AVAILABLE AT
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT

PERGAMON

Neural Networks 16 (2003) 755–761

Neural Networks

www.elsevier.com/locate/neunet

2003 Special issue

Stochastic resonance in noisy threshold neurons

Bart Kosko[a,*], Sanya Mitaim[b]

[a]Department of Electrical Engineering, Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089-2564, USA
[b]Department of Electrical Engineering, Faculty of Engineering, Thammasat University, Klong Luang, Pathumthani 12121, Thailand

Abstract

Stochastic resonance occurs when noise improves how a nonlinear system performs. This paper presents two general stochastic-resonance theorems for threshold neurons that process noisy Bernoulli input sequences. The performance measure is Shannon mutual information. The theorems show that small amounts of independent additive noise can increase the mutual information of threshold neurons if the neurons detect subthreshold signals. The first theorem shows that this stochastic-resonance effect holds for all finite-variance noise probability density functions that obey a simple mean constraint that the user can control. A corollary shows that this stochastic-resonance effect occurs for the important family of (right-sided) gamma noise. The second theorem shows that this effect holds for all infinite-variance noise types in the broad family of stable distributions. Stable bell curves can model extremely impulsive noise environments. So the second theorem shows that this stochastic-resonance effect is robust against violent fluctuations in the additive noise process.

Keywords: Stochastic resonance; Noise processing; Threshold neurons; Infinite-variance noise; Mutual information

1. The benefits of noise

Noise can sometimes help neural or other nonlinear systems. Fig. 1 shows that small amounts of Gaussian pixel noise improves the standard 'baboon' image while too much noise degrades the image.

Small amounts of additive noise can also improve the performance of threshold neurons or of neurons with steep signal functions when the neurons process noisy Bernoulli sequences. Several researchers have found some form of this "stochastic resonance" (SR) effect (Bulsara & Zador, 1996; Collins, Chow, Capela, & Imhoff, 1996; Collins, Chow, & Imhoff, 1995; Douglass, Wilkens, Pantazelou, & Moss, 1993; Gammaitoni, 1995; Godivier & Chapeau-Blondeau, 1998; Hess & Albano, 1998; Jung, 1995; Jung & Mayer-Kress, 1995; Stocks, 2001) when either mutual information or input–output correlation (or signal-to-noise ratio) measures a neuron's response to a pulse stream of noisy subthreshold signals. But these studies have all used simple *finite*-variance noise types such as Gaussian or uniform noise. They further assume that the noise is both symmetric and two-sided (hence zero mean). We show that SR still occurs if the noise violates these assumptions.

The two theorems below establish that the mutual-information form of the SR effect occurs for almost all noisy threshold neurons. The first theorem holds for any finite-variance noise type that obeys a simple mean condition. A corollary shows that the SR effect still occurs for right-sided noise from the popular family of gamma probability density functions. Fig. 3 shows some simulation instances of this corollary. The second theorem holds for any infinite-variance noise type from the broad family of stable distributions. All signals are subthreshold.

Infinite variance does not imply infinite dispersion. Stable probability densities have finite dispersions but have infinite variances and infinite higher-order moments. The dispersion controls the width of the bell curve for symmetric stable densities (see Fig. 4). Fig. 2 shows a simulation instance of the second theorem. Infinite-variance Cauchy noise corrupts the subthreshold signal stream but still produces the characteristic nonmonotonic signature of SR. The theorem on infinite-variance noise implies that the SR effect is robust against impulsive noise: a threshold neuron can extract some information-theoretic gain even from noise streams that contain occasional violent spikes of noise. The noise stream itself is a local form of free energy that neurons can exploit.

The combined results support Linsker's hypothesis (Linsker, 1988, 1997) that neurons have evolved to

* Corresponding author.

0893-6080/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0893-6080(03)00128-X

