.

Casella, G., & Berger, R. (2001). *Statistical inference*. Duxbury Resource Center.

Caticha, N., Palo Tejada, J. E., Lancet, D., & Domany, E. (2002). Computational capacity of an odorant discriminator: The linear separability of curves. *Neural Computation*, *14*(9), 2201–2220.

Chapeau-Blondeau, F., & Rousseau, D. (2004). Noise-enhanced performance for an optimal Bayesian estimator. *Signal Processing, IEEE Transactions on*, *52*(5), 1327–1334.

Deco, G., & Schürmann, B. (1998). Stochastic resonance in the mutual information between input and output spike trains of noisy central neurons. *Physica D*, *117*(1–4), 276–282.

Faisal, A. A., Selen, L. P., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, *9*(4), 292–303.

Folland, G. B. (1999). *Real analysis: Modern techniques and their applications* (second ed.). Wiley-Interscience.

Freund, Y., & Schapire, R. E. (1999). *Large margin classification using the perceptron algorithm*: *Vol. 37* (pp. 277–296).

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

Gazor, S., & Zhang, W. (2003). Speech probability distribution. *Signal Processing Letters, IEEE*, *10*(7), 204–207.

Grigoriu, M. (1995). *Applied non-Gaussian processes*. Prentice Hall.

Hänggi, P. (2002). Stochastic resonance in biology. *ChemPhysChem*, *3*, 285–290.

Hoch, T., Wenning, G., & Obermayer, K. (2003). Optimal noise-aided signal transmission through populations of neurons. *Physical Review E*, *68*, 011911–1–11.

Jones, D. (1982). *The theory of generalized functions* (second ed.). Cambridge University Press.

Kay, S. (2000). Can detectability be improved by adding noise? *Signal Processing Letters, IEEE*, *7*(1), 8–10.

Kay, S., Michels, J., Chen, H., & Varshney, P. (2006). Reducing probability of decision error using stochastic resonance. *Signal Processing Letters, IEEE*, *13*(11), 695–698.

Kosko, B. (2006). *Noise*. Viking/Penguin.

Kosko, B., & Mitaim, S. (2003). Stochastic resonance in noisy threshold neurons. *Neural Networks*, *16*(5–6), 755–761.

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

Krupiński, R., & Purczyński, J. (2006). Approximated fast estimator for the shape parameter of generalized Gaussian distribution. *Signal Processing*, *86*(2), 205–211.

Läer, L., Kloppstech, M., Schöfl, C., Sejnowski, T. J., Brabant, G., & Prank, K. (2001). Noise enhanced hormonal signal transduction through intracellular calcium oscillations. *Biophysical Chemistry*, *91*, 157–166.

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

Li, H., Hou, Z., & Xin, H. (2005). Internal noise enhanced detection of hormonal signal through intracellular calcium oscillations. *Chemical Physics Letters*, *402*(4-6), 444–449.

Manwani, A., & Koch, C. (1999). Detecting and estimating signals in noisy cable structures, i: Neuronal noise sources. *Neural Computation*, *11*(9), 1797–1829.

Markram, H., & Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, *382*, 807–810.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, (5), 115–133. Reprinted in *Neurocomputing: Foundations of Research, ed. by J. A. Anderson and E Rosenfeld*. MIT Press 1988.

McDonnell, M., Stocks, N., Pearce, C., & Abbott, D. (2006). Optimal information transmission in nonlinear arrays through suprathreshold stochastic resonance. *Physics Letters A*, *352*, 183–189.

McDonnell, M. D., Stocks, N. G., Pearce, C. E. M., & Abbott, D. (2008). *Stochastic resonance: From suprathreshold stochastic resonance to stochastic signal quantization*. Cambridge University Press.

Min, M. Y., & Appenteng, K. (1996). Multimodal distribution of amplitudes of miniature and spontaneous epsps recorded in rat trigeminal motoneurones. *Journal Physiology*, *494*, 171–182.

Minsky, M., & Papert, S. (1988). *Perceptrons: An introduction to computational geometry* (expanded ed.). Cambridge, MA: MIT Press.

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

Mitaim, S., & Kosko, B. (2004). Adaptive stochastic resonance in noisy neurons based on mutual information. *IEEE Transactions on Neural Networks*, *15*(6), 1526–1540.

Moss, F., Ward, L. M., & Sannita, W. G. (2004). Stochastic resonance and sensory information processing: A tutorial and review of application. *Clin Neurophysiol*, *115*(2), 267–281.

Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics*, *32*(7), 685–694.

Nikias, C. L., & Shao, M. (1995). *Signal processing with alpha-stable distributions and applications*. New York, NY, USA: Wiley-Interscience.

Pantazelou, E., Dames, C., Moss, F., Douglass, J., & Wilkens, L. (1995). Temperature dependence and the role of internal noise in signal transduction efficiency of crayfish mechanoreceptors. *International Journal of Bifurcation and Chaos*, *5*, 101–108.

Patel, A., & Kosko, B. (2005). Stochastic resonance in noisy spiking retinal and sensory neuron models. *Neural Networks*, *18*(5–6), 467–478.

Patel, A., & Kosko, B. (2008). Stochastic resonance in continuous and spiking neuron models with levy noise. *Neural Networks, IEEE Transactions on*, *19*(12), 1993–2008.

Patel, A., & Kosko, B. (2009a). Neural signal-detection noise benefits based on error probability. In *Proceedings of the international joint conference on neural networks* (pp. 2423–2430).

Patel, A., & Kosko, B. (2009b). Optimal noise benefits in neyman-pearson and inequality-constrained statistical signal detection. *IEEE Transactions on Signal Processing*, *57*(5), 1655–1669.

Patel, A., & Kosko, B. (2009c). Quantizer noise benefits in nonlinear signal detection with alpha-stable channel noise. In *IEEE international conference on acoustic, speech and signal processing* (pp. 3269–3272).

Proakis, J., & Salehi, M. (2008). *Digital communications* (fifth ed.). McGraw Hill.

Rao, C. V., Wolf, D. M., & Arkin, A. P. (2002). Control, exploitation and tolerance of intracellular noise. *Nature*, *420*(6912), 231–237.

Rousseau, D., Anand, G. V., & Chapeau-Blondeau, F. (2006). Noise-enhanced nonlinear detector to improve signal detection in non-Gaussian noise. *Signal Processing*, *86*(11), 3456–3465.

Rousseau, D., & Chapeau-Blondeau, F. (2005a). Constructive role of noise in signal detection from parallel array of quantizers. *Signal Processing*, *85*(3), 571–580.

Rousseau, D., & Chapeau-Blondeau, F. (2005b). Stochastic resonance and improvement by noise in optimal detection strategies. *Digital Signal Processing*, *15*(1), 19–32.

Saha, A. A., & Anand, G. V. (2003). Design of detectors based on stochastic resonance. *Signal Processing*, *83*(6), 1193–1212.

Sasaki, H., Sakane, S., Ishida, T., Todorokihara, M., Kitamura, T., & Aoki, R. (2008). Suprathreshold stochastic resonance in visual signal detection. *Behavioural Brain Research*, *193*(1), 152–155.

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*(7), 1679–1703.

Stacey, W. C., & Durand, D. M. (2000). Stochastic resonance improves signal detection in hippocampal ca1 neurons. *Journal Neurophysiology*, *83*, 1394–1402.

Stocks, N., Appligham, D., & Morse, R. (2002). The application of suprathreshold stochastic resonance to cochlear implant coding. *Fluctuation and Noise Letters*, *2*, L169–L181.

Stocks, N. G. (2001). Information transmission in parallel threshold arrays: Suprathreshold stochastic resonance. *Physical Review E*, *63*(4), 041114.

Stocks, N. G., & Mannella, R. (2001). Generic noise-enhanced coding in neuronal arrays. *Physical Review E*, *64*(3), 030902–1–4.

Wang, W., & Wang, Z. D. (1997). Internal-noise-enhanced signal transduction in neuronal systems. *Physical Review E*, *55*(6), 7379–7384.

White, J. A., Rubinstein, J. T., & Kay, A. R. (2000). Channel noise in neurons. *Trends in Neuroscience*, *23*(3), 131–137.

Wiesenfeld, K., & Moss, F. (1995). Stochastic resonance and the benefits of noise: From ice ages to crayfish and squid. *Nature (London)*, *373*, 33.

Wimmer, K., Hildebrandt, K., Hennig, R., & Obermayer, K. (2008). Adaptation and selective information transmission in the cricket auditory neuron AN2. *PLoS Computational Biology*, *4*(9), e1000182+.

Zemanian, A. H. (1987). *Distribution theory and transform analysis: An introduction to generalized functions, with applications*. New York, NY, USA: Dover Publications, Inc.