



รายงานวิจัยฉบับสมบูรณ์

การกำหนดทางสถิติและสัญญาณรบกวนในระบบการประมวลสัญญาณ

นายสัญญา มิตรเอม

กันยายน 2548

รายงานวิจัยฉบับสมบูรณ์

การกำหนดทางสถิติและสัญญาณรบกวนในระบบการประมวลสัญญาณ

นายสัญญา มิตรเอม
ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์
มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สำนักงานกองทุนสนับสนุนการวิจัยไม่จำเป็นต้องเห็นด้วยเสมอไป)

รหัสโครงการ: RSA4680001

โครงการ: การกำหนดทางสถิติและสัญญาณรบกวนในระบบการประมวลสัญญาณ

หัวหน้าโครงการ: นายสัญญา มิตรเอม

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์

มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต ปทุมธานี 12120

E-mail Address: msanya@engr.tu.ac.th

ระยะเวลาโครงการ: 1 ตุลาคม 2545 – 30 กันยายน 2548

งานวิจัยนี้ศึกษาการกำหนดทางสถิติ (Stochastic Resonance) และสัญญาณรบกวนในระบบการประมวลสัญญาณแบบไม่เชิงเส้น การกำหนดทางสถิติคือปรากฏการณ์ที่สัญญาณรบกวนในปริมาณที่พอเหมาะทำให้ประสิทธิภาพของการทำงานของระบบประมวลสัญญาณดีขึ้น เช่น signal-to-noise ratio หรือ cross-correlation measure หรือ mutual information งานวิจัยนี้ได้ศึกษาระบบ threshold system ที่ทำการประมวลผลสัญญาณ random binary signals โดยการวัดประสิทธิภาพด้วย mutual information ผลของการวิจัยในส่วนนี้คือทฤษฎีและบทพิสูจน์ที่แสดงถึงเงื่อนไขจำเป็นและเงื่อนไขพอเพียงของความสัมพันธ์ระหว่างตัวแปรของสัญญาณรบกวน สัญญาณอินพุตและค่า threshold ของระบบในการเกิดการกำหนดทางสถิติ ทฤษฎีนี้แสดงให้เห็นว่าสัญญาณรบกวนที่ทำให้เกิดการกำหนดทางสถิติมีได้หลายรูปแบบ โดยสัญญาณรบกวนอาจมีการแจกแจงทั้งแบบที่มีค่าความแปรปรวนเป็นค่าจำกัดหรือเป็นอนันต์ (เช่น การแจกแจงในตระกูล alpha-stable) หรือสัญญาณรบกวนที่การแจกแจงมีรูปแบบไม่สมมาตร (unsymmetric density function) การวิเคราะห์ความสัมพันธ์ถดถอยแสดงให้เห็นถึงความสัมพันธ์แบบเอ็กซ์โปเนนเชียล ของค่าความเบี่ยงเบนที่ดีที่สุด (optimal dispersion) ที่สัมพันธ์ต่อพารามิเตอร์อัลฟาของการแจกแจงของสัญญาณรบกวน alpha-stable แบบสมมาตร และความสัมพันธ์ที่เกือบจะเป็นเส้นตรงของค่า mutual information ที่สูงที่สุด (เมื่อเกิดการกำหนดทางสถิติ) ที่สัมพันธ์ต่อพารามิเตอร์อัลฟาของการแจกแจงของสัญญาณรบกวนแบบ alpha-stable นอกจากการพิจารณาการกำหนดทางสถิติในระบบ threshold system แล้ว งานวิจัยนี้ยังได้ศึกษาปรากฏการณ์นี้ในแบบจำลองหน่วยย่อยในระบบประสาท (neuron models) ที่เป็นระบบพลวัต (dynamical system) อีกด้วย ซึ่งแสดงให้เห็นว่าสัญญาณรบกวนสามารถช่วยให้หน่วยย่อยๆ ในระบบประสาทเหล่านี้ทำงานได้อย่างมีประสิทธิภาพมากขึ้นในบางสภาวะแวดล้อม และงานวิจัยนี้ได้เสนอแนวทางการใช้สัญญาณรบกวนในการตรวจจับวัตถุในภาพระบบดิจิทัล ซึ่งผลการทดลองได้แสดงให้เห็นว่าสัญญาณรบกวนสามารถช่วยให้การตรวจจับวัตถุโดยใช้การแยกวัตถุ (object segmentation) ด้วยวิธีการ color thresholding มีความแม่นยำมากขึ้น โดยตัวชี้วัดประสิทธิภาพของการแยกวัตถุคือ mutual information จำนวนพิกเซลที่ผิดพลาด (error pixel counts) และความผิดพลาดของตำแหน่ง (position error) ผลการวิจัยต่างๆ นี้แสดงให้เห็นถึงความเป็นไปได้ในการออกแบบระบบต่างๆ ให้ใช้สัญญาณรบกวนที่ไม่สามารถหลีกเลี่ยงได้ให้เป็นประโยชน์ ควบคู่ไปกับการที่จะกำจัดสัญญาณรบกวนนั้นออกไป

Project Code: RSA4680001
Project Title: Stochastic Resonance and Noise in Signal Processing Systems
Principal Investigator: Mr. Sanya Mitaim
Department of Electrical Engineering, Faculty of Engineering
Thammasat University, Rangsit Campus
Pathumthani 12120
E-mail Address: msanya@engr.tu.ac.th
Project Duration: 1 October 2002 - 30 September 2005

This research explores the stochastic resonance effect in several nonlinear signal processing systems. Stochastic resonance or SR occurs when noise enhances the output of the system that processes input signals. The performance measures can be an output signal-to-noise ratio (output SNR), cross-correlation measure, or mutual information of input and output. The effort studies in more depth a threshold system that processes random binary input signals and uses mutual information as a performance measure. The research effort results in theorems that provide both necessary and sufficient conditions for the SR effect. The results suggest how the noise mean or location parameter should relate to the input signals and the system's threshold. The result holds for broad classes of noise distributions that include all finite-variance noise with both symmetric-shaped and unsymmetric-shaped distributions as well as infinite-variance (impulsive) noise in alpha-stable distributions. Regression analysis reveals both an exponential relationship for the optimal noise dispersion as a function of the alpha bell-curve tail thickness and an approximate linear relationship for the SR-maximal mutual information as a function of the alpha-bell curve tail thickness. The research also explores several models of continuous neuron widely used in neural network applications. Simulations show that these neuron models also exhibit the SR effect when they process small random binary signals. The neurons' mutual information measures are maximal when the system's noise is nonzero. The SR effect also occurs in these feedback nonlinear dynamical systems even when the noise is impulsive. Finally the research reports the benefits of noise in color object segmentation algorithm used in image processing and computer vision. Experiments show that the proposed noise-added color thresholding scheme can help improve several measures used in object segmentation. These measures include mutual information, error pixel counts, and position error. The results suggest that scientists and engineers should consider exploiting the benefits of noise as well as canceling or filtering it out.

1 Summary: Stochastic Resonance and Noise-Enhanced Systems

Noise has been an unwanted signal or source of energy. Scientists and engineers have largely tried to filter noise or cancel it or mask it out of existence. But noise can sometimes enhance a signal as well as hurt it. The fact that “noise can help” may seem at odds with almost a century of effort in signal processing to filter noise or to mask or cancel it. But noise is itself a signal and a free source of energy. Noise can amplify a faint signal in some feedforward and feedback nonlinear systems even though too much noise can swamp the signal. This implies that a system’s optimal noise level need not be zero noise. It also suggests that nonlinear signal systems with nonzero-noise optima may be the rule rather than the exception.

The new field of *stochastic resonance* or SR [7, 8, 13, 32, 37, 66, 67, 76, 90] rests on an exception to this undeclared war on noise. SR occurs when noise enhances a faint signal in a nonlinear system. It occurs when the addition of a small amount of noise increases a nonlinear system’s performance measure such as its signal-to-noise ratio, cross-correlation, or mutual information. The nonlinearity is often as simple as a memoryless threshold. So a great deal of SR research has focused on how dither-like noise can help spiking neurons process data streams [16, 38, 46]. SR occurs in physical systems such as ring lasers [70], threshold hysteretic Schmitt triggers [33], Chua’s electrical circuit [5], superconducting quantum interference devices (SQUIDs) [44], Josephson junctions [11], chemical systems [31], and quantum-mechanical systems [40]. SR also occurs in biological systems such as the rat [23], crayfish [29], cricket [62], river paddlefish [82], and in many types of model neurons [12, 14, 21, 22, 78].

Figure 1 shows that a small amount of Gaussian pixel noise can improve our subjective perception of an image. The system quantizes the original gray-scale “baboon” image into a binary image of black and white pixels. It emits a white pixel as output if the input gray-scale pixel equals or exceeds a threshold. It emits a black pixel as output if the input gray-scale pixel falls below the threshold. This quantizer is biased because it does not set the threshold at the midpoint of the gray scale. So the quantized version of the original image contains almost no information. A small level of noise sharpens the image contours and helps fill in features when it adds to the original image before the system applies the threshold. Too much noise swamps the image and degrades its contours. Gammaitoni [35] and others [87] have proposed a dithering argument for this SR effect and still others [69] have applied this argument

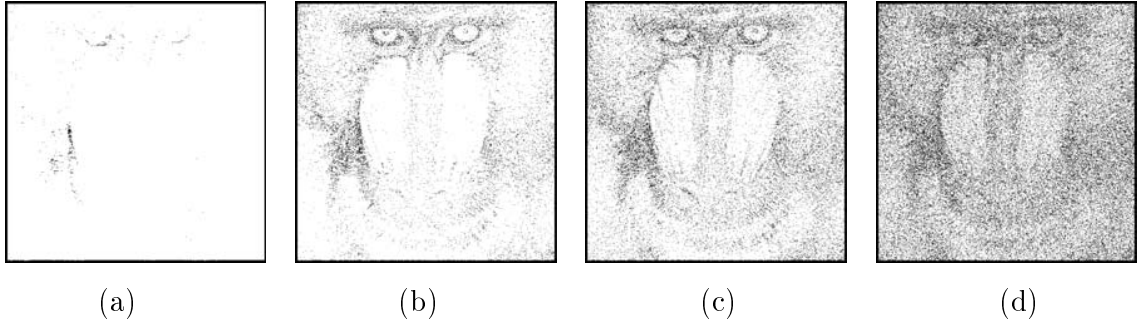


Figure 1: Gaussian pixel noise can improve subjective quality of an image through a stochastic-resonance or dithering process [35, 87]. The noise produces a nonmonotonic response: A small level of noise sharpens the image features while too much noise degrades them. These noisy images result when we apply a pixel threshold to the ‘baboon’ image: $y = g((x + n) - \theta)$ 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 $\theta = 0.06$. 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^2 grows from (b)-(d): $\sigma^2 = 1.00 \times 10^{-2}$ in (b), $\sigma^2 = 2.25 \times 10^{-2}$ in (c), and $\sigma^2 = 9.00 \times 10^{-2}$ in (d).

to still images. The argument involves adding dither noise to a signal before quantization. Consider gray-scale pixel $x \in [0, 1]$ and binary output pixel $y \in \{0, 1\}$ with threshold $\theta = \frac{1}{2}$. Then the dithered quantizer gives $E[Y|x] = 1 - \Pr\{n < \theta - x\} = x$ if and only if the noise is uniform on $(-\frac{1}{2}, \frac{1}{2})$. But the subjective SR result in 1 holds for Gaussian noise and for many other types of nonuniform noise distributions that also includes the infinite-variance noise types in the broad family of alpha-stable distributions [72, 59]. So the dithering argument only partially explains this subjective SR effect.

Simulation and theoretical results show that memoryless threshold neurons benefit from small amounts of almost all types of additive noise and so produce the stochastic-resonance or SR effect. Figure 2 shows a typical simulation confirmation of this SR result for additive Gaussian noise. Input-output mutual information measures the performance of such threshold systems that use subthreshold signals. The SR result holds for all possible noise probability density functions with finite variance. The only constraint is that the noise mean must fall outside a “forbidden” threshold-related interval that the user can control. The theorem shows that this condition is also necessary. A corollary and simulations show that the SR effect occurs for

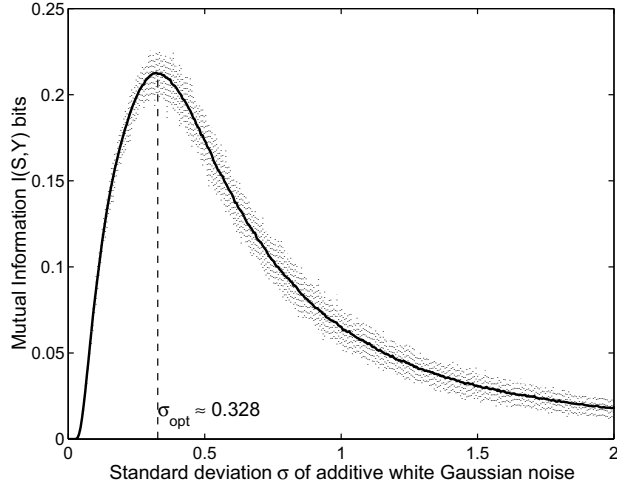


Figure 2: The non-monotonic signature of stochastic resonance. The graph shows the smoothed input-output mutual information of a threshold system as a function of the standard deviation of additive white Gaussian noise n_t . The vertical dashed lines show the absolute deviation between the smallest and largest outliers in each sample average of 100 outcomes. The system has a nonzero noise optimum at $\sigma_{opt} \approx 0.328$ and thus shows the SR effect. The noisy signal-forced threshold system has the form (6). The Gaussian noise n_t adds to the external forcing bipolar signal s_t .

right-sided beta, gamma, and Weibull noise as well. These SR results further hold for the entire uncountably infinite class of alpha-stable probability density functions. Alpha-stable noise densities have infinite variance and infinite higher-order moments and often model impulsive noise environments. The stable noise densities include the special case of symmetric bell-curve densities with thick tails such as the Cauchy probability density. The SR result for alpha-stable noise densities shows that the SR effect in threshold and threshold-like systems is robust against occasional or even frequent violent fluctuations in noise. Regression analysis reveals both an exponential relationship for the optimal noise dispersion as a function of the alpha bell-curve tail thickness and an approximate linear relationship for the SR-maximal mutual information as a function of the alpha bell-curve tail thickness.

The report next shows extensive simulations that confirms the SR effect in several standard continuous sigmoidal neurons and for Gaussian radial basis functions. It then reviews a robust learning law can find the optimal noise variance and dispersion for both threshold and continuous neurons and for both finite-variance and infinite-variance noise. It derives and tests a new robustified learning law that finds the

entropically optimal noise level given histogram estimates of the underlying marginal and conditional probability density functions. This statistically robust algorithm uses only the sign of the noise gradient rather than the gradient itself.

The results show that model neurons can exploit low levels of crosstalk or other forms of noise in their local environment. Even highly impulsive noise can help neurons maximize their throughput information. Such noise-based information maximization is consistent with Linsker's principle of information maximization in neural networks [63, 64]. These findings support the implicit SR conjecture that biological neurons have evolved to computationally exploit their noisy environments [15, 23, 24, 29, 62, 74, 79, 86]. Further support is that these adaptive SR effects still hold for other sigmoidal and nonsigmoidal (Gaussian) neurons as Figure 20 shows. These results suggest that biological neurons should experience less mutual information if they do not use their local noise.

Then the report presents an SR effect in object segmentation with color thresholding algorithm in image processing. This preliminary result shows that an addition of a small amount of noise can improve the accuracy of color object segmentation. It shows that this SR effect does occur for various performance indices that measure how well an object is segmented from the background. The performance measures are mutual information, error pixels count, and object position error. These measures compare the segmented images obtained from the original thresholding algorithm with the proposed SR-extended algorithm. The study confirms by examples that addition of noise can robustify the color thresholding algorithm and thus provides an alternative for engineers when they need to detect color objects in noisy input images.

2 Mutual Information and SR in Neuron Models

This section reviews Shannon's measure of mutual information between two random variables. Then it reviews the simple nonlinear threshold model of a neuron and the continuous neuron model that show the SR effect for bipolar signals.

Mutual information [25] can measure the stochastic resonance (SR) effect [16, 27, 38, 52, 84]. The discrete Shannon mutual information of the input S and output Y has the form

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

$$= - \sum_y P_Y(y) \log P_Y(y) + \sum_s \sum_y P_{SY}(s, y) \log P_{Y|S}(y|s) \quad (2)$$

$$= - \sum_y P_Y(y) \log P_Y(y) + \sum_s P(s) \sum_y P(y|s) \log P(y|s) \quad (3)$$

$$= \sum_{s,y} P_{SY}(s,y) \log \frac{P_{SY}(s,y)}{P_S(s)P_Y(y)} \quad (4)$$

We can view the mutual information in the form of expectation of a random variable $\log \frac{P_{SY}(s,y)}{P_S(s)P_Y(y)}$:

$$I(S,Y) = E \left[\log \frac{P_{SY}(s,y)}{P_S(s)P_Y(y)} \right] \quad (5)$$

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 joint density of the input S and the output Y .

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 [25] pseudo-distance measure $I(S,Y) = \sum_s \sum_y P_{SY}(s,y) \log \frac{P_{SY}(s,y)}{P_S(s)P_Y(y)}$. 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.

2.1 Noisy Threshold Neuron (Threshold System)

We use the discrete-time threshold neuron model [16, 35, 47, 54, 57]

$$y_t = \text{sgn}(s_t + n_t - \theta) = \begin{cases} 1 & \text{if } s_t + n_t \geq \theta \\ -1 & \text{if } s_t + n_t < \theta \end{cases} \quad (6)$$

where $\theta > 0$ is the neuron's threshold, s_t is the bipolar input Bernoulli signal (with success probability $\frac{1}{2}$) with amplitude $A > 0$, and n_t is the additive white noise with probability density $p(n)$. Experiments with other success probabilities near $\frac{1}{2}$ did not produce substantially different simulation results.

2.2 Noisy Continuous Neuron

We use the additive continuous neuron model with a neuronal signal function $S(x)$ [57]

$$\dot{x} = -x + S(x) + s(t) + n(t) \quad (7)$$

$$y(t) = \text{sgn}(x(t)). \quad (8)$$

Here $s(t)$ and $n(t)$ are the input and additive noise of the neuron and $y(t)$ is the binary output. The neuron feeds its output signal $S(x)$ back to itself and emits the threshold bipolar signal $y(t)$ as output.

- **Hyperbolic Tangent.** This signal function gives an additive neuron model that is bistable [4, 14, 20, 47, 48, 57]

$$S(x) = 2 \tanh x \quad (9)$$

- **Linear-Threshold.** This simple linear-threshold signal function [57] also gives the SR effect in the neuron:

$$S(x) = \begin{cases} cx & |cx| < 1 \\ 1 & cx > 1 \\ -1 & cx < -1 \end{cases} \quad (10)$$

for a constant $c > 0$. We use $c = 2$.

- **Exponential.** This signal function is asymmetric with the form [57]

$$S(x) = \begin{cases} 1 - \exp\{-cx\} & x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

for a constant $c > 0$. We use $c = 10$.

- **Gaussian.** The Gaussian or “radial basis” signal function [57] differs from other signal functions above because it is nonmonotonic:

$$S(x) = \exp\{-cx^2\} \quad (12)$$

for a constant $c > 0$. We use $c = 100$.

3 Mutual Information of the Threshold Neuron with Bipolar Input Signals

This section derives analytical SR results for the noisy threshold neuron based on the marginal probability density function of the output $P_Y(y)$ and the conditional density $P_{Y|S}(y|s)$. The system is the binary neuron with a fixed threshold θ . The bipolar (Bernoulli with success probability p) input signal s_t has amplitude A : $s_t \in \{-A, A\}$ with probability density $P_S(s)$. The noise n_t adds to the signal s_t before it enters the neuron. So the neuron’s output y_t has the form (6). Figure 5 plots the mutual information $I(S, Y)$ for four standard closed-form noise probability density functions

(19)-(38). The central result is a theorem that holds for almost all noise probability densities so long as the mean noise falls outside a user-controlled interval that depends on the threshold θ .

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

$$\begin{aligned} P_{Y|S}(0|0) &= \Pr\{s + n < \theta\} \Big|_{s=-A} \\ &= \Pr\{n < \theta + A\} = \int_{-\infty}^{\theta+A} p(n)dn \end{aligned} \quad (13)$$

$$P_{Y|S}(1|0) = 1 - P_{Y|S}(0|0) \quad (14)$$

$$\begin{aligned} P_{Y|S}(0|1) &= \Pr\{s + n < \theta\} \Big|_{s=A} \\ &= \Pr\{n < \theta - A\} = \int_{-\infty}^{\theta-A} p(n)dn \end{aligned} \quad (15)$$

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

and the marginal density is

$$P_Y(y) = \sum_s P_{Y|S}(y|s)P_S(s) \quad (17)$$

Researchers have derived the conditional probabilities $P_{Y|S}(y|s)$ of the threshold system with *Gaussian* noise with bipolar inputs [16] and Gaussian inputs [84]. We next derive $P_{Y|S}(y|s)$ for uniform, Laplace, and (infinite-variance) Cauchy noise as well. Figure 3 shows four examples of the unimodal noise densities and their realizations. Then we introduce stable distributions to model a spectrum of impulsive noise types.

• **Gaussian Noise.** The Gaussian density with zero mean and variance $\sigma_n^2 = \sigma^2$ has the form

$$p(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n^2}{2\sigma^2}\right\} \quad (18)$$

Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$\begin{aligned} P_{Y|S}(0|0) &= \int_{-\infty}^{\theta+A} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n^2}{2\sigma^2}\right\} dn \\ &= \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{\theta+A}{\sigma\sqrt{2}}\right) \end{aligned} \quad (19)$$

$$P_{Y|S}(1|0) = \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{\theta+A}{\sigma\sqrt{2}}\right) \quad (20)$$

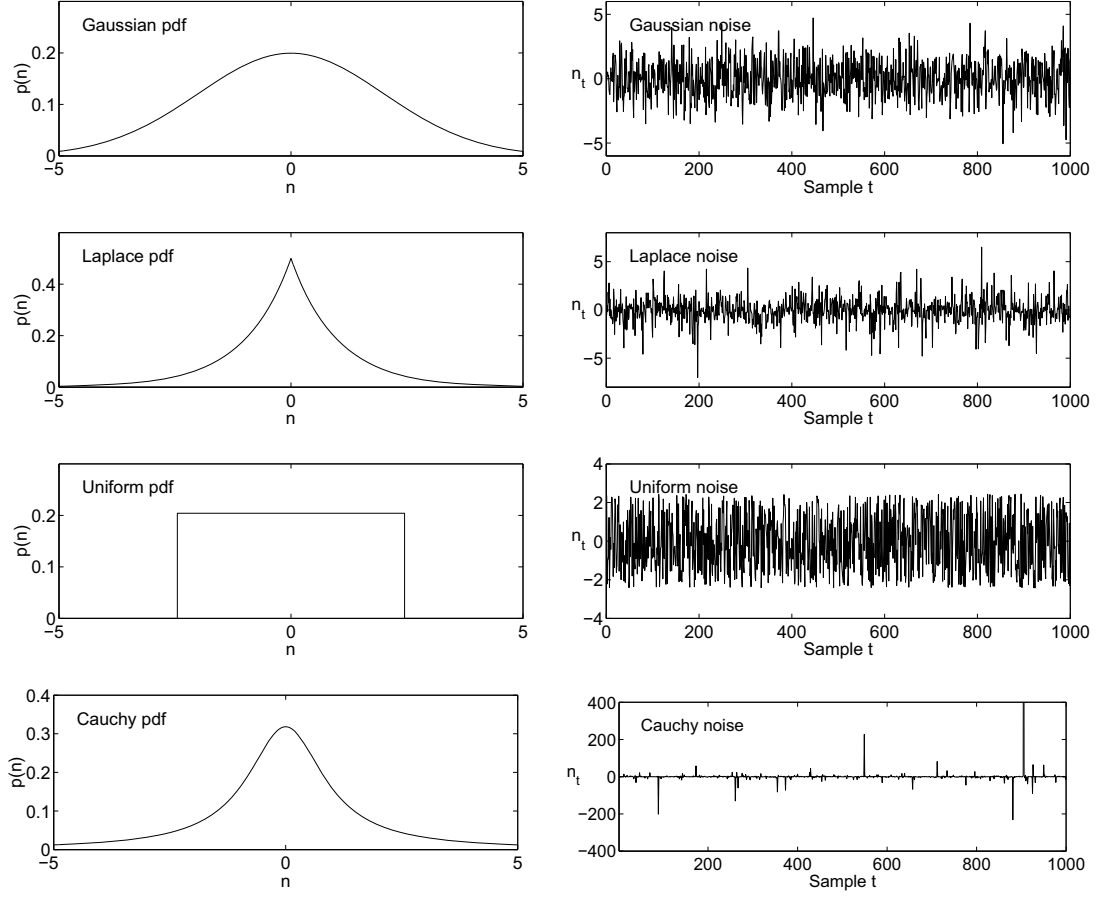


Figure 3: Probability density functions and sample realizations. The figure shows Gaussian, Laplace, and uniform random variables n with zero mean and variance of two: $E[n] = 0$ and $E[n^2] = \sigma^2 = 2$. The Cauchy density function has zero location and unit dispersion but infinite variance. The pseudo-random number generators in [65] act as noise sources for these probability densities.

$$P_{Y|S}(0|1) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{\theta - A}{\sigma\sqrt{2}} \right) \quad (21)$$

$$P_{Y|S}(1|1) = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\frac{\theta - A}{\sigma\sqrt{2}} \right) \quad (22)$$

The error function erf is

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\{-t^2\} dt \quad (23)$$

• **Uniform Noise.** The uniform density with zero mean and variance $\sigma_n^2 = \frac{a^2}{12}$ has the form

$$p(n) = \begin{cases} \frac{1}{a} & \text{if } -\frac{a}{2} < n < \frac{a}{2} \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$\begin{aligned} P_{Y|S}(0|0) &= \begin{cases} 1 & \text{if } \frac{a}{2} < \theta + A \\ \frac{1}{2} + \frac{A+\theta}{a} & \text{otherwise} \end{cases} \\ &= \min\{1, \frac{1}{2} + \frac{\theta + A}{a}\} \end{aligned} \quad (25)$$

$$P_{Y|S}(1|0) = \max\{0, \frac{1}{2} - \frac{\theta + A}{a}\} \quad (26)$$

$$P_{Y|S}(0|1) = \min\{1, \frac{1}{2} + \frac{\theta - A}{a}\} \quad (27)$$

$$P_{Y|S}(1|1) = \max\{0, \frac{1}{2} - \frac{\theta - A}{a}\} \quad (28)$$

• **Laplace Noise.** The Laplace density with zero mean and variance $\sigma_n^2 = 2\beta^2$ has the form

$$p(n) = \frac{1}{2\beta} \exp\{-|\frac{n}{\beta}|\} \quad (29)$$

Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$P_{Y|S}(0|0) = 1 - \frac{1}{2} \exp\{-\frac{\theta + A}{\beta}\} \quad (30)$$

$$P_{Y|S}(1|0) = \frac{1}{2} \exp\{-\frac{\theta + A}{\beta}\} \quad (31)$$

$$P_{Y|S}(0|1) = 1 - \frac{1}{2} \exp\{-\frac{\theta - A}{\beta}\} \quad (32)$$

$$P_{Y|S}(1|1) = \frac{1}{2} \exp\{-\frac{\theta - A}{\beta}\} \quad (33)$$

• **Cauchy Noise.** The Cauchy density with zero location and finite dispersion γ (but infinite variance) has the form

$$p(n) = \frac{1}{\pi} \frac{\gamma}{n^2 + \gamma^2}. \quad (34)$$

Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$P_{Y|S}(0|0) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \frac{\theta + A}{\gamma} \quad (35)$$

$$P_{Y|S}(1|0) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{\theta + A}{\gamma} \quad (36)$$

$$P_{Y|S}(0|1) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \frac{\theta - A}{\gamma} \quad (37)$$

$$P_{Y|S}(1|1) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{\theta - A}{\gamma} \quad (38)$$

• Symmetric Alpha-Stable Noise: Thick-Tailed Bell Curves

We 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 [10, 34, 41, 77]. The parameter α controls tail thickness and lies in $0 < \alpha \leq 2$. 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 [10] 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. 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 [58, 77]. 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.

Figure 4 shows realizations of four symmetric alpha-stable random variables. A general alpha-stable probability density function f has characteristic function φ [3, 9, 41, 77]:

$$\varphi(\omega) = \exp\left\{ia\omega - \gamma|\omega|^\alpha \left(1 + i\beta\text{sign}(\omega) \tan \frac{\alpha\pi}{2}\right)\right\} \quad \text{for } \alpha \neq 1 \quad (39)$$

and

$$\varphi(\omega) = \exp\{ia\omega - \gamma|\omega|(1 - 2i\beta \ln |\omega|\text{sign}(\omega)/\pi)\} \quad \text{for } \alpha = 1 \quad (40)$$

where

$$\text{sign}(\omega) = \begin{cases} 1 & \text{if } \omega > 0 \\ 0 & \text{if } \omega = 0 \\ -1 & \text{if } \omega < 0 \end{cases} \quad (41)$$

and $i = \sqrt{-1}$, $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, 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 $\beta = 0$. The theorem below still holds even when $\beta \neq 0$. 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. Numerical integration of φ gives the probability densities in Figure 4.

The main results are theorems that show the SR effect in simple (memoryless) threshold neurons as often found in literature of neural networks [42, 43, 57]. The first theorem (Theorem 1.1) shows that threshold neurons exhibit the SR effect for all finite-variance noise densities if the system performance measure is Shannon’s mutual information and if the mean or location parameter falls outside a “forbidden” interval that one can often pick in advance. A corollary shows that this SR effect still occurs for right-sided gamma, beta, and Weibull noise. Traditional SR research has focused almost exclusively on two-sided noise. The second theorem (Theorem 2.1) shows that this also holds for all infinite-variance densities that belong to the large class of stable distributions. Both theorems assume that all signals are subthreshold signals. The other two theorems (Theorems 1.2 and 2.2) show that there is no SR effect if the mean or location parameters fall within the forbidden threshold interval. Figure 11 shows a simulation instance of this predicted forbidden-interval effect for Gaussian and Cauchy noise.

We can derive more specific results for closed-form noise densities. Figure 5 shows I -versus- σ profiles of a threshold system with four kinds of noise: Gaussian, uniform, Laplace, and Cauchy. The I profile of the uniform noise has the highest peak among the four noise densities for the same system (same threshold θ and same input amplitude A). And the I profile has a distinct shape: it drops sharply after it reaches its peak as σ grows. Gaussian noise gives the second highest I while Cauchy gives the lowest. The threshold system requires different optimal standard deviations (or dispersions) for different kinds of noise.

The closed form of the I -versus- σ profiles in Figure 5 also allows a direct analysis of how the optimal noise depends on the signal amplitude A for Gaussian, uniform, Laplace, and Cauchy noise. Suppose the signal amplitude A is a subthreshold input in a noisy threshold neuron with threshold θ : $A < \theta$. Then will the optimal noise σ_{opt} (or γ_{opt}) decrease as the signal amplitude A moves closer to the threshold θ ?

Intuition might suggest that the threshold system should need less noise to produce the entropic SR effect as the amplitude moves closer to the threshold θ . But the results

in Figure 6 show that the compound nonlinearities involved produce no such simple relationship. The different noise types produce different SR optimality schedules. Figure 6 shows four optimal noise schedules for the threshold value $\theta = 0.5$. Other threshold values produced similar results. Only optimal Laplace and Cauchy noise produce the more intuitive monotone decrease in the optimal noise level with rising signal amplitude A . Optimal uniform noise grows linearly with signal amplitude while optimal Gaussian noise defines a nonmonotonic schedule.

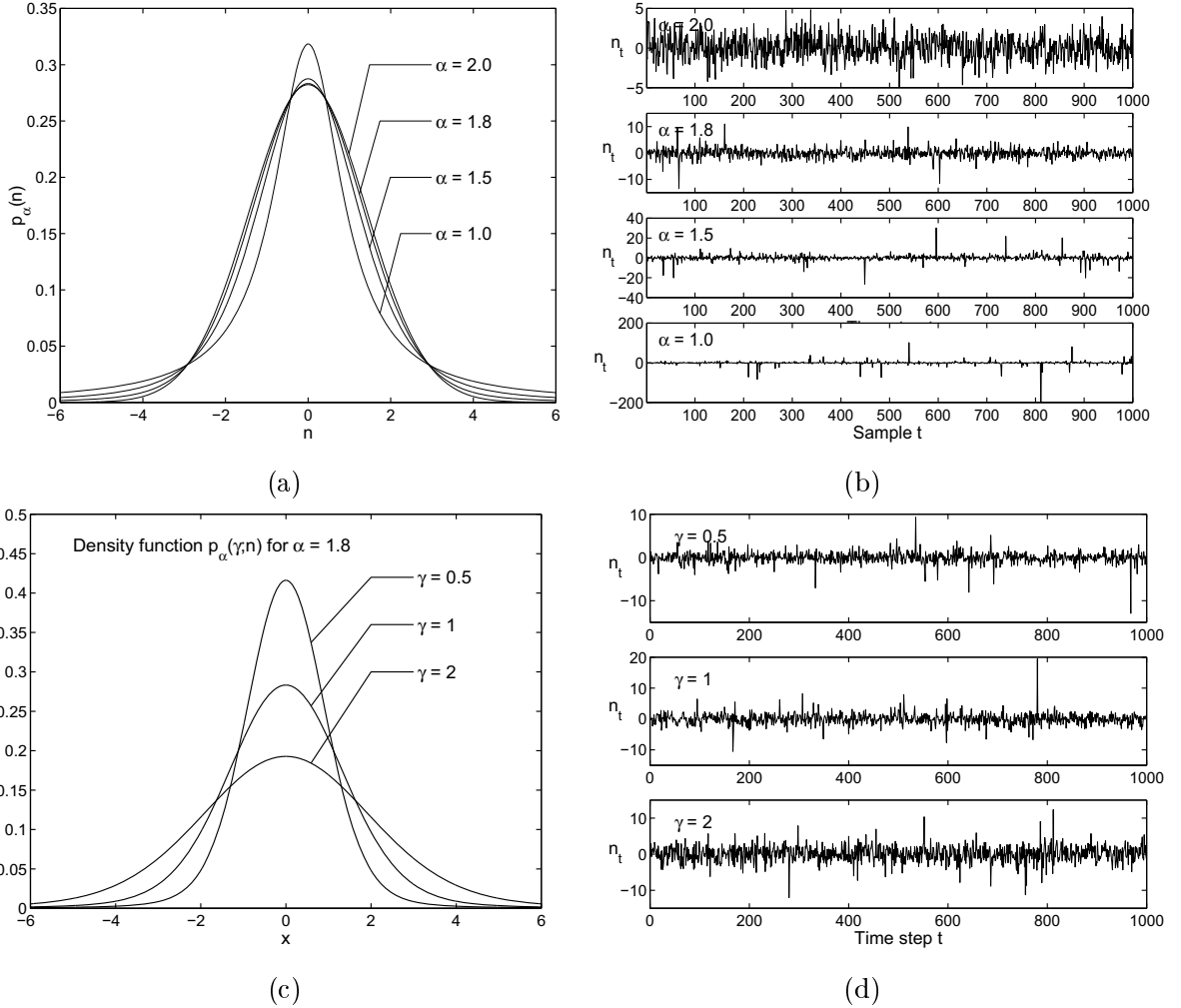


Figure 4: Samples of standard symmetric alpha-stable probability densities and their realizations. (a) Density functions with zero location ($a = 0$) and unit dispersion ($\gamma = 1$) for $\alpha = 2, 1.8, 1.5$, and 1 . The densities are bell curves that have thicker tails as α decreases and thus that 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. (b) Samples of alpha-stable random variables with zero location and unit dispersion. The plots show realizations when $\alpha = 2, 1.8, 1.5$, and 1 . Note the scale differences on the y -axes. The alpha-stable noise n becomes more impulsive as the parameter α falls. The algorithm in [17, 85] generates these realizations. (c) Density functions for $\alpha = 1.8$ with dispersions $\gamma = 0.5, 1$, and 2 . (d) Samples of alpha-stable noise n for $\alpha = 1.8$ with dispersions $\gamma = 0.5, 1$, and 2 .

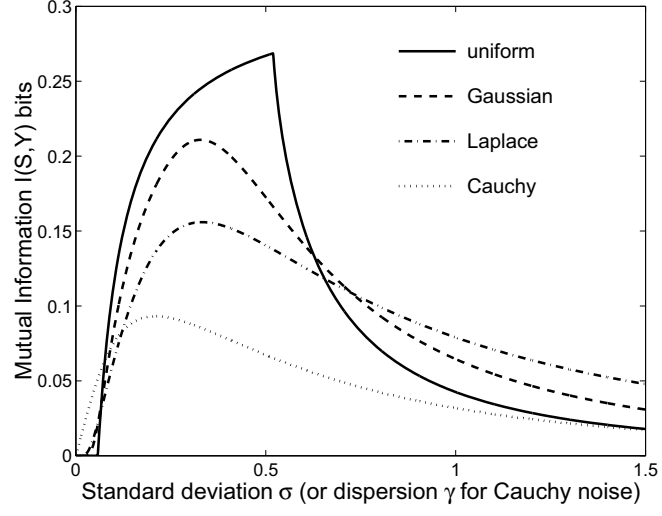


Figure 5: Mutual information I profiles of a threshold system with bipolar input for four kinds of noise. The system has threshold $\theta = 0.5$. The input Bernoulli signal is bipolar with amplitude $A = 0.4$.