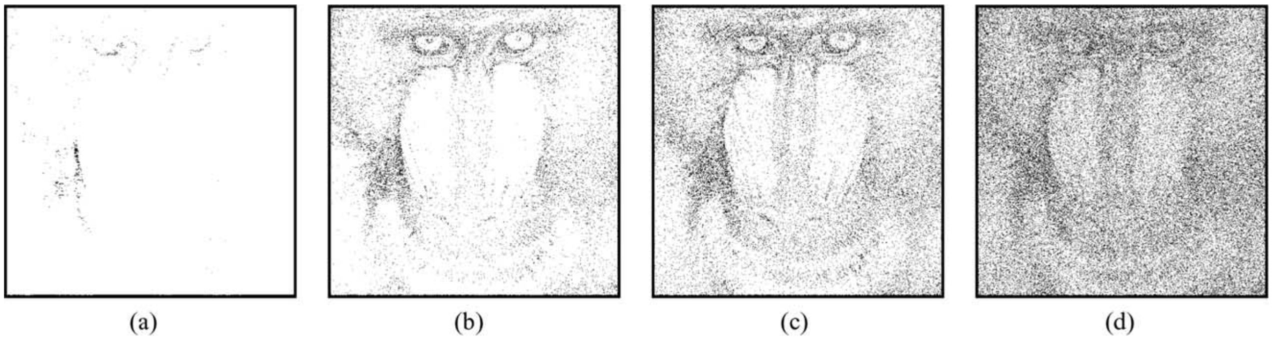

Fig. 1. Gaussian pixel noise can improve the quality of an image through a stochastic-resonance or dithering process (Gaimmaitoni, 1995; Wannamaker, Lipshitz & Vanderkooy, 2000). The noise produces a nonmonotonic response: A small level of noise sharpens the image features while too much noise degrades them. These noisy images result when we apply a pixel threshold to the 'baboon' image. The system first quantizes the original gray-scale baboon image into a binary image of black and white pixels. It gives a white pixel as output if the input gray-scale pixel equals or exceeds a threshold θ. It gives a black pixel as output if the input gray-scale pixel falls below the threshold $\theta : y = g((x + n) - \theta)$ where $g(x) = 1$ if $x \geq 0$ and $g(x) = 0$ if $x < 0$ for an input pixel value$x \in [0, 1]$ and output pixel value $y \in \{0, 1\}$. The input image's gray-scale pixels vary from 0 (black) to 1 (white). The threshold is $\theta = 0.04$. Thresholding the original baboon image gives the faint image in (a). The Gaussian noise n has zero mean for images (b)–(d). The noise variance σ_n^2 grows from (b) to (d): $\sigma_n^2 = 1.00 \times 10^{-2}$ in (b), $\sigma_n^2 = 2.25 \times 10^{-2}$ in (c), and $\sigma_n^2 = 9.00 \times 10^{-2}$ in (d).

maximize the information content of their local environment. The new twist to the hypothesis is that maximizing a threshold neuron's mutual information requires deliberate use of environmental noise.

2. Threshold neurons and Shannon's mutual information

We use the standard discrete-time threshold neuron model (Bulsara & Zador, 1996; Gammaitoni, 1995; Hopfield, 1982; Jung, 1995; Kosko, 1991; Kosko & Mitaim, 2001)

$$y_t = \mathrm{sgn}(s_t + n_t - \theta) = \begin{cases} 1 & \text{if } s_t + n_t \geq \theta \\ 0 & \text{if } s_t + n_t < \theta \end{cases} \tag{1}$$

where $\theta > 0$ is the neuron's threshold, s_t is the bipolar input Bernoulli signal (with arbitrary success probability p such that $0 < p < 1$) with amplitude $A > 0$, and n_t is the additive white noise with probability density $p(n)$.

The threshold neuron uses subthreshold binary signals. The symbol '0' denotes the input signal $s = -A$ and output signal

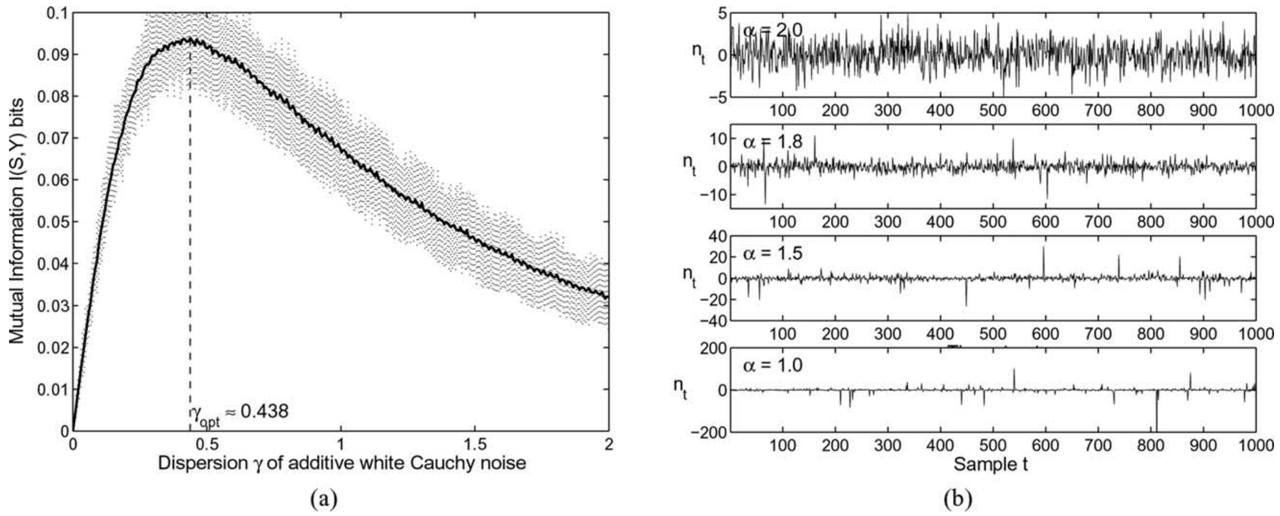

Fig. 2. SR with infinite-variance Cauchy noise. (a) The graph shows the smoothed input-output mutual information of a threshold neuron as a function of the dispersion of additive white Cauchy noise n_t. The dispersion γ controls the width of the Cauchy bell curve. The vertical dashed lines show the absolute deviation between the smallest and largest outliers in each sample average of 100 outcomes. The neuron has a nonzero noise optimum at $\gamma_{\text{opt}} \approx 0.438$ and thus shows the SR effect. The noisy signal-forced threshold neuron has the form of Eq. (1). The Cauchy noise n_t adds to the bipolar input Bernoulli signal s_t. The neuron has threshold $\theta = 1$. The input Bernoulli signal has amplitude $A = 0.8$ with success probability $p = \frac{1}{2}$. Each trial produced 10,000 input–output samples $\{s_t, y_t\}$ that estimated the probability densities to obtain the mutual information. (b) Sample realizations of symmetric (bell-curve) alpha-stable random variables with zero location ($a = 0$) and unit dispersion ($\gamma = 1$). The plots show realizations when $\alpha = 2,1.8,1.5$, and 1. Note the scale differences on the y-axes. The alpha-stable variable n becomes more impulsive as the parameter α falls. The algorithm in (Chambers, Mallows, & Stuck, 1976; Tsakalides & Nikias, 1996) generated these realizations.

$y = 0$. The symbol '1' denotes the input signal $s = A$ and output signal $y = 1$. We assume subthreshold input signals: $A < \theta$. Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$P_{Y|S}(0|0) = Pr\{s + n < \theta\}|_{s=-A} = Pr\{n < \theta + A\}$$

$$= \int_{-\infty}^{\theta+A} p(n)\mathrm{d}n \quad (2)$$

$$P_{Y|S}(1|0) = 1 - P_{Y|S}(0|0) \quad (3)$$

$$P_{Y|S}(0|1) = Pr\{s + n < \theta\}|_{s=A} = Pr\{n < \theta - A\}$$

$$= \int_{-\infty}^{\theta-A} p(n)\mathrm{d}n \quad (4)$$

$$P_{Y|S}(1|1) = 1 - P_{Y|S}(0|1) \quad (5)$$

and the marginal density is

$$P_Y(y) = \sum_s P_{Y|S}(y|s)P_S(s) \quad (6)$$

Other researchers have derived the conditional probabilities $P_{Y|S}(y|s)$ of the threshold system with Gaussian noise with bipolar inputs (Bulsara & Zador, 1996) and Gaussian inputs (Stocks, 2001). We neither restrict the noise density to be Gaussian nor require that the density have finite variance even if the density has a bell-curve shape.

We use Shannon mutual information (Cover & Thomas, 1991) to measure the noise enhancement or SR effect (Bulsara & Zador, 1996; Deco & Schürmann, 1998; Godivier & Chapeau-Blondeau, 1998; Inchiosa, Robinson, & Bulsara, 2000; Stocks, 2001). The discrete Shannon mutual information of the input S and output Y is the difference between the output unconditional entropy $H(Y)$ and the output conditional entropy $H(Y|X)$:

$$I(S,Y) = H(Y) - H(Y|S) \quad (7)$$

$$= -\sum_y P_Y(y)\log P_Y(y) + \sum_s \sum_y P_{SY}(s,y)\log P_{Y|S}(y|s) \quad (8)$$

$$= -\sum_y P_Y(y)\log P_Y(y) + \sum_s P_S(s)\sum_y P_{Y|S}(y|s) \times \log P_{Y|S}(y|s) \quad (9)$$

$$= \sum_{s,y} P_{SY}(s,y)\log\frac{P_{SY}(s,y)}{P_S(s)P_Y(y)} \quad (10)$$

So the mutual information is the expectation of the random variable $\log[P_{SY}(s,y)/(P_S(s)P_Y(y))]$

$$I(S,Y) = E\left[\log\frac{P_{SY}(s,y)}{P_S(s)P_Y(y)}\right] \quad (11)$$

Here $P_S(s)$ is the probability density of the input S, $P_Y(y)$ is the probability density of the output Y, $P_{Y|S}(y|s)$ is the conditional density of the output Y given the input S, and $P_{SY}(s,y)$ is the joint density of the input S and the output Y. Simple bipolar histograms of samples can estimate these densities in practice.

Mutual information also measures the pseudo-distance between the joint probability density $P_{SY}(s,y)$ and the product density $P_S(s)P_Y(y)$. This holds for the Kullback (Cover & Thomas, 1991) pseudo-distance measure

$$I(S,Y) = \sum_s \sum_y P_{SY}(s,y)\log\frac{P_{SY}(s,y)}{P_S(s)P_Y(y)} \quad (12)$$

Then Jensen's inequality implies that $I(S,Y) \geq 0$. Random variables S and Y are statistically independent if and only if $I(S,Y) = 0$. Hence $I(S,Y) > 0$ implies some degree of dependence. We use this fact in the following proofs.