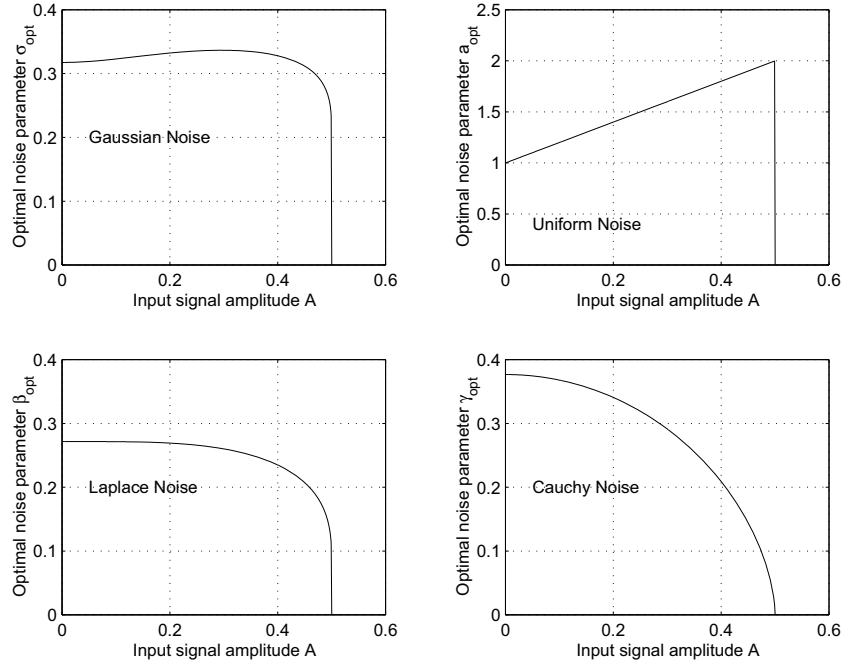


Figure 6: Optimal SR noise schedules for a noisy threshold neuron with threshold $\theta = 0.5$. The schedules show how optimal noise variance or dispersion depends on signal amplitude A for the four closed-form noise results in Figure 5.

4 SR for Threshold Systems with Finite-Variance Noise

Almost all finite-variance noise densities produce the SR effect in threshold neurons with subthreshold signals. This holds for all probability density functions defined on a two-symbol alphabet. The proof of Theorem 1.1 in [60] 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 “forbidden” signal-threshold interval $(\theta - A, \theta + A)$.

Theorem 1.1. *Suppose that the threshold signal system (6) 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 system (6) exhibits the nonmonotone SR effect in the sense that $I(S, Y) \rightarrow 0$ as $\sigma \rightarrow 0$.*

Corollary 1.1. *The threshold neuron (6) exhibits stochastic resonance for the additive gamma, beta, and Weibull noise densities under the hypotheses of Theorem 1.1.*

The gamma density has the form

$$p(n) = \begin{cases} \frac{n^{\alpha-1} e^{-n/\beta}}{\Gamma(\alpha) \beta^\alpha} & n \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (42)$$

Parameters α and β are positive constants and Γ is the gamma function

$$\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy \quad x > 0. \quad (43)$$

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. Figure 7 shows simulation realizations of this corollary for the gamma noise density.

The generalized beta probability density function has the form

$$p(n) = \begin{cases} \frac{1}{b-a} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{n-a}{b-a}\right)^{\alpha-1} \left(\frac{b-n}{b-a}\right)^{\beta-1} & a \leq n \leq b \\ 0 & \text{otherwise} \end{cases} \quad (44)$$

Parameters α and β are positive shape constants, parameters a and b are constants $-\infty < a < b < \infty$, and Γ is the gamma function

$$\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy \quad x > 0. \quad (45)$$

The mean and variance of the beta density are

$$m_n = a + (b-a) \frac{\alpha}{\alpha+\beta} \quad (46)$$

$$\sigma_n^2 = \frac{(b-a)^2 \alpha \beta}{(\alpha+\beta)^2 (\alpha+\beta+1)} \quad (47)$$

So the beta density is right-sided for $a \geq 0$. We used $a = 0$ and $b = 10$ and so defined the beta density in the interval $[0, 10]$ for the SR simulation instance in Figure 8. The algorithm in [18] generated the beta noise. Bayesian statisticians often use a beta density to encode prior information about a parameter (such as a binomial success parameter p) over a fixed-length interval [81]. The beta density can also model the semblance or the ratio of stacked energy to total energy across a signal array [56], fluctuations of the radar-scattering cross-sections of targets [61], the self-similar process of video traffic [6], and the variation of the narrow-band vector channels or spatial signature variations due to movement [55].

The Weibull probability density function has the form

$$p(n) = \begin{cases} \alpha n^{\beta-1} e^{-\alpha n^\beta / \beta} & n \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (48)$$

for positive shape parameters α and β . The mean and variance of the Weibull density are

$$m_n = \left(\frac{\beta}{\alpha}\right)^{1/\beta} \Gamma\left(1 + \frac{1}{\beta}\right) \quad (49)$$

$$\sigma_n^2 = \left(\frac{\beta}{\alpha}\right)^{2/\beta} \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \left\{ \Gamma\left(1 + \frac{1}{\beta}\right) \right\}^2 \right] \quad (50)$$

Figure 9 shows simulation realizations of this corollary for the Weibull noise density. Matlab 6.5 [28] generated the Weibull noise. Weibull [89] first proposed this parametric probability density function to model the fracture of materials under repetitive

stress. This density has become a standard model of multi-part system reliability [71]. It can also effectively model signals and noise in many systems such as radar clutter [83] and confocal laser scanning microscope data [49].

Proof of Theorem 1.1.

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: $I(\sigma) \rightarrow 0$ as $\sigma \rightarrow 0$. Recall that $I(S, Y) = 0$ if and only if S and Y are statistically independent [25]. 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 $\sigma \rightarrow 0$ for some signal symbols $s \in S$ and $y \in Y$. The two-symbol alphabet set S gives

$$P_Y(y) = \sum_s P_{Y|S}(y|s)P_S(s) \quad (51)$$

$$= P_{Y|S}(y|0)P_S(0) + P_{Y|S}(y|1)P_S(1) \quad (52)$$

$$= P_{Y|S}(y|0)P_S(0) + P_{Y|S}(y|1)(1 - P_S(0)) \quad (53)$$

$$= (P_{Y|S}(y|0) - P_{Y|S}(y|1))P_S(0) + P_{Y|S}(y|1) \quad (54)$$

So we need to show only that $P_{Y|S}(y|0) - P_{Y|S}(y|1) = 0$ as $\sigma \rightarrow 0$. 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 (13) and (15) imply that

$$\begin{aligned} P_{Y|S}(0|0) - P_{Y|S}(0|1) \\ = \int_{-\infty}^{\theta+A} p(n)dn - \int_{-\infty}^{\theta-A} p(n)dn \end{aligned} \quad (55)$$

$$= \int_{\theta-A}^{\theta+A} p(n)dn \quad (56)$$

Similarly for $y = '1'$:

$$P_{Y|S}(1|0) = \int_{\theta+A}^{\infty} p(n)dn \quad (57)$$

$$P_{Y|S}(1|1) = \int_{\theta-A}^{\infty} p(n)dn \quad (58)$$

Then

$$P_{Y|S}(1|0) - P_{Y|S}(1|1) = - \int_{\theta-A}^{\theta+A} p(n)dn \quad (59)$$

The result now follows if

$$\int_{\theta-A}^{\theta+A} p(n)dn \rightarrow 0 \quad \text{as } \sigma \rightarrow 0 \quad (60)$$

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 $\varepsilon = \frac{1}{2}d(\theta - A, m) = \frac{1}{2}(\theta - A - m) > 0$. So $\theta - A - \varepsilon = \theta - A - \varepsilon + m - m = m + (\theta - A - m) - \varepsilon = m + 2\varepsilon - \varepsilon = m + \varepsilon$. Then

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) = \int_{\theta-A}^{\theta+A} p(n)dn \quad (61)$$

$$\leq \int_{\theta-A}^{\infty} p(n)dn \quad (62)$$

$$\leq \int_{\theta-A-\varepsilon}^{\infty} p(n)dn \quad (63)$$

$$= \int_{m+\varepsilon}^{\infty} p(n)dn \quad (64)$$

$$= Pr\{n \geq m + \varepsilon\} \quad (65)$$

$$= Pr\{n - m \geq \varepsilon\} \quad (66)$$

$$\leq Pr\{|n - m| \geq \varepsilon\} \quad (67)$$

$$\leq \frac{\sigma^2}{\varepsilon^2} \quad \text{by Chebyshev's inequality} \quad (68)$$

$$\rightarrow 0 \quad \text{as } \sigma \rightarrow 0 \quad (69)$$

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

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) \leq \frac{\sigma^2}{\varepsilon^2} \rightarrow 0 \quad \text{as } \sigma \rightarrow 0 \quad (70)$$

Q.E.D.

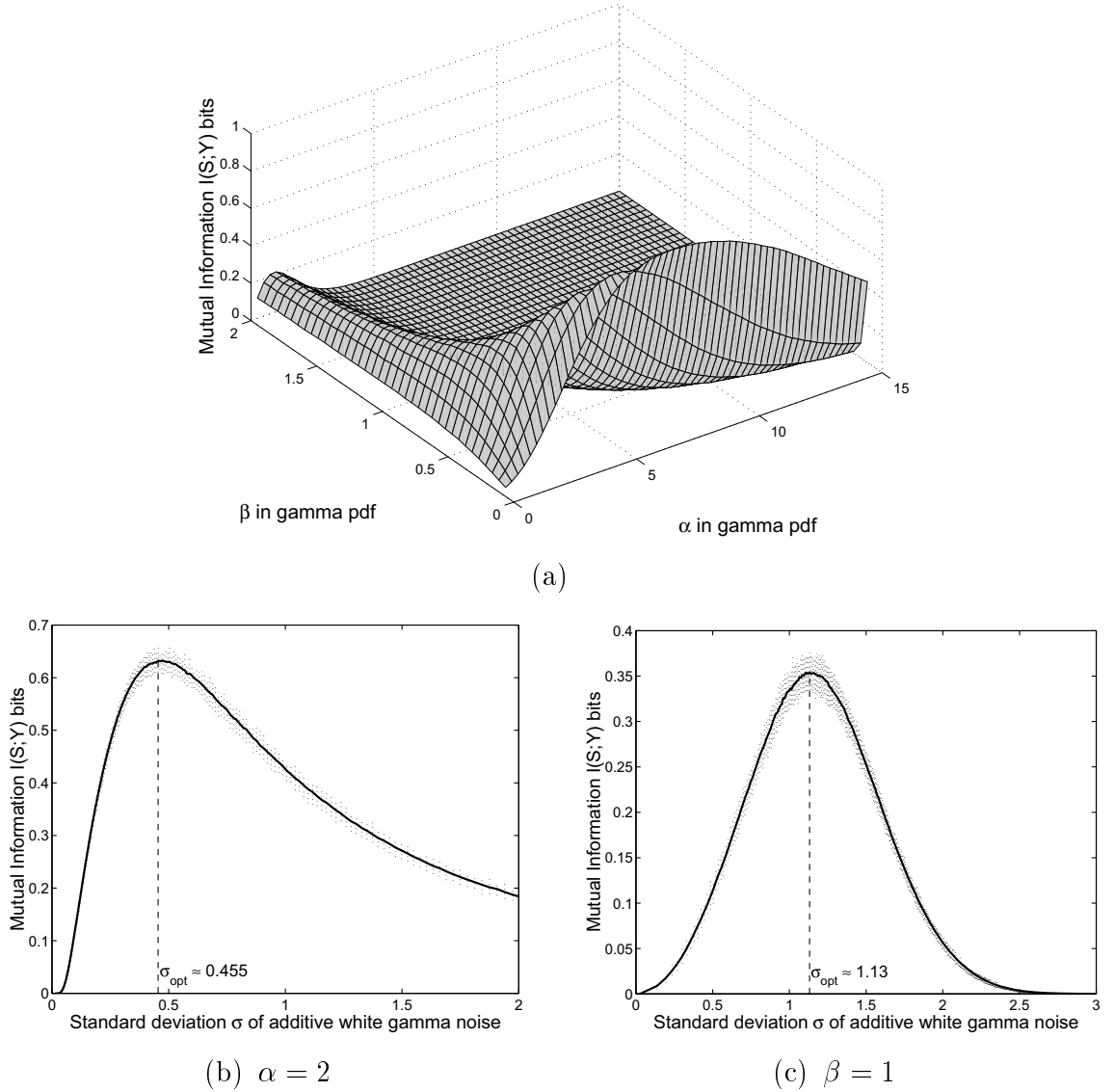


Figure 7: Stochastic resonance with (finite-variance) gamma noise. The noisy signal-forced threshold neuron has the form (6). The gamma noise n_t adds to the bipolar input Bernoulli signal s_t . The neuron has threshold $\theta = 1$. The input Bernoulli signal has amplitude $A = 0.8$ with success probability $p = \frac{1}{2}$. 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 [1, 2] 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 $\sigma_{opt} > 0$ for each $\alpha > 0$. It also has a nonzero noise optimum $\sigma_{opt} > 0$ for each $\beta > 0$. Figure (b) shows the cross section of the mutual-information surface for $\alpha = 2$. Figure (c) shows the cross section for $\beta = 1$. Note that the mean and variance of the gamma noise are $m = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$.

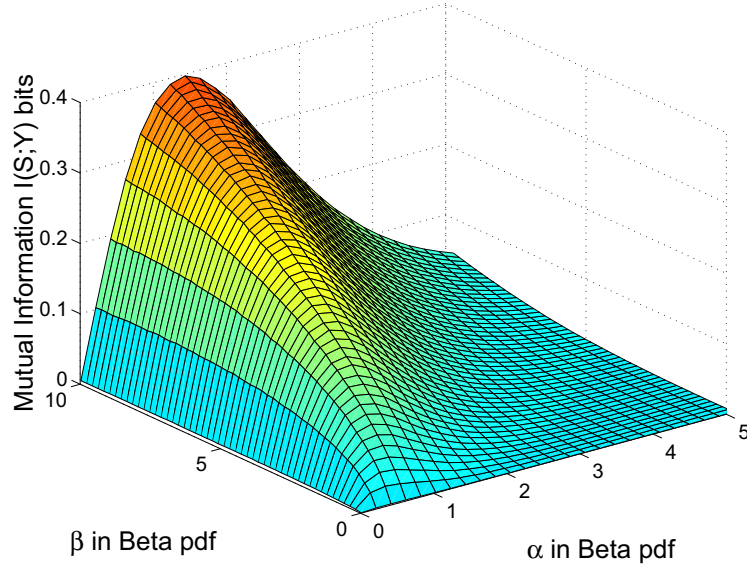


Figure 8: Stochastic resonance with (right-sided) beta noise. The noisy signal-forced threshold neuron has the form (6). The beta noise n_t adds to the bipolar input Bernoulli signal s_t . The parametrized interval $[a, b]$ of the beta density (44) has $a = 0$ and $b = 10$. The neuron has threshold $\theta = 1$. 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 graph shows the smoothed input-output mutual information of a threshold neuron as a function of the parameters α and β of additive white beta noise n_t . The neuron's mutual information has a nonzero noise optimum $\sigma_{opt} > 0$ where the variance has the form $\sigma_n^2 = \frac{(b-a)^2 \alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$.

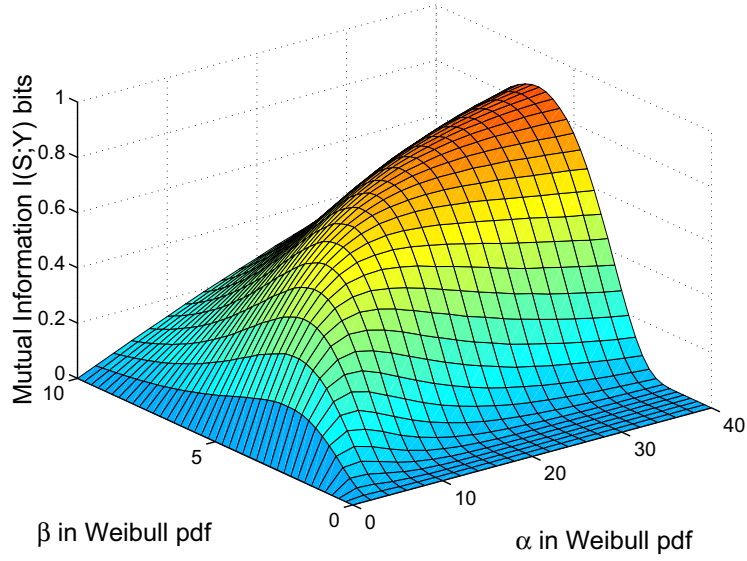


Figure 9: Stochastic resonance with (right-sided) Weibull noise. The noisy signal-forced threshold neuron has the form (6). The Weibull noise n_t adds to the bipolar input Bernoulli signal s_t . The neuron has threshold $\theta = 0.5$. The input Bernoulli signal has amplitude $A = 0.2$ with success probability $p = \frac{1}{2}$. Each trial produced 10,000 input-output samples $\{s_t, y_t\}$ that estimated the probability densities to obtain the mutual information. The graph shows the smoothed input-output mutual information of a threshold neuron as a function of the parameters α and β of additive white Weibull noise n_t . The neuron's mutual information has a nonzero noise optimum $\sigma_{opt} > 0$ where the variance has the form $\sigma_n^2 = \left(\frac{\beta}{\alpha}\right)^{2/\beta} \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \left\{ \Gamma\left(1 + \frac{1}{\beta}\right) \right\}^2 \right]$.

5 SR for Threshold Systems with Infinite-Variance Noise

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 [59, 72]. We can model many types of impulsive noise with *symmetric* alpha-stable bell-curve probability density functions with parameter α in the characteristic function $\phi(\omega) = \exp\{-\gamma|\omega|^\alpha\}$. Here γ is the *dispersion* parameter [10, 34, 41, 77]. Figure 4 shows examples of symmetric (bell-curve) alpha-stable probability density functions with different α tail thicknesses and different bell-curve dispersions γ .

Theorem 2.1 applies to *any* alpha-stable noise model. The density need not be symmetric. The proof of Theorem 2.1 is simpler than the proof in the finite-variance case because all stable noise densities have a characteristic function with the exponential form in (39)-(40). 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 2.1. Suppose $I(S, Y) > 0$ and the threshold neuron (6) uses alpha-stable noise with location parameter $a \notin (\theta - A, \theta + A)$. Then the neuron (6) exhibits the nonmonotone SR effect if the input signal is subthreshold.

Proof of Theorem 2.1. Again the result follows if

$$\int_{\theta-A}^{\theta+A} p(n)dn \rightarrow 0 \quad \text{as } \gamma \rightarrow 0 \quad (71)$$

The characteristic function $\varphi(\omega)$ of alpha-stable noise density $p(n)$ has the exponential form (39)-(40). This reduces to a simple complex exponential in the zero-dispersion limit:

$$\lim_{\gamma \rightarrow 0} \varphi(\omega) = \exp\{ia\omega\} \quad (72)$$

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

$$\lim_{\gamma \rightarrow 0} p(n) = \delta(n - a) \quad (73)$$

Then

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) = \int_{\theta-A}^{\theta+A} p(n)dn \quad (74)$$

$$= \int_{\theta-A}^{\theta+A} \delta(n-a) dn \quad (75)$$

$$= 0 \quad (76)$$

because $a \notin [\theta - A, \theta + A]$. Q.E.D.

We next state two theorems that show that we cannot in general omit the threshold-interval condition in the hypothesis of Theorem 1.1. Noise does not help a threshold θ that already lies between $\theta - A$ and $\theta + A$.

Theorem 1.2. *Suppose that the threshold signal system (6) has noise probability density function $p(n)$ and that the input signal S is subthreshold ($A < \theta$). Suppose that the noise mean $E[n]$ lies in the signal-threshold interval $(\theta - A, \theta + A)$ if $p(n)$ has finite variance. Then the threshold system (6) does not exhibit the nonmonotone SR effect in the sense that $I(S, Y)$ is maximum as $\sigma \rightarrow 0$:*

$$I(S, Y) = H(Y) = H(S) \quad \text{as } \sigma \rightarrow 0 \quad (77)$$

Theorem 2.2. *Suppose that the threshold signal system (6) has subthreshold input signal and use alpha-stable noise with location parameter $a \in (\theta - A, \theta + A)$. Then the threshold system (6) does not exhibit the nonmonotone SR effect: $I(S, Y)$ is maximum as $\gamma \rightarrow 0$:*

$$I(S, Y) = H(Y) = H(S) \quad \text{as } \gamma \rightarrow 0 \quad (78)$$

The two proofs below use the same idea as do the proofs for Theorems 1.1 and 2.1. Assume $0 < P_S(s) < 1$ to avoid triviality when $P_S(s) = 0$ or 1 . We show that $H(Y) \rightarrow H(S)$ and $H(Y|S) \rightarrow 0$ as $\sigma \rightarrow 0$ or $\gamma \rightarrow 0$. So $I(S, Y) \rightarrow H(S)$ as $\sigma \rightarrow 0$ or $\gamma \rightarrow 0$ and is maximum since $I(S, Y) = H(Y) - H(Y|S)$ and $I(S, Y) \leq H(S)$ by the Data Processing Inequality: $I(S, S) \geq I(S, g(S)) = I(S, Y)$ for a Markov chain $S \rightarrow S \rightarrow Y$ [25]. The boundary case $I(S, S) = H(S)$ implies $I(S, Y) \leq H(S)$.

Proof of Theorem 1.2: Finite-variance noise case. Now we show that $P_{Y|S}(y|s)$ is either 1 or 0 as $\sigma \rightarrow 0$ or $\gamma \rightarrow 0$. Let the mean of the noise be $m = E[n]$ and the variance be $\sigma^2 = E[(x - m)^2]$. Then $m \in (\theta - A, \theta + A)$ by hypothesis.

Consider $P_{Y|S}(0|0)$. Pick $\varepsilon = \frac{1}{2}d(\theta + A, m) = \frac{1}{2}(\theta + A - m) > 0$. So $\theta + A - \varepsilon = \theta + A - \varepsilon + m - m = m + (\theta + A - m) - \varepsilon = m + 2\varepsilon - \varepsilon = m + \varepsilon$. Then

$$P_{Y|S}(0|0) = \int_{-\infty}^{\theta+A} p(n)dn \quad (79)$$

$$\geq \int_{-\infty}^{\theta+A-\varepsilon} p(n)dn \quad (80)$$

$$= \int_{-\infty}^{m+\varepsilon} p(n)dn \quad (81)$$

$$= 1 - \int_{m+\varepsilon}^{\infty} p(n)dn \quad (82)$$

$$= 1 - \Pr\{n \geq m + \varepsilon\} = 1 - \Pr\{n - m \geq \varepsilon\} \quad (83)$$

$$\geq 1 - \Pr\{|n - m| \geq \varepsilon\} \quad (84)$$

$$\geq 1 - \frac{\sigma^2}{\varepsilon^2} \quad \text{by Chebyshev's inequality} \quad (85)$$

$$\rightarrow 1 \quad \text{as } \sigma \rightarrow 0 \quad (86)$$

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

Similarly for $P_{Y|S}(1|1)$. Pick $\varepsilon = \frac{1}{2}d(\theta - A, m) = \frac{1}{2}(m - \theta + A) > 0$. So $\theta - A + \varepsilon = \theta - A + \varepsilon + m - m = m + (\theta - A - m) + \varepsilon = m - 2\varepsilon + \varepsilon = m - \varepsilon$. Then

$$P_{Y|S}(1|1) = \int_{\theta+A}^{\infty} p(n)dn \quad (87)$$

$$\geq \int_{\theta-A+\varepsilon}^{\infty} p(n)dn \quad (88)$$

$$= \int_{m-\varepsilon}^{\infty} p(n)dn \quad (89)$$

$$= 1 - \int_{-\infty}^{m-\varepsilon} p(n)dn \quad (90)$$

$$= 1 - \Pr\{n \leq m - \varepsilon\} = 1 - \Pr\{n - m \leq -\varepsilon\} \quad (91)$$

$$\geq 1 - \Pr\{|n - m| \geq \varepsilon\} \quad (92)$$

$$\geq 1 - \frac{\sigma^2}{\varepsilon^2} \quad \text{by Chebyshev's inequality} \quad (93)$$

$$\rightarrow 1 \quad \text{as } \sigma \rightarrow 0 \quad (94)$$

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

Proof of Theorem 2.2: Alpha-stable noise case. The characteristic function $\varphi(\omega)$ of alpha-stable noise density $p(n)$ has the exponential form (39)-(40). This reduces to a simple complex exponential in the zero-dispersion limit:

$$\lim_{\gamma \rightarrow 0} \varphi(\omega) = \exp\{i a \omega\} \quad (95)$$

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

$$\lim_{\gamma \rightarrow 0} p(n) = \delta(n - a) \quad (96)$$

Then

$$P_{Y|S}(0|0) = \int_{-\infty}^{\theta+A} p(n) dn \quad (97)$$

$$\rightarrow \int_{-\infty}^{\theta+A} \delta(n - a) dn = 1 \quad \text{as } \gamma \rightarrow 0 \quad (98)$$

Similarly

$$P_{Y|S}(1|1) = \int_{\theta-A}^{\infty} p(n) dn \quad (99)$$

$$\rightarrow \int_{\theta-A}^{\infty} \delta(n - a) dn = 1 \quad \text{as } \gamma \rightarrow 0 \quad (100)$$

The two conditional probabilities for both finite-variance and infinite-variance cases imply that $P_{Y|S}(0|1) = P_{Y|S}(1|0) = 0$ as $\sigma \rightarrow 0$ and $\gamma \rightarrow 0$. (We can proceed in a similar manner to obtain these two probabilities). These four probabilities further imply that

$$H(Y|S) = - \sum_s \sum_y P_{SY}(s, y) \log_2 P_{Y|S}(y|s) \quad (101)$$

$$= \sum_s P_S(s) \sum_y P_{Y|S}(y|s) \log_2 P_{Y|S}(y|s) \quad (102)$$

$$= 0 \quad (103)$$

where we use the convention that $0 \log_2 0 = 0$. The entropy $H(Y)$ becomes

$$H(Y) = - \sum_y P_Y(y) \log_2 P_Y(y) \quad (104)$$

$$= - \sum_s P_S(s) \log_2 P_S(s) \quad (105)$$

$$= H(S) \quad (106)$$

since

$$P_Y(y) = \sum_s P_{Y|S}(y|s) P_S(s) \quad (107)$$

$$= P_{Y|S}(y|0) P_S(0) + P_{Y|S}(y|1) P_S(1) \quad (108)$$

$$= P_{Y|S}(y|0) P_S(0) + P_{Y|S}(y|1) (1 - P_S(0)) \quad (109)$$

$$= (P_{Y|S}(y|0) - P_{Y|S}(y|1)) P_S(0) + P_{Y|S}(y|1) \quad (110)$$

$$= (P_{Y|S}(y|1) - P_{Y|S}(y|0))P_S(1) + P_{Y|S}(y|0) \quad (111)$$

$$= \begin{cases} P_S(1) & \text{if } y = 1 \\ P_S(0) & \text{if } y = 0 \end{cases} \quad (112)$$

Q.E.D.

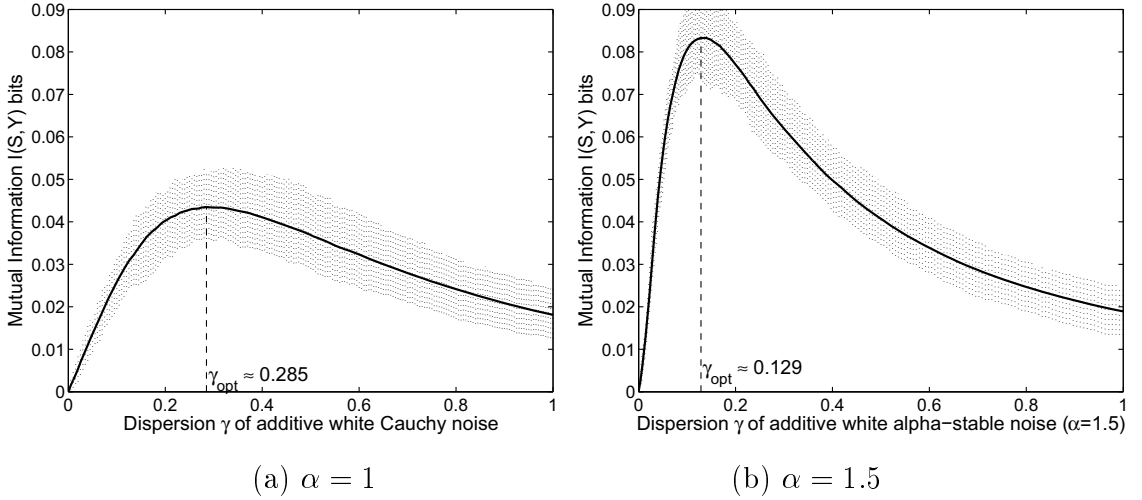


Figure 10: Stochastic resonance with highly impulsive (infinite-variance) alpha-stable noise. The graphs show the smoothed input-output mutual information of a threshold system as a function of the dispersion of additive white alpha-stable noise n_t with $\alpha = 1$ (Cauchy noise) in (a) and $\alpha = 1.5$ in (b). The vertical dashed lines show the absolute deviation between the smallest and largest outliers in each sample average of 100 outcomes. The system has a nonzero noise optimum at $\gamma_{opt} \approx 0.285$ for $\alpha = 1$ and $\gamma_{opt} \approx 0.129$ for $\alpha = 1.5$ and thus shows the SR effect. The noisy signal-forced threshold system has the form (6). The alpha-stable noise n_t adds to the bipolar input Bernoulli signal s_t . The system has threshold $\theta = 0.5$. The input Bernoulli signal has amplitude $A = 0.3$ 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. Note that decreasing the tail-thickness parameter α increases the optimal noise dispersion γ_{opt} as in Figure 12 and decreases the SR-maximal mutual information $I_{max}(S, Y)$ as in Figure 13.

Figure 10 gives a typical example of the SR effect for highly impulsive noise with infinite variance. The alpha-stable noises have $\alpha = 1$ (Cauchy) and $\alpha = 1.5$. So frequent and violent noise spikes interfere with the signal. Figure 10 also illustrates the empirical trends in Figures 12 and 13: A falling tail-thickness parameter α produces an increasing optimal noise dispersion γ_{opt} but a decreasing SR-maximal mutual in-

formation $I_{max}(S, Y)$. We next state a new sufficient condition for SR not to occur in an impulsive threshold system.

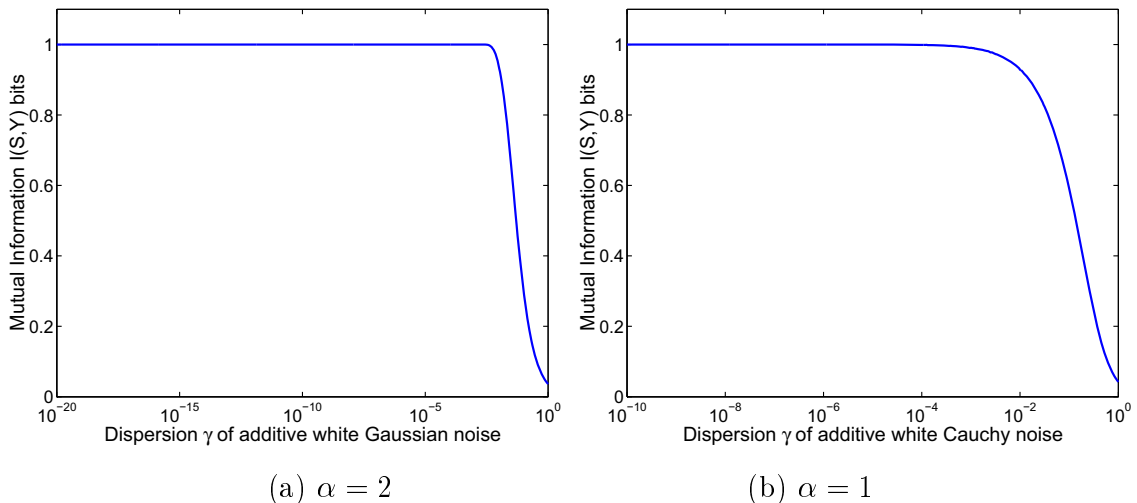


Figure 11: No SR in the “forbidden” interval (per Theorems 1.2 and 2.2)—mutual information versus alpha-stable noise dispersion when the noise mean (location) lies in the “forbidden” signal-threshold interval: $a \in (\theta - A, \theta + A)$. The graphs show the smoothed input-output mutual information of 100 trials of a threshold system as a function of the dispersion of additive white alpha-stable noise n_t with $\alpha = 2$ (Gaussian) in (a) and $\alpha = 1$ (Cauchy noise) in (b). The system is optimal when $\gamma \rightarrow 0$ and thus *does not* show the SR effect: The mutual information $I(S, Y)$ is maximum as it equals the input entropy $H(S)$. The noisy signal-forced threshold system has the form (6). The alpha-stable noise n_t has location $a = 0.4$ and adds to the bipolar input Bernoulli signal s_t . The system has threshold $\theta = 0.5$. The input Bernoulli signal has amplitude $A = 0.4$ 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.

Statistical regression confirmed an exponential relationship between the optimal noise dispersion γ_{opt} and the bell-curve tail-thickness parameter α : $\gamma_{opt}(\alpha) = 10^{\beta_0 + \beta_1 \alpha}$ for parameters β_0 and β_1 that depend on the signal amplitude A . Then the log-transformation of the optimal dispersion gives the linear model $\log_{10} \gamma_{opt}(\alpha) = \beta_0 + \beta_1 \alpha$. Table 1 shows the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ and the coefficient of determination r_l^2 for 20 signal amplitudes in the threshold neuron using SPSS software. All observed significance levels or p -values were less than 10^{-4} . The p -values measure the credibility of the null hypothesis that the regression lines have zero slope

or other coefficients. The exponential trend's exponent is linear for most amplitudes but becomes quadratic for very small amplitudes and for amplitudes close to the threshold $\theta = \frac{1}{2}$ (or $\gamma_{opt}(\alpha) = 10^{\beta_0 + \beta_1\alpha + \beta_2\alpha^2}$ for a quadratic fit to the data). Figure 12 shows 6 of the 20 log-linear plots. We also found an approximate linear relationship $I_{max}(S, Y; \alpha) = \beta_0 + \beta_1\alpha$ for the SR-maximal mutual information $I_{max}(S, Y)$ as a function of the tail-thickness parameter α . Table 2 shows the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ and the coefficient of determination r_l^2 for 20 signal amplitudes in the threshold neuron. All observed significance levels or p -values were less than 10^{-4} . There is a clear linear trend for most amplitudes A . The trend becomes quadratic for very small amplitudes and for amplitudes close to the threshold $\theta = \frac{1}{2}$. Figure 13 shows 6 of the 20 linear plots.

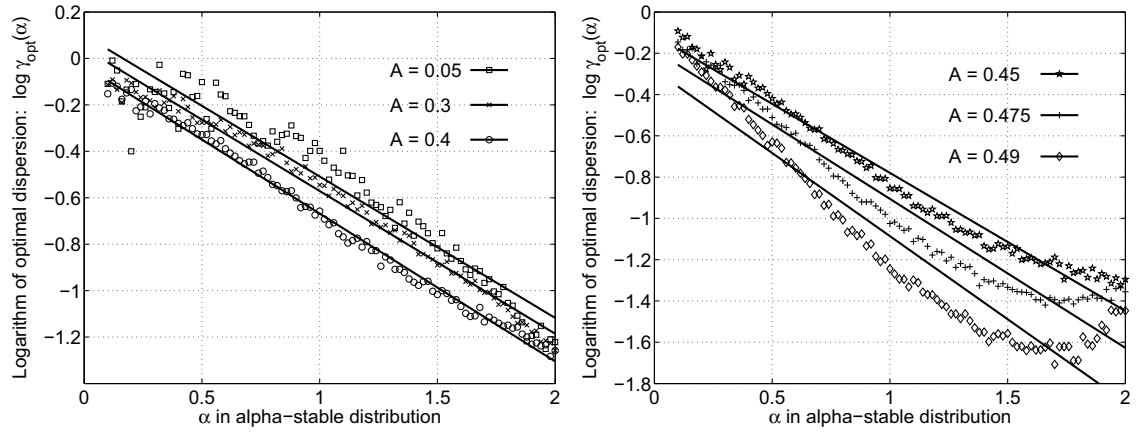


Figure 12: Exponential law for optimal noise dispersion γ_{opt} as a function of bell-curve thickness parameter α for the mutual-information performance measure and for different signal amplitudes A . The optimal noise dispersion γ_{opt} depends on the parameter α through the exponential relation $\gamma_{opt}(\alpha) = 10^{\beta_0 + \beta_1\alpha}$ for parameters β_0 and β_1 (or $\gamma_{opt}(\alpha) = 10^{\beta_0 + \beta_1\alpha + \beta_2\alpha^2}$ for a quadratic fit to the data). Table 1 shows the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ for 20 input Bernoulli signal amplitudes A . The exponential trend's exponent is linear for most amplitudes but becomes quadratic for very small amplitudes and for amplitudes close to the threshold $\theta = \frac{1}{2}$. All observed significance levels or p -values were less than 10^{-4} .