3. Proof of stochastic resonance for threshold neurons

We now prove that almost all finite-variance noise densities produce the SR effect in threshold neurons with subthreshold signals. This holds for all probability distributions on a two-symbol alphabet. The proof shows that if $I(S,Y) > 0$ then eventually the mutual information $I(S,Y)$ tends toward zero as the noise variance tends toward zero. So the mutual information $I(S,Y)$ must *increase* as the noise variance increases from zero. The only limiting assumption is that the noise mean $E[n]$ does not lie in the signal-threshold interval $[\theta - A, \theta + A]$.

Theorem 1. *Suppose that the threshold neuron* (1) *has noise probability density function $p(n)$ and that the input signal S is subthreshold* ($A < \theta$). *Suppose that there is some statistical dependence between input random variable S and output random variable Y (so that* $I(S,Y) > 0$). *Suppose that the noise mean $E[n]$ does not lie in the signal-threshold interval $[\theta - A, \theta + A]$ if $p(n)$ has finite variance. Then the threshold neuron* (1) *exhibits the nonmonotone SR effect in the sense that $I(S,Y) \to 0$ as $\sigma \to 0$.*

Proof. Assume $0 < P_S(s) < 1$ to avoid triviality when $P_S(s) = 0$ or 1. We show that S and Y are asymptotically independent: $I(\sigma) \to 0$ as $\sigma \to 0$. Recall that $I(S,Y) = 0$ if and only if S and Y are statistically independent (Cover & Thomas, 1991). So we need to show only that $P_{SY}(s,y) = P_S(s)P_Y(y)$ or $P_{Y|S}(y|s) = P_Y(y)$ as $\sigma \to 0$ for some signal symbols $s \in \mathcal{S}$ and $y \in \mathcal{Y}$. The two-symbol alphabet set $\mathcal{S}$ gives

$$P_Y(y) = \sum_s P_{Y|S}(y|s)P_S(s) \quad (13)$$

$$= P_{Y|S}(y|0)P_S(0) + P_{Y|S}(y|1)P_S(1) \quad (14)$$

$$= P_{Y|S}(y|0)P_S(0) + P_{Y|S}(y|1)(1 - P_S(0)) \quad (15)$$

$$= (P_{Y|S}(y|0) - P_{Y|S}(y|1))P_S(0) + P_{Y|S}(y|1) \quad (16)$$

So we need to show only that $P_{Y|S}(y|0) - P_{Y|S}(y|1) = 0$ as $\sigma \to 0$. This condition implies that $P_Y(y) = P_{Y|S}(y|1)$ and $P_Y(y) = P_{Y|S}(y|0)$. We assume for simplicity that the noise density $p(n)$ is integrable. The argument below still holds if $p(n)$ is discrete and if we replace integrals with appropriate sums.

Consider $y =$ '0'. Then Eqs. (2) and (4) imply that

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) = \int_{-\infty}^{\theta+A} p(n)\mathrm{d}n - \int_{-\infty}^{\theta-A} p(n)\mathrm{d}n \quad (17)$$

$$= \int_{\theta-A}^{\theta+A} p(n)\mathrm{d}n \quad (18)$$

Similarly for $y =$ '1':

$$P_{Y|S}(1|0) = \int_{\theta+A}^{\infty} p(n)\mathrm{d}n \quad (19)$$

$$P_{Y|S}(1|1) = \int_{\theta-A}^{\infty} p(n)\mathrm{d}n \quad (20)$$

Then

$$P_{Y|S}(1|0) - P_{Y|S}(1|1) = -\int_{\theta-A}^{\theta+A} p(n)\mathrm{d}n \quad (21)$$

The result now follows if

$$\int_{\theta-A}^{\theta+A} p(n)\mathrm{d}n \to 0 \text{ as } \sigma \to 0 \quad (22)$$

Let the mean of the noise be $m = E[n]$ and the variance be $\sigma^2 = E[(x-m)^2]$. Then $m \notin [\theta - A, \theta + A]$ by hypothesis.

Now suppose that $m < \theta - A$. Pick $\epsilon = \frac{1}{2}\mathrm{d}(\theta - A, m) = \frac{1}{2}(\theta - A - m) > 0$. So $\theta - A - \epsilon = \theta - A - \epsilon + m - m = m + (\theta - A - m) - \epsilon = m + 2\epsilon - \epsilon = m + \epsilon$. Then

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) = \int_{\theta-A}^{\theta+A} p(n)\mathrm{d}n \quad (23)$$

$$\le \int_{\theta-A}^{\infty} p(n)\mathrm{d}n \quad (24)$$

$$\le \int_{\theta-A-\epsilon}^{\infty} p(n)\mathrm{d}n \quad (25)$$

$$= \int_{m+\epsilon}^{\infty} p(n)\mathrm{d}n \quad (26)$$

$$= Pr\{n \ge m + \epsilon\} \quad (27)$$

$$= Pr\{n - m \ge \epsilon\} \quad (28)$$

$$\le Pr\{|n - m| \ge \epsilon\} \quad (29)$$

$$\le \frac{\sigma^2}{\epsilon^2} \text{ by Chebyshev inequality} \quad (30)$$

$$\to 0 \text{ as } \sigma \to 0 \quad (31)$$

A symmetric argument shows that for $m > \theta + A$

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) \le \frac{\sigma^2}{\epsilon^2} \to 0 \text{ as } \sigma \to 0 \qquad \square \quad (32)$$

Corollary. *The threshold neuron Eq.* (1) *exhibits SR for the additive gamma noise density*

$$p(n) = \begin{cases} \dfrac{n^{\alpha-1}e^{-n/\beta}}{\Gamma(\alpha)\beta^\alpha} & n \ge 0 \\ 0 & \text{otherwise} \end{cases} \quad (33)$$

under the hypotheses of Theorem 1. Parameters α and β are positive constants and Γ is the gamma function

$$\Gamma(x) = \int_0^\infty y^{x-1}e^y \,\mathrm{d}y \quad x > 0 \quad (34)$$

Gamma random variables have finite mean $\alpha\beta$ and functionally related finite variance $\alpha\beta^2$. Gamma family of random variables includes the popular special cases of exponential, Erlang, and chi-square random variables. All these random variables are right-sided. Fig. 3 shows simulation realizations of this corollary. This appears to be the first demonstration of the SR effect for *right-sided* noise processes.

We now proceed to the more general (and more realistic) case where infinite-variance noise interferes with the threshold neuron. The SR effect also occurs in other systems with impulsive infinite-variance noise (Kosko & Mitaim, 2001; Mitaim & Kosko, 1998). We can model many types of impulsive noise with *symmetric* alpha-stable bell-curve probability density functions with parameter α in the characteristic function $\varphi(\omega) = \exp\{-\gamma|\omega|^\alpha\}$. Here γ is the *dispersion* parameter (Breiman, 1968; Feller, 1966; Grigoriu, 1995; Nikias & Shao, 1995). Fig. 4 shows examples of symmetric (bell-curve) alpha-stable probability density functions with different α tail thicknesses and different bell-curve dispersions γ.

The parameter α controls tail thickness and lies in $0 < \alpha \le 2$. Noise grows more impulsive as α falls and the bell-curve tails grow thicker. The (thin-tailed) Gaussian density results when $\alpha = 2$ or when $\varphi(\omega) = \exp\{-\gamma\omega^2\}$. So the standard Gaussian random variable has zero mean and variance $\sigma^2 = 2$ (when $\gamma = 1$). The parameter α gives the thicker-tailed Cauchy bell curve when $\alpha = 1$ or $\varphi(\omega) = \exp\{-|\omega|\}$ for a zero *location* ($a = 0$) and unit dispersion ($\gamma = 1$) Cauchy random variable. The moments of stable distributions with $\alpha < 2$ are finite only up to the order k for $k < \alpha$. The Gaussian density alone has finite variance and higher moments. Alpha-stable random variables characterize the class of normalized sums of independent random variables that converge in distribution to a random variable (Breiman, 1968) as in the famous Gaussian special case called the "central limit theorem."

Alpha-stable models tend to work well when the noise or signal data contains 'outliers'—and all do to some degree.