Signal Amplitude A	Linear model Regression coefficients			Quadratic model r_q^2
	$\hat{\beta}_0$	$\hat{\beta}_1$	r_l^2	
0.025	0.0701	-0.5944	0.9003	0.9444
0.050	0.1002	-0.6087	0.9321	0.9723
0.075	0.1124	-0.6192	0.9490	0.9842
0.100	0.1180	-0.6261	0.9558	0.9888
0.125	0.1090	-0.6228	0.9594	0.9910
0.150	0.1078	-0.6251	0.9679	0.9921
0.175	0.1026	-0.6273	0.9672	0.9933
0.200	0.0915	-0.6214	0.9699	0.9942
0.225	0.0810	-0.6161	0.9737	0.9950
0.250	0.0694	-0.6172	0.9781	0.9959
0.275	0.0595	-0.6149	0.9826	0.9964
0.300	0.0439	-0.6148	0.9869	0.9961
0.325	0.0290	-0.6184	0.9903	0.9962
0.350	0.0116	-0.6211	0.9935	0.9961
0.375	-0.0134	-0.6215	0.9957	0.9960
0.400	-0.0313	-0.6367	0.9947	0.9951
0.425	-0.0705	-0.6432	0.9903	0.9950
0.450	-0.1107	-0.6688	0.9757	0.9944
0.475	-0.1837	-0.7217	0.9408	0.9911
0.490	-0.2805	-0.8053	0.8987	0.9863

Table 1: Linear regression estimates of the SR-optimal log dispersion γ_{opt} as a function of the bell-curve tail-thickness parameter α from a symmetric alpha-stable noise density. The parameters β_0 and β_1 relate $\log_{10} \gamma_{opt}$ and α through a linear relationship: $\log_{10} \gamma_{opt}(\alpha) = \beta_0 + \beta_1 \alpha$. The coefficient of determination r_l^2 shows how well the linear model fits the log-transformed data. The last column shows the coefficient of determination r_q^2 for the quadratic model $\log_{10} \gamma_{opt}(\alpha) = \beta_0 + \beta_1 \alpha + \beta_2 \alpha^2$. All observed significance levels or p -values were less than 10^{-4} .

Signal Amplitude A	Linear model Regression coefficients			Quadratic model r_q^2
	$\hat{\beta}_0$	$\hat{\beta}_1$	r_l^2	
0.025	-0.0001	0.0006	0.9312	0.9907
0.050	-0.0008	0.0022	0.9370	0.9972
0.075	-0.0018	0.0049	0.9401	0.9985
0.100	-0.0031	0.0086	0.9440	0.9990
0.125	-0.0048	0.0134	0.9477	0.9993
0.150	-0.0068	0.0190	0.9521	0.9995
0.175	-0.0090	0.0256	0.9558	0.9997
0.200	-0.0113	0.0329	0.9612	0.9998
0.225	-0.0138	0.0411	0.9658	0.9998
0.250	-0.0161	0.0500	0.9715	0.9997
0.275	-0.0185	0.0596	0.9764	0.9995
0.300	-0.0207	0.0698	0.9816	0.9993
0.325	-0.0224	0.0807	0.9866	0.9990
0.350	-0.0236	0.0920	0.9913	0.9987
0.375	-0.0240	0.1039	0.9951	0.9984
0.400	-0.0229	0.1161	0.9976	0.9981
0.425	-0.0196	0.1286	0.9972	0.9977
0.450	-0.0120	0.1408	0.9905	0.9975
0.475	0.0058	0.1513	0.9655	0.9973
0.490	0.0336	0.1527	0.9145	0.9959

Table 2: Linear regression of the SR-maximal mutual information $I_{max}(S, Y)$ as a function of the bell-curve tail-thickness parameter α from a symmetric alpha-stable noise density. The parameters β_0 and β_1 relate $I_{max}(S, Y)$ and α through a linear relationship: $I_{max}(S, Y; \alpha) = \beta_0 + \beta_1 \alpha$. The coefficient of determination r_l^2 shows how well the linear model fits the data. The last column shows the coefficient of determination r_q^2 for the quadratic model $I_{max}(S, Y; \alpha) = \beta_0 + \beta_1 \alpha + \beta_2 \alpha^2$. All observed significance levels or p -values were less than 10^{-4} .

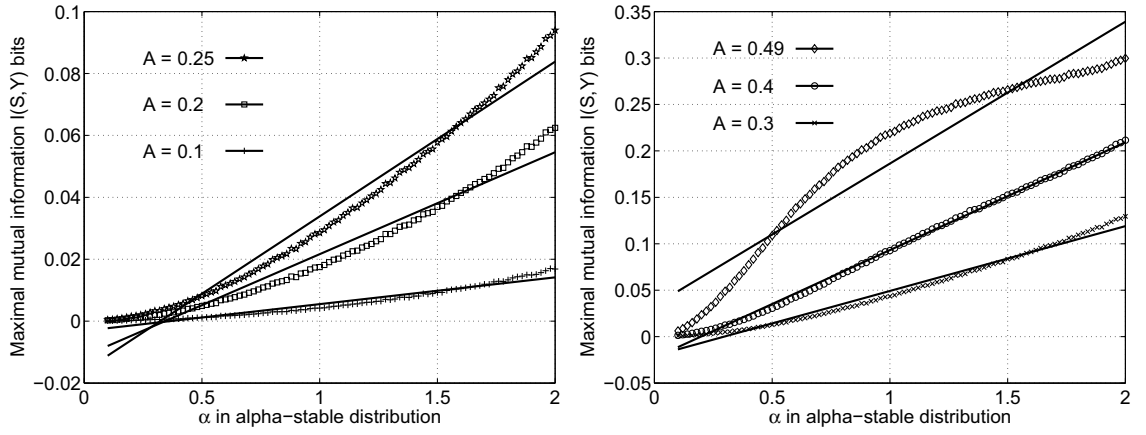


Figure 13: Linear regression for maximal mutual information $I_{max}(S, Y)$ as a function of bell-curve thickness parameter α for different signal amplitudes A . The maximal mutual information $I_{max}(S, Y)$ depends on the parameter α through the linear relationship $I_{max}(S, Y; \alpha) = \beta_0 + \beta_1 \alpha$ for parameters β_0 and β_1 (or $I_{max}(\alpha) = \beta_0 + \beta_1 \alpha + \beta_2 \alpha^2$ for a quadratic fit to the data). Table 2 shows the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ for 20 input Bernoulli signal amplitudes A . The linear trend is strong for most amplitudes A . The trend becomes quadratic for very small amplitudes and for amplitudes close to the threshold $\theta = \frac{1}{2}$. All observed significance levels or p -values were less than 10^{-4} .

6 Computer Simulation for the SR Effect in Continuous Neurons

Discrete simulations can model continuous-time nonlinear dynamical systems if a stochastic numerical scheme approximates the system dynamics and its signal and noise response. We used a simple stochastic version of the Euler scheme to model a nonlinear system with input forcing signal and noise. We measured how the system performed based on only the system's input-output samples.

Consider the forced dynamical system with additive forcing input signal s and “white” noise n

$$\dot{x} = f(x) + s(t) + n(t) \quad (113)$$

$$y(t) = g(x(t)). \quad (114)$$

These models simply add a noise term to a differential equation rather than use formal Ito or Stratonovich stochastic differentials [19, 30, 36]. “Whiteness” of a random variable n here means that n is white only over some large but finite frequency bandwidth interval $[-B, B]$ for some large $B > 0$. Random numbers from the algorithms in [17, 80, 85] act as noise from various probability densities in our simulations. The next sections show how discretized continuous-time systems produced the discrete-time systems we used for computer simulations.

6.1 Nonlinear Systems with White Gaussian Noise

Consider the dynamical system (113) with initial condition $x(t_0) = x_0$. Here the white Gaussian noise w has zero mean and unit variance so that $n = \sigma w$ has zero mean and variance σ^2 . This system corresponds to the stochastic initial value problem [36]

$$dX = \tilde{f}(t, X) + \sigma(t, X)dW \quad (115)$$

for initial condition $X(t_0) = X_0$. Here $\tilde{f}(t, X) = f(X) + s(t)$, $\sigma(t, X) = \sigma$, and W is the standard Wiener process [36]. We used Euler's method (the Euler-Maruyama scheme) [26, 36, 51] to obtain the discrete form for computer simulation:

$$x_{t+1} = x_t + \Delta T \left(f(x_t) + s_t \right) + \sigma \sqrt{\Delta T} w_t \quad (116)$$

$$y_t = g(x_t) \quad (117)$$

for $t = 0, 1, 2, \dots$ and initial condition x_0 . The input sample s_t has the value of the signal $s(t\Delta T)$ at time step t . The zero-mean white Gaussian noise sequence $\{w_t\}$ has

unit variance $\sigma_w^2 = 1$. The term $\sqrt{\Delta T}$ scales w_t so that $\sqrt{\Delta T}w_t$ conforms with the Wiener increment [36, 51, 75]. The output sample y_t is some transformation g of the system's state x_t .

This simple algorithm gives fairly accurate results for moderate nonlinear systems [36, 51, 65, 75]. Other algorithms may give more accurate numerical solutions of the stochastic differential equations for more complicated system dynamics [36, 68]. All of our simulations used the Euler's scheme in (116)-(117).

The numerical algorithm in [80] generates a sequence of pseudo-random numbers from a Gaussian density with zero mean and unit variance for $\{w_t\}$ in (116). Figure 3 shows the Gaussian and other densities that have zero mean and a variance of two.

6.2 Nonlinear Systems with Other Finite-Variance Noise

We next consider a system (113) with finite-variance noise n . Suppose the noise n has variance σ^2 and again apply the above Euler's method:

$$x_{t+1} = x_t + \Delta T (f(x_t) + s_t) + \sigma \sqrt{\Delta T} w_t \quad (118)$$

$$y_{t+1} = g(x_{t+1}). \quad (119)$$

Here the random sequence $\{w_t\}$ has density function $p(w)$ with zero mean and unit variance. The numerical algorithms in [80] generate sequences of random variables for Laplace and uniform density functions. Figure 3 plots these probability density functions and their realizations with mean zero and variance of two: $E[x] = 0$ and $E[x^2] = 2$.

6.3 Nonlinear Systems with Alpha-Stable Noise

Figure 3-4 show realizations of the symmetric alpha-stable random variable when $\alpha = 1$ (Cauchy density). Again we assume that the Euler's method above applies to this class of random variables with infinite variance. Let w be a standard alpha-stable random variable with parameter α and zero location and unit dispersion: $a = 0$ and $\gamma = 1$. Let $\kappa = \gamma^{1/\alpha}$ denote a "scale" factor of a random variable. Then $n = \kappa w$ has zero location and dispersion $\gamma = \kappa^\alpha$. This leads to the Euler's numerical solution

$$x_{t+1} = x_t + \Delta T (f(x_t) + s_t) + \kappa \sqrt{\Delta T} w_t \quad (120)$$

$$y_t = g(x_t). \quad (121)$$

The algorithm in [17, 85] generates a standard alpha-stable random variable w .

7 SR and Adaptive SR in Continuous Neurons

This section shows the SR effect in several models of continuous neurons shown in Section 2. Then it shows that the robustified gradient ascent algorithm can also learn the optimal noise levels of these neuron models.

7.1 Derivation of SR Learning Law

The stochastic gradient ascent has the form [72, 73]:

$$\sigma_{k+1} = \sigma_k + \mu_k \frac{\partial I}{\partial \sigma} \quad (122)$$

We assume that $P(s)$ does not depend on σ and we use the natural logarithm. Then the learning term $\frac{\partial I}{\partial \sigma}$ has the form

$$\frac{\partial I}{\partial \sigma} = \frac{\partial}{\partial \sigma} \left(- \sum_y P(y) \log P(y) + \sum_s P(s) \sum_y P(y|s) \log P(y|s) \right) \quad (123)$$

$$= - \sum_y \left(P(y) \frac{1}{P(y)} \frac{\partial P(y)}{\partial \sigma} + \log P(y) \frac{\partial P(y)}{\partial \sigma} \right) + \sum_s \sum_y \left(P(s) P(y|s) \frac{1}{P(y|s)} \frac{\partial P(y|s)}{\partial \sigma} + P(s) \log P(y|s) \frac{\partial P(y|s)}{\partial \sigma} \right) \quad (124)$$

$$= - \sum_y \left(\frac{\partial P(y)}{\partial \sigma} + \log P(y) \frac{\partial P(y)}{\partial \sigma} \right) + \sum_s \sum_y \left(P(s) \frac{\partial P(y|s)}{\partial \sigma} + P(s) \log P(y|s) \frac{\partial P(y|s)}{\partial \sigma} \right) \quad (125)$$

The sum $\sum_y P(y) = 1$ implies $\sum_y \frac{\partial P(y)}{\partial \sigma} = \frac{\partial}{\partial \sigma} \sum_y P(y) = 0$. And $\sum_s \sum_y \frac{\partial P(y|s)}{\partial \sigma} = 0$ because $\sum_y P(y|s) = 1$. So

$$\frac{\partial I}{\partial \sigma} = - \sum_y \log P(y) \frac{\partial P(y)}{\partial \sigma} + \sum_s \sum_y P(s) \log P(y|s) \frac{\partial P(y|s)}{\partial \sigma} \quad (126)$$

We estimate the partial derivative with a ratio of time differences and replace the denominator with the signum function to avoid numerical instability:

$$\begin{aligned} \frac{\partial P(y)}{\partial \sigma} &\approx \frac{P_k(y) - P_{k-1}(y)}{\sigma_k - \sigma_{k-1}} \\ &\approx \text{sgn}(\sigma_k - \sigma_{k-1}) [P_k(y) - P_{k-1}(y)] \end{aligned} \quad (127)$$

$$\begin{aligned} \frac{\partial P(y|s)}{\partial \sigma} &\approx \frac{P_k(y|s) - P_{k-1}(y|s)}{\sigma_k - \sigma_{k-1}} \\ &\approx \text{sgn}(\sigma_k - \sigma_{k-1}) [P_k(y|s) - P_{k-1}(y|s)] \end{aligned} \quad (128)$$

where $P_k(y)$ is the marginal density function of the output Y at time t and $P_k(y|s)$ is the conditional density function at time t . Then the learning term becomes

$$\begin{aligned} \frac{\partial I}{\partial \sigma} \approx & \operatorname{sgn}(\sigma_k - \sigma_{k-1}) \left(- \sum_y [P_k(y) - P_{k-1}(y)] \log P_k(y) \right. \\ & \left. + \sum_s \sum_y P_k(s) [P_k(y|s) - P_{k-1}(y|s)] \log P_k(y|s) \right) \end{aligned} \quad (129)$$

Our previous work [59, 72] on adaptive SR found through statistical tests that the random learning term $\frac{\partial P}{\partial \sigma}$ had an approximately Cauchy distribution for the spectral signal-to-noise and cross-correlation performance measures P . These frequent and energetic Cauchy impulse spikes destabilized the stochastic learning process. So we “robustified” the learning term with the standard Cauchy error suppressor $\phi(z_k) = 2z_k/(1 + z_k^2)$ [45, 50]. This included the threshold neuron given a periodic input sequence.

But detailed simulations revealed a special pattern in the case of mutual information: The density $P_k(y)$ tends to stay close to the past density $P_{k-1}(y)$ if the values of σ_k and σ_{k-1} are close. This causes the learning paths σ_k to converge quickly near the initial conditions. So we can replace the learning term $\frac{\partial I}{\partial \sigma}$ with its sign $\operatorname{sgn}(\frac{\partial I}{\partial \sigma})$ and the learning law simplifies to

$$\sigma_{k+1} = \sigma_k + \mu_k \operatorname{sgn}\left(\frac{\partial I}{\partial \sigma}\right) \quad (130)$$

The signum is a simple robustifier and formally consistent with a two-sided Laplacian distribution [45].

7.2 Simulation Results

We tested the robust learning law in (130) with the approximation of the learning term in (129). We needed to estimate the marginal and conditional probability densities $P_k(y)$, $P_k(s)$, and $P_k(y|s)$ at each iteration k . So at each k we collected 1000 input-output samples $\{s_t, y_t\}$ and used them to estimate the densities with histograms for the threshold system. We used 500 of the input-output symbols to estimate the probability densities for the continuous neuron model. We chose the neurons’ and signals’ parameters below to demonstrate the algorithm. Other parameters gave similar results.

7.2.1 Noisy Threshold Neuron

The threshold neuron had a fixed threshold $\theta = 0.5$. The bipolar input Bernoulli signal has probability $P_S(-A) = P_S(A) = \frac{1}{2}$ where the amplitude A varied from $A = 0.1$ to $A = 0.4$ (subthreshold inputs). We tried several noise densities that included the Gaussian, uniform, Laplace, and the impulsive Cauchy density. All noise densities had zero mean (zero location for Cauchy). We tried to learn the optimal standard deviation σ_{opt} (or optimal dispersion γ_{opt} for Cauchy noise). We used constant learning rates $\mu_k = 0.01$ for Gaussian and uniform noise, $\mu_k = 0.02$ for Laplace and Cauchy noise, and $\mu_k = 0.02$ for alpha-stable noise with $\alpha = 1.9$ and $\alpha = 1.5$. We started the learning from several initial conditions with different noise seeds.

Figures 14-16 show the adapted SR profiles and the σ_{opt} learning paths for different noise types. The learning paths converged to the optimal standard deviation σ_{opt} (or dispersion γ_{opt}) if the initial value was near σ_{opt} . The learning paths tended to stay nearer the optimal values for larger input amplitudes.

7.2.2 Noisy Continuous Neuron

We used the discrete model in Section 6 for simulations. We used $dt = 0.01$ s and let each input symbol stay for 50 s. So for each input symbol we presented the corresponding “spikes” (plus noise) 5000 times to the neuron. And we collected 5000 discrete-time output “spikes” and averaged them to get the output symbol. This procedure applied to all types of signal functions and for all types of noise.

• Continuous Neurons with Hyperbolic Tangent Signal Function

We tested the continuous neuron model with hyperbolic tangent signal function with several noise densities such as the Gaussian, uniform, Laplace, and alpha-stable (which included the Cauchy density). All noise densities had zero mean (zero location for Cauchy). The bipolar input Bernoulli signal had success probability $P_S(-A) = P_S(A) = \frac{1}{2}$ where the amplitude A varied from $A = 0.1$ to $A = 0.4$ (subthreshold inputs). We used constant learning rates $\mu_k = 0.03$ for Gaussian, uniform, and Laplace noise. We used the smaller learning rates $\mu_k = 0.02$ for alpha-stable noise with $\alpha = 1.9$ and $\alpha = 1.5$ and used the still smaller learning rate $\mu_k = 0.005$ for Cauchy noise. We started the learning from several initial conditions with different noise seeds. Figures 17-19 show the adapted SR profiles and the σ_{opt} learning paths

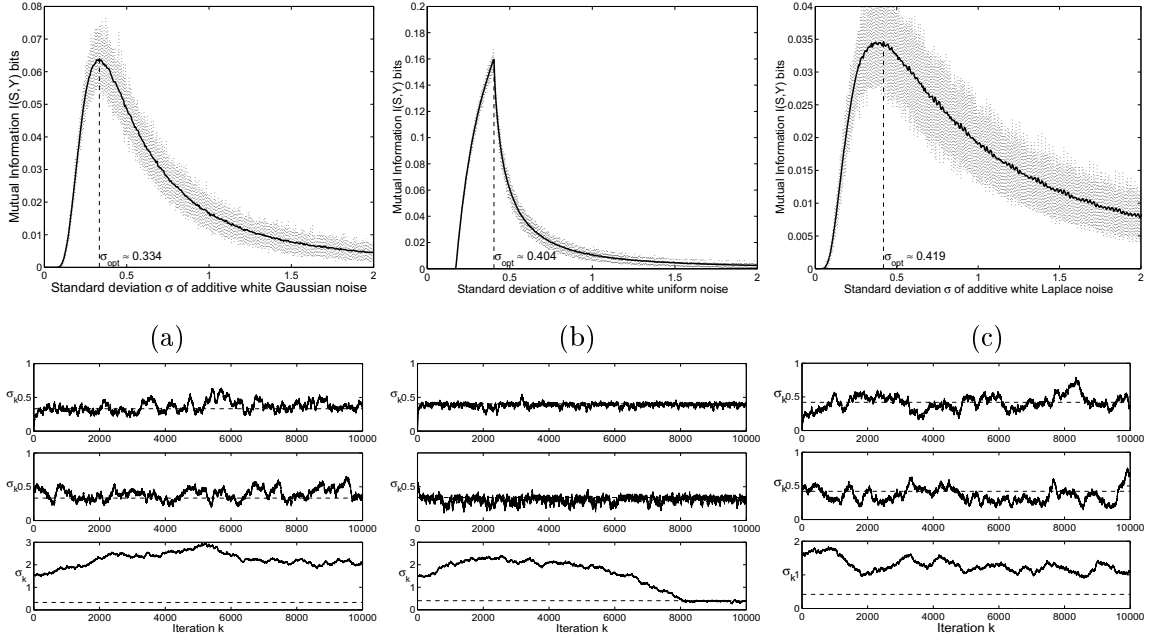


Figure 14: Finite-variance noise cases: Adaptive stochastic resonance for the noisy threshold neuron (6) with bipolar input signal s_t , amplitude $A = 0.2$, and threshold $\theta = 0.5$. The additive noise are (a) Gaussian, (b) uniform, and (c) Laplace. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence if the initial condition σ_0 is close to the optimal noise level σ_{opt} . Distant initial conditions may lead to divergence as the third learning path in (a) shows. The constant learning rates are $\mu_k = 0.01$ for Gaussian and uniform noise and $\mu_k = 0.02$ for Laplace noise.

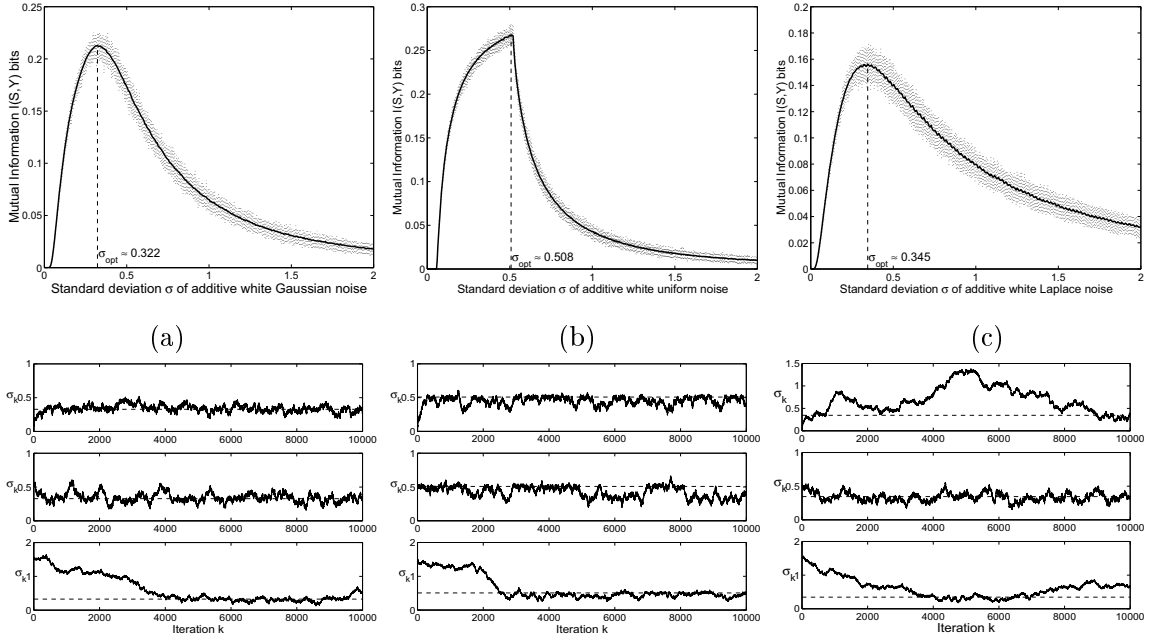


Figure 15: Finite-variance noise cases: Adaptive stochastic resonance for the noisy threshold neuron (6) with bipolar input signal s_t , amplitude $A = 0.4$, and threshold $\theta = 0.5$. The additive noise are (a) Gaussian, (b) uniform, and (c) Laplace. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each noise density. The constant learning rates are $\mu_k = 0.01$ for Gaussian and uniform noise and $\mu_k = 0.02$ for Laplace noise.

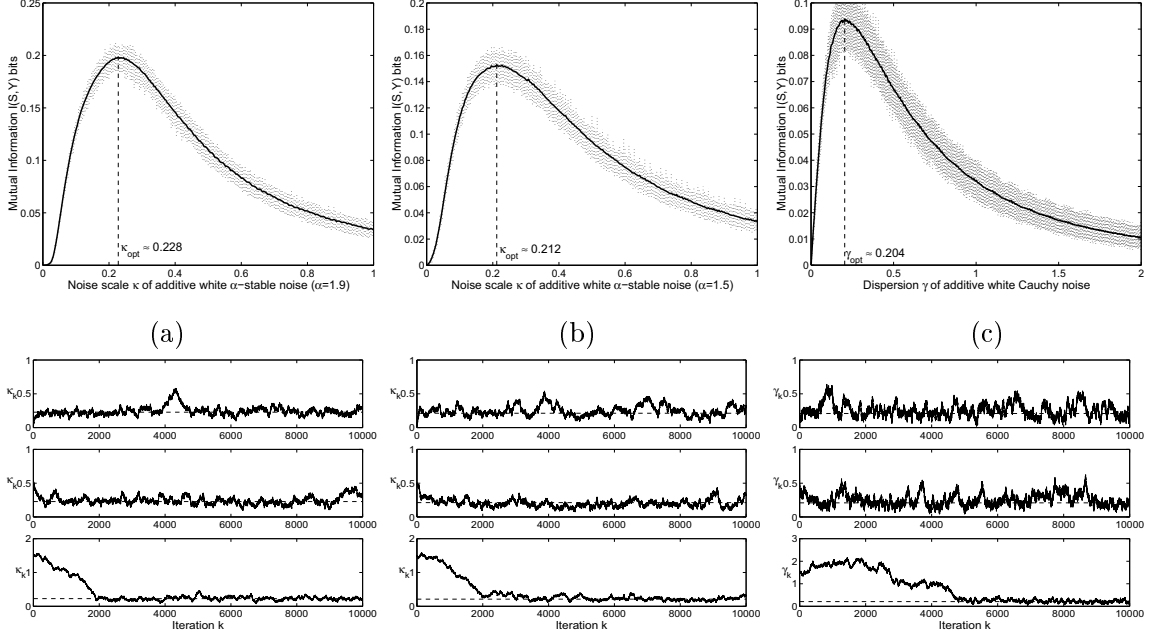


Figure 16: Impulsive noise cases: Adaptive stochastic resonance for the noisy threshold neuron (6) with bipolar input signal s_t , amplitude $A = 0.4$, and threshold $\theta = 0.5$. The additive noise are α -stable distributed with the parameter (a) $\alpha = 1.9$, (b) $\alpha = 1.5$, and (c) $\alpha = 1$ or Cauchy density. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence of the noise scale κ_k to the noise optimum κ_{opt} for each noise density. The corresponding dispersions are $\gamma = \kappa^\alpha$ for each α -stable noise. The constant learning rates are $\mu_k = 0.01$ for $\alpha = 1.9$ and $\alpha = 1.5$ noise and $\mu_k = 0.02$ for Cauchy noise.

for different noise types. The learning paths converged near the optimal standard deviation σ_{opt} (or dispersion γ_{opt}) if the initial value was near σ_{opt} .

• Continuous Neurons with Linear-Threshold, Exponential, and Gaussian (Radial Basis) Signal Functions

We further tested the continuous neuron model with linear-threshold, exponential, and Gaussian (radial basis) signal functions in Gaussian noise to show the generality of the SR effect. We used the same bipolar input Bernoulli signal with success probability $P_S(-A) = P_S(A) = \frac{1}{2}$ where the amplitude is $A = 0.4$ for the linear-threshold and Gaussian signal functions and $A = 0.6$ for the exponential signal function. The input amplitudes were “subthreshold” for the neuron models with these signal functions. We used constant learning rates $\mu_k = 0.02$ for the exponential and Gaussian

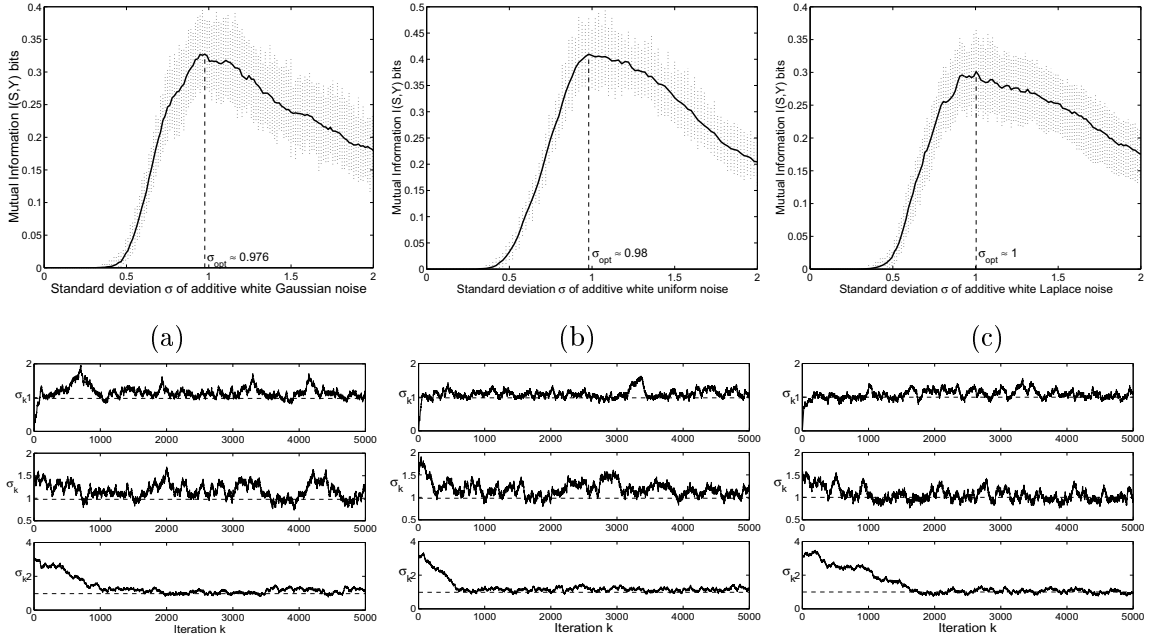


Figure 17: Finite-variance noise cases: Adaptive stochastic resonance for the noisy continuous neuron (7) with hyperbolic signal function (9) and bipolar input signal s_t with amplitude $A = 0.2$. The additive noise are (a) Gaussian, (b) uniform, and (c) Laplace. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each noise density. The constant learning rates are $\mu_k = 0.03$ for all cases.

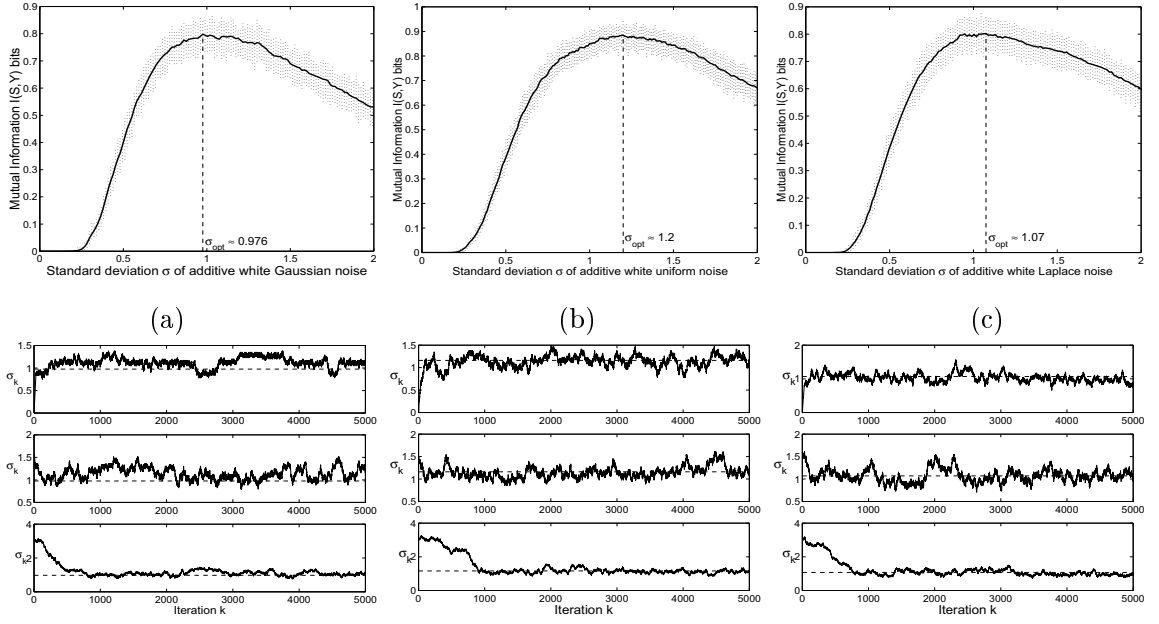


Figure 18: Finite-variance noise cases: Adaptive stochastic resonance for the noisy continuous neuron (7) with hyperbolic signal function (9) and bipolar input signal s_t with amplitude $A = 0.4$. The additive noise are (a) Gaussian, (b) uniform, and (c) Laplace. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each noise density. The constant learning rates are $\mu_k = 0.03$ for all cases.

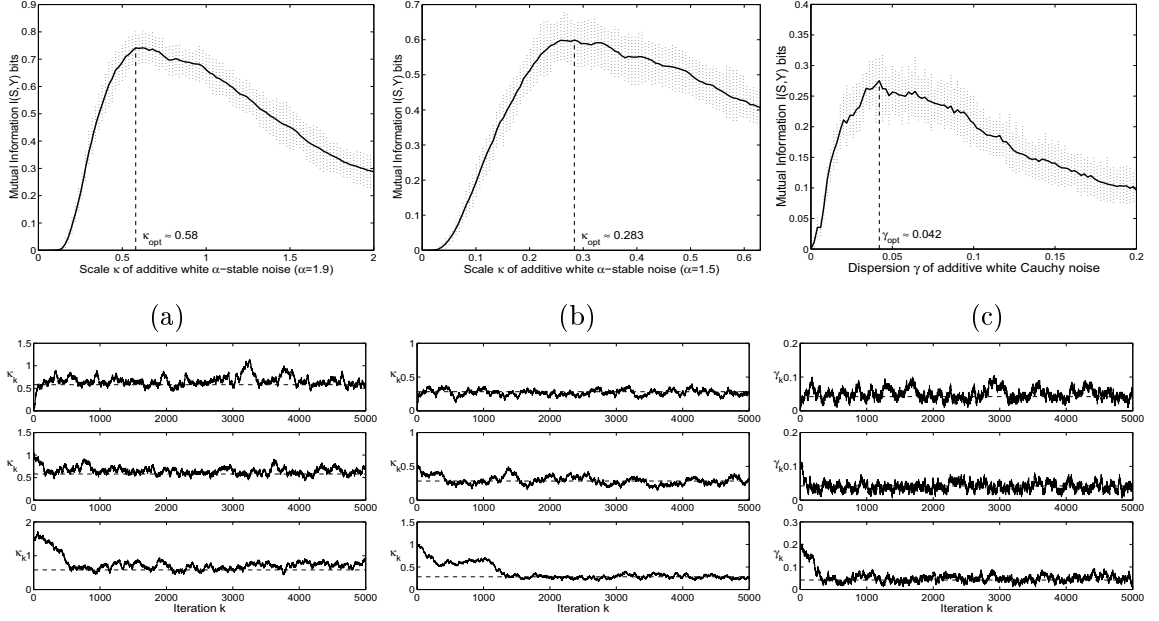


Figure 19: Impulsive noise cases: Adaptive stochastic resonance for the noisy continuous neuron (7) with hyperbolic signal function (9) and bipolar input signal s_t with amplitude $A = 0.4$. The additive noise are α -stable distributed with the parameter (a) $\alpha = 1.9$, (b) $\alpha = 1.5$, and (c) $\alpha = 1$ or Cauchy density. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each noise density. The constant learning rates are $\mu_k = 0.02$ for $\alpha = 1.9$, $\mu_k = 0.01$ for $\alpha = 1.5$, and $\mu_k = 0.005$ for $\alpha = 1$.

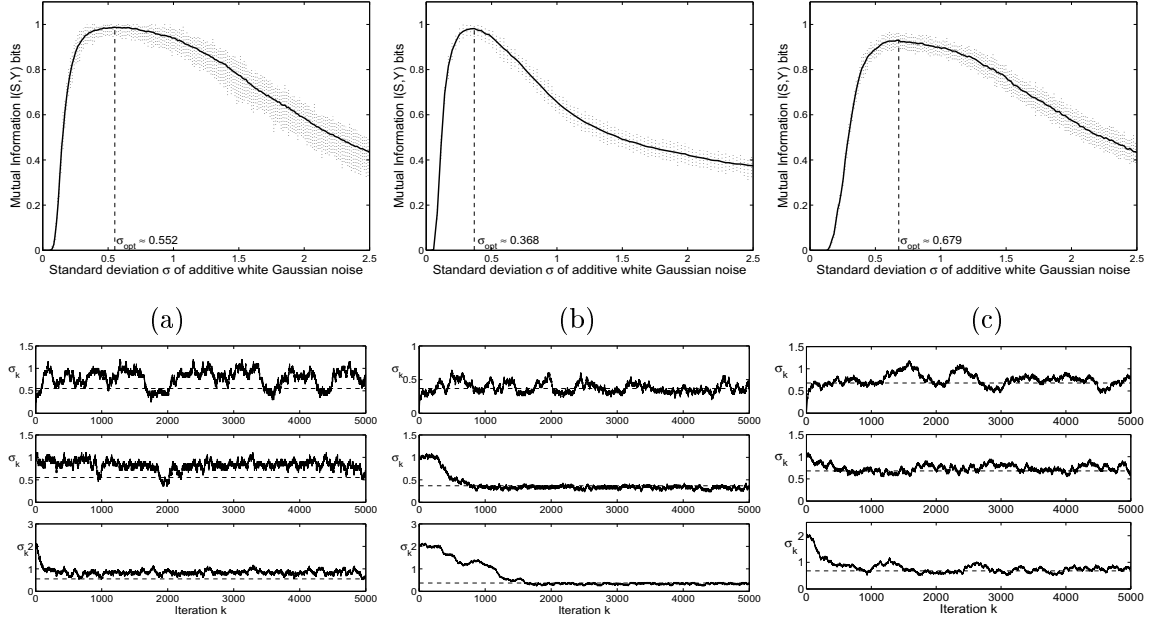


Figure 20: Adaptive stochastic resonance for continuous neurons with linear-threshold, exponential, and Gaussian (radial basis) signal functions. The bipolar input signal s_t has amplitude $A = 0.4$ for the linear-threshold and Gaussian signal functions and $A = 0.6$ for the exponential signal function. The additive noise n_t is Gaussian. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each case of signal functions: (a) linear-threshold, (b) exponential, and (c) Gaussian. The constant learning rates are $\mu_k = 0.05$ for the linear-threshold signal function and $\mu_k = 0.02$ for the exponential and Gaussian signal functions.

signal functions and $\mu_k = 0.05$ for the linear-threshold signal functions. We started the learning from several initial conditions with different noise seeds. Figure 20 shows the adapted SR profiles and the σ_{opt} learning paths for the three other signal functions. The learning paths converged near the optimal standard deviation σ_{opt} if the initial value was near σ_{opt} .

8 Stochastic Resonance Effect in Object Segmentation with Color Thresholding

Object segmentation is one of the most important tasks in image analysis and computer vision. Color thresholding provides a fast and simple scheme for such task. But it is sensitive to lighting conditions and other noise effect in images obtained from the real world applications. This section shows that addition of a small amount of noise can improve the accuracy of such color object segmentation and thus this SR effect offers an alternative to detecting color objects in noisy input images.

8.1 Color Image Thresholding

The thresholding method described here can be used with general multidimensional, color spaces that have discrete component color levels [53, 39] such as RGB, HIS. This paper uses the RGB color space. An color object is segmented with a set of six threshold values, two for each color dimension in the RGB color space. A pixel which has color component value in the interval of the two thresholds, lower and upper thresholds, is represented as one or otherwise is zero:

$$g(y) = \begin{cases} 1 & \theta_{\min} \leq y \leq \theta_{\max} \\ 0 & \text{otherwise} \end{cases} \quad (131)$$

where y is the image pixel value, $\theta_{\min}$ is the lower threshold and $\theta_{\max}$ is the upper threshold. Red, Green, and Blue color components of the input image are compared and classified to three binary images of each color space by the set of RGB thresholds. The AND operation of the pixels of these three binary images gives a target object region.

8.2 Performance Measures

8.2.1 Mutual Information Measure

We use Shannon mutual information $I(S, Y)$ to measure the SR effect in Section 2 for image segmentation task. The input S is the correctly segmented (binary) image and the output Y is the output segmented (binary) images using the color thresholding algorithm and its modified (SR) version.

8.2.2 Error Pixels Count

This measure directly describes mistaken pixels between two binary images. The bit-wise XOR operation shows the pixels in which their binary values do not match the pixels in another binary image at the same locations. So we can obtain the amount of the error pixels by counting the results of 1 of the bit-wise XOR. The error pixels count C_e between a correctly segmented binary image S and an output binary image Y with $m \times n$ dimensions has the form

$$C_e = \sum_{i=1}^n \sum_{j=1}^n S_{ij} \oplus Y_{ij} \quad (132)$$

where

$$S_{ij} \oplus Y_{ij} = \begin{cases} 0 & \text{if } S_{ij} = Y_{ij} \\ 1 & \text{if } S_{ij} \neq Y_{ij} \end{cases} \quad (133)$$

8.2.3 Position Error

In many applications of image processing [53, 39, 88] an object positioning has been a useful function. But noise sensitivity in the image segmentation process causes error in estimating the object region and so error of the estimated position. Thus the position error is also a necessary measure for showing the SR effect of our approach. We apply this measure to the binary images of classified target objects using the color image thresholding. We determine the position of an object by calculate-ing the centroid of the classified region (region classified as an object):

$$\begin{bmatrix} c_x \\ c_y \end{bmatrix} = \frac{1}{N} \sum_{i=1}^n \begin{bmatrix} x_i \\ y_i \end{bmatrix} \quad (134)$$

where c_x and c_y are row and column coordinates of the centroid. Then the position error P_e is

$$P_e = d(c_{in}, c_{out}) = \sqrt{(c_{x,in} - c_{x,out})^2 + (c_{y,in} - c_{y,out})^2} \quad (135)$$

where $[x_i \ y_i]^T$ is the coordinate vector of a pixel which would be a target object, N is the number of pixels of an object region, c_{in} and c_{out} are centroids of the target object of original image and noisy image: $c_{in} = [c_{x,in} \ c_{y,in}]^T$ and $c_{out} = [c_{x,out} \ c_{y,out}]^T$.

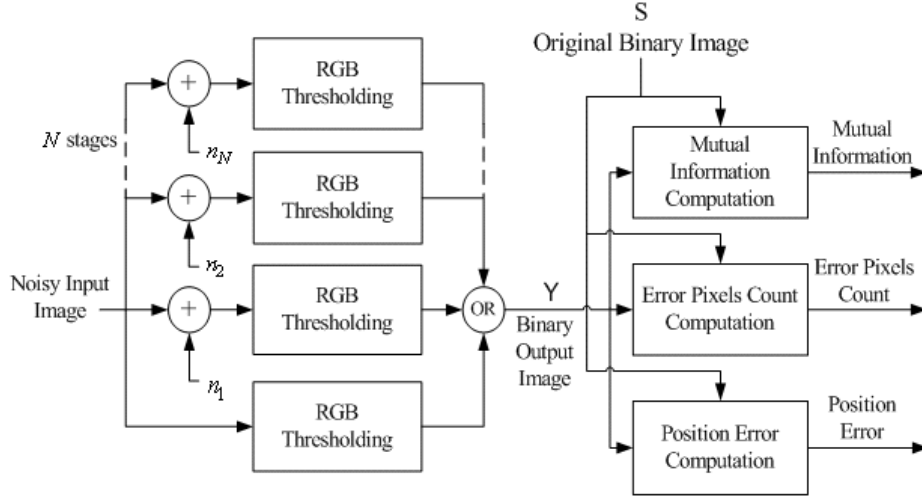


Figure 21: The SR segmentation system consists of the multiple-stage RGB color thresholdings where N is the number of stages and n_i is independent Gaussian noise for stage i . S and Y are correctly segmented and output segmented (binary) images.

8.3 Proposed SR Segmentation System and Experimental Results

8.3.1 SR Segmentation System

We propose a new image segmentation technique using N stages of noisy RGB color thresholding system. Each stage simply adds independent white Gaussian noise to a noisy input image before performing the usual color thresholding. Binary output images of all stages are combined with an OR operation to obtain a binary output image of the SR segmentation system. We measure the performance of the SR segmentation system using mutual information $I(S, Y)$, error pixels count C_e , and position error P_e to determine how the output segmented (binary) image Y matches the correctly segmented (binary) image S as shown in Figure 21.

8.3.2 Experimental Results: Synthetic Images

We first tested the SR segmentation system with synthetic images. The original image consisted of an orange circle (as an object) on the green background. Then we added Gaussian noise to the original image and performed the blurring and shadowing operations on it to produce noisy test images shown in Figure 22. By using this synthetic image we could precisely determine the pixels that actually belong to the

object. The SR segmentation system in Figure 21 1 has $N = 1$ stage of RGB color thresholding system. The additive noise is Gaussian. The segmentation system will separate the target object (the orange circle) from the noisy image. In this experiment the RGB color thresholding algorithm uses the following thresholds (based on the 0-255 levels of intensity):

$$\begin{aligned} \text{Red : } & \theta_{R\min} = 120 \quad \theta_{R\max} = 255 \\ \text{Green : } & \theta_{G\min} = 50 \quad \theta_{G\max} = 120 \\ \text{Blue : } & \theta_{B\min} = 0 \quad \theta_{B\max} = 40 \end{aligned} \tag{136}$$

The images that we tested are the noisy images (with 0%, 10%, 20% and 50% brightness) shown in Figure 22. We did not show the test image with 20% brightness here. Figure 23 shows the performances when the noise standard deviation increases. The system has nonzero optimal noise level and thus shows the SR effect for the synthetic test images. The effect is more pronounced for very noisy images.

Object positioning error can be a good indicator as well. Figure 26 (a)-(d) graphically show the positions obtained from SR and non-SR segmentation systems comparing to the correct position in all four test images in Figurefg:SyntheticImages. The Figure shows that the positions obtained from the SR-system (with optimal amount of noise) are more accurate than the ones obtained from the original (non-SR) segmentation system. A specific case of Figure fg:SyntheticImages (d) shows that conventional color thresholding cannot detect any pixels as an object. So we do not have an estimate of the position in this case (and so there is no white circle shown). But the SR segmentation algorithm can find some pixels that belong to the object and gives an estimate of the object position.

8.3.3 Experimental Results: Real Images

Here we tested the SR segmentation system on images taken from the real world using a digital camera. The images show an orange golf ball on the green carpet as a background. We measured the position of the golf ball against the known mark on the carpet. Three images shown in Figure 24 are taken in three different illuminations (different light settings).

The SR segmentation used in this system is the same as in Section 8.3.2 (the system in Figure 21) but now with $N = 50$ stages. The noise in each stage is independent Gaussian noise. The thresholds used for the RGB color thresholding

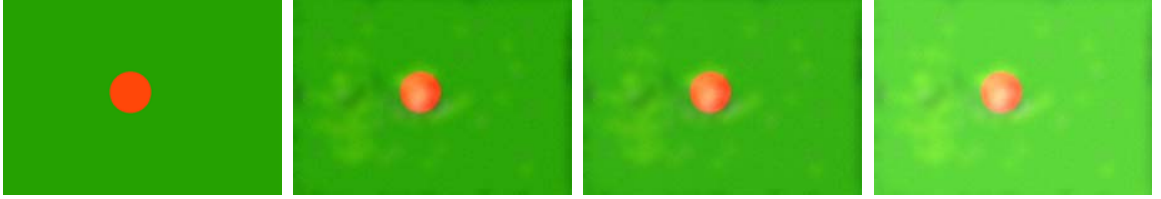


Figure 22: Synthetic images of an orange object on a green background. The first (far-left) image is the original image and the second image is a noisy, blurred, and shadowed version of the original one. The third and fourth images are the brightness-increased versions of the noisy image (the second one). We label the second image as “Brightness 0%,” the third one as “Brightness 10%,” and the last one (far-right) as “Brightness 50%”. We do not show the “Brightness 20%” image here.

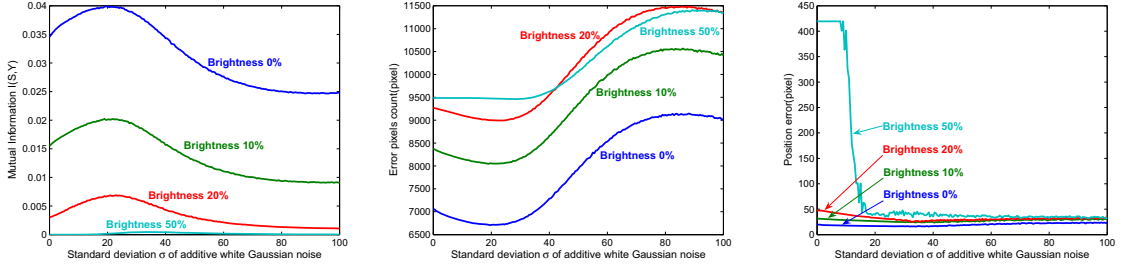


Figure 23: SR effect in image segmentation for synthetic images using SR segmentation system in Figurefg:SRsystem with Gaussian noise n_i . The performance measures are mutual information, error pixels count, and position error. The graphs show the results of the SR segmentation system with $N = 1$ for the four synthetic images.



Figure 24: Real images of an orange golf ball on the green carpet. The three images were taken from three different illuminations (different light settings). From left to right: “Image1,” “Image2,” and “Image3.”

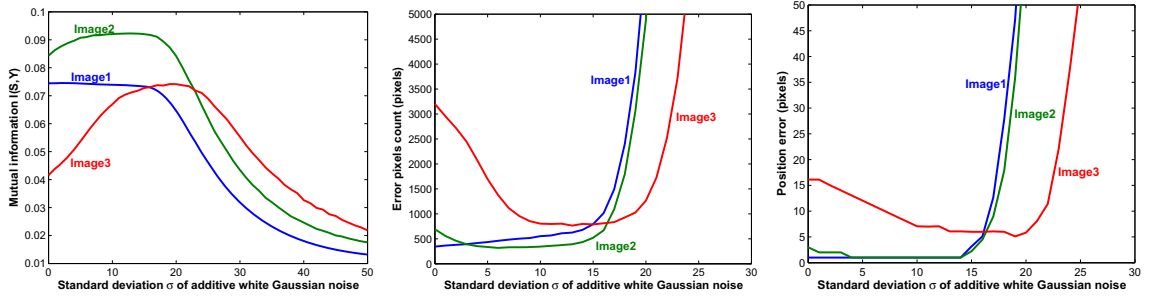


Figure 25: SR effect in image segmentation for real images using the SR segmentation system in Figure 21 with Gaussian noise n_i . The performance measures are mutual information, error pixels count, and position error. The graphs show the results of SR segmentation system with $N = 50$ for the real images (Image1 - Image3).

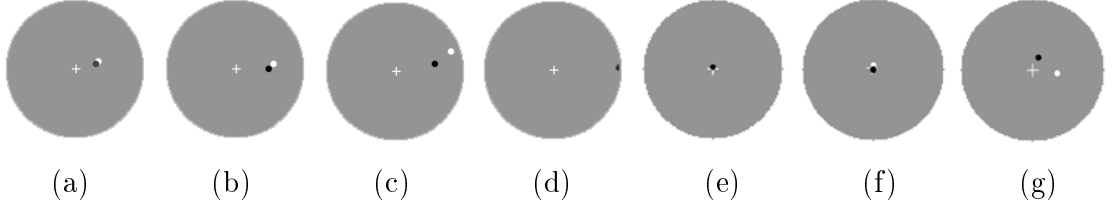


Figure 26: Actual and estimated positions of the objects. The white ‘+’ denotes the actual positions of the objects (at the center). The white circles are positions obtained from non-SR segmentation systems (original RGB color thresholding algorithm). The dark circles represent positions from the optimal SR effect in segmentation. The SR system gives estimated position with less error. The four panels (a)-(d) are results of the four synthetic images where the Gaussian noise has SR-optimal standard deviations around 22, 24, 24, and 30 for (a)-(d). The three panels (e)-(g) are results of the three real images with SR-optimal standard deviations of Gaussian noise around 13, 13, and 19. These positions refer to the SR optimal results of the synthetic and real images in Figures 23 and 25.

scheme in each stage are (based on the 0-255 levels of intensity):

$$\begin{aligned}
 Red : \quad \theta_{Rmin} &= 150 \quad \theta_{Rmax} = 255 \\
 Green : \quad \theta_{Gmin} &= 110 \quad \theta_{Gmax} = 205 \\
 Blue : \quad \theta_{Bmin} &= 0 \quad \theta_{Bmax} = 50
 \end{aligned} \tag{137}$$

These set of thresholds are manually set for optimal segmentation of Image1. So they are different from the optimal thresholds for the synthetic image “Brightness 0 and Figure 26 (e)-(g) show how white Gaussian pixel noise can improve our image information and segmentation in terms of mutual information, error pixels count, and position error. The results show that noise can improve the segmentation performances when the preset RGB thresholds do not match the lighting conditions. The *perfect* threshold case of “Image1” also shows that a small amount of noise does not ruin the accuracy of segmentation while it can increase the accuracy of the noisy images (“Image2” and “Image3”).

The results show that the stochastic resonance or SR effect occurs when we use this algorithm to segment a color object from a plane background in an image that is not perfectly captured or that the preset thresholds do not perfectly match the color distribution. They suggest that addition of white Gaussian noise can enhance the mutual information between the correct image and the image acquired from the actual environment. The number of error pixels and the object’s position error also decrease

when we add the right amount of noise to the noisy input images. These results confirm that noise can robustify segmentation using RGB thresholding algorithm. For an image with good lighting condition (so the RGB thresholds perfectly match its histograms) and with no other distortion, a little amount of noise only worsens the performance of the segmentation a little white noise can significantly improve the performance of the noisy ones. So for real-world applications in which captured images contain a lot of interferences, engineers might consider using optimal noise to design a more robust image segmentation algorithm.

9 Conclusions

The counter-intuitive SR phenomenon is more general than most people have realized. The results show that various models of signal processing systems and applications (such as threshold systems, continuous neurons, and color thresholding scheme in this report) exhibit the SR effect. The results confirm that this SR effect does not depend only on a few popular noise density functions such as uniform noise that audio engineers use in dithering the analog signal before they digitize it or Gaussian noise that system engineers often use to model interferences in communications systems. The effect also persists for impulsive noise. This implies that noise should be equally treated as a working source of energy to help boost the system's performance as well as a source of bad effect that destroys the system's output.

Future research in this direction should examine in details the effect of noise in connected neurons that form a network or other more complex systems that include the threshold systems as their building blocks. This includes the regular neural networks that we use in various applications in communications, control, signal processing, or pattern recognition. The evidence of the benefits of noise implies that we always need to consider two possibilities: try to eliminate the noise or try to make the most out of it.

References

- [1] J. H. Ahrens and U. Dieter, “Computer Methods for Sampling from Gamma, Beta, Poisson and Binomial Distributions,” *Computing*, vol. 12, pp. 223–246, 1974.
- [2] J. H. Ahrens and U. Dieter, “Generating Gamma Variates by a Modified Rejection Technique,” *Communications of the ACM*, vol. 25, no. 1, pp. 47–54, January 1982.
- [3] V. Akgiray and C. G. Lamoureux, “Estimation of Stable-Law Parameters: A Comparative Study,” *Journal of Business and Economic Statistics*, vol. 7, pp. 85–93, January 1989.
- [4] S. Amari, “Neural Theory of Association and Concept Formation,” *Biological Cybernetics*, vol. 26, pp. 175–185, 1977.
- [5] V. S. Anishchenko, M. A. Safonova, and L. O. Chua, “Stochastic Resonance in Chua’s Circuit,” *International Journal of Bifurcation and Chaos*, vol. 2, no. 2, pp. 397–401, 1992.
- [6] N. Ansari, H. Liu, Y. Q. Shi, and H. Zhao, “On Modeling MPEG Video Traffics,” *IEEE Transactions on Broadcasting*, vol. 48, no. 4, pp. 337–347, December 2002.
- [7] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, “Stochastic Resonance in Climatic Change,” *Tellus*, vol. 34, pp. 10–16, 1982.
- [8] R. Benzi, A. Sutera, and A. Vulpiani, “The Mechanism of Stochastic Resonance,” *Journal of Physics A: Mathematical and General*, vol. 14, pp. L453–L457, 1981.
- [9] H. Bergstrom, “On Some Expansions of Stable Distribution Functions,” *Arkiv för Matematik*, vol. 2, pp. 375–378, 1952.
- [10] L. Breiman, *Probability*, Addison-Wesley, 1968.
- [11] P. Bryant, K. Wiesenfeld, and B. McNamara, “The Nonlinear Effects of Noise on Parametric Amplification: An Analysis of Noise Rise in Josephson Junctions and Other Systems,” *Journal of Applied Physics*, vol. 62, no. 7, pp. 2898–2913, October 1987.
- [12] A. R. Bulsara, T. C. Elston, C. R. Doering, S. B. Lowen, and K. Lindenberg, “Cooperative Behavior in Periodically Driven Noisy Integrate-Fire Models of Neuronal Dynamics,” *Physical Review E*, vol. 53, no. 4, pp. 3958–3969, April 1996.
- [13] A. R. Bulsara and L. Gammaitoni, “Tuning in to Noise,” *Physics Today*, pp. 39–45, March 1996.
- [14] A. R. Bulsara, E. W. Jacobs, T. Zhou, F. Moss, and L. Kiss, “Stochastic Resonance in a Single Neuron Model: Theory and Analog Simulation,” *Journal of Theoretical Biology*, vol. 152, pp. 531–555, 1991.
- [15] A. R. Bulsara, A. J. Maren, and G. Schmera, “Single Effective Neuron: Dendritic Coupling Effects and Stochastic Resonance,” *Biological Cybernetics*, vol. 70, pp. 145–156, 1993.
- [16] A. R. Bulsara and A. Zador, “Threshold Detection of Wideband Signals: A Noise-Induced Maximum in the Mutual Information,” *Physical Review E*, vol. 54, no. 3, pp. R2185–R2188, September 1996.

- [17] J. M. Chambers, C. L. Mallows, and B. W. Stuck, "A Method for Simulating Stable Random Variables," *Journal of the American Statistical Association*, vol. 71, no. 354, pp. 340–344, 1976.
- [18] R. C. H. Cheng, "Generating Beta Variates with Nonintegral Shape Parameters," *Communications of the ACM*, vol. 21, no. 4, pp. 317–322, 1978.
- [19] K. L. Chung and R. J. Williams, *Introduction to Stochastic Integration*, Birkhäuser, second edition, 1990.
- [20] M. A. Cohen and S. Grossberg, "Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, pp. 815–826, September 1983.
- [21] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, "Aperiodic Stochastic Resonance," *Physical Review E*, vol. 54, no. 5, pp. 5575–5584, November 1996.
- [22] J. J. Collins, C. C. Chow, and T. T. Imhoff, "Stochastic Resonance without Tuning," *Nature*, vol. 376, pp. 236–238, July 1995.
- [23] J. J. Collins, T. T. Imhoff, and P. Grigg, "Noise-Enhanced Information Transmission in Rat SA1 Cutaneous Mechanoreceptors via Aperiodic Stochastic Resonance," *Journal of Neurophysiology*, vol. 76, no. 1, pp. 642–645, July 1996.
- [24] J. J. Collins, T. T. Imhoff, and P. Grigg, "Noise-Enhanced Tactile Sensation," *Nature*, vol. 383, pp. 770, October 1996.
- [25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.
- [26] G. Dahlquist and Å. Björck, *Numerical Methods*, Prentice Hall, 1974.
- [27] G. Deco and B. Schürmann, "Stochastic Resonance in the Mutual Information between Input and Output Spike Trains of Noisy Central Neurons," *Physica D*, vol. 117, pp. 276–282, 1998.
- [28] L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, 1986.
- [29] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, "Noise Enhancement of Information Transfer in Crayfish Mechanoreceptors by Stochastic Resonance," *Nature*, vol. 365, pp. 337–340, September 1993.
- [30] R. Durrett, *Stochastic Calculus: A Practical Introduction*, CRC Press, 1996.
- [31] M. I. Dykman, T. Horita, and J. Ross, "Statistical Distribution and Stochastic Resonance in a Periodically Driven Chemical System," *Journal of Chemical Physics*, vol. 103, no. 3, pp. 966–972, July 1995.
- [32] J.-P. Eckmann and L. E. Thomas, "Remarks on Stochastic Resonances," *Journal of Physics A: Mathematical and General*, vol. 15, pp. L261–L266, 1982.
- [33] S. Fauve and F. Heslot, "Stochastic Resonance in a Bistable System," *Physics Letters A*, vol. 97, no. 1,2, pp. 5–7, August 1983.

- [34] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley & Sons, 1966.
- [35] L. Gammaitoni, "Stochastic Resonance and the Dithering Effect in Threshold Physical Systems," *Physical Review E*, vol. 52, no. 5, pp. 4691–4698, November 1995.
- [36] T. C. Gard, *Introduction to Stochastic Differential Equations*, Marcel Dekker, Inc., 1988.
- [37] J. Glanz, "Mastering the Nonlinear Brain," *Science*, vol. 277, pp. 1758–1760, September 1997.
- [38] X. Godivier and F. Chapeau-Blondeau, "Stochastic Resonance in the Information Capacity of a Nonlinear Dynamic System," *International Journal of Bifurcation and Chaos*, vol. 8, no. 3, pp. 581–589, 1998.
- [39] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, 2002.
- [40] M. Grifoni, M. Sassetti, P. Hänggi, and U. Weiss, "Cooperative Effects in the Nonlinearly Driven Spin-Boson System," *Physical Review E*, vol. 52, no. 4, pp. 3596–3607, October 1995.
- [41] M. Grigoriu, *Applied Non-Gaussian Processes*, Prentice Hall, 1995.
- [42] S. Grossberg, *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control*, Kluwer Academic Publishers, 1982.
- [43] S. Haykin, *Neural Networks: A Comprehensive Foundation*, McMillan, 1994.
- [44] A. Hibbs, E. W. Jacobs, J. Bekkedahl, A. R. Bulsara, and F. Moss, "Signal Enhancement in a r.f. SQUID Using Stochastic Resonance," *Il Nuovo Cimento*, vol. 17 D, no. 7-8, pp. 811–817, Luglio-Agosto 1995.
- [45] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*, Prentice Hall, fifth edition, 1995.
- [46] N. Hohn and A. N. Burkitt, "Enhanced Stochastic Resonance in Threshold Detectors," in *Proceedings of the 2001 IEEE International Joint Conference on Neural Networks (IJCNN '01)*, 2001, pp. 644–647.
- [47] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proceedings of the National Academy of Science*, vol. 79, pp. 2554–2558, 1982.
- [48] J. J. Hopfield, "Neural Networks with Graded Resonse Have Collective Computational Properties Like Those of Two-State Neurons," *Proceedings of the National Academy of Science*, vol. 81, pp. 3088–3092, 1984.
- [49] J. Hu, A. Razden, G. M. Nielson, G. E. Farin, D. P. Baluch, and D. G. Capco, "Volumetric Segmentation Using Weibull E-SD Fields," *IEEE Transactions on Visualization and Computer Graphics*, vol. 9, no. 3, pp. 320–328, July-September 2003.
- [50] P. J. Huber, *Robust Statistics*, John Wiley & Sons, 1981.

- [51] M. E. Inchiosa and A. R. Bulsara, "Nonlinear Dynamic Elements with Noisy Sinusoidal Forcing: Enhancing Response via Nonlinear Coupling," *Physical Review E*, vol. 52, no. 1, pp. 327–339, July 1995.
- [52] M. E. Inchiosa, J. W. C. Robinson, and A. R. Bulsara, "Information-Theoretic Stochastic Resonance in Noise-Floor Limited Systems: The Case for Adding Noise," *Physical Review Letters*, vol. 85, pp. 3369–3372, October 2000.
- [53] T. Balch J. Bruce and M. Veloso, "Fast and Inexpensive Color Image Segmentation for Interactive Robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000)*, 31 October - 5 November 2000, pp. 2061–2066.
- [54] P. Jung, "Stochastic Resonance and Optimal Design of Threshold Detectors," *Physics Letters A*, vol. 207, pp. 93–104, October 1995.
- [55] A. Kavak, M. Torlak, W. J. Vogel, and G. Xu, "Vector Channels for Smart Antennas—Measurements, Statistical Modeling, and Directional Properties in Outdoor Environments," *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, no. 6, pp. 930–937, June 2000.
- [56] C. V. Kimball and D. J. Scheibner, "Error Bars for Sonic Slowness Measurements," *Geophysics*, vol. 63, no. 2, pp. 345–353, March-April 1998.
- [57] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice Hall, 1992.
- [58] B. Kosko, *Fuzzy Engineering*, Prentice Hall, 1996.
- [59] B. Kosko and S. Mitaim, "Robust Stochastic Resonance: Signal Detection and Adaptation in Impulsive Noise," *Physical Review E*, vol. 64, no. 051110, October 2001.
- [60] B. Kosko and S. Mitaim, "Stochastic Resonance in Noisy Threshold Neurons," *Neural Networks*, vol. 16, no. 5-6, pp. 755–761, June-July 2003.
- [61] R. L. Kulp, "Detection Probabilities for Beta-Distributed Scattering Cross Sections," *Electromagnetics*, vol. 4, no. 2-3, pp. 165–183, 1984.
- [62] J. E. Levin and J. P. Miller, "Broadband Neural Encoding in the Cricket Cercal Sensory System Enhanced by Stochastic Resonance," *Nature*, vol. 380, pp. 165–168, March 1996.
- [63] R. Linsker, "Self-Organization in a Perceptual Network," *Computer*, vol. 21, no. 3, pp. 105–117, March 1988.
- [64] R. Linsker, "A Local Learning Rule That Enables Information Maximization for Arbitrary Input Distributions," *Neural Computation*, vol. 9, no. 8, pp. 1661–1665, November 1997.
- [65] A. Longtin, "Autonomous Stochastic Resonance in Bursting Neurons," *Physical Review E*, vol. 55, no. 1, pp. 868–876, January 1997.
- [66] J. Maddox, "Towards the Brain-Computer's Code?," *Nature*, vol. 352, pp. 469, August 1991.

- [67] J. Maddox, “Bringing More Order out of Noisiness,” *Nature*, vol. 369, pp. 271, May 1994.
- [68] R. Mannella and V. Palleschi, “Fast and Precise Algorithm for Computer Simulation of Stochastic Differential Equations,” *Physical Review A*, vol. 40, no. 6, pp. 3381–3386, September 1989.
- [69] R. J. Marks II, B. Thompson, M. A. El-Sharkawi, W. L. J. Fox, and R. T. Miyamoto, “Stochastic Resonance of a Threshold Detector: Image Visualization and Explanation,” in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS 2002)*, 2002, vol. 4, pp. IV–521 –IV–523.
- [70] B. McNamara, K. Wiesenfeld, and R. Roy, “Observation of Stochastic Resonance in a Ring Laser,” *Physical Review Letters*, vol. 60, no. 25, pp. 2626–2629, June 1988.
- [71] W. Mendenhall and T. Sincich, *Statistics for Engineering and the Sciences*, Prentice Hall, fourth edition, 1995.
- [72] S. Mitaim and B. Kosko, “Adaptive Stochastic Resonance,” *Proceedings of the IEEE: Special Issue on Intelligent Signal Processing*, vol. 86, no. 11, pp. 2152–2183, November 1998.
- [73] S. Mitaim and B. Kosko, “Adaptive Stochastic Resonance in Noisy Neurons Based on Mutual Information,” *IEEE Transactions on Neural Networks*, vol. 15, no. 6, pp. 1526–1540, November 2004.
- [74] R. P. Morse and E. F. Evans, “Enhancement of Vowel Coding for Cochlear Implants by Addition of Noise,” *Nature Medicine*, vol. 2, no. 8, pp. 928–932, August 1996.
- [75] F. Moss and P. V. E. McClintock, Eds., *Noise in Nonlinear Dynamical Systems*, vol. I-III, Cambridge University Press, 1989.
- [76] C. Nicolis, “Stochastic Aspects of Climatic Transitions–Response to a Periodic Forcing,” *Tellus*, vol. 34, pp. 1–9, 1982.
- [77] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, John Wiley & Sons, 1995.
- [78] X. Pei, K. Bachmann, and F. Moss, “The Detection Threshold, Noise and Stochastic Resonance in the Fitzhugh-Nagumo Neuron Model,” *Physics Letters A*, vol. 206, pp. 61–65, October 1995.
- [79] X. Pei, L. Wilkens, and F. Moss, “Light Enhances Hydrodynamic Signaling in the Multimodal Caudal Photoreceptor Interneurons of the Crayfish,” *Journal of Neurophysiology*, vol. 76, no. 5, pp. 3002–3011, November 1996.
- [80] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, second edition, 1993.
- [81] E. A. Tanis R. V. Hogg, *Probability and Statistical Inference*, Prentice Hall, sixth edition, 2001.
- [82] D. F. Russell, L. A. Willkens, and F. Moss, “Use of Behavioural Stochastic Resonance by Paddle Fish for Feeding,” *Nature*, vol. 402, pp. 291–294, November 1999.

- [83] M. A. Sletten, "Multipath Scattering in Ultrawide-Band Radar Sea Spikes," *IEEE Transactions on Antennas and Propagation*, vol. 46, no. 1, pp. 45–56, January 1998.
- [84] N. G. Stocks, "Information Transmission in Parallel Threshold Arrays," *Physical Review E*, vol. 63, no. 041114, 2001.
- [85] P. Tsakalides and C. L. Nikias, "The Robust Covariation-Based MUSIC (ROC-MUSIC) Algorithm for Bearing Estimation in Impulsive Noise Environments," *IEEE Transactions on Signal Processing*, vol. 44, no. 7, pp. 1623–1633, July 1996.
- [86] M. Usher and M. Feingold, "Stochastic Resonance in the Speed of Memory Retrieval," *Biological Cybernetics*, vol. 83, pp. L11–L16, 2000.
- [87] R. A. Wannamaker, S. P. Lipshitz, and J. Vanderkooy, "Stochastic Resonance as Dithering," *Physical Review E*, vol. 61, no. 1, pp. 233–236, January 2000.
- [88] Z. Wasik and A. Saffiotti, "Robust Color Segmentation for the RoboCup Domain," in *Proceedings of the International Conference on Pattern Recognition (ICPR 2002)*, August 2002, pp. 2780–2785.
- [89] W. Weibull, "A Statistical Distribution Function of Wide Applicability," *Journal of Applied Mechanics*, vol. 18, pp. 293–297, 1951.
- [90] K. Wiesenfeld and F. Moss, "Stochastic Resonance and the Benefits of Noise: From Ice Ages to Crayfish and SQUIDS," *Nature*, vol. 373, pp. 33–36, January 1995.

10 Research Output and Deliverables

Part of this research effort produced the following journal and conference papers. The Appendix includes the reprints of the journal and conference papers.

Journal Articles

- [1] B. Kosko and S. Mitaim, “Stochastic Resonance in Noisy Threshold Neurons,” *Neural Networks*, vol. 16, pp. 755–761, 2003.
- [2] S. Mitaim and B. Kosko, “Adaptive Stochastic Resonance in Noisy Neurons Based on Mutual Information,” *IEEE Transactions on Neural Networks*, vol. 15, no. 6, pp. 1526–1540, November 2004.
- [3] B. Kosko and S. Mitaim, “Robust Stochastic Resonance for Simple Threshold Neurons,” *Physical Review E*, vol. 70, no. 031911, 27 September 2004.

Conference Proceedings

- [1] B. Kosko and S. Mitaim, “Almost All Noise Types Can Improve the Mutual Information of Threshold Neurons That Detect Subthreshold Signals,” *Proceedings of the International Joint Conference on Neural Networks (IJCNN’03)*, USA, pp. 2740–2745, July 2003.
- [2] S. Mitaim and B. Kosko, “Stochastic Resonance in Threshold Neurons: Noise Enhances Mutual Information,” *Proceedings of the 26th Electrical Engineering Conference (EECON-26)*, Petchaburi, Thailand, pp. 1020–1027, November 2003.
- [3] S. Turmchokkasam and S. Mitaim, “Weighted Fuzzy C-Means Algorithm for Room Equalization at Multiple Locations,” *Proceedings of the 28th Electrical Engineering Conference (EECON-28)*, Phuket, Thailand, pp. 913–916 October 2005.
- [4] S. Janpaiboon and S. Mitaim, “A Study of Stochastic Resonance Effect in Object Segmentation with Color Thresholding,” *Proceedings of the 28th Electrical Engineering Conference (EECON-28)*, Phuket, Thailand, pp. 941–944 October 2005.

11 Appendix



PERGAMON

AVAILABLE AT
www.ComputerScienceWeb.com

POWERED BY SCIENCE @ DIRECT®

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

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

^bDepartment 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.

© 2003 Elsevier Science Ltd. All rights reserved.

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.