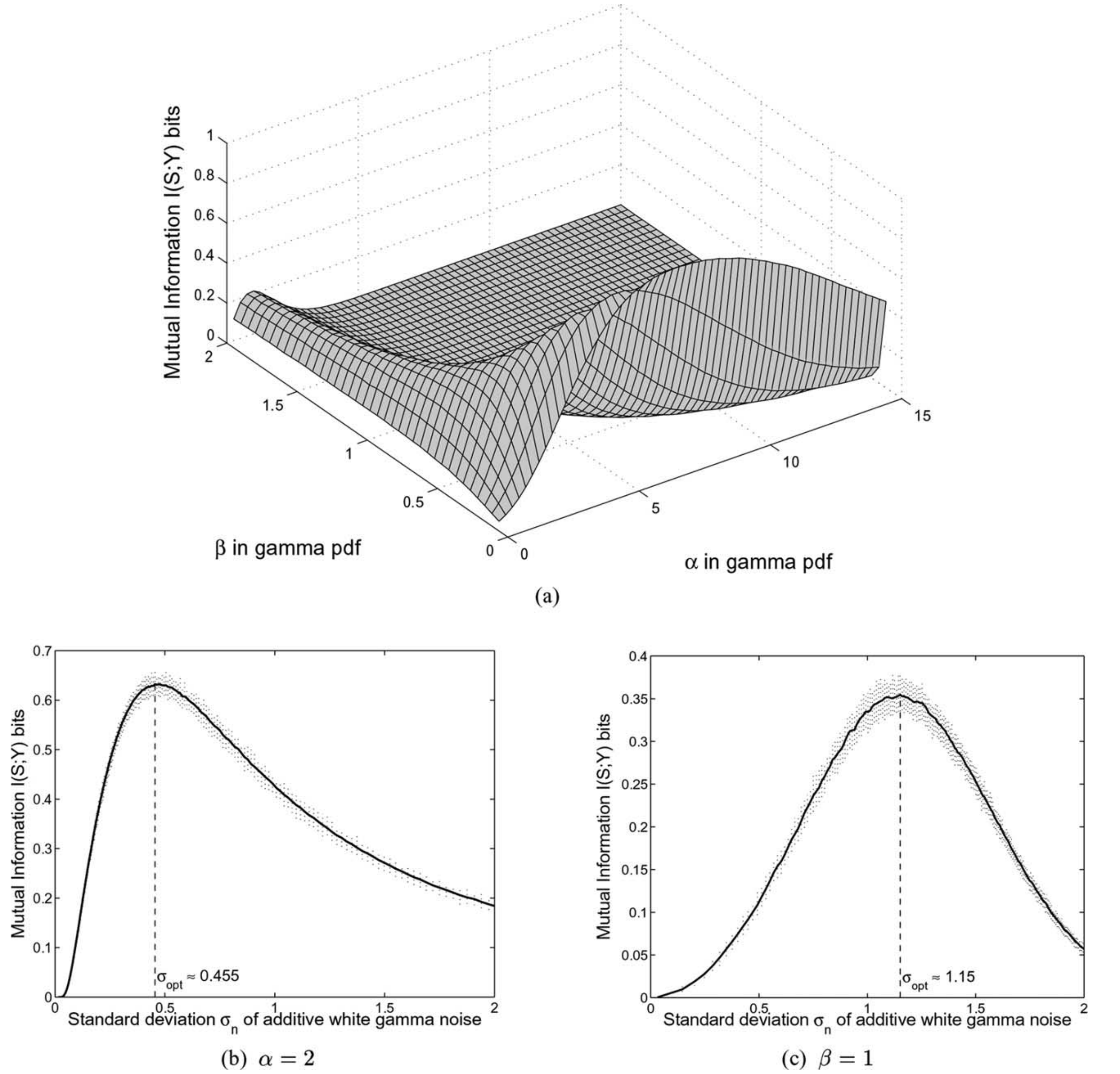

Fig. 3. SR with (finite-variance) gamma noise. The noisy signal-forced threshold neuron has the form of Eq. (1). The gamma noise n_t adds to the bipolar input Bernoulli signal s_t. The neuron has threshold $\theta = 1$. The input Bernoulli signal has amplitude $A = 0.8$ with success probability $p = \frac{1}{2}$. Each trial produced 10,000 input–output samples $\{s_t, y_t\}$ that estimated the probability densities to obtain the mutual information. The algorithm in (Ahrens & Dieter, 1974, 1982) generated realizations of the gamma random variable. (a) The graph shows the smoothed input–output mutual information of a threshold neuron as a function of the parameters α and β of additive white gamma noise n_t. The neuron's mutual information has a nonzero noise optimum $\sigma_{\text{opt}} > 0$ for each $\alpha > 0$. It also has a nonzero noise optimum $\sigma_{\text{opt}} > 0$ for each $\beta > 0$. (b) The graph shows the cross-section of the mutual-information surface for $\alpha = 2$. (c) The graph shows the cross-section for $\beta = 1$. Note that the mean and variance of the gamma noise are $m_n = \alpha\beta$ and $\sigma_n^2 = \alpha\beta^2$.

Models with $\alpha < 2$ can accurately describe impulsive noise in telephone lines, underwater acoustics, low-frequency atmospheric signals, fluctuations in gravitational fields and financial prices, and many other processes (Kosko, 1996; Nikias & Shao, 1995). Note that the best choice of α is an *empirical* question for bell-curve phenomena. Bell-curve behavior alone does not justify the (extreme) assumption of the Gaussian bell curve.

Theorem 2 applies to *any* alpha-stable noise model. The density need not be symmetric. A general alpha-stable probability density function f has characteristic function φ (Akgiray & Lamoureux, 1989; Bergstrom, 1952; Grigoriu, 1995; Nikias & Shao, 1995):

$$\varphi(\omega) = \exp\left\{ia\omega - \gamma|\omega|^{\alpha}\left(1 + i\beta\,\text{sign}(\omega)\tan\frac{\alpha\pi}{2}\right)\right\} \text{ for } \alpha \neq 1 \tag{35}$$