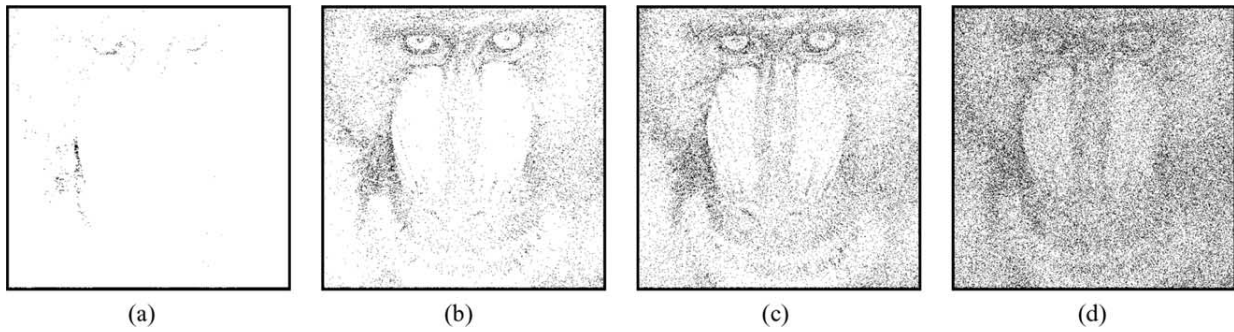


Fig. 1. Gaussian pixel noise can improve the quality of an image through a stochastic-resonance or dithering process (Gammaitoni, 1995; Wannamaker, Lipshitz & Vanderkooy, 2000). The noise produces a nonmonotonic response: A small level of noise sharpens the image features while too much noise degrades 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 θ : $y = g((x + n) - \theta)$ 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)

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

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

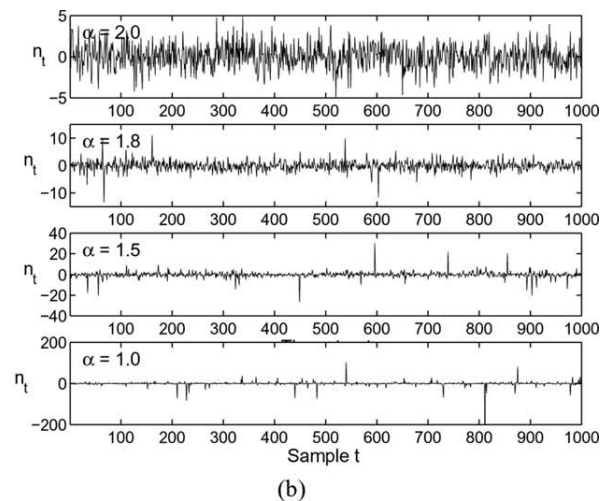
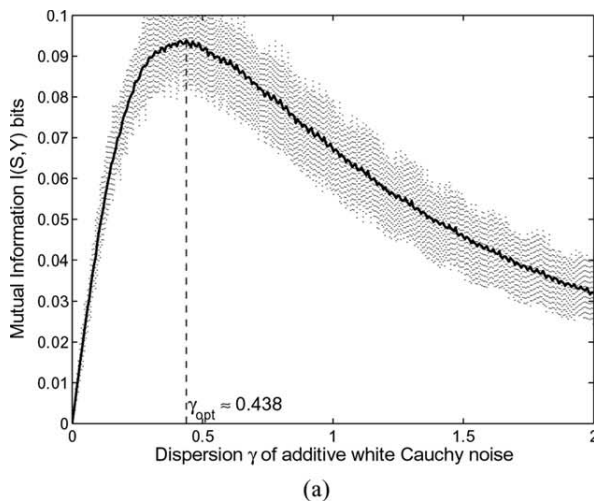


Fig. 2. SR with infinite-variance Cauchy noise. (a) The graph shows the smoothed input-output mutual information of a threshold neuron as a function of the dispersion of additive white Cauchy noise n_t . The dispersion γ controls the width of the Cauchy bell curve. The vertical dashed lines show the 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

$$P_{Y|S}(0|0) = Pr\{s + n < \theta\}_{s=-A} = Pr\{n < \theta + A\} \\ = \int_{-\infty}^{\theta+A} p(n)dn \quad (2)$$

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

$$P_{Y|S}(0|1) = Pr\{s + n < \theta\}_{s=A} = Pr\{n < \theta - A\} \\ = \int_{-\infty}^{\theta-A} p(n)dn \quad (4)$$

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

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 $I(S, Y) \rightarrow 0$ as $\sigma \rightarrow 0$.

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: $I(\sigma) \rightarrow 0$ as $\sigma \rightarrow 0$. 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 $\sigma \rightarrow 0$ 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 $\sigma \rightarrow 0$. 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

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

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

Similarly for $y = '1'$:

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

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

Then

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

The result now follows if

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

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 $\epsilon = \frac{1}{2}d(\theta - A, m) = \frac{1}{2}(\theta - A - m) > 0$. So $\theta - A - \epsilon = \theta - A - \epsilon + m - m = m + (\theta - A - m) - \epsilon = m + 2\epsilon - \epsilon = m + \epsilon$. Then

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

$$\leq \int_{\theta-A}^{\infty} p(n)dn \quad (24)$$

$$\leq \int_{\theta-A-\epsilon}^{\infty} p(n)dn \quad (25)$$

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

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

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

$$\leq Pr\{|n - m| \geq \epsilon\} \quad (29)$$

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

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

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

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

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

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

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

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

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 $0 < \alpha \leq 2$. 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.

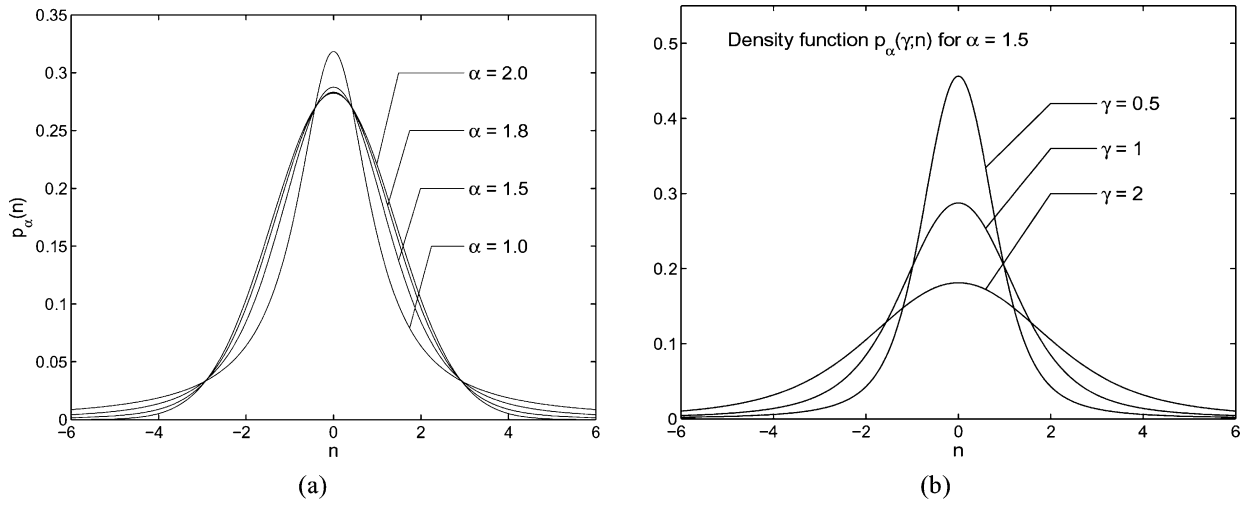


Fig. 4. Samples of standard symmetric ($\beta = 0$) alpha-stable probability densities. (a) Density functions with zero location ($\alpha = 0$) and unit dispersion ($\gamma = 1$) for $\alpha = 2, 1.8, 1.5$, and 1 . The densities are bell curves that have thicker tails as α decreases and thus that model 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

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

where

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

and $i = \sqrt{-1}$, $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, 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 $\beta \neq 0$. 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

$$\int_{\theta-A}^{\theta+A} p(n)dn \rightarrow 0 \text{ as } \gamma \rightarrow 0 \quad (38)$$

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:

$$\lim_{\gamma \rightarrow 0} \varphi(\omega) = \exp\{ia\omega\} \quad (39)$$

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

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

Then

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

$$= \int_{\theta-A}^{\theta+A} \delta(n - a)dn \quad (42)$$

$$= 0 \quad (43)$$

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

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.

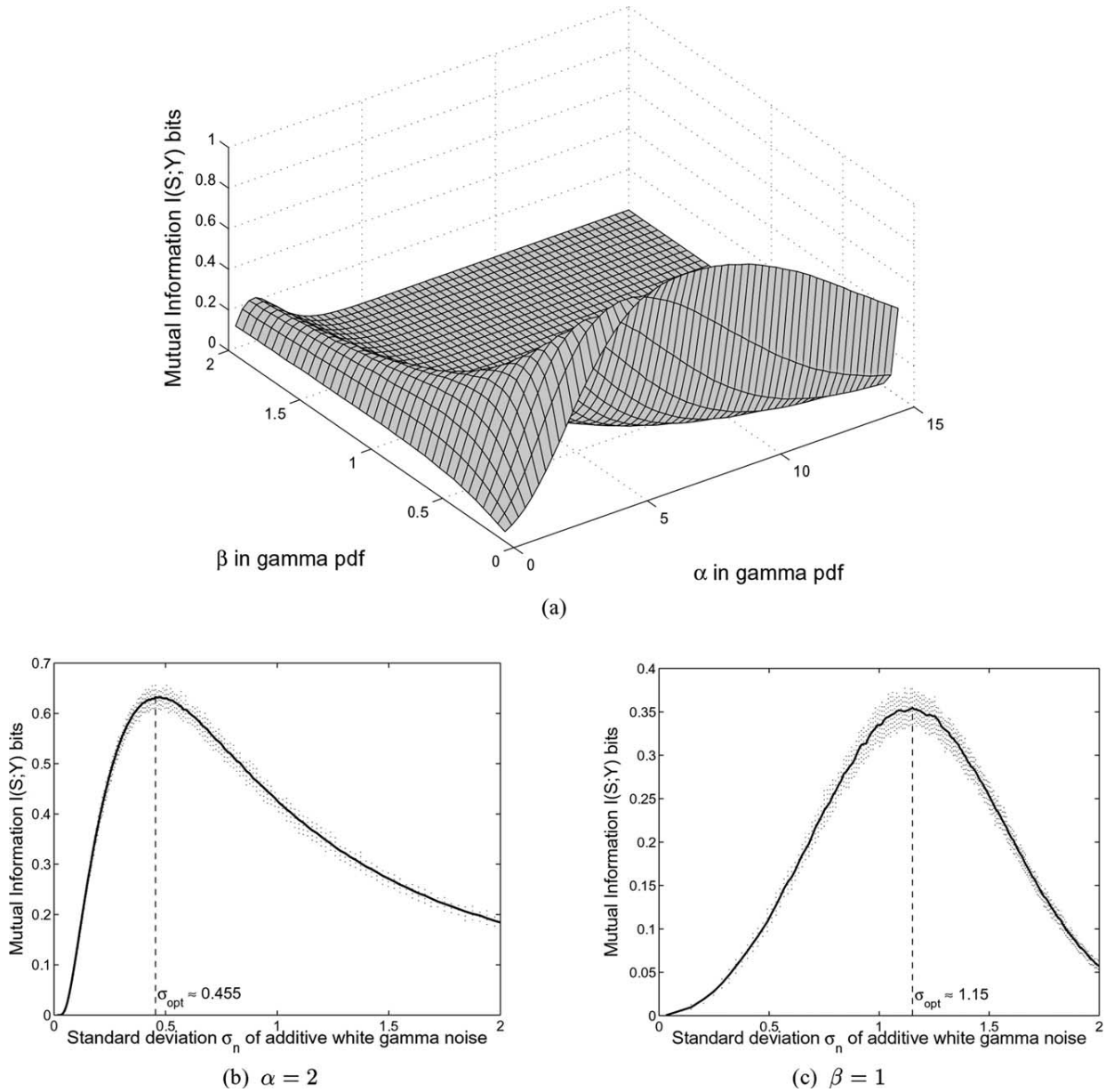


Fig. 3. SR with (finite-variance) gamma noise. The noisy signal-forced threshold neuron has the form of Eq. (1). The gamma noise n_t adds to the bipolar input Bernoulli signal s_t . The neuron has threshold $\theta = 1$. The input Bernoulli signal has amplitude $A = 0.8$ with success probability $p = \frac{1}{2}$. 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 (Akgriray & Lamoureux, 1989; Bergstrom, 1952; Grigoriu, 1995; Nikias & Shao, 1995):

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

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. *Arkiviv 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., Pantazeliou, 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.

Robust stochastic resonance for simple threshold neurons

Bart Kosko¹ and Sanya Mitaim²

¹*Department of Electrical Engineering, Signal and Image Processing Institute, University of Southern California, Los Angeles, California 90089-2564, USA*

²*Department of Electrical Engineering, Faculty of Engineering, Thammasat University, Pathumthani 12120, Thailand*

(Received 27 January 2003; revised manuscript received 6 May 2004; published 27 September 2004)

Simulation and theoretical results show that memoryless threshold neurons benefit from small amounts of almost all types of additive noise and so produce the stochastic-resonance or SR effect. Input-output mutual information measures the performance of such threshold systems that use subthreshold signals. The SR result holds for all possible noise probability density functions with finite variance. The only constraint is that the noise mean must fall outside a “forbidden” threshold-related interval that the user can control—a new theorem shows that this condition is also necessary. A corollary and simulations show that the SR effect occurs for *right*-sided beta and Weibull noise as well. These SR results further hold for the entire uncountably infinite class of alpha-stable probability density functions. Alpha-stable noise densities have infinite variance and infinite higher-order moments and often model impulsive noise environments. The stable noise densities include the special case of symmetric bell-curve densities with thick tails such as the Cauchy probability density. The SR result for alpha-stable noise densities shows that the SR effect in threshold and thresholdlike systems is robust against occasional or even frequent violent fluctuations in noise. Regression analysis reveals both an exponential relationship for the optimal noise dispersion as a function of the alpha bell-curve tail thickness and an approximate linear relationship for the SR-maximal mutual information as a function of the alpha bell-curve tail thickness.

DOI: 10.1103/PhysRevE.70.031911

PACS number(s): 87.10.+e

I. ALMOST ALL THRESHOLD SYSTEMS EXHIBIT STOCHASTIC RESONANCE

Several researchers have found that threshold neurons and other threshold systems exhibit stochastic resonance [1–8]: Small amounts of noise improve the threshold neuron’s input-output correlation measure [9,10] or mutual information [1,8,11]. All of these simulations and analyses assume a noise probability density function that has finite variance. Most further assume that the noise is simply Gaussian or uniform. Yet the statistics of real-world noise can differ substantially from these simple and finite-variance probability descriptions. The noise can be impulsive and irregular and have infinite variance and infinite higher-order moments. Computer simulations alone cannot decide whether this uncountable class of noise densities produces the SR effect in threshold systems. Theoretical techniques can decide the issue and we show that the answer is positive: Almost all threshold systems exhibit the SR effect in terms of mutual information or a bit-based measure of system performance.

The two theorems in [12] show the SR effect in simple (memoryless) threshold neurons as often found in the literature of neural networks [13–15]. We state these two theorems below (Theorems 1.1 and 2.1) without proof and derive a corollary and two new related theorems. The first theorem (Theorem 1.1) shows that threshold neurons exhibit the SR effect for all finite-variance noise densities if the system performance measure is Shannon’s mutual information and if the mean or location parameter falls outside a “forbidden” interval that one can often pick in advance. A corollary shows that this SR effect still occurs for right-sided beta and Weibull noise. Traditional SR research has focused almost exclusively on two-sided noise. The second theorem (Theo-

rem 2.1) shows that this also holds for all infinite-variance densities that belong to the large class of stable distributions. Both theorems assume that all signals are subthreshold signals. The two new theorems (Theorems 1.2 and 2.2) show that there is no SR effect if the mean or location parameters fall within the forbidden threshold interval. Figure 4 shows a simulation instance of this predicted forbidden-interval effect for Gaussian and Cauchy noise.

The paper then presents several regression analyses of simulation experiments that confirm and extend the exponential relationship between the optimal noise dispersion and alpha bell-curve tail thickness [16]. This exponential relationship corresponds to a similar one for infinite-variance SR systems that use a signal-to-noise ratio or a cross correlation for the system performance measure [16]. Regression also shows that the SR-maximal mutual information in noisy threshold neurons depends approximately linearly on the bell-curve tail thickness for symmetric alpha-stable noise.

Figure 1 shows the system-flow diagram of a noisy threshold neuron system that processes subthreshold signals. Figure 2 shows the first use of (right-sided) beta noise for

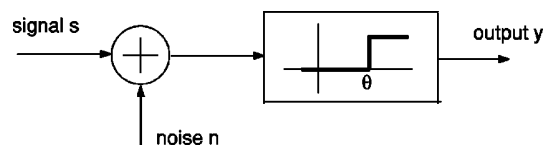


FIG. 1. System-flow diagram of a noisy threshold neuron. The neuron’s signal function has the form (1) with threshold parameter θ , where s is the input signal and n is the input additive noise. We assume subthreshold signals: $A < \theta$, where A is the amplitude of the Bernoulli input s .

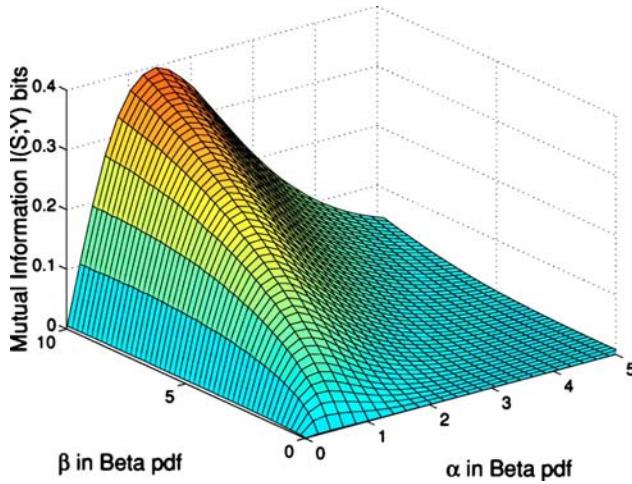


FIG. 2. (Color online) Stochastic resonance with (right-sided) beta noise. The noisy signal-forced threshold neuron has the form (1). The beta noise n_t adds to the bipolar input Bernoulli signal s_t . The parametrized interval $[a, b]$ of the beta density (14) has $a=0$ and $b=10$. The neuron has threshold $\theta=1$. 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 graph shows the smoothed input-output mutual information of a threshold neuron as a function of the parameters α and β of additive white beta noise n_t . The neuron's mutual information has a nonzero noise optimum $\sigma_{\text{opt}} > 0$ where the variance has the form $\sigma_n^2 = [(b-a)^2 \alpha \beta] / [(\alpha + \beta)^2 (\alpha + \beta + 1)]$.

SR. The beta density generalizes the uniform density and is popular in Bayesian statistics [17] because it allows analysts to control the shape of the density with two parameters and scale or translate the finite-length domain. Figure 3 shows the first use of (right-sided) Weibull noise for SR. The Weibull density generalizes the exponential and Rayleigh densities and has an infinite-length domain. Figure 4 shows several symmetrical alpha-stable noise densities whose bell curves have thick tails that produce infinite variance and often highly impulsive noise spikes. Figure 5 shows a simulation instance of both Theorem 2.1 and the empirical trends in Figs. 7 and 8. Infinite-variance Cauchy noise produces the SR effect when plotted against the Shannon mutual information of the threshold system. The linear regression results in Table I and Fig. 7 reveal the exponential relationship between the optimal noise dispersion and the alpha bell-curve tail thickness. The linear dependence of the log-transformed optimal noise dispersion on the bell-curve thickness becomes quadratic when the signal amplitude is too small or too close to the neuron's threshold. The regression results in Table II and Fig. 8 show a similar pattern. The linear dependence of the SR-maximal mutual information on the bell-curve thickness also becomes quadratic when the signal amplitude is too small or too close to the neuron's threshold.

II. THRESHOLD NEURONS AND SHANNON'S MUTUAL INFORMATION

We use the standard discrete-time threshold neuron model [1,2,6,15,16,18]

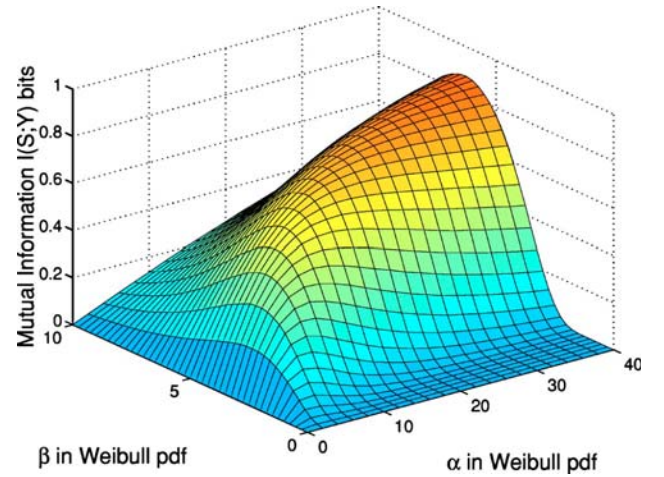


FIG. 3. (Color online) Stochastic resonance with (right-sided) Weibull noise. The noisy signal-forced threshold neuron has the form (1). The Weibull noise n_t adds to the bipolar input Bernoulli signal s_t . The neuron has threshold $\theta=0.5$. The input Bernoulli signal has amplitude $A=0.2$ with 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 graph shows the smoothed input-output mutual information of a threshold neuron as a function of the parameters α and β of additive white Weibull noise n_t . The neuron's mutual information has a nonzero noise optimum $\sigma_{\text{opt}} > 0$ where the variance has the form $\sigma_n^2 = (\beta/\alpha)^{2/\beta} [\Gamma(1+2/\beta) - \{\Gamma(1+1/\beta)\}^2]$.

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

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 $p(n)$. Figure 1 shows the system flow of the threshold system.

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

$$P_{Y|S}(0|0) = \Pr\{s + n < \theta\}_{s=-A} \quad (2)$$

$$\begin{aligned} &= \Pr\{n < \theta + A\} \\ &= \int_{-\infty}^{\theta+A} p(n) dn \end{aligned} \quad (3)$$

$$P_{Y|S}(1|0) = 1 - P_{Y|S}(0|0) \quad (4)$$

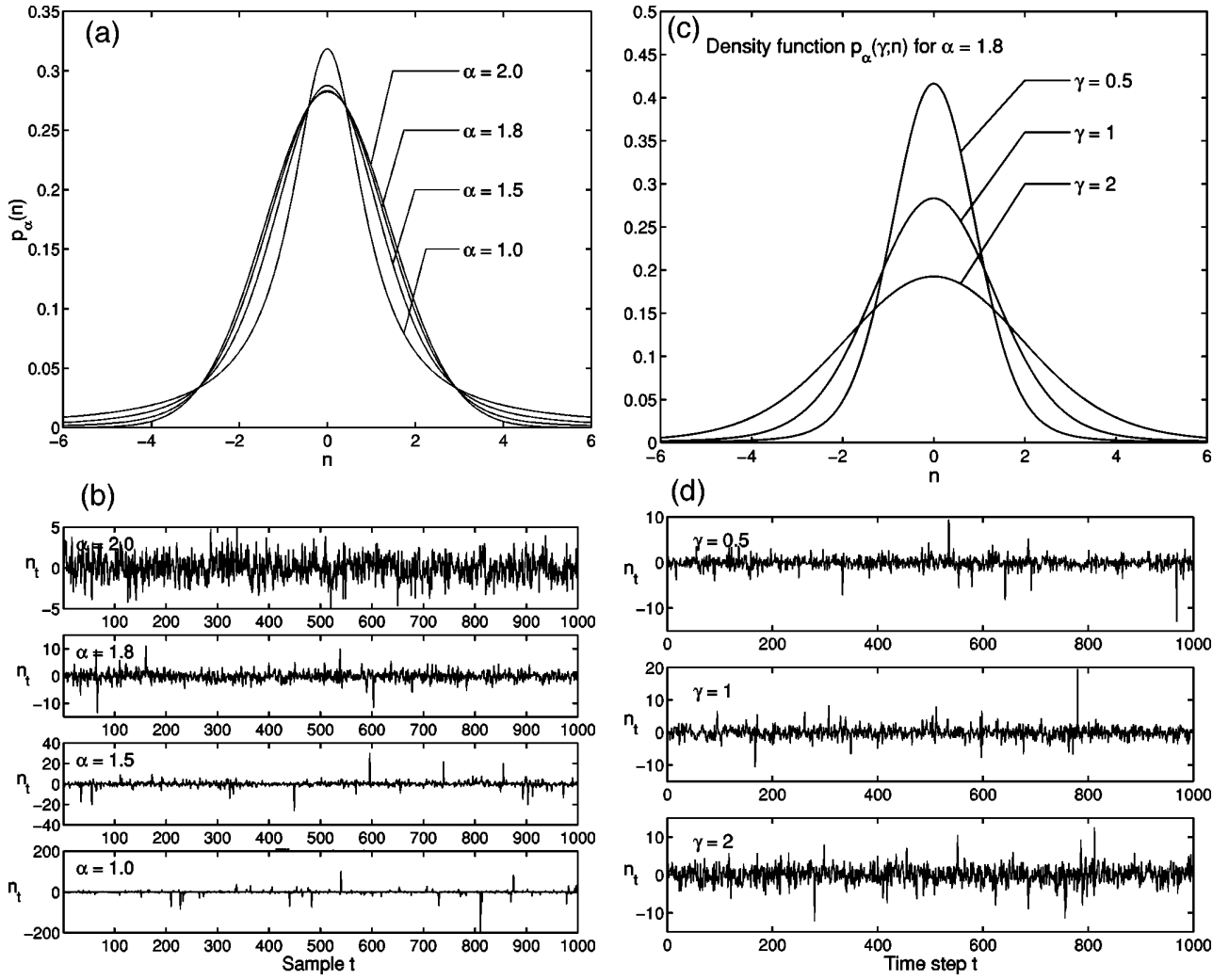


FIG. 4. Samples of standard symmetric alpha-stable probability densities and their realizations. (a) Density functions with zero location ($a=0$) and unit dispersion ($\gamma=1$) for $\alpha=2, 1.8, 1.5$, and 1 . The densities are bell curves that have thicker tails as α decreases and thus that model increasingly impulsive noise as α decreases. The case $\alpha=2$ gives a Gaussian density with variance 2 (or unit dispersion). The parameter $\alpha=1$ gives the Cauchy density with infinite variance. (b) Samples of alpha-stable random variables with zero location and unit dispersion. 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 [39,40] generated these realizations. (c) Density functions for $\alpha=1.8$ with dispersions $\gamma=0.5, 1$, and 2 . (d) Samples of alpha-stable noise n for $\alpha=1.8$ with dispersions $\gamma=0.5, 1$, and 2 .

$$P_{Y|S}(0|1) = \Pr\{s + n < \theta\}_{s=A} \quad (5)$$

$$\begin{aligned} &= \Pr\{n < \theta - A\} \\ &= \int_{-\infty}^{\theta-A} p(n) dn, \end{aligned} \quad (6)$$

$$P_{Y|S}(1|1) = 1 - P_{Y|S}(0|1) \quad (7)$$

and the marginal density is

$$P_Y(y) = \sum_s P_{Y|S}(y|s) P_S(s). \quad (8)$$

Other researchers have derived the conditional probabilities $P_{Y|S}(y|s)$ of the threshold system with Gaussian noise with bipolar inputs [1] and Gaussian inputs [8]. 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 [19] to measure the noise enhancement or “stochastic resonance” (SR) effect [1,3,8,20,21]. 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 $H(Y|X)$:

$$I(S, Y) = H(Y) - H(Y|S) \quad (9)$$

$$= - \sum_y P_Y(y) \log_2 P_Y(y) + \sum_s \sum_y P_{SY}(s, y) \log_2 P_{Y|S}(y|s) \quad (10)$$

$$= - \sum_y P_Y(y) \log_2 P_Y(y) + \sum_s P(s) \sum_y P(y|s) \log_2 P(y|s) \quad (11)$$

$$= \sum_{s,y} P_{SY}(s, y) \log_2 \frac{P_{SY}(s, y)}{P_S(s)P_Y(y)}. \quad (12)$$

So the mutual information is the expectation of the random variable $\log_2\{[P_{SY}(s, y)]/[P_S(s)P_Y(y)]\}$:

$$I(S, Y) = E \left[\log_2 \frac{P_{SY}(s, y)}{P_S(s)P_Y(y)} \right]. \quad (13)$$

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 pseudodistance between the joint probability density $P_{SY}(s, y)$ and the product density $P_S(s)P_Y(y)$. This holds for the Kullback [19] pseudodistance measure

$$I(S, Y) = \sum_s \sum_y P_{SY}(s, y) \log_2 \frac{P_{SY}(s, y)}{P_S(s)P_Y(y)}.$$

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. The proofs in [12] and the Appendix use this fact.

III. SR FOR THRESHOLD SYSTEMS WITH FINITE-VARIANCE NOISE

Almost all finite-variance noise densities produce the SR effect in threshold neurons with subthreshold signals. This holds for all probability density functions defined on a two-symbol alphabet. The proof of Theorem 1.1 in [12] 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 "forbidden" signal-threshold interval $(\theta - A, \theta + A)$.

Theorem 1.1. Suppose that the threshold signal system (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 system (1) exhibits the nonmonotone SR effect in the sense that $I(S, Y) \rightarrow 0$ as $\sigma \rightarrow 0$.

Corollary 1.1. The threshold neuron (1) exhibits stochastic resonance for the additive beta and Weibull noise densities under the hypotheses of Theorem 1.1.

The generalized beta probability density function has the form

$$p(n) = \begin{cases} \frac{1}{b-a} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{n-a}{b-a}\right)^{\alpha-1} \left(\frac{b-n}{b-a}\right)^{\beta-1} & \text{if } a \leq n \leq b \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

Parameters α and β are positive shape constants, parameters a and b are constants $-\infty < a < b < \infty$, and Γ is the gamma function

$$\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy, \quad x > 0. \quad (15)$$

The mean and variance of the beta density are

$$m_n = a + (b-a) \frac{\alpha}{\alpha+\beta}, \quad (16)$$

$$\sigma_n^2 = \frac{(b-a)^2 \alpha \beta}{(\alpha+\beta)^2 (\alpha+\beta+1)}. \quad (17)$$

So the beta density is right-sided for $a \geq 0$. We used $a=0$ and $b=10$ and so defined the beta density in the interval $[0, 10]$ for the SR simulation instance in Fig. 2. The algorithm in [22] generated the beta noise. Bayesian statisticians often use a beta density to encode prior information about a parameter (such as a binomial success parameter p) over a fixed-length interval [23]. The beta density can also model the semblance or the ratio of stacked energy to total energy across a signal array [24], fluctuations of the radar-scattering cross sections

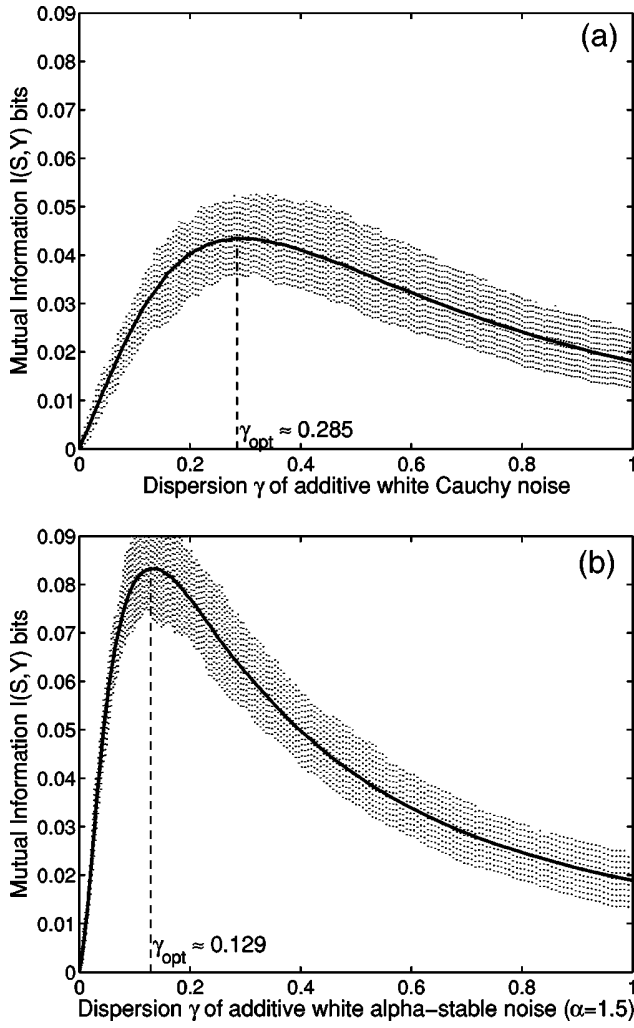


FIG. 5. Stochastic resonance with highly impulsive (infinite-variance) alpha-stable noise. The graphs show the smoothed input-output mutual information of a threshold system as a function of the dispersion of additive white alpha-stable noise n_t with $\alpha=1$ (Cauchy noise) in (a) and $\alpha=1.5$ in (b). The vertical dashed lines show the absolute deviation between the smallest and largest outliers in each sample average of 100 outcomes. The system has a nonzero noise optimum at $\gamma_{\text{opt}} \approx 0.285$ for $\alpha=1$ and $\gamma_{\text{opt}} \approx 0.129$ for $\alpha=1.5$ and thus shows the SR effect. The noisy signal-forced threshold system has the form (1). The alpha-stable noise n_t adds to the bipolar input Bernoulli signal s_t . The system has threshold $\theta=0.5$. The input Bernoulli signal has amplitude $A=0.3$ 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. Note that decreasing the tail-thickness parameter α increases the optimal noise dispersion γ_{opt} as in Fig. 7 and decreases the SR-maximal mutual information $I_{\text{max}}(S, Y)$ as in Fig. 8.

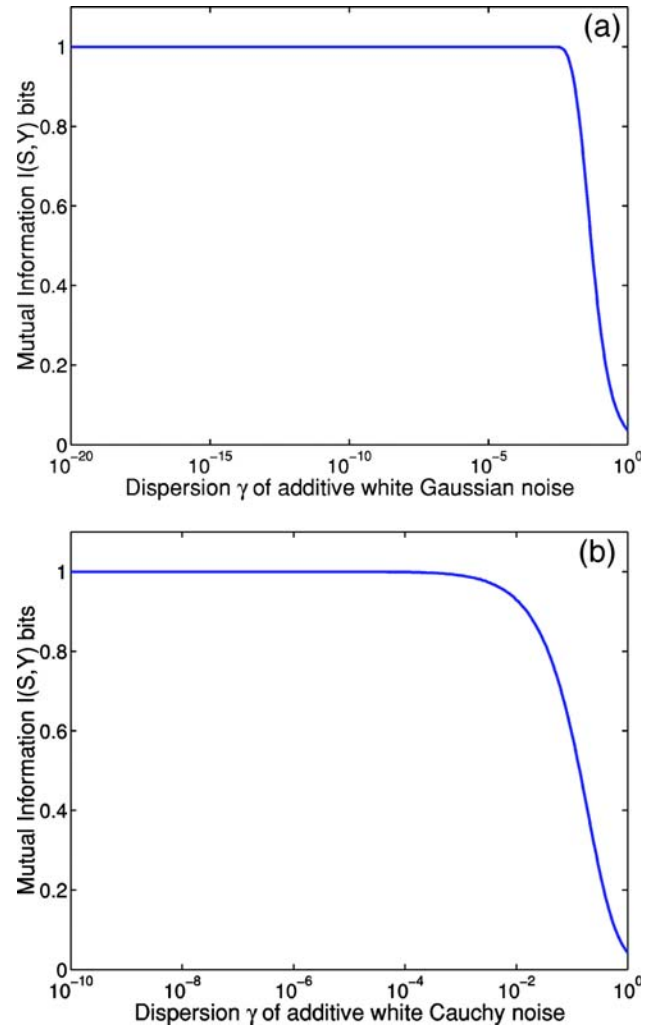


FIG. 6. No SR in the “forbidden” interval (per Theorems 1.2 and 2.2)—mutual information versus alpha-stable noise dispersion when the noise mean (location) lies in the “forbidden” signal-threshold interval: $a \in (\theta-A, \theta+A)$. The graphs show the smoothed input-output mutual information of 100 trials of a threshold system as a function of the dispersion of additive white alpha-stable noise n_t with $\alpha=2$ (Gaussian) in (a) and $\alpha=1$ (Cauchy noise) in (b). The system is optimal when $\gamma \rightarrow 0$ and thus *does not* show the SR effect: The mutual information $I(S, Y)$ is maximum when it equals the input entropy $H(S)$. The noisy signal-forced threshold system has the form (1). The alpha-stable noise n_t has location $a=0.4$ and adds to the bipolar input Bernoulli signal s_t . The system has threshold $\theta=0.5$. The input Bernoulli signal has amplitude $A=0.4$ 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.

of targets [25], the self-similar process of video traffic [26], and the variation of the narrowband vector channels or spatial signature variations due to movement [27].

The Weibull probability density function has the form

$$p(n) = \begin{cases} \alpha n^{\beta-1} e^{-\alpha n^{\beta}} & \text{if } n \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

for positive shape parameters α and β . The mean and variance of the Weibull density are

$$m_n = \left(\frac{\beta}{\alpha}\right)^{1/\beta} \Gamma\left(1 + \frac{1}{\beta}\right), \quad (19)$$

$$\sigma_n^2 = \left(\frac{\beta}{\alpha}\right)^{2/\beta} \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \left\{ \Gamma\left(1 + \frac{1}{\beta}\right) \right\}^2 \right]. \quad (20)$$

Figure 3 shows simulation realizations of this corollary for the Weibull noise density. MATLAB 6.5 [28] generated the Weibull noise. Weibull [29] first proposed this parametric probability density function to model the fracture of materials under repetitive stress. This density has become a standard model of multipart system reliability [30]. It can also effectively model signals and noise in many data-rich systems such as radar clutter [31] and confocal laser scanning microscopy [32].

We next state a result that shows that we cannot in general omit the threshold-interval condition in the hypothesis of Theorem 1.1. Noise does not help a threshold θ that already lies between $\theta-A$ and $\theta+A$.

Theorem 1.2. Suppose that the threshold signal system (1) has noise probability density function $p(n)$ and that the input signal S is subthreshold ($A < \theta$). Suppose that the noise mean $E[n]$ lies in the signal-threshold interval ($\theta-A, \theta+A$) if $p(n)$ has finite variance. Then the threshold system (1) does not exhibit the nonmonotone SR effect in the sense that $I(S, Y)$ is maximum as $\sigma \rightarrow 0$:

$$I(S, Y) = H(Y) = H(S) \quad \text{as} \quad \sigma \rightarrow 0. \quad (21)$$

The Appendix gives the proof.

IV. SR FOR THRESHOLD SYSTEMS WITH INFINITE-VARIANCE NOISE

We now proceed to the more general (and more realistic) case where infinite-variance noise interferes with the threshold system. The SR effect also occurs in other systems with impulsive infinite-variance noise [16,33]. 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 [34–37].

The parameter α controls tail thickness and lies in $0 < \alpha \leq 2$. 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 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 [34] 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 contain “outliers”—and all do to some degree.

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 [37,38]. 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. Figure 4 shows realizations of four symmetric alpha-stable noise random variables.

Theorem 2.1 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 [36,37,41,42]

$$\varphi(\omega) = \exp\left\{ i a \omega - \gamma |\omega|^\alpha \left(1 + i \beta \operatorname{sgn}(\omega) \tan \frac{\alpha \pi}{2} \right) \right\} \quad \text{for } \alpha \neq 1 \quad (22)$$

and

$$\varphi(\omega) = \exp\{ i a \omega - \gamma |\omega| [1 - 2i\beta \ln|\omega| \operatorname{sgn}(\omega)/\pi] \} \quad \text{for } \alpha = 1, \quad (23)$$

where

$$\operatorname{sgn}(\omega) = \begin{cases} 1 & \text{if } \omega > 0 \\ 0 & \text{if } \omega = 0 \\ -1 & \text{if } \omega < 0 \end{cases} \quad (24)$$

and $i = \sqrt{-1}$, $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, 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 $\beta=0$. The theorem below still holds even when $\beta \neq 0$. 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. Numerical integration of φ produced the simulation results in Fig. 4.

The proof of Theorem 2.1 in [12] 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. (22) and (23). 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 2.1. Suppose $I(S, Y) > 0$ and the threshold system (1) uses alpha-stable noise with location parameter $a \notin (\theta-A, \theta+A)$. Then the system (1) exhibits the nonmonotone SR effect if the input signal is subthreshold.

Figure 5 gives a typical example of the SR effect for highly impulsive noise with infinite variance. The alpha-stable noises have $\alpha=1$ (Cauchy) and $\alpha=1.5$. So frequent and violent noise spikes interfere with the signal. Figure 5 also illustrates the empirical trends in Figs. 7 and 8: A falling tail-thickness parameter α produces an increasing optimal noise dispersion γ_{opt} but a decreasing SR-maximal mutual information $I_{\text{max}}(S, Y)$. We next state a new sufficient condition for SR not to occur in an impulsive threshold system.

TABLE I. Linear regression estimates of the SR-optimal log dispersion γ_{opt} as a function of the bell-curve tail-thickness parameter α from a symmetric alpha-stable noise density. The parameters β_0 and β_1 relate $\log_{10}\gamma_{\text{opt}}$ and α through a linear relationship: $\log_{10}\gamma_{\text{opt}}(\alpha) = \beta_0 + \beta_1\alpha$. The coefficient of determination r_l^2 shows how well the linear model fits the log-transformed data. The last column shows the coefficient of determination r_q^2 for the quadratic model $\log_{10}\gamma_{\text{opt}}(\alpha) = \beta_0 + \beta_1\alpha + \beta_2\alpha^2$. All observed significance levels or p -values were less than 10^{-4} .

Signal amplitude A	Linear model Regression coefficients			Quadratic model r_q^2
	$\hat{\beta}_0$	$\hat{\beta}_1$	r_l^2	
0.025	0.0701	-0.5944	0.9003	0.9444
0.050	0.1002	-0.6087	0.9321	0.9723
0.075	0.1124	-0.6192	0.9490	0.9842
0.100	0.1180	-0.6261	0.9558	0.9888
0.125	0.1090	-0.6228	0.9594	0.9910
0.150	0.1078	-0.6251	0.9679	0.9921
0.175	0.1026	-0.6273	0.9672	0.9933
0.200	0.0915	-0.6214	0.9699	0.9942
0.225	0.0810	-0.6161	0.9737	0.9950
0.250	0.0694	-0.6172	0.9781	0.9959
0.275	0.0595	-0.6149	0.9826	0.9964
0.300	0.0439	-0.6148	0.9869	0.9961
0.325	0.0290	-0.6184	0.9903	0.9962
0.350	0.0116	-0.6211	0.9935	0.9961
0.375	-0.0134	-0.6215	0.9957	0.9960
0.400	-0.0313	-0.6367	0.9947	0.9951
0.425	-0.0705	-0.6432	0.9903	0.9950
0.450	-0.1107	-0.6688	0.9757	0.9944
0.475	-0.1837	-0.7217	0.9408	0.9911
0.490	-0.2805	-0.8053	0.8987	0.9863

Theorem 2.2. Suppose that the threshold signal system (1) has subthreshold input signal and use alpha-stable noise with location parameter $a \in (\theta - A, \theta + A)$. Then the threshold system (1) does not exhibit the nonmonotone SR effect: $I(S, Y)$ is maximum as $\gamma \rightarrow 0$:

$$I(S, Y) = H(Y) = H(S) \quad \text{as } \gamma \rightarrow 0. \quad (25)$$

The Appendix gives the proof. Figure 6 shows the noise-mutual information profile of the subthreshold signal system with noise location (mean) in the “forbidden” signal-threshold interval.

Statistical regression confirmed an exponential relationship between the optimal noise dispersion γ_{opt} and the bell-curve tail-thickness parameter α : $\gamma_{\text{opt}}(\alpha) = 10^{\beta_0 + \beta_1\alpha}$ for parameters β_0 and β_1 that depend on the signal amplitude A . Then the log-transformation of the optimal dispersion gives the linear model $\log_{10}\gamma_{\text{opt}}(\alpha) = \beta_0 + \beta_1\alpha$. Table I shows the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ and the coefficient of determination r_l^2 for 20 signal amplitudes in the threshold neuron using SPSS software. All observed significance levels or

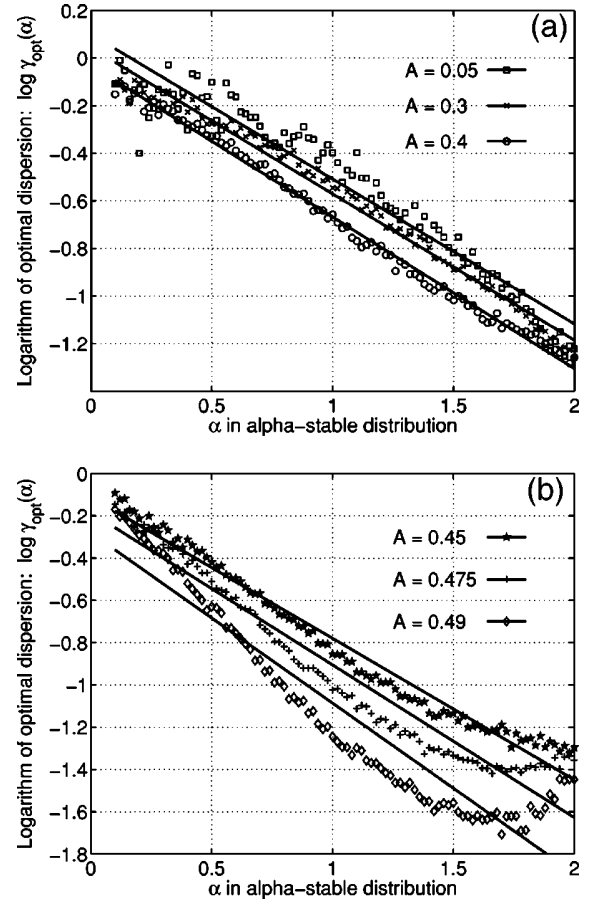


FIG. 7. Exponential law for optimal noise dispersion γ_{opt} as a function of bell-curve thickness parameter α for the mutual-information performance measure and for different signal amplitudes A . The optimal noise dispersion γ_{opt} depends on the parameter α through the exponential relation $\gamma_{\text{opt}}(\alpha) = 10^{\beta_0 + \beta_1\alpha}$ for parameters β_0 and β_1 [or $\gamma_{\text{opt}}(\alpha) = 10^{\beta_0 + \beta_1\alpha + \beta_2\alpha^2}$ for a quadratic fit to the data]. Table I shows the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ for 20 input Bernoulli signal amplitudes A . The exponential trend's exponent is linear for most amplitudes but becomes quadratic for very small amplitudes and for amplitudes close to the threshold $\theta = \frac{1}{2}$. All observed significance levels or p -values were less than 10^{-4} .

p -values were less than 10^{-4} . The p -values measure the credibility of the null hypothesis that the regression lines have zero slope or other coefficients. The exponential trend's exponent is linear for most amplitudes but becomes quadratic for very small amplitudes and for amplitudes close to the threshold $\theta = \frac{1}{2}$ [or $\gamma_{\text{opt}}(\alpha) = 10^{\beta_0 + \beta_1\alpha + \beta_2\alpha^2}$ for a quadratic fit to the data]. Figure 7 shows 6 of the 20 log-linear plots.

We also found an approximate linear relationship $I_{\text{max}}(S, Y; \alpha) = \beta_0 + \beta_1\alpha$ for the SR-maximal mutual information $I_{\text{max}}(S, Y)$ as a function of the tail-thickness parameter α . Table II shows the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ and the coefficient of determination r_l^2 for 20 signal amplitudes in the threshold neuron. All observed significance levels or p -values were less than 10^{-4} . There is a clear linear trend for most amplitudes A . The trend becomes quadratic for very

TABLE II. Linear regression of the SR-maximal mutual information $I_{\max}(S, Y)$ as a function of the bell-curve tail-thickness parameter α from a symmetric alpha-stable noise density. The parameters β_0 and β_1 relate $I_{\max}(S, Y)$ and α through a linear relationship: $I_{\max}(S, Y; \alpha) = \beta_0 + \beta_1 \alpha$. The coefficient of determination r_l^2 shows how well the linear model fits the data. The last column shows the coefficient of determination r_q^2 for the quadratic model $I_{\max}(S, Y; \alpha) = \beta_0 + \beta_1 \alpha + \beta_2 \alpha^2$. All observed significance levels or p -values were less than 10^{-4} .

Signal amplitude A	Linear model Regression coefficients			Quadratic model r_q^2
	$\hat{\beta}_0$	$\hat{\beta}_1$	r_l^2	
0.025	-0.0001	0.0006	0.9312	0.9907
0.050	-0.0008	0.0022	0.9370	0.9972
0.075	-0.0018	0.0049	0.9401	0.9985
0.100	-0.0031	0.0086	0.9440	0.9990
0.125	-0.0048	0.0134	0.9477	0.9993
0.150	-0.0068	0.0190	0.9521	0.9995
0.175	-0.0090	0.0256	0.9558	0.9997
0.200	-0.0113	0.0329	0.9612	0.9998
0.225	-0.0138	0.0411	0.9658	0.9998
0.250	-0.0161	0.0500	0.9715	0.9997
0.275	-0.0185	0.0596	0.9764	0.9995
0.300	-0.0207	0.0698	0.9816	0.9993
0.325	-0.0224	0.0807	0.9866	0.9990
0.350	-0.0236	0.0920	0.9913	0.9987
0.375	-0.0240	0.1039	0.9951	0.9984
0.400	-0.0229	0.1161	0.9976	0.9981
0.425	-0.0196	0.1286	0.9972	0.9977
0.450	-0.0120	0.1408	0.9905	0.9975
0.475	0.0058	0.1513	0.9655	0.9973
0.490	0.0336	0.1527	0.9145	0.9959

small amplitudes and for amplitudes close to the threshold $\theta = \frac{1}{2}$. Figure 8 shows 6 of the 20 linear plots.

V. CONCLUSIONS

Both theory and detailed simulations show that almost all noise types produce stochastic resonance in threshold systems that use subthreshold signals. These results help explain the widespread occurrence of the SR effect in mechanical and biological threshold systems [43–49]. The broad generality of the results suggests that SR should occur in any nonlinear system whose input-output structure approximates a threshold system as in the many models of continuous neurons [50–52]. The infinite-variance theoretical and simulation results further imply that such widespread SR effects should be robust against violent noise impulses.

ACKNOWLEDGMENTS

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

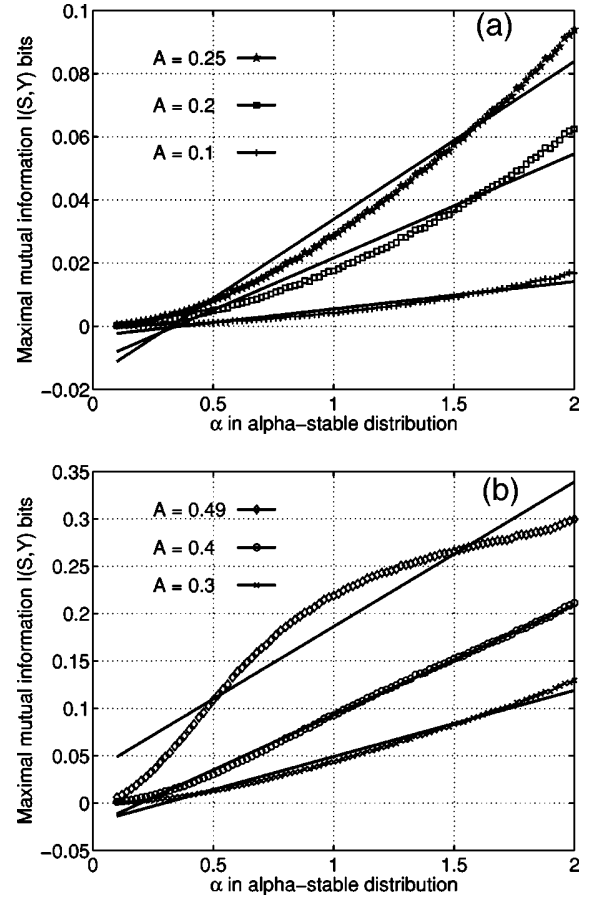


FIG. 8. Linear regression for maximal mutual information $I_{\max}(S, Y)$ as a function of bell-curve thickness parameter α for different signal amplitudes A . The maximal mutual information $I_{\max}(S, Y)$ depends on the parameter α through the linear relationship $I_{\max}(S, Y; \alpha) = \beta_0 + \beta_1 \alpha$ for parameters β_0 and β_1 [or $I_{\max}(S, Y; \alpha) = \beta_0 + \beta_1 \alpha + \beta_2 \alpha^2$ for a quadratic fit to the data]. Table II shows the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ for 20 input Bernoulli signal amplitudes A . The linear trend is strong for most amplitudes A . The trend becomes quadratic for very small amplitudes and for amplitudes close to the threshold $\theta = \frac{1}{2}$. All observed significance levels or p -values were less than 10^{-4} .

APPENDIX: PROOFS OF THEOREMS 1.2 AND 2.2

The two proofs below use the same idea as do the proofs for Theorems 1.1 and 2.1 [12]. Assume $0 < P_S(s) < 1$ to avoid triviality when $P_S(s) = 0$ or 1. We show that $H(Y) \rightarrow H(S)$ and $H(Y|S) \rightarrow 0$ as $\sigma \rightarrow 0$ or $\gamma \rightarrow 0$. So $I(S, Y) \rightarrow H(S)$ as $\sigma \rightarrow 0$ or $\gamma \rightarrow 0$ and is maximum since $I(S, Y) = H(Y) - H(Y|S)$ and $I(S, Y) \leq H(S)$ by the data processing inequality: $I(S, S) \geq I(S, g(S)) = I(S, Y)$ for a Markov chain $S \rightarrow S \rightarrow Y$ [19]. The boundary case $I(S, S) = H(S)$ implies $I(S, Y) \leq H(S)$.

Finite-variance noise case (Theorem 1.2)

Now we show that $P_{Y|S}(y|s)$ is either 1 or 0 as $\sigma \rightarrow 0$ or $\gamma \rightarrow 0$. Let the mean of the noise be $m = E[n]$ and the variance be $\sigma^2 = E[(n - m)^2]$. Then $m \in (\theta - A, \theta + A)$ by hypothesis.

Consider first $P_{Y|S}(0|0)$. Pick $\varepsilon = \frac{1}{2}d(\theta+A, m) = \frac{1}{2}(\theta+A-m) > 0$. So $\theta+A-\varepsilon = \theta+A-\varepsilon+m-m = m+(\theta+A-m)-\varepsilon = m+2\varepsilon-\varepsilon = m+\varepsilon$. Then

$$P_{Y|S}(0|0) = \int_{-\infty}^{\theta+A} p(n)dn \quad (A1)$$

$$\geq \int_{-\infty}^{\theta+A-\varepsilon} p(n)dn \quad (A2)$$

$$= \int_{-\infty}^{m+\varepsilon} p(n)dn \quad (A3)$$

$$= 1 - \int_{m+\varepsilon}^{\infty} p(n)dn \quad (A4)$$

$$= 1 - \Pr\{n \geq m + \varepsilon\} = 1 - \Pr\{n - m \geq \varepsilon\} \quad (A5)$$

$$\geq 1 - \Pr\{|n - m| \geq \varepsilon\} \quad (A6)$$

$$\geq 1 - \frac{\sigma^2}{\varepsilon^2} \text{ by Chebyshev's inequality} \quad (A7)$$

$$\rightarrow 1 \text{ as } \sigma \rightarrow 0. \quad (A8)$$

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

Similarly for $P_{Y|S}(1|1)$: Pick $\varepsilon = \frac{1}{2}d(\theta-A, m) = \frac{1}{2}(m-\theta+A) > 0$. So $\theta-A+\varepsilon = \theta-A+\varepsilon+m-m = m+(\theta-A-m)+\varepsilon = m-2\varepsilon+\varepsilon = m-\varepsilon$. Then

$$P_{Y|S}(1|1) = \int_{\theta-A}^{\infty} p(n)dn \quad (A9)$$

$$\geq \int_{\theta-A+\varepsilon}^{\infty} p(n)dn \quad (A10)$$

$$= \int_{m-\varepsilon}^{\infty} p(n)dn \quad (A11)$$

$$= 1 - \int_{-\infty}^{m-\varepsilon} p(n)dn \quad (A12)$$

$$= 1 - \Pr\{n \leq m - \varepsilon\} = 1 - \Pr\{n - m \leq -\varepsilon\} \quad (A13)$$

$$\geq 1 - \Pr\{|n - m| \geq \varepsilon\} \quad (A14)$$

$$\geq 1 - \frac{\sigma^2}{\varepsilon^2} \text{ by Chebyshev's inequality} \quad (A15)$$

$$\rightarrow 1 \text{ as } \sigma \rightarrow 0. \quad (A16)$$

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

Alpha-stable noise case (Theorem 2.2)

The characteristic function $\varphi(\omega)$ of alpha-stable noise density $p(n)$ has the exponential form (22) and (23). This reduces to a simple complex exponential in the zero-dispersion limit:

$$\lim_{\gamma \rightarrow 0} \varphi(\omega) = \exp\{i a \omega\} \quad (A17)$$

for all characteristic exponents α , skewnesses β , and locations a . So Fourier transformation gives the corresponding density function in the limiting case ($\gamma \rightarrow 0$) as a translated delta function δ :

$$\lim_{\gamma \rightarrow 0} p(n) = \delta(n - a). \quad (A18)$$

Then $a \in (\theta-A, \theta+A)$ gives

$$P_{Y|S}(0|0) = \int_{-\infty}^{\theta+A} p(n)dn \quad (A19)$$

$$\rightarrow \int_{-\infty}^{\theta+A} \delta(n - a)dn = 1 \text{ as } \gamma \rightarrow 0. \quad (A20)$$

Similarly

$$P_{Y|S}(1|1) = \int_{\theta-A}^{\infty} p(n)dn \quad (A21)$$

$$\rightarrow \int_{\theta-A}^{\infty} \delta(n - a)dn = 1 \text{ as } \gamma \rightarrow 0. \quad (A22)$$

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

$$H(Y|S) = - \sum_s \sum_y P_{SY}(s, y) \log_2 P_{Y|S}(y|s) \quad (A23)$$

$$= \sum_s P_S(s) \sum_y P_{Y|S}(y|s) \log_2 P_{Y|S}(y|s) \quad (A24)$$

$$= 0, \quad (A25)$$

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

$$H(Y) = - \sum_y P_Y(y) \log_2 P_Y(y) \quad (A26)$$

$$= - \sum_s P_S(s) \log_2 P_S(s) \quad (A27)$$

$$= H(S) \quad (A28)$$

because

$$P_Y(y) = \sum_s P_{Y|S}(y|s)P_S(s) \quad (\text{A29})$$

$$= P_{Y|S}(y|0)P_S(0) + P_{Y|S}(y|1)P_S(1) \quad (\text{A30})$$

$$= P_{Y|S}(y|0)P_S(0) + P_{Y|S}(y|1)[1 - P_S(0)] \quad (\text{A31})$$

$$= [P_{Y|S}(y|0) - P_{Y|S}(y|1)]P_S(0) + P_{Y|S}(y|1) \quad (\text{A32})$$

$$= [P_{Y|S}(y|1) - P_{Y|S}(y|0)]P_S(1) + P_{Y|S}(y|0) \quad (\text{A33})$$

$$= \begin{cases} P_S(1) & \text{if } y = 1 \\ P_S(0) & \text{if } y = 0. \end{cases} \quad (\text{A34})$$

-
- [1] A. R. Bulsara and A. Zador, *Phys. Rev. E* **54**, R2185 (1996).
[2] L. Gammaitoni, *Phys. Lett. A* **208**, 315 (1995).
[3] X. Godivier and F. Chapeau-Blondeau, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **8**, 581 (1998).
[4] S. M. Hess and A. M. Albano, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **8**, 395 (1998).
[5] N. Hohn and A. N. Burkitt, in *Proceedings of the 2001 IEEE International Joint Conference on Neural Networks (IJCNN '01)* (IEEE, Washington, D.C., 2001), pp. 644–647.
[6] P. Jung, *Phys. Lett. A* **207**, 93 (1995).
[7] P. Jung and G. Mayer-Kress, *Nuovo Cimento D* **17**, 827 (1995).
[8] N. G. Stocks, *Phys. Rev. E* **63** 041114 (2001).
[9] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, *Phys. Rev. E* **54**, 5575 (1996).
[10] J. J. Collins, C. C. Chow, and T. T. Imhoff, *Nature (London)* **376**, 236 (1995).
[11] S. Mitaim and B. Kosko, in *Proceedings of the 2002 IEEE International Joint Conference on Neural Networks (IJCNN '02), Honolulu* (IEEE, Honolulu, 2002), pp. 1980–1985.
[12] B. Kosko and S. Mitaim, *Neural Networks* **16**, 755 (2003).
[13] S. Grossberg, *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control* (Kluwer Academic Publishers, Dordrecht, 1982).
[14] S. Haykin, *Neural Networks: A Comprehensive Foundation* (McMillan, 1994).
[15] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence* (Prentice Hall, Englewood Cliffs, NJ, 1992).
[16] B. Kosko and S. Mitaim, *Phys. Rev. E* **64**, 051110 (2001).
[17] I. Miller and M. Miller, *Mathematical Statistics*, 6th ed. (Prentice Hall, Englewood Cliffs, NJ, 1999).
[18] J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554 (1982).
[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, 1991).
[20] G. Deco and B. Schürmann, *Physica D* **117**, 276 (1998).
[21] M. E. Inchiosa, J. W. C. Robinson, and A. R. Bulsara, *Phys. Rev. Lett.* **85**, 3369 (2000).
[22] R. C. H. Cheng, *Commun. ACM* **21**, 317 (1978).
[23] E. A. Tanis and R. V. Hogg, *Probability and Statistical Inference*, 6th ed. (Prentice Hall, Englewood Cliffs, NJ, 2001).
[24] C. V. Kimball and D. J. Scheibner, *Geophysics* **63**, 345 (1998).
[25] R. L. Kulp, *Electromagnetics* **4**, 165 (1984).
[26] N. Ansari, H. Liu, Y. Q. Shi, and H. Zhao, *IEEE Trans. Broadcasting* **48**, 337 (2002).
[27] A. Kavak, M. Torlak, W. J. Vogel, and G. Xu, *IEEE Trans. Microwave Theory Tech.* **46**, 930 (2000).
[28] L. Devroye, *Non-Uniform Random Variate Generation* (Springer-Verlag, Berlin, 1986).
[29] W. Weibull, *J. Appl. Mech.* **18**, 293 (1951).
[30] W. Mendenhall and T. Sincich, *Statistics for Engineering and the Sciences*, 4th ed. (Prentice Hall, Englewood Cliffs, NJ, 1995).
[31] M. A. Sletten, *IEEE Trans. Antennas Propag.* **46**, 45 (1998).
[32] J. Hu, A. Razden, G. M. Nielson, G. E. Farin, D. P. Baluch, and D. G. Capco, *IEEE Trans. Vis. Comput. Graph.* **9**, 320 (2003).
[33] S. Mitaim and B. Kosko, *Proc. IEEE* **86**, 2152 (1998), special issue on intelligent signal processing.
[34] L. Breiman, *Probability* (Addison-Wesley, Reading, MA, 1968).
[35] W. Feller, *An Introduction to Probability Theory and Its Applications* (John Wiley & Sons, Reading, MA, 1966) Vol. II.
[36] M. Grigoriu, *Applied Non-Gaussian Processes* (Prentice Hall, Englewood Cliffs, NJ, 1995).
[37] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications* (John Wiley & Sons, 1995).
[38] B. Kosko, *Fuzzy Engineering* (Prentice Hall, Englewood Cliffs, NJ, 1996).
[39] J. M. Chambers, C. L. Mallows, and B. W. Stuck, *J. Am. Stat. Assoc.* **71**, 340 (1976).
[40] P. Tsakalides and C. L. Nikias, *IEEE Trans. Signal Process.* **44**, 1623 (1996).
[41] V. Akgiray and C. G. Lamoureux, *J. Bus. Econ. Stat.* **7**, 85 (1989).
[42] H. Bergstrom, *Arkiv Föur Matematik* **2**, 375 (1952).
[43] H. A. Braun, H. Wissing, K. Schäfer, and M. C. Hirsch, *Nature (London)* **367**, 270 (1994).
[44] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, *Nature (London)* **365**, 337 (1993).
[45] S. Fauve and F. Heslot, *Phys. Lett.* **97A**, 5 (1983).
[46] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, *Rev. Mod. Phys.* **70**, 223 (1998).
[47] J. E. Levin and J. P. Miller, *Nature (London)* **380**, 165 (1996).
[48] V. I. Melnikov, *Phys. Rev. E* **48**, 2481 (1993).
[49] D. F. Russell, L. A. Willkens, and F. Moss, *Nature (London)* **402**, 291 (1999).
[50] A. R. Bulsara, A. J. Maren, and G. Schnera, *Biol. Cybern.* **70**, 145 (1993).
[51] A. Longtin, *J. Stat. Phys.* **70**, 309 (1993).
[52] H. E. Plesser and S. Tanaka, *Phys. Lett. A* **225**, 228 (1997).

Adaptive Stochastic Resonance in Noisy Neurons Based on Mutual Information

Sanya Mitaim and Bart Kosko

Abstract—Noise can improve how memoryless neurons process signals and maximize their throughput information. Such favorable use of noise is the so-called “stochastic resonance” or SR effect at the level of threshold neurons and continuous neurons. This paper presents theoretical and simulation evidence that 1) lone noisy threshold and continuous neurons exhibit the SR effect in terms of the mutual information between random input and output sequences, 2) a new statistically robust learning law can find this entropy-optimal noise level, and 3) the adaptive SR effect is robust against highly impulsive noise with infinite variance. Histograms estimate the relevant probability density functions at each learning iteration. A theorem shows that almost all noise probability density functions produce some SR effect in threshold neurons even if the noise is impulsive and has infinite variance. The optimal noise level in threshold neurons also behaves nonlinearly as the input signal amplitude increases. Simulations further show that the SR effect persists for several sigmoidal neurons and for Gaussian radial-basis-function neurons.

Index Terms—Alpha-stable noise, impulsive noise, infinite-variance statistics, mutual information, noise processing, sigmoidal neurons and radial basis functions, stochastic gradient learning, stochastic resonance (SR), threshold neurons.

I. NOISE AND ADAPTIVE STOCHASTIC RESONANCE

NOISE is an unwanted signal or source of energy. Scientists and engineers have largely tried to filter noise or cancel it or mask it out of existence. The Noise Pollution Clearinghouse condemns noise outright: “Noise is unwanted sound. It is derived from the Latin word ‘nausea’ meaning seasickness. Noise is among the most pervasive pollutants today. Noise from road traffic, jet planes, jet skis, garbage trucks, construction equipment, manufacturing processes, lawn mowers, leaf blowers, and boom boxes, to name a few, are among the unwanted sounds that are routinely broadcast into the air.”

The new field of *stochastic resonance* or SR [3], [4], [9], [26], [32], [52], [53], [61], [71] rests on an exception to this undeclared war on noise. SR occurs when noise enhances a faint signal in a nonlinear system. It occurs when the addition of a small amount of noise increases a nonlinear system’s performance measure such as its signal-to-noise ratio (SNR), cross-correlation, or mutual information. The nonlinearity is

often as simple as a memoryless threshold. So a great deal of SR research has focused on how dither-like noise can help spiking neurons process data streams [12], [33], [38]. SR occurs in physical systems such as ring lasers [56], threshold hysteretic Schmitt triggers [27], superconducting quantum interference devices (SQUIDS) [36], Josephson junctions [7], chemical systems [25], and quantum-mechanical systems [34]. SR also occurs in biological systems such as the rat [18], crayfish [23], cricket [48], river paddlefish [66], and in many types of model neurons [8], [10], [16], [17], [63].

Fig. 1 shows how uniform pixel noise can improve our subjective perception of an image. The system quantizes the original gray-scale “Lena” image into a binary image of black and white pixels. It emits a white pixel as output if the input gray-scale pixel equals or exceeds a threshold. It emits a black pixel as output if the input gray-scale pixel falls below the threshold. This quantizer is biased because it does not set the threshold at the midpoint of the gray scale. So the quantized version of the original image contains almost no information. A small level of noise sharpens the image contours and helps fill in features when it adds to the original image before the system applies the threshold. Too much noise swamps the image and degrades its contours. Gammaitoni [29] and others [70] have proposed a dithering argument for this SR effect and still others [55] have applied this argument to still images. The argument involves adding dither noise to a signal before quantization. Consider gray-scale pixel $x \in [0, 1]$ and binary output pixel $y \in \{0, 1\}$ with threshold $\theta = 1/2$. Then the dithered quantizer gives $E[Y|x] = 1 - \Pr\{n < \theta - x\} = x$ if and only if the noise is uniform on $(-1/2, 1/2)$. But the subjective SR result in Fig. 1 holds for nonuniform *infinite*-variance Cauchy noise and for many other types of nonuniform noise. So the dithering argument only partially explains this subjective SR effect.

We first show that noise added to a memoryless threshold neuron produces the SR effect in terms of the Shannon mutual information $I(S, Y)$ between realizations of a random (Bernoulli) bipolar input signal S and realizations of the thresholded output random variable Y . Fig. 2 shows a typical simulation confirmation of this SR result for additive Gaussian noise. The theorem holds for more general bell curves that have thicker tails and thus that have infinite variance and can produce impulsive noise. Extensive simulations reproduce these SR effects for several standard continuous sigmoidal neurons and for Gaussian radial basis functions (see Fig. 13).

We next show that a new robust learning law can find the optimal noise variance and dispersion for both threshold and continuous neurons and for both finite-variance and infinite-variance noise. We introduced adaptive stochastic resonance in [57]

Manuscript received May 11, 2002; revised May 21, 2003 and January 16, 2004. This work was supported by the National Science Foundation Grant ECS-0070284, and Thailand Research Fund Grants PDF/29/2543 and RSA4680001.

S. Mitaim is with the Department of Electrical Engineering, Faculty of Engineering, Thammasat University, Rangsit Campus, Klong Luang, Pathumthani 12120, Thailand (e-mail:msanya@engr.tu.ac.th).

B. Kosko is with the Department of Electrical Engineering, Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089 USA (e-mail:kosko@sipi.usc.edu).

Digital Object Identifier 10.1109/TNN.2004.826218

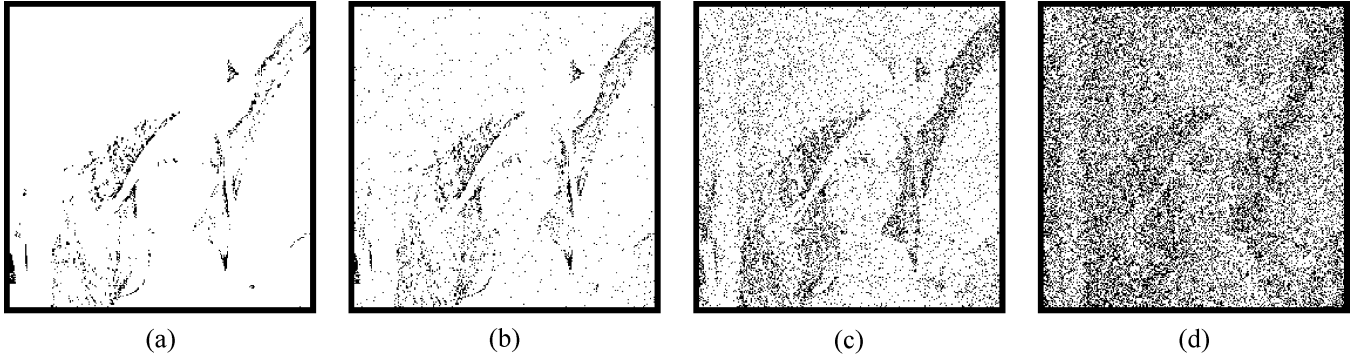


Fig. 1. A “dithering” Cauchy pixel noise can improve subjective image quality. The noise produces a nonmonotonic response: A small level of noise sharpens the image features while too much noise degrades them. These noisy images result when we apply a pixel threshold to the popular “Lena” image used in signal processing [60]: $y = g((x + n) - \theta)$ 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 $\theta = 0.06$. Thresholding the original “Lena” image gives the faint image in (a). The Cauchy noise n has zero location and its dispersion γ_n grows from (b)–(d): $\gamma_n = 0.01$ in (b), $\gamma_n = 0.08$ in (c), and $\gamma_n = 0.50$ in (d).

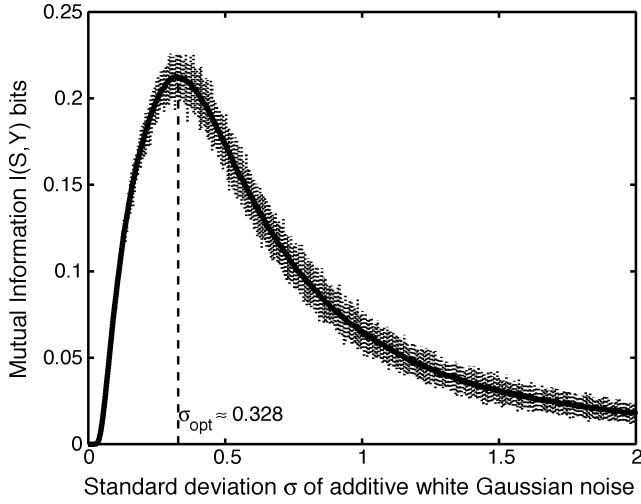


Fig. 2. The nonmonotonic signature of stochastic resonance. The graph shows the smoothed input-output mutual information of a threshold system as a function of the standard deviation of additive white Gaussian noise n_t . The vertical dashed lines show the absolute deviation between the smallest and largest outliers in each sample average of 100 outcomes. The system has a nonzero noise optimum at $\sigma_{\text{opt}} \approx 0.328$ and thus shows the SR effect. The noisy signal-forced threshold system has the form (6). The Gaussian noise n_t adds to the external forcing bipolar signal s_t .

and [47] as a robustified stochastic gradient ascent algorithm that slowly finds the optimal noise variance or dispersion given thousands of joint samples of the noise input and the nonlinear system’s spectral SNR or its cross correlation. This paper extends adaptive SR to the mutual-information performance measure. The last section derives and tests a new robustified learning law that finds the entropically optimal noise level given histogram estimates of the underlying marginal and conditional probability density functions. This statistically robust algorithm uses only the sign of the noise gradient rather than the gradient itself.

The results show that model neurons can exploit low levels of crosstalk or other forms of noise in their local environment. Even highly impulsive noise can help neurons maximize their throughput information. Such noise-based information maximization is consistent with Linsker’s principle of information maximization in neural networks [49], [50]. These findings support the implicit SR conjecture that biological neurons have

evolved to computationally exploit their noisy environments [11], [18], [19], [23], [48], [58], [64], [69]. Further support is that these adaptive SR effects still hold for other sigmoidal and nonsigmoidal (Gaussian) neurons as Fig. 13 shows. These results suggest that biological neurons should experience less mutual information if they do not use their local noise.

II. MUTUAL INFORMATION AND SR IN NEURON MODELS

This section reviews Shannon’s measure of mutual information between two random variables. Then it reviews the simple nonlinear threshold model of a neuron and the continuous neuron model that show the SR effect for bipolar signals.