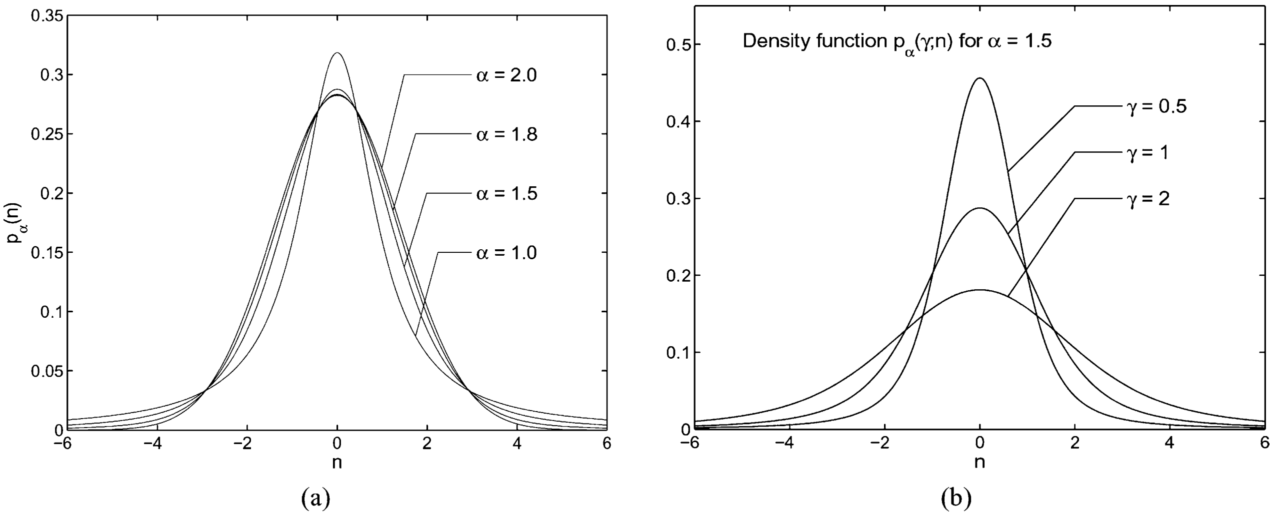

Fig. 4. Samples of standard symmetric ($\beta = 0$) alpha-stable probability densities. (a) Density functions with zero location ($\alpha = 0$) and unit dispersion ($\gamma = 1$) for $\alpha = 2$,1.8,1.5, and 1. The densities are bell curves that have thicker tails as α decreases and thus that model increasingly impulsive noise as α decreases. The case $\alpha = 2$ gives a Gaussian density with variance two (or unit dispersion). The parameter $\alpha = 1$ gives the Cauchy density with infinite variance. (b) Density functions for $\alpha = 1.5$ with dispersions $\gamma = 0.5$,1, and 2.

and

$$\varphi(\omega) = \exp\{ia\omega - \gamma|\omega|(1 - 2i\beta \ln|\omega| \mathrm{sign}(\omega)/\pi)\} \text{ for } \alpha = 1 \tag{36}$$

where

$$\mathrm{sign}(\omega) = \begin{cases} 1 & \text{if } \omega > 0 \\ 0 & \text{if } \omega = 0 \\ -1 & \text{if } \omega < 0 \end{cases} \tag{37}$$

and $i = \sqrt{-1}$, $0 < \alpha \le 2$, $-1 \le \beta \le 1$, and $\gamma > 0$. The parameter α is the characteristic exponent. Again the variance of an alpha-stable density does not exist if $\alpha < 2$. The location parameter a is the "mean" of the density when $\alpha > 1$. β is a skewness parameter. The density is symmetric about a when $\beta = 0$. Theorem 2 still holds even when $\beta \ne 0$. The dispersion parameter γ acts like a variance because it controls the width of a symmetric alpha-stable bell curve. There are no known closed forms of the α-stable densities for most α's.

The proof of Theorem 2 is simpler than the proof in the finite-variance case because all stable noise densities have a characteristic function with the exponential form in Eqs. (35) and (36). So zero noise dispersion gives φ as a simple complex exponential and hence gives the corresponding density as a delta spike that can fall outside the interval $[\theta - A, \theta + A]$.

Theorem 2. *Suppose $I(S, Y) > 0$ and the threshold neuron* Eq (1) *uses alpha-stable noise with location parameter $a \notin [\theta - A, \theta + A]$. Then the neuron* (1) *exhibits the nonmonotone SR effect if the input signal is subthreshold.*

Proof. Again the result follows if

$$\int_{\theta-A}^{\theta+A} p(n)\mathrm{d}n \to 0 \text{ as } \gamma \to 0 \tag{38}$$

The characteristic function $\varphi(\omega)$ of alpha-stable noise density, $p(n)$ has the exponential form Eqs. (35) and (36). This reduces to a simple complex exponential in the zero-dispersion limit:

$$\lim_{\gamma \to 0} \varphi(\omega) = \exp\{ia\omega\} \tag{39}$$

for *all* α's, skewness β's, and location a's. So Fourier transformation gives the corresponding density function in the limiting case ($\gamma \to 0$) as a translated delta function

$$\lim_{\gamma \to 0} p(n) = \delta(n - a) \tag{40}$$

Then

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) = \int_{\theta-A}^{\theta+A} p(n)\mathrm{d}n \tag{41}$$

$$= \int_{\theta-A}^{\theta+A} \delta(n - a)\mathrm{d}n \tag{42}$$

$$= 0 \tag{43}$$

because $a \notin [\theta - A, \theta + A]$. □

Fig. 2 gives a typical example of the SR effect for highly impulsive noise with infinite variance. Here the noise type is Cauchy ($\alpha = 1$) and thus frequent and violent noise spikes interfere with the signal.