A. Mutual Information Measure

Mutual information [20] can measure the SR effect [12], [22], [33], [43], [67]. The discrete Shannon mutual information of the input S and output Y has the form

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

$$= - \sum_y P_Y(y) \log P_Y(y) + \sum_s \sum_y P_{SY}(s, y) \log P_{Y|S}(y|s) \quad (2)$$

$$= - \sum_y P_Y(y) \log P_Y(y) + \sum_s P(s) \sum_y P(y|s) \log P(y|s) \quad (3)$$

$$= \sum_{s,y} P_{SY}(s, y) \log \frac{P_{SY}(s, y)}{P_S(s)P_Y(y)}. \quad (4)$$

We can view the mutual information in the form of expectation of a random variable $\log(P_{SY}(s, y)/P_S(s)P_Y(y))$:

$$I(S, Y) = E \left[\log \frac{P_{SY}(s, y)}{P_S(s)P_Y(y)} \right]. \quad (5)$$

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 joint density of the input S and the output Y .

Mutual information also measures the pseudodistance between the joint probability density $P_{SY}(s, y)$ and the product density $P_S(s)P_Y(y)$. This holds for the Kullback [20] pseudodistance measure $I(S, Y) = \sum_s \sum_y P_{SY}(s, y) \log(P_{SY}(s, y)/P_S(s)P_Y(y))$. 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.

B. Noisy Threshold Neuron

We use the discrete-time threshold neuron model [12], [29], [39], [44], [45]

$$y_t = \text{sgn}(s_t + n_t - \theta) = \begin{cases} 1 & \text{if } s_t + n_t \geq \theta \\ -1 & \text{if } s_t + n_t < \theta \end{cases} \quad (6)$$

where $\theta > 0$ is the neuron's threshold, s_t is the bipolar input Bernoulli signal (with success probability 1/2) with amplitude $A > 0$, and n_t is the additive white noise with probability density $p(n)$. Experiments with other success probabilities near 1/2 did not produce substantially different simulation results.

C. Noisy Continuous Neurons

We use the additive continuous neuron model with a neuronal signal function $S(x)$ [45]

$$\dot{x} = -x + S(x) + s(t) + n(t) \quad (7)$$

$$y(t) = \text{sgn}(x(t)). \quad (8)$$

Here $s(t)$ and $n(t)$ are the input and additive noise of the neuron and $y(t)$ is the binary output. The neuron feeds its output signal $S(x)$ back to itself and emits the threshold bipolar signal $y(t)$ as output.

- **Hyperbolic Tangent** This signal function gives an additive neuron model that is bistable [2], [10], [15], [39], [40], [45]

$$S(x) = 2 \tanh x. \quad (9)$$

- **Linear-Threshold** This simple linear-threshold signal function [45] also gives the SR effect in the neuron

$$S(x) = \begin{cases} cx & |cx| < 1 \\ 1 & cx > 1 \\ -1 & cx < -1 \end{cases} \quad (10)$$

for a constant $c > 0$. We use $c = 2$.

- **Exponential** This signal function is asymmetric with the form [45]

$$S(x) = \begin{cases} 1 - \exp\{-cx\} & x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

for a constant $c > 0$. We use $c = 10$.

- **Gaussian.** The Gaussian or "radial basis" signal function [45] differs from the other signal functions above because it is nonmonotonic

$$S(x) = \exp\{-cx^2\} \quad (12)$$

for a constant $c > 0$. We use $c = 100$.

III. MUTUAL INFORMATION OF THE THRESHOLD NEURON WITH BIPOLAR INPUT SIGNALS

A. SR in Memoryless Threshold Neurons

This section derives analytical SR results for the noisy threshold neuron based on the marginal probability density function of the output $P_Y(y)$ and the conditional density $P_{Y|S}(y|s)$. The system is the binary neuron with a fixed threshold θ . The bipolar (Bernoulli with success probability p) input signal s_t has amplitude A : $s_t \in \{-A, A\}$ with probability density $P_S(s)$. The noise n_t adds to the signal s_t before it enters the neuron. So the neuron's output y_t has the form (6). Fig. 5 plots the mutual information $I(S, Y)$ for four standard closed-form noise probability density functions 18, 24, 29, and 38. The central result is a theorem that holds for almost all noise probability densities so long as the mean noise falls outside a user-controlled interval that depends on the threshold θ .

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

$$\begin{aligned} P_{Y|S}(0|0) &= \Pr\{s + n < \theta\}|_{s=-A} \\ &= \Pr\{n < \theta + A\} = \int_{-\infty}^{\theta+A} p(n)dn \end{aligned} \quad (13)$$

$$P_{Y|S}(1|0) = 1 - P_{Y|S}(0|0) \quad (14)$$

$$\begin{aligned} P_{Y|S}(0|1) &= \Pr\{s + n < \theta\}|_{s=A} \\ &= \Pr\{n < \theta - A\} = \int_{-\infty}^{\theta-A} p(n)dn \end{aligned} \quad (15)$$

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

and the marginal density is

$$P_Y(y) = \sum_s P_{Y|S}(y|s)P_S(s). \quad (17)$$

Researchers have derived the conditional probabilities $P_{Y|S}(y|s)$ of the threshold system with *Gaussian* noise with bipolar inputs [12] and Gaussian inputs [67]. We next derive $P_{Y|S}(y|s)$ for uniform, Laplace, and (infinite-variance) Cauchy noise as well. Fig. 3 shows four examples of the unimodal noise densities and their realizations. Then we introduce stable distributions to model a spectrum of impulsive noise types.

- **Gaussian Noise** The Gaussian density with zero mean and variance $\sigma_n^2 = \sigma^2$ has the form

$$p(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n^2}{2\sigma^2}\right\}. \quad (18)$$

Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$\begin{aligned} P_{Y|S}(0|0) &= \int_{-\infty}^{\theta+A} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n^2}{2\sigma^2}\right\} dn \\ &= \frac{1}{2} + \frac{1}{2} \text{erf}\left(\frac{\theta+A}{\sigma\sqrt{2}}\right) \end{aligned} \quad (19)$$

$$P_{Y|S}(1|0) = \frac{1}{2} - \frac{1}{2} \text{erf}\left(\frac{\theta+A}{\sigma\sqrt{2}}\right) \quad (20)$$

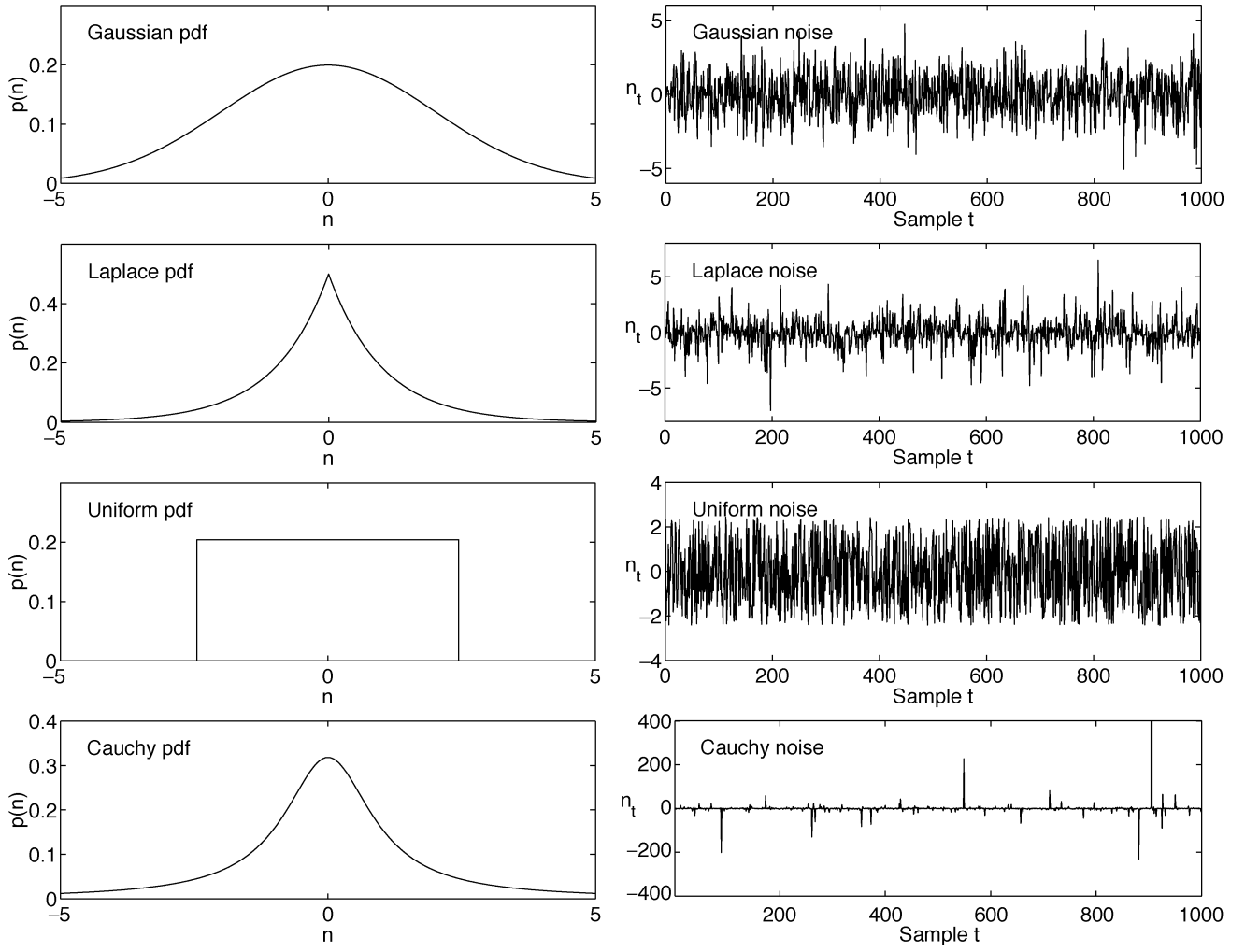


Fig. 3. Probability density functions and sample realizations. The figure shows Gaussian, Laplace, and uniform random variables n with zero mean and variance of two: $E[n] = 0$ and $E[n^2] = \sigma^2 = 2$. The Cauchy density function has zero location and unit dispersion but infinite variance. The pseudorandom number generators in [65] act as noise sources for these probability densities.

$$P_{Y|S}(0|1) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{\theta - A}{\sigma\sqrt{2}} \right) \quad (21)$$

$$P_{Y|S}(1|1) = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\frac{\theta - A}{\sigma\sqrt{2}} \right). \quad (22)$$

The error function erf is

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\{-t^2\} dt. \quad (23)$$

- **Uniform Noise** The uniform density with zero mean and variance $\sigma_n^2 = a^2/12$ has the form

$$p(n) = \begin{cases} \frac{1}{a} & \text{if } -\frac{a}{2} < n < \frac{a}{2} \\ 0 & \text{otherwise.} \end{cases} \quad (24)$$

Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$\begin{aligned} P_{Y|S}(0|0) &= \begin{cases} 1 & \text{if } \frac{a}{2} < \theta + A \\ \frac{1}{2} + \frac{A+\theta}{a} & \text{otherwise} \end{cases} \\ &= \min \left\{ 1, \frac{1}{2} + \frac{\theta + A}{a} \right\} \end{aligned} \quad (25)$$

$$P_{Y|S}(1|0) = \max \left\{ 0, \frac{1}{2} - \frac{\theta + A}{a} \right\} \quad (26)$$

$$P_{Y|S}(0|1) = \min \left\{ 1, \frac{1}{2} + \frac{\theta - A}{a} \right\} \quad (27)$$

$$P_{Y|S}(1|1) = \max \left\{ 0, \frac{1}{2} - \frac{\theta - A}{a} \right\}. \quad (28)$$

- **Laplace Noise** The Laplace density with zero mean and variance $\sigma_n^2 = 2\beta^2$ has the form

$$p(n) = \frac{1}{2\beta} \exp \left\{ -\left| \frac{n}{\beta} \right| \right\}. \quad (29)$$

Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$P_{Y|S}(0|0) = 1 - \frac{1}{2} \exp \left\{ -\frac{\theta + A}{\beta} \right\} \quad (30)$$

$$P_{Y|S}(1|0) = \frac{1}{2} \exp \left\{ -\frac{\theta + A}{\beta} \right\} \quad (31)$$

$$P_{Y|S}(0|1) = 1 - \frac{1}{2} \exp \left\{ -\frac{\theta - A}{\beta} \right\} \quad (32)$$

$$P_{Y|S}(1|1) = \frac{1}{2} \exp \left\{ -\frac{\theta - A}{\beta} \right\}. \quad (33)$$

- **Cauchy Noise** The Cauchy density with zero location and finite dispersion γ (but infinite variance) has the form

$$p(n) = \frac{1}{\pi} \frac{\gamma}{n^2 + \gamma^2}. \quad (34)$$

Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$P_{Y|S}(0|0) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \frac{\theta + A}{\gamma} \quad (35)$$

$$P_{Y|S}(1|0) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{\theta + A}{\gamma} \quad (36)$$

$$P_{Y|S}(0|1) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \frac{\theta - A}{\gamma} \quad (37)$$

$$P_{Y|S}(1|1) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{\theta - A}{\gamma}. \quad (38)$$

- **Symmetric Alpha-Stable Noise: Thick-Tailed Bell Curves**

We 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 [6], [28], [35], [62]. The parameter α controls tail thickness and lies in $0 < \alpha \leq 2$. 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 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 [6] 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. 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 [46], [62]. Note that the best choice of α is an *empirical* question for bell-curve phenomena. Bell-curve behavior alone does not justify the assumption of the Gaussian bell curve.

Fig. 4 shows realizations of four symmetric alpha-stable random variables. A general alpha-stable probability density function f has characteristic function φ [1], [5], [35], [62]

$$\varphi(\omega) = \exp\left\{i a \omega - \gamma |\omega|^\alpha \left(1 + i \beta \text{sign}(\omega) \tan \frac{\alpha \pi}{2}\right)\right\} \quad \text{for } \alpha \neq 1 \quad (39)$$

and

$$\varphi(\omega) = \exp\left\{i a \omega - \gamma |\omega| \left(1 - \frac{2i \beta \ln |\omega| \text{sign}(\omega)}{\pi}\right)\right\} \quad \text{for } \alpha = 1 \quad (40)$$

where

$$\text{sign}(\omega) = \begin{cases} 1 & \text{if } \omega > 0 \\ 0 & \text{if } \omega = 0 \\ -1 & \text{if } \omega < 0 \end{cases} \quad (41)$$

and $i = \sqrt{-1}$, $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, 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 $\beta = 0$. The theorem shown still holds even when $\beta \neq 0$. 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 alpha-stable densities for most α 's. Numerical integration of φ gives the probability densities in Fig. 4.

The following theorem shows that noisy threshold neurons produce some SR effect for almost all noise probability descriptions. The proof shows that if $I(S, Y) > 0$ then eventually the mutual information $I(S, Y)$ tends toward zero as the noise variance or dispersion tends toward zero. So the mutual information $I(S, Y)$ must increase as the noise variance increases from zero. The crucial assumption is that the noise mean $E[n]$ (or location parameter) not lie in the signal-threshold interval $(\theta - A, \theta + A)$.

Theorem: Suppose that the threshold signal system (6) 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. Suppose that $a \notin (\theta - A, \theta + A)$ for the location parameter a of an alpha-stable noise density with characteristic function (39), (40). Then the threshold system (6) exhibits the nonmonotone SR effect in the sense that $I(S, Y) \rightarrow 0$ as $\sigma \rightarrow 0$ or $\gamma \rightarrow 0$.

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: $I(S, Y) \rightarrow 0$ as $\sigma \rightarrow 0$ (or as $\gamma \rightarrow 0$). Recall that $I(S, Y) = 0$ if and only if S and Y are statistically independent [20]. 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 $\sigma \rightarrow 0$ (or as $\gamma \rightarrow 0$) for all signal symbols $s \in S$ and $y \in Y$. The two-symbol alphabet set S gives

$$P_Y(y) = \sum_s P_{Y|S}(y|s) P_S(s) \quad (42)$$

$$= P_{Y|S}(y|0) P_S(0) + P_{Y|S}(y|1) P_S(1) \quad (43)$$

$$= P_{Y|S}(y|0) P_S(0) + P_{Y|S}(y|1) (1 - P_S(0)) \quad (44)$$

$$= (P_{Y|S}(y|0) - P_{Y|S}(y|1)) P_S(0) + P_{Y|S}(y|1). \quad (45)$$

So we need to show only that $P_{Y|S}(y|0) - P_{Y|S}(y|1) = 0$ as $\sigma \rightarrow 0$ (or as $\gamma \rightarrow 0$). 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 (13) and (15) imply that

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) = \int_{-\infty}^{\theta+A} p(n) dn - \int_{-\infty}^{\theta-A} p(n) dn \quad (46)$$

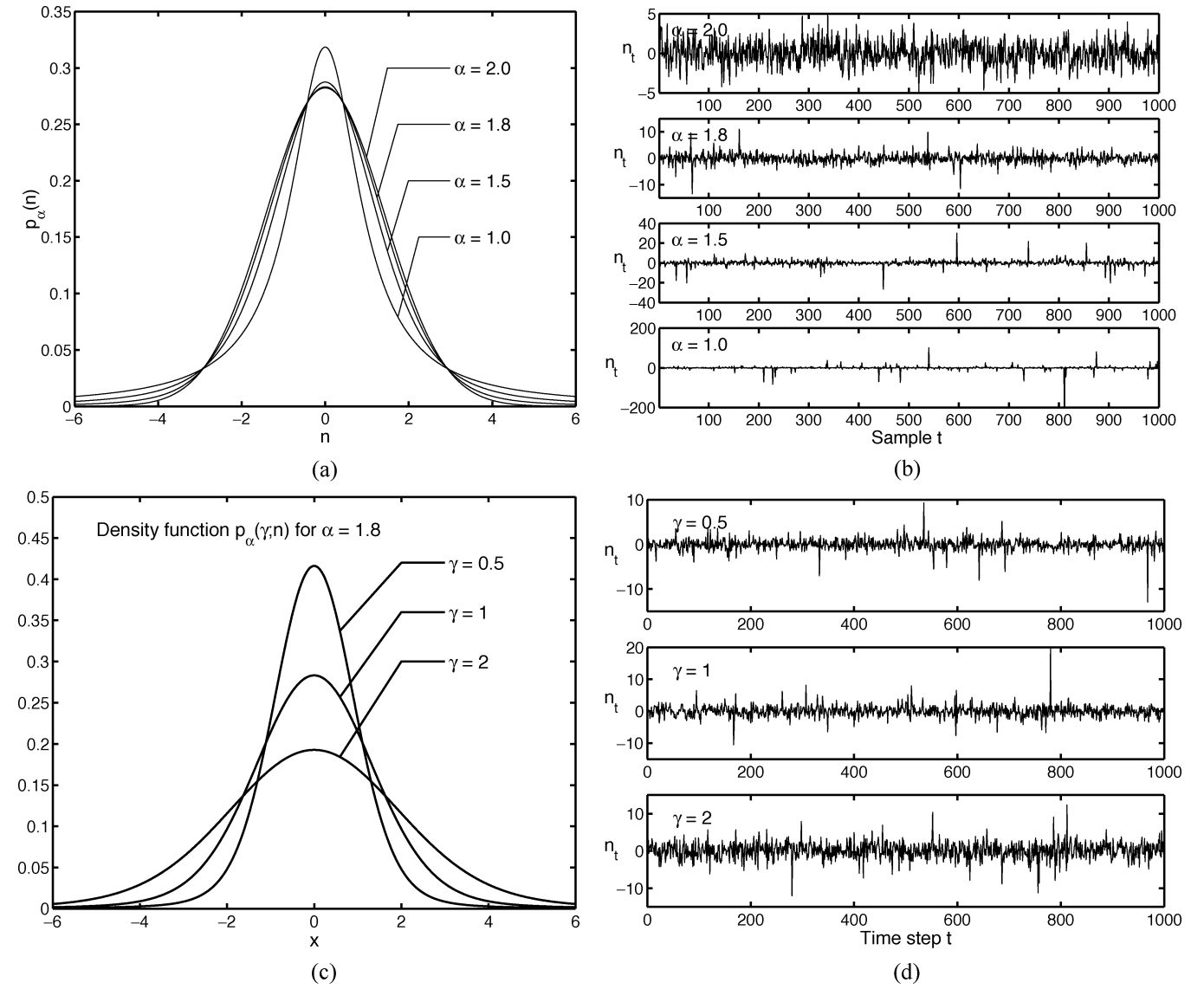


Fig. 4. Samples of standard symmetric alpha-stable probability densities and their realizations. (a) Density functions with zero location ($a = 0$) and unit dispersion ($\gamma = 1$) for $\alpha = 2, 1.8, 1.5$, and 1 . The densities are bell curves that have thicker tails as α decreases and thus that 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. (b) Samples of alpha-stable random variables with zero location and unit dispersion. The plots show realizations when $\alpha = 2, 1.8, 1.5$, and 1 . Note the scale differences on the y -axes. The alpha-stable noise n becomes more impulsive as the parameter α falls. The algorithm in [13], [68] generates these realizations. (c) Density functions for $\alpha = 1.8$ with dispersions $\gamma = 0.5, 1$, and 2 . (d) Samples of alpha-stable noise n for $\alpha = 1.8$ with dispersions $\gamma = 0.5, 1$, and 2 .

$$= \int_{\theta-A}^{\theta+A} p(n)dn. \quad (47)$$

Similarly for $y = '1'$

$$P_{Y|S}(1|0) = \int_{\theta-A}^{\theta+A} p(n)dn \quad (48)$$

$$P_{Y|S}(1|1) = \int_{\theta-A}^{\infty} p(n)dn. \quad (49)$$

Then

$$P_{Y|S}(1|0) - P_{Y|S}(1|1) = - \int_{\theta-A}^{\theta+A} p(n)dn. \quad (50)$$

The result now follows if we can show that

$$\int_{\theta-A}^{\theta+A} p(n)dn \rightarrow 0 \quad \text{as } \sigma \rightarrow 0 \text{ or } \gamma \rightarrow 0. \quad (51)$$

Case 1) Finite-variance noise. Let the mean of the noise be $m = E[n]$ and the variance be $\sigma^2 = E[(n - m)^2]$. Then $m \notin (\theta - A, \theta + A)$ by hypothesis.

Now suppose that $m < \theta - A$. Pick $\varepsilon = (1/2)d(\theta - A, m) = (1/2)(\theta - A - m) > 0$. So $\theta - A - \varepsilon = \theta - A - \varepsilon + m - m = m + (\theta - A - m) - \varepsilon = m + 2\varepsilon - \varepsilon = m + \varepsilon$. Then

$$\begin{aligned} P_{Y|S}(0|0) - P_{Y|S}(0|1) \\ = \int_{\theta-A}^{\theta+A} p(n)dn \end{aligned} \quad (52)$$

$$\leq \int_{\theta-A}^{\infty} p(n)dn \quad (53)$$

$$\leq \int_{\theta-A-\varepsilon}^{\infty} p(n)dn \quad (54)$$

$$= \int_{m+\varepsilon}^{\infty} p(n)dn \quad (55)$$

$$= \Pr\{n \geq m + \varepsilon\} = \Pr\{n - m \geq \varepsilon\} \quad (56)$$

$$\leq \Pr\{|n - m| \geq \varepsilon\} \quad (57)$$

$$\leq \frac{\sigma^2}{\varepsilon^2} \quad \text{by Chebyshev's inequality} \quad (58)$$

$$\rightarrow 0 \quad \text{as } \sigma \rightarrow 0. \quad (59)$$

Suppose next that $m > \theta + A$. Then pick $\varepsilon = (1/2)d(\theta + A, m) = (1/2)(m - \theta - A) > 0$ and so $\theta + A + \varepsilon = \theta + A + \varepsilon + m - m = m - (m - \theta - A) + \varepsilon = m - 2\varepsilon + \varepsilon = m - \varepsilon$. Then

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) = \int_{\theta-A}^{\theta+A} p(n)dn \quad (60)$$

$$\leq \int_{-\infty}^{\theta+A} p(n)dn \quad (61)$$

$$\leq \int_{-\infty}^{\theta+A+\varepsilon} p(n)dn \quad (62)$$

$$= \int_{-\infty}^{m-\varepsilon} p(n)dn \quad (63)$$

$$= \Pr\{n \leq m - \varepsilon\} = \Pr\{n - m \leq -\varepsilon\} \quad (64)$$

$$\leq \Pr\{|n - m| \geq \varepsilon\} \quad (65)$$

$$\leq \frac{\sigma^2}{\varepsilon^2} \quad \text{by Chebyshev's inequality} \quad (66)$$

$$\rightarrow 0 \quad \text{as } \sigma \rightarrow 0. \quad (67)$$

Case 2) Impulsive noise: Alpha-stable noise. The characteristic function $\varphi(\omega)$ of alpha-stable density $p(n)$ has the exponential form (39), (40). This reduces to a simple complex exponential in the zero-dispersion limit

$$\lim_{\gamma \rightarrow 0} \varphi(\omega) = \exp\{ia\omega\} \quad (68)$$

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

$$\lim_{\gamma \rightarrow 0} p(n) = \delta(n - a). \quad (69)$$

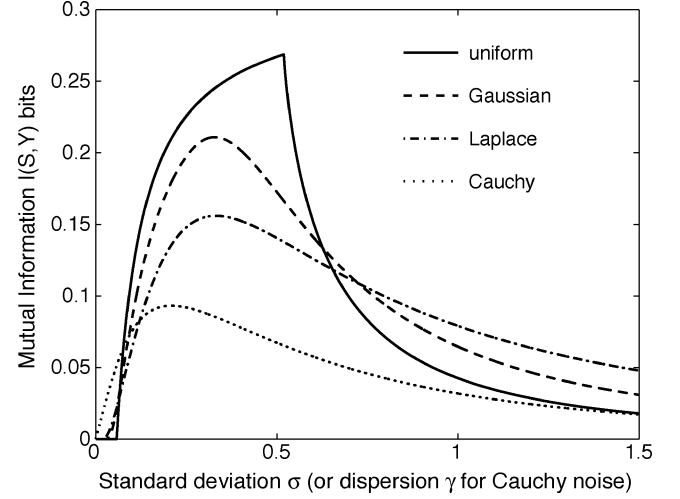


Fig. 5. Mutual information I profiles of a threshold system with bipolar input for four kinds of noise. The system has threshold $\theta = 0.5$. The input Bernoulli signal is bipolar with amplitude $A = 0.4$.

Then

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) = \int_{\theta-A}^{\theta+A} p(n)dn \quad (70)$$

$$= \int_{\theta-A}^{\theta+A} \delta(n - a)dn \quad (71)$$

$$= 0 \quad (72)$$

because $a \notin (\theta - A, \theta + A)$.

Then $P_Y(y) = P_{Y|S}(y|s)$ as $\gamma \rightarrow 0$. So Cases 1 and 2 imply that $I(S, Y) \rightarrow 0$ as $\sigma \rightarrow 0$ for finite-variance noise or as $\gamma \rightarrow 0$ for alpha-stable noise. Q.E.D.

B. Theoretical Results for Closed-Form Noise Densities

Inserting Gaussian or other specific closed-form conditional probability densities $P_{Y|S}(y|s)$ from (19)–(38) into (1)–(4) gives exact solutions of the mutual information $I(S, Y)$ as a function of the noise parameter σ . Fig. 5 shows I -versus- σ profiles of a threshold system with four kinds of noise: Gaussian, uniform, Laplace, and Cauchy. The I profile of the uniform noise has the highest peak among the four noise densities for the same system (same threshold θ and same input amplitude A). And the I profile has a distinct shape: it drops sharply after it reaches its peak as σ grows. Gaussian noise gives the second highest I while Cauchy gives the lowest. The threshold system requires different optimal standard deviations (or dispersions) for different kinds of noise.

The closed form of the I versus σ profiles in Fig. 5 also allows a direct analysis of how the optimal noise depends on the signal amplitude A for Gaussian, uniform, Laplace, and Cauchy noise. Suppose the signal amplitude A is a subthreshold input in a noisy threshold neuron with threshold θ : $A < \theta$. Then will the optimal noise σ_{opt} (or γ_{opt}) decrease as the signal amplitude A moves closer to the threshold θ ?

Intuition might suggest that the threshold system should need less noise to produce the entropic SR effect as the amplitude

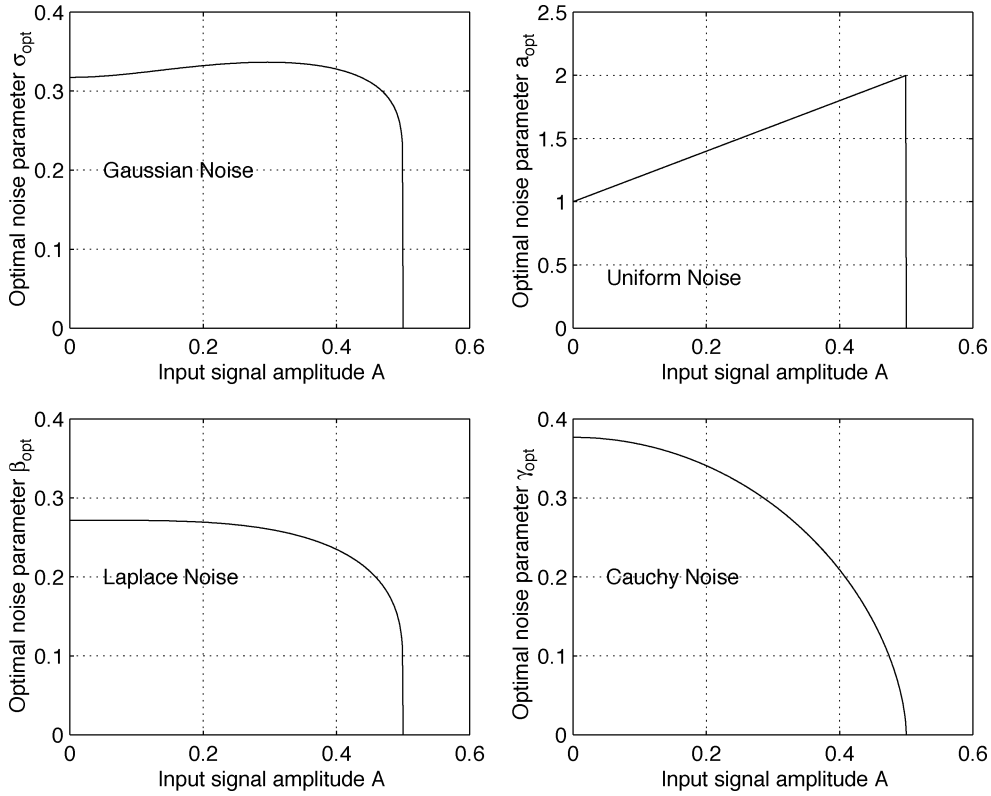


Fig. 6. Optimal SR noise schedules for a noisy threshold neuron with threshold $\theta = 0.5$. The schedules show how optimal noise variance or dispersion depends on signal amplitude A for the four closed-form noise results in Fig. 5.

moves closer to the threshold θ . But the results in Fig. 6 show that the compound nonlinearities involved produce no such simple relationship. The different noise types produce different SR optimality schedules. Fig. 6 shows four optimal noise schedules for the threshold value $\theta = 0.5$. Other threshold values produced similar results. Only optimal Laplace and Cauchy noise produce the more intuitive monotone decrease in the optimal noise level with rising signal amplitude A . Optimal uniform noise grows linearly with signal amplitude while optimal Gaussian noise defines a nonmonotonic schedule.

IV. STOCHASTIC RESONANCE IN COMPUTER SIMULATIONS

Discrete simulations can model continuous-time nonlinear dynamical systems if a stochastic numerical scheme approximates the system dynamics and its signal and noise response. We used a simple stochastic version of the Euler scheme to model a nonlinear system with input forcing signal and noise. We measured how the system performed based on only the system's input-output samples.

Consider the forced dynamical system with additive forcing input signal s and “white” noise n

$$\dot{x} = f(x) + s(t) + n(t) \quad (73)$$

$$y(t) = g(x(t)). \quad (74)$$

These models simply add a noise term to a differential equation rather than use formal Ito or Stratonovich stochastic differentials [14], [24], [31]. “Whiteness” of a random variable n here means that n is white only over some large but finite frequency bandwidth interval $[-B, B]$ for some large $B > 0$. Random num-

bers from the algorithms in [13], [65], [68] act as noise from various probability densities in our simulations. The next sections show how discretized continuous-time systems produced the discrete-time systems we used for computer simulations.

A. Nonlinear Systems With White Gaussian Noise

Consider the dynamical system (73) with initial condition $x(t_0) = x_0$. Here the white Gaussian noise w has zero mean and unit variance so that $n = \sigma w$ has zero mean and variance σ^2 . This system corresponds to the stochastic initial value problem [31]

$$dX = \tilde{f}(t, X) + \sigma(t, X)dW \quad (75)$$

for initial condition $X(t_0) = X_0$. Here $\tilde{f}(t, X) = f(X) + s(t)$, $\sigma(t, X) = \sigma$, and W is the standard Wiener process [31]. We used Euler's method (the Euler-Maruyama scheme) [21], [31], [42] to obtain the discrete form for computer simulation

$$x_{t+1} = x_t + \Delta T (f(x_t) + s_t) + \sigma \sqrt{\Delta T} w_t \quad (76)$$

$$y_t = g(x_t) \quad (77)$$

for $t = 0, 1, 2, \dots$ and initial condition x_0 . The input sample s_t has the value of the signal $s(t\Delta T)$ at time step t . The zero-mean white Gaussian noise sequence $\{w_t\}$ has unit variance $\sigma_w^2 = 1$. The term $\sqrt{\Delta T}$ scales w_t so that $\sqrt{\Delta T}w_t$ conforms with the Wiener increment [31], [42], [59]. The output sample y_t is some transformation g of the system's state x_t .

This simple algorithm gives fairly accurate results for moderate nonlinear systems [31], [42], [51], [59]. Other algorithms may give more accurate numerical solutions of the stochastic

differential equations for more complicated system dynamics [31], [54]. All of our simulations used the Euler's scheme in (76), (77).

The numerical algorithm in [65] generates a sequence of pseudo-random numbers from a Gaussian density with zero mean and unit variance for $\{w_t\}$ in (76). Fig. 3 shows the Gaussian and other densities that have zero mean and a variance of two.

B. Nonlinear Systems With Other Finite-Variance Noise

We next consider a system (73) with finite-variance noise n . Suppose the noise n has variance σ^2 and again apply Euler's method

$$x_{t+1} = x_t + \Delta T (f(x_t) + s_t) + \sigma \sqrt{\Delta T} w_t \quad (78)$$

$$y_{t+1} = g(x_{t+1}). \quad (79)$$

Here the random sequence $\{w_t\}$ has density function $p(w)$ with zero mean and unit variance. The numerical algorithms in [65] generate sequences of random variables for Laplace and uniform density functions. Fig. 3 plots these probability density functions and their realizations with mean zero and variance of two: $E[n] = 0$ and $E[n^2] = 2$.

C. Nonlinear Systems With Alpha-Stable Noise

Figs. 3 and 4 show realizations of the symmetric alpha-stable random variable for several characteristic exponents α . Again we assume that the Euler's method above applies to this class of random variables with infinite variance. Let w be a standard alpha-stable random variable with parameter α and zero location and unit dispersion: $a = 0$ and $\gamma = 1$. Let $\kappa = \gamma^{1/\alpha}$ denote a "scale" factor of a random variable. Then $n = \kappa w$ has zero location and dispersion $\gamma = \kappa^\alpha$. This leads to the Euler's numerical solution

$$x_{t+1} = x_t + \Delta T (f(x_t) + s_t) + \kappa \sqrt{\Delta T} w_t \quad (80)$$

$$y_t = g(x_t). \quad (81)$$

The algorithm in [13], [68] generates a standard alpha-stable random variable w .

V. DERIVATION OF SR LEARNING LAW

We show that a memoryless neuron can use stochastic gradient ascent to learn the SR effect [47], [57]

$$\sigma_{k+1} = \sigma_k + \mu_k \frac{\partial I}{\partial \sigma}. \quad (82)$$

We assume that $P(s)$ does not depend on σ and we use the natural logarithm. Then the learning term $\partial I / \partial \sigma$ has the form

$$\begin{aligned} \frac{\partial I}{\partial \sigma} = \frac{\partial}{\partial \sigma} & \left(- \sum_y P(y) \log P(y) \right. \\ & \left. + \sum_s P(s) \sum_y P(y|s) \log P(y|s) \right) \end{aligned} \quad (83)$$

$$\begin{aligned} = & - \sum_y \left(P(y) \frac{1}{P(y)} \frac{\partial P(y)}{\partial \sigma} + \log P(y) \frac{\partial P(y)}{\partial \sigma} \right) \\ & + \sum_s \sum_y \left(P(s) P(y|s) \frac{1}{P(y|s)} \frac{\partial P(y|s)}{\partial \sigma} \right. \\ & \left. + P(s) \log P(y|s) \frac{\partial P(y|s)}{\partial \sigma} \right) \end{aligned} \quad (84)$$

$$\begin{aligned} = & - \sum_y \left(\frac{\partial P(y)}{\partial \sigma} + \log P(y) \frac{\partial P(y)}{\partial \sigma} \right) \\ & + \sum_s \sum_y \left(P(s) \frac{\partial P(y|s)}{\partial \sigma} \right. \\ & \left. + P(s) \log P(y|s) \frac{\partial P(y|s)}{\partial \sigma} \right). \end{aligned} \quad (85)$$

The sum $\sum_y P(y) = 1$ implies $\sum_y (\partial P(y) / \partial \sigma) = (\partial / \partial \sigma) \sum_y P(y) = 0$. And $\sum_s \sum_y (\partial P(y|s) / \partial \sigma) = 0$ because $\sum_y P(y|s) = 1$. So

$$\frac{\partial I}{\partial \sigma} = - \sum_y \log P(y) \frac{\partial P(y)}{\partial \sigma} + \sum_s \sum_y P(s) \log P(y|s) \frac{\partial P(y|s)}{\partial \sigma}. \quad (86)$$

We estimate the partial derivative with a ratio of time differences and replace the denominator with the signum function to avoid numerical instability

$$\begin{aligned} \frac{\partial P(y)}{\partial \sigma} & \approx \frac{P_k(y) - P_{k-1}(y)}{\sigma_k - \sigma_{k-1}} \\ & \approx \text{sgn}(\sigma_k - \sigma_{k-1}) [P_k(y) - P_{k-1}(y)] \end{aligned} \quad (87)$$

$$\begin{aligned} \frac{\partial P(y|s)}{\partial \sigma} & \approx \frac{P_k(y|s) - P_{k-1}(y|s)}{\sigma_k - \sigma_{k-1}} \\ & \approx \text{sgn}(\sigma_k - \sigma_{k-1}) [P_k(y|s) - P_{k-1}(y|s)] \end{aligned} \quad (88)$$

where $P_k(y)$ is the marginal density function of the output Y at time k and $P_k(y|s)$ is the conditional density function at time k . Then the learning term becomes

$$\begin{aligned} \frac{\partial I}{\partial \sigma} & \approx \text{sgn}(\sigma_k - \sigma_{k-1}) \left(- \sum_y [P_k(y) - P_{k-1}(y)] \log P_k(y) \right. \\ & \left. + \sum_s \sum_y P_k(s) [P_k(y|s) - P_{k-1}(y|s)] \log P_k(y|s) \right). \end{aligned} \quad (89)$$

Our previous work [47], [57] on adaptive SR found through statistical tests that the random learning term $\partial P / \partial \sigma$ had an approximately Cauchy distribution for the spectral signal-to-noise and cross-correlation performance measures P . These frequent and energetic Cauchy impulse spikes destabilized the stochastic learning process. So we "robustified" the learning term with the standard Cauchy error suppressor $\phi(z_k) = 2z_k / (1 + z_k^2)$ [37], [41]. This included the threshold neuron given a periodic input sequence.

But detailed simulations revealed a special pattern in the case of mutual information: The density $P_k(y)$ tends to stay close to the past density $P_{k-1}(y)$ if the values of σ_k and σ_{k-1} are close. This causes the learning paths σ_k to converge quickly near the

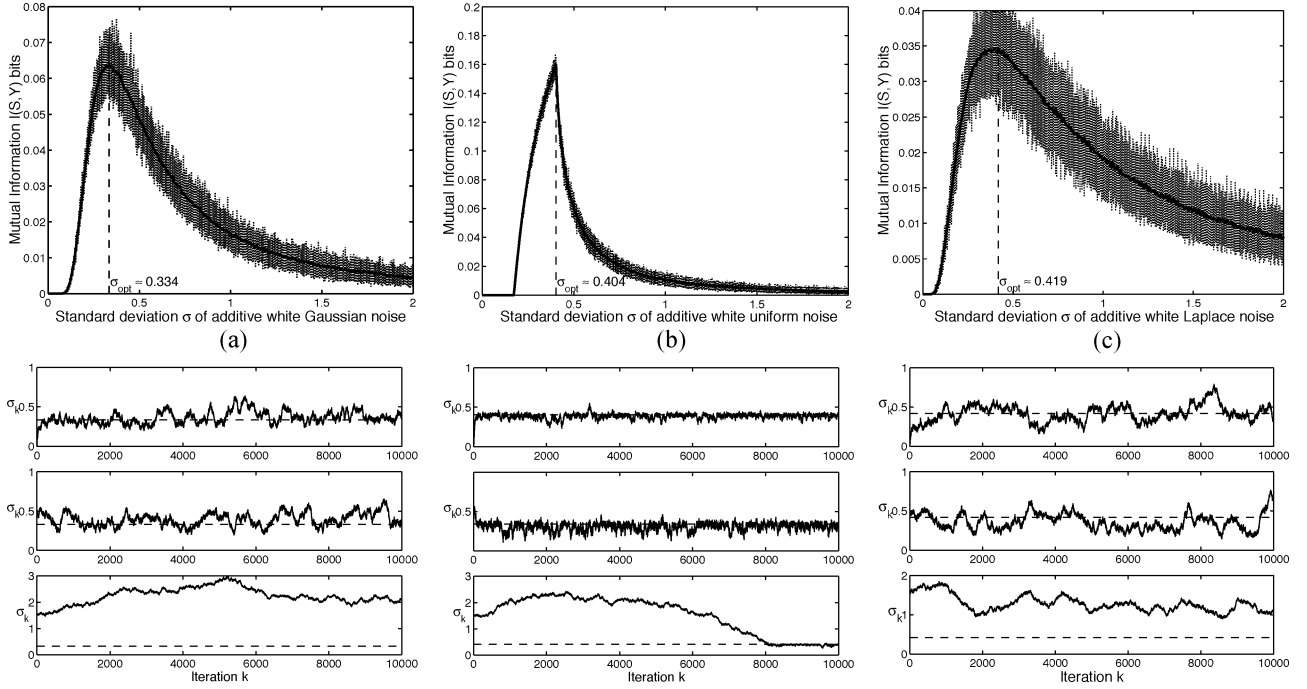


Fig. 7. Finite-variance noise cases: adaptive stochastic resonance for the noisy threshold neuron (6) with bipolar input signal s_t , amplitude $A = 0.2$, and threshold $\theta = 0.5$. The additive noise are (a) Gaussian, (b) uniform, and (c) Laplace. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence if the initial condition σ_0 is close to the optimal noise level σ_{opt} . Distant initial conditions may lead to divergence as the third learning path in (a) shows. The constant learning rates are $\mu_k = 0.01$ for Gaussian and uniform noise and $\mu_k = 0.02$ for Laplace noise.

initial conditions. So we can replace the learning term $\partial I / \partial \sigma$ with its sign $\text{sgn}(\partial I / \partial \sigma)$ and the learning law simplifies to

$$\sigma_{k+1} = \sigma_k + \mu_k \text{sgn} \left(\frac{\partial I}{\partial \sigma} \right). \quad (90)$$

The signum is a simple robustifier and formally consistent with a two-sided Laplacian distribution [37].

VI. SIMULATION RESULTS

We tested the robust learning law in (90) with the approximation of the learning term in (89). We needed to estimate the marginal and conditional probability densities $P_k(y)$, $P_k(s)$, and $P_k(y|s)$ at each iteration k . So at each k we collected 1000 input-output samples $\{s_t, y_t\}$ and used them to estimate the densities with histograms for the threshold system. We used 500 of the input-output symbols to estimate the probability densities for the continuous neuron model. We chose the neurons' and signals' parameters below to demonstrate the algorithm. Other parameters gave similar results.

A. Noisy Threshold Neuron

The threshold neuron had a fixed threshold $\theta = 0.5$. The bipolar input Bernoulli signal has probability $P_S(-A) = P_S(A) = 1/2$ where the amplitude A varied from $A = 0.1$ to $A = 0.4$ (subthreshold inputs). We tried several noise densities that included the Gaussian, uniform, Laplace, and the impulsive alpha-stable densities that include the Cauchy density. All noise densities had zero mean (zero location for Cauchy). We tried to learn the optimal standard deviation σ_{opt} (or optimal dispersion γ_{opt} for alpha-stable noise). We used constant learning rates

$\mu_k = 0.01$ for Gaussian and uniform noise, $\mu_k = 0.02$ for Laplace and Cauchy noise, and $\mu_k = 0.02$ for alpha-stable noise with $\alpha = 1.9$ and $\alpha = 1.5$. We started the learning from several initial conditions with different noise seeds.

Figs. 7–9 show the adapted SR profiles and the σ_{opt} learning paths for different noise types. The learning paths converged to the optimal standard deviation σ_{opt} (or dispersion γ_{opt}) if the initial value was near σ_{opt} . The learning paths tended to stay nearer the optimal values for larger input amplitudes.

B. Noisy Continuous Neuron

We used the discrete model in Section IV for simulations. We used $dt = 0.01$ s and let each input symbol stay for 50 s. So for each input symbol we presented the corresponding “spikes” (plus noise) 5000 times to the neuron. And we collected 5000 discrete-time output “spikes” and averaged them to get the output symbol. This procedure applied to all types of signal functions and for all types of noise.

1) *Continuous Neurons with Hyperbolic Tangent Signal Function:* We tested the continuous neuron model with hyperbolic tangent signal function with several noise densities such as the Gaussian, uniform, Laplace, and alpha-stable (which included the Cauchy density). All noise densities had zero mean (zero location for Cauchy). The bipolar input Bernoulli signal had success probability $P_S(-A) = P_S(A) = 1/2$ where the amplitude A varied from $A = 0.1$ to $A = 0.4$ (subthreshold inputs). We used constant learning rates $\mu_k = 0.03$ for Gaussian, uniform, and Laplace noise. We used the smaller learning rates $\mu_k = 0.02$ for alpha-stable noise with $\alpha = 1.9$ and $\alpha = 1.5$ and used the still smaller learning rate $\mu_k = 0.005$ for Cauchy noise. We started the learning from several initial

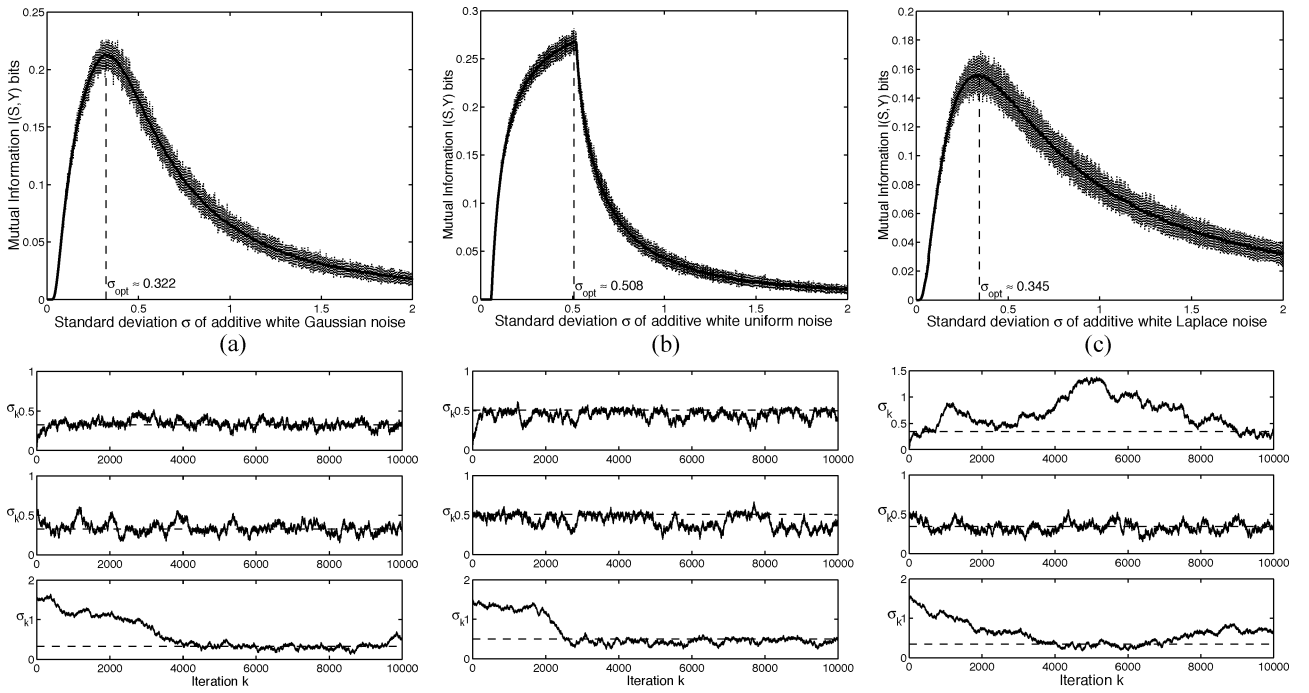


Fig. 8. Finite-variance noise cases: adaptive stochastic resonance for the noisy threshold neuron (6) with bipolar input signal s_t , amplitude $A = 0.4$, and threshold $\theta = 0.5$. The additive noise are (a) Gaussian, (b) uniform, and (c) Laplace. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each noise density. The constant learning rates are $\mu_k = 0.01$ for Gaussian and uniform noise and $\mu_k = 0.02$ for Laplace noise.

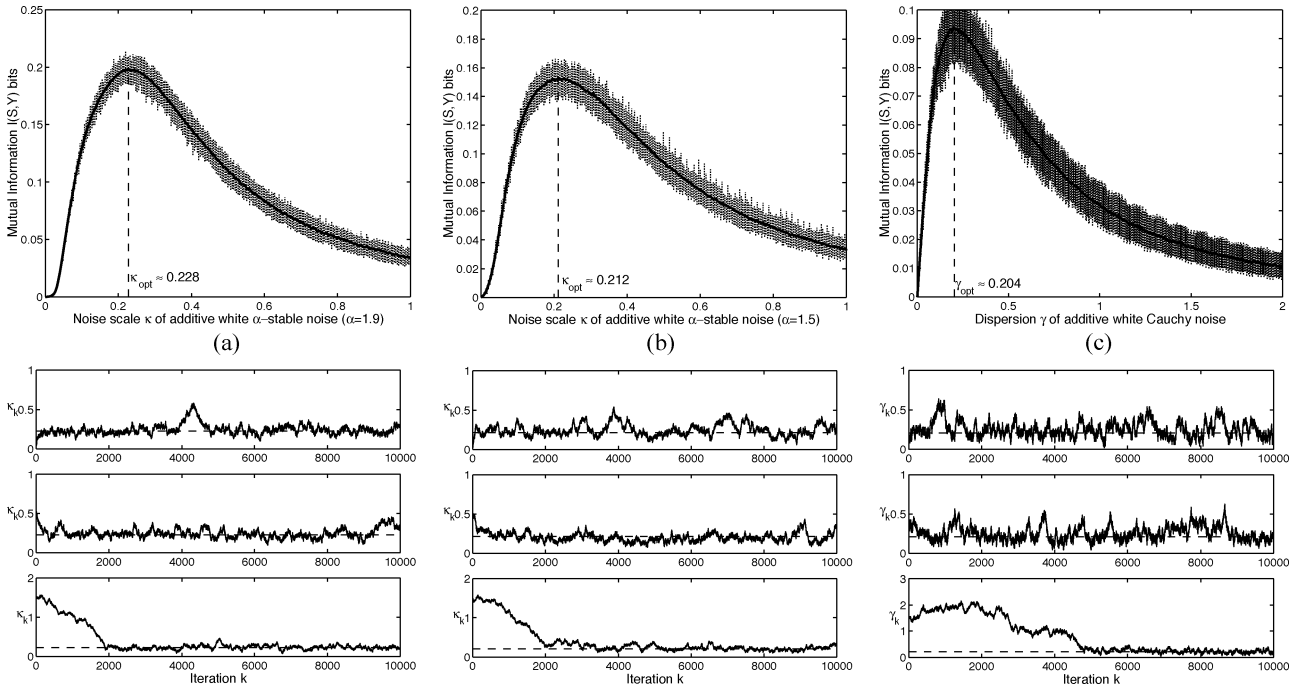


Fig. 9. Impulsive noise cases: Adaptive stochastic resonance for the noisy threshold neuron (6) with bipolar input signal s_t , amplitude $A = 0.4$, and threshold $\theta = 0.5$. The additive noise are α -stable distributed with the parameter (a) $\alpha = 1.9$, (b) $\alpha = 1.5$, and (c) $\alpha = 1$ or Cauchy density. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence of the noise scale κ_k to the noise optimum κ_{opt} for each noise density. The corresponding dispersions are $\gamma = \kappa^\alpha$ for each α -stable noise. The constant learning rates are $\mu_k = 0.01$ for $\alpha = 1.9$ and $\alpha = 1.5$ noise and $\mu_k = 0.02$ for Cauchy noise $\alpha = 1$.

conditions with different noise seeds. Figs. 10–12 show the adapted SR profiles and the σ_{opt} learning paths for different noise types. The learning paths converged near the optimal standard deviation σ_{opt} (or dispersion γ_{opt}) if the initial value was near σ_{opt} .

2) *Continuous Neurons with Linear-Threshold, Exponential, and Gaussian (Radial Basis) Signal Functions:* We further tested the continuous neuron model with linear-threshold, exponential, and Gaussian (radial basis) signal functions in Gaussian noise to show the generality of the SR effect. We

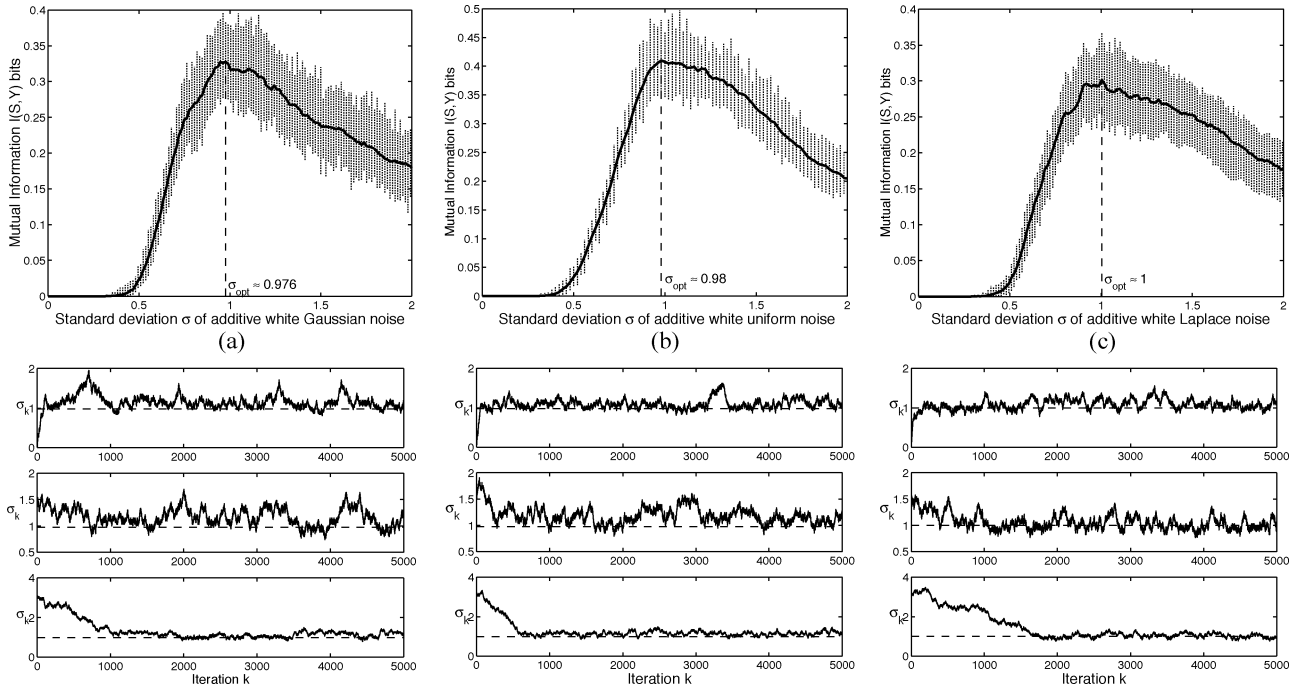


Fig. 10. Finite-variance noise cases: Adaptive stochastic resonance for the noisy continuous neuron (7) with hyperbolic signal function (9) and bipolar input signal s_t with amplitude $A = 0.2$. The additive noise are (a) Gaussian, (b) uniform, and (c) Laplace. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each noise density. The constant learning rates are $\mu_k = 0.03$ for all cases.

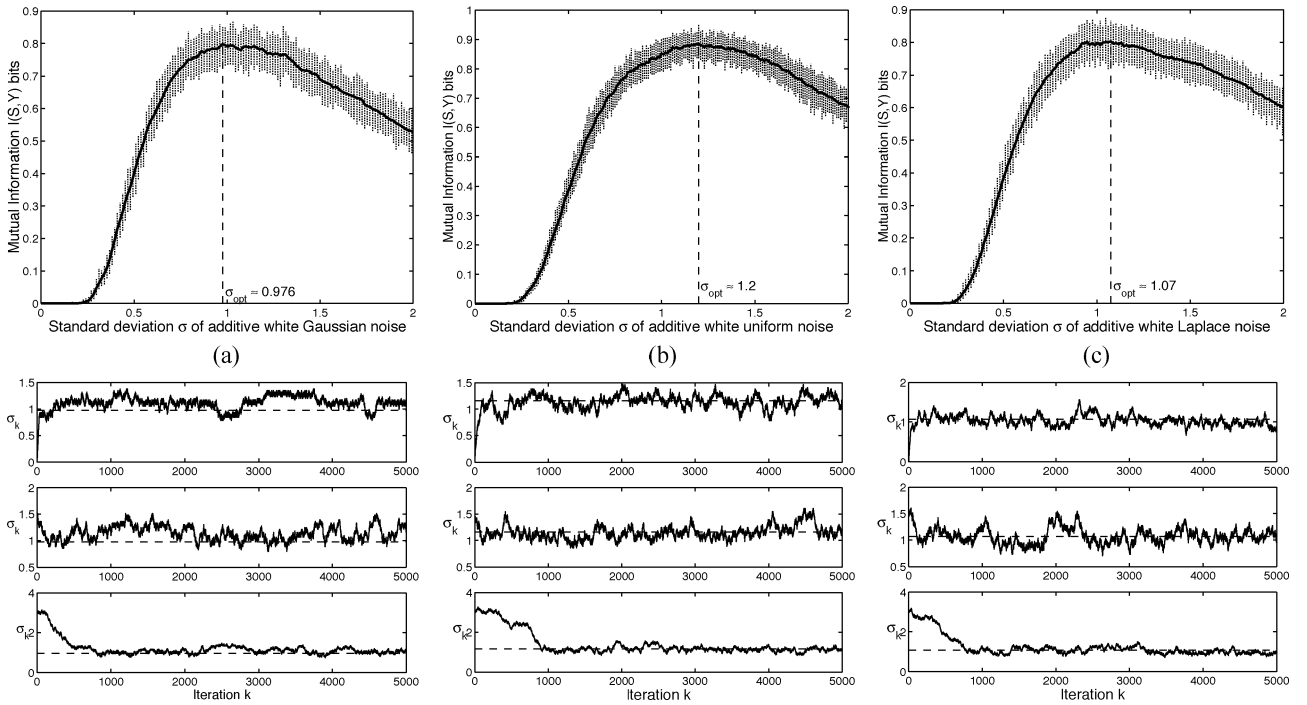


Fig. 11. Finite-variance noise cases: Adaptive stochastic resonance for the noisy continuous neuron (7) with hyperbolic signal function (9) and bipolar input signal s_t with amplitude $A = 0.4$. The additive noise are (a) Gaussian, (b) uniform, and (c) Laplace. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each noise density. The constant learning rates are $\mu_k = 0.03$ for all cases.

used the same bipolar input Bernoulli signal with success probability $P_S(-A) = P_S(A) = (1/2)$ where the amplitude is $A = 0.4$ for the linear-threshold and Gaussian signal functions and $A = 0.6$ for the exponential signal function. The input amplitudes were “subthreshold” for the neuron models

with these signal functions. We used constant learning rates $\mu_k = 0.02$ for the exponential and Gaussian signal functions and $\mu_k = 0.05$ for the linear-threshold signal functions. We started the learning from several initial conditions with different noise seeds. Fig. 13 shows the adapted SR profiles and the σ_{opt}

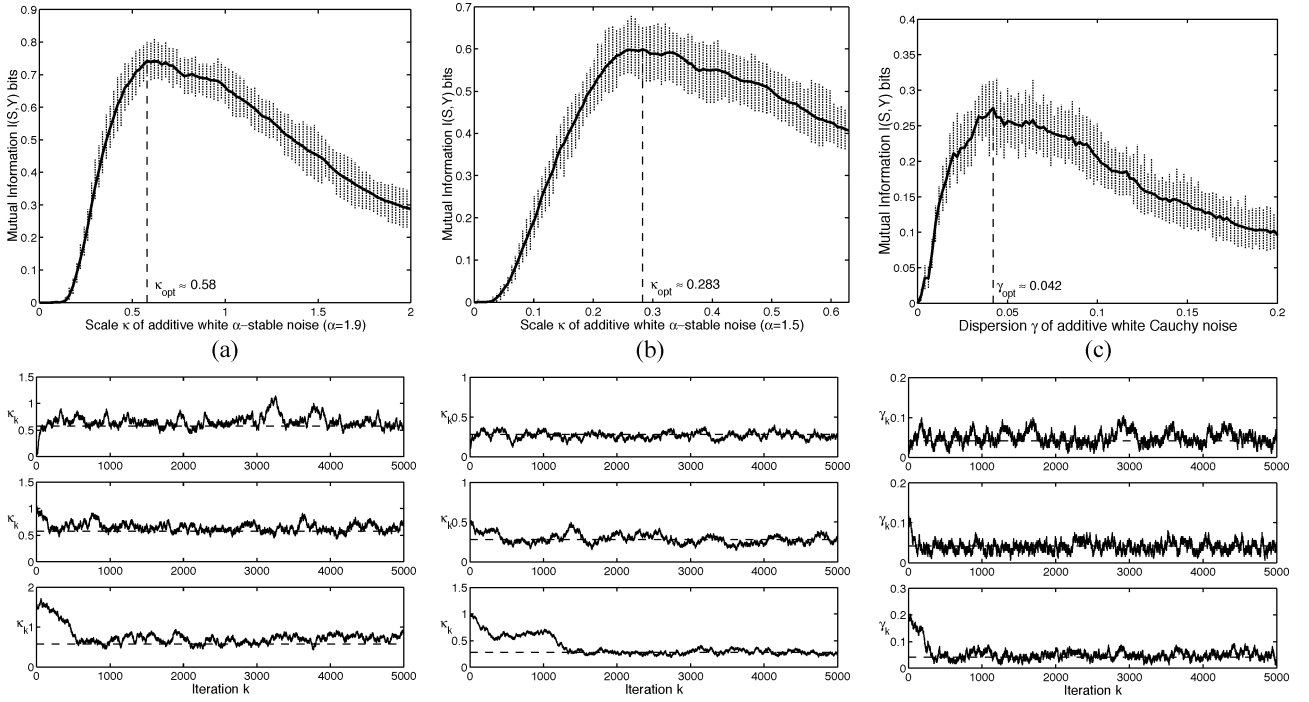


Fig. 12. Impulsive noise cases: Adaptive stochastic resonance for the noisy continuous neuron (7) with hyperbolic signal function (9) and bipolar input signal s_t with amplitude $A = 0.4$. The additive noise are α -stable distributed with the parameter (a) $\alpha = 1.9$, (b) $\alpha = 1.5$, and (c) $\alpha = 1$ or Cauchy density. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each noise density. The constant learning rates are $\mu_k = 0.02$ for $\alpha = 1.9$, $\mu_k = 0.01$ for $\alpha = 1.5$, and $\mu_k = 0.005$ for $\alpha = 1$.

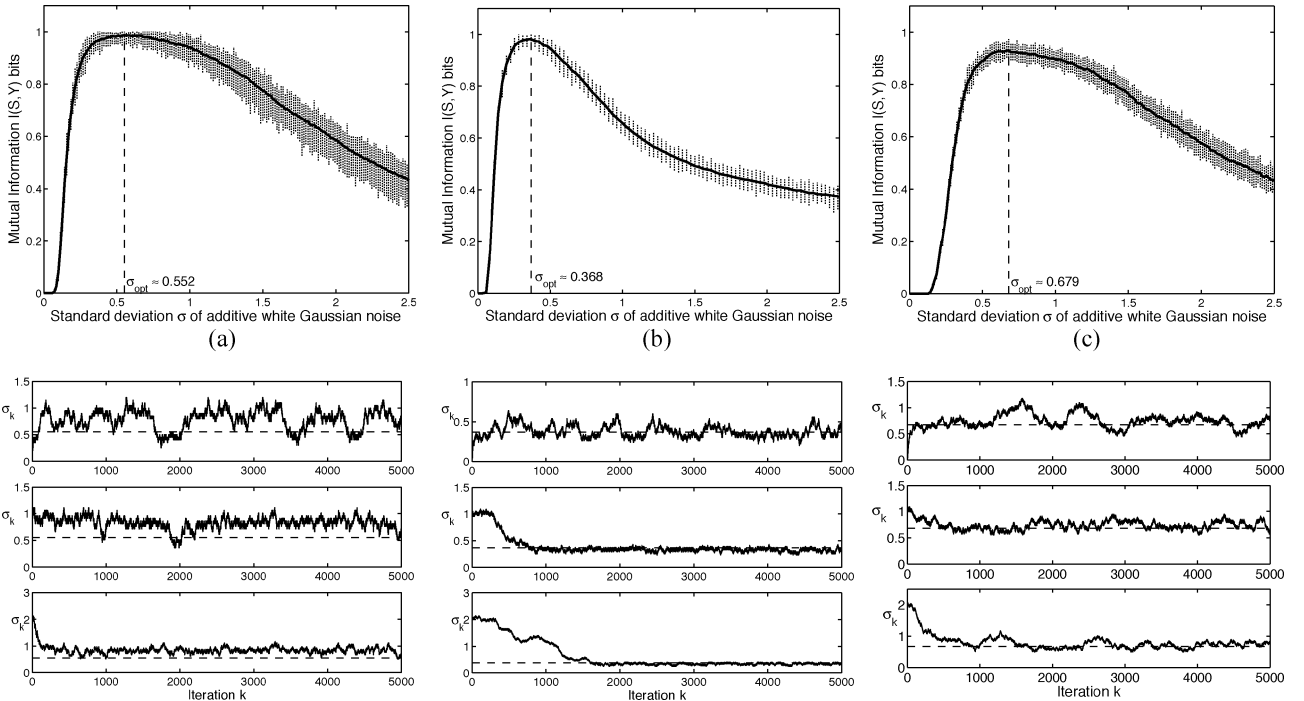


Fig. 13. Adaptive stochastic resonance for continuous neurons with linear-threshold, exponential, and Gaussian (radial basis) signal functions. The bipolar input signal s_t has amplitude $A = 0.4$ for the linear-threshold and Gaussian signal functions and $A = 0.6$ for the exponential signal function. The additive noise n_t is Gaussian. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each case of signal functions: (a) linear-threshold, (b) exponential, and (c) Gaussian. The constant learning rates are $\mu_k = 0.05$ for the linear-threshold signal function and $\mu_k = 0.02$ for the exponential and Gaussian signal functions.

learning paths for the three other signal functions. The learning paths converged near the optimal standard deviation σ_{opt} if the initial value was near σ_{opt} .

VII. CONCLUSION

Threshold neurons exhibit stochastic resonance—they increase their throughput mutual information when faint input

noise increases in intensity. A theorem shows that this holds for almost all noise densities. Such noise-based information maximization is consistent with Linsker's principle of information maximization in neural networks [49], [50]. Closed-form noise densities allow us to derive the exact dependence of mutual information on noise dispersion and to observe the nonlinear relationships between the optimal noise level and the magnitude of the input signal amplitude. Extensive simulations confirmed this entropic SR effect for noisy threshold (memoryless) neurons and for simple continuous neurons.

A simple robust stochastic learning law can find the entropically optimal noise level for both threshold and continuous neurons that process noisy bipolar input signals. This result holds for many types of finite-variance and infinite-variance (impulsive) noise. These noise types can model energetic disturbances that range from thermal jitter to unmodeled environmental effects to the random crosstalk of neurons in large neural networks. This robust finding supports the implicit SR conjecture that biological neurons [11], [18], [19], [23], [48], [58], [64], [69] have evolved over genetic eons to exploit the noise energy freely available in their local environment.

REFERENCES

- [1] V. Akgiray and C. G. Lamoureux, "Estimation of stable-law parameters: a comparative study," *J. Bus. Econ. Statist.*, vol. 7, pp. 85–93, Jan. 1989.
- [2] S. Amari, "Neural theory of association and concept formation," *Biol. Cybern.*, vol. 26, pp. 175–185, 1977.
- [3] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, "Stochastic resonance in climatic change," *Tellus*, vol. 34, pp. 10–16, 1982.
- [4] R. Benzi, A. Sutera, and A. Vulpiani, "The mechanism of stochastic resonance," *J. Phys. A, Math. Gen.*, vol. 14, pp. L453–L457, 1981.
- [5] H. Bergstrom, "On some expansions of stable distribution functions," *Arkiv för Matematik*, vol. 2, pp. 375–378, 1952.
- [6] L. Breiman, *Probability*. Reading, MA: Addison-Wesley, 1968.
- [7] P. Bryant, K. Wiesenfeld, and B. McNamara, "The nonlinear effects of noise on parametric amplification: an analysis of noise rise in Josephson junctions and other systems," *J. Appl. Phys.*, vol. 62, no. 7, pp. 2898–2913, Oct. 1987.
- [8] A. R. Bulsara, T. C. Elston, C. R. Doering, S. B. Lowen, and K. Lindenberg, "Cooperative behavior in periodically driven noisy integrate-fire models of neuronal dynamics," *Phys. Rev. E*, vol. 53, no. 4, pp. 3958–3969, Apr. 1996.
- [9] A. R. Bulsara and L. Gammaitoni, "Tuning in to noise," *Phys. Today*, pp. 39–45, Mar. 1996.
- [10] A. R. Bulsara, E. W. Jacobs, T. Zhou, F. Moss, and L. Kiss, "Stochastic resonance in a single neuron model: theory and analog simulation," *J. Theoret. Biol.*, vol. 152, pp. 531–555, 1991.
- [11] A. R. Bulsara, A. J. Maren, and G. Schmera, "Single effective neuron: dendritic coupling effects and stochastic resonance," *Biolog. Cybern.*, vol. 70, pp. 145–156, 1993.
- [12] A. R. Bulsara and A. Zador, "Threshold detection of wideband signals: a noise-induced maximum in the mutual information," *Phys. Rev. E*, vol. 54, no. 3, pp. R2185–R2188, Sept. 1996.
- [13] J. M. Chambers, C. L. Mallows, and B. W. Stuck, "A method for simulating stable random variables," *J. Amer. Statist. Assoc.*, vol. 71, no. 354, pp. 340–344, 1976.
- [14] K. L. Chung and R. J. Williams, *Introduction to Stochastic Integration*, 2nd ed. Germany: Birkhäuser, 1990.
- [15] M. A. Cohen and S. Grossberg, "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, pp. 815–826, Sept. 1983.
- [16] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, "A periodic stochastic resonance," *Phys. Rev. E*, vol. 54, no. 5, pp. 5575–5584, Nov. 1996.
- [17] J. J. Collins, C. C. Chow, and T. T. Imhoff, "Stochastic resonance without tuning," *Nature*, vol. 376, pp. 236–238, July 1995.
- [18] J. J. Collins, T. T. Imhoff, and P. Grigg, "Noise-enhanced information transmission in Rat SA1 cutaneous mechanoreceptors via aperiodic stochastic resonance," *J. Neurophysiol.*, vol. 76, no. 1, pp. 642–645, July 1996.
- [19] —, "Noise-enhanced tactile sensation," *Nature*, vol. 383, p. 770, Oct. 1996.
- [20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [21] G. Dahlquist and Å. Björck, *Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
- [22] G. Deco and B. Schürmann, "Stochastic resonance in the mutual information between input and output spike trains of noisy central neurons," *Phys. D*, vol. 117, pp. 276–282, 1998.
- [23] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, "Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance," *Nature*, vol. 365, pp. 337–340, Sept. 1993.
- [24] R. Durrett, *Stochastic Calculus: A Practical Introduction*. CRC Press, 1996.
- [25] M. I. Dykman, T. Horita, and J. Ross, "Statistical distribution and stochastic resonance in a periodically driven chemical system," *J. Chem. Phys.*, vol. 103, no. 3, pp. 966–972, July 1995.
- [26] J.-P. Eckmann and L. E. Thomas, "Remarks on stochastic resonances," *J. Phys. A: Math. General*, vol. 15, pp. L261–L266, 1982.
- [27] S. Fauve and F. Heslot, "Stochastic resonance in a bistable system," *Phys. Lett. A*, vol. 97, no. 1, 2, pp. 5–7, Aug. 1983.
- [28] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York: Wiley, 1966, vol. II.
- [29] L. Gammaitoni, "Stochastic resonance and the dithering effect in threshold physical systems," *Phys. Rev. E*, vol. 52, no. 5, pp. 4691–4698, Nov. 1995.
- [30] —, "Stochastic resonance in multi-threshold systems," *Phys. Lett. A*, vol. 208, pp. 315–322, Dec. 1995.
- [31] T. C. Gard, *Introduction to Stochastic Differential Equations*. New York: Marcel Dekker, 1988.
- [32] J. Glanz, "Mastering the nonlinear brain," *Science*, vol. 277, pp. 1758–1760, Sept. 1997.
- [33] X. Godivier and F. Chapeau-Blondeau, "Stochastic resonance in the information capacity of a nonlinear dynamic system," *Int. J. Bifurc. Chaos*, vol. 8, no. 3, pp. 581–589, 1998.
- [34] M. Grifoni, M. Sasseti, P. Hänggi, and U. Weiss, "Cooperative effects in the nonlinearly driven spin-boson system," *Phys. Rev. E*, vol. 52, no. 4, pp. 3596–3607, Oct. 1995.
- [35] M. Grigoriu, *Applied Non-Gaussian Processes*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [36] A. Hibbs, E. W. Jacobs