4. Conclusions

Noise affects neural systems in complex ways. The above theorems show that almost all noise types produce SR in threshold neurons that use subthreshold signals and small amounts of noise. This includes right-sided finite-variance noise such as gamma noise. The theorems do not guarantee that the predicted increase in mutual information will be significant. They guarantee only that some increase will occur. Other work (Kosko & Mitaim, 2001) suggests that the increase will decrease in significance as the impulsiveness of the noise process increases. All our simulations showed a significant and visible SR effect.

These results help explain the widespread occurrence of the SR effect in mechanical and biological threshold systems (Braun, Wissing, Schäfer, & Hirsch, 1994; Douglass et al., 1993; Fauve & Heslot, 1983; Melnikov, 1993; Levin & Miller, 1996; Russell, Willkens, & Moss, 1999). The broad generality of the results suggests that SR should occur in any nonlinear system whose input–output structure approximates a threshold system and that includes most model neurons. The infinite-variance result further implies that such widespread SR effects should be robust against violent noise impulses. The combined results support the hypothesis (Linsker, 1988, 1997) that neurons have evolved to maximize their local information if they process subthreshold signals in the presence of noise. This need not hold for suprathreshold signals.

Acknowledgements

National Science Foundation Grant No. ECS-0070284 and Thailand Research Fund Grant No. RSA4680001 supported this research.

References

Ahrens, J. H., & Dieter, U. (1974). Computer methods for sampling from gamma, beta, poisson and binomial distributions. *Computing*, *12*, 223–246.

Ahrens, J. H., & Dieter, U. (1982). Generating gamma variates by a modified rejection technique. *Communications of the ACM*, *25*(1), 47–54.

Akgiray, V., & Lamoureux, C. G. (1989). Estimation of stable-law parameters: A comparative study. *Journal of Business and Economic Statistics*, *7*, 85–93.

Bergstrom, H. (1952). On some expansions of stable distribution functions. *Arkhiv Mathematics*, *2*, 375–378.

Braun, H. A., Wissing, H., Schäfer, K., & Hirsch, M. C. (1994). Oscillation and noise determine signal transduction in shark multimodal sensory cells. *Nature*, *367*, 270–273.

Breiman, L. (1968). *Probability*. Reading, MA: Addison-Wesley.

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.

Chambers, J. M., Mallows, C. L., & Stuck, B. W. (1976). A method for simulating stable random variables. *Journal of the American Statistical Association*, *71*(354), 340–344.

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). Stochastic resonance without tuning. *Nature*, *376*, 236–238.

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

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

Douglass, J. K., Wilkens, L., Pantazelou, E., & Moss, F. (1993). Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. *Nature*, *365*, 337–340.

Fauve, S., & Heslot, F. (1983). Stochastic resonance in a bistable system. *Physics Letters A*, *97*(1–2), 5–7.

Feller, W. (1966) (*Vol. II*). *An introduction to probability theory and its applications*. New York: Wiley.

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

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.

Hess, S. M., & Albano, A. M. (1998). Minimum requirements for stochastic resonance in threshold systems. *International Journal of Bifurcation and Chaos*, *8*(2), 395–400.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science*, *79*, 2554–2558.

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.

Jung, P., & Mayer-Kress, G. (1995). Stochastic resonance in threshold devices. *Il Nuovo Cimento*, *17D*(7–8), 827–834.

Kosko, B. (1991). *Neural networks and fuzzy systems: A dynamical systems approach to machine intelligence*. Englewood Cliffs, NJ: Prentice Hall.

Kosko, B. (1996). *Fuzzy engineering*. Englewood Cliffs, NJ: Prentice Hall.

Kosko, B., & Mitaim, S. (2001). Robust stochastic resonance: signal detection and adaptation in impulsive noise. *Physical Review E*, *64*(051110).

Levin, J. E., & Miller, J. P. (1996). Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance. *Nature*, *380*, 165–168.

Linsker, R. (1988). Self-organization in a perceptual network. *Computer*, *21*(3), 105–117.

Linsker, R. (1997). A local learning rule that enables information maximization for arbitrary input distributions. *Neural Computation*, *9*(8), 1661–1665.

Melnikov, V. I. (1993). Schmitt trigger: A solvable model of stochastic resonance. *Physical Review E*, *48*(4), 2481–2489.

Mitaim, S., & Kosko, B. (1998). Adaptive stochastic resonance. *Proceedings of the IEEE: Special issue on intelligent signal processing*, *86*(11), 2152–2183.

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

Russell, D. F., Willkens, L. A., & Moss, F. (1999). Use of behavioural stochastic resonance by paddle fish for feeding. *Nature*, *402*, 291–294.

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

Tsakalides, P., & Nikias, C. L. (1996). The robust covariation-based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive noise environments. *IEEE Transactions on Signal Processing*, *44*(7), 1623–1633.

Wannamaker, R.A., Lipshitz, S.P., & Vanderkooy, J. (2000). Stochastic resonance as dithering. *Physical Review E*, *61*(1), 233–236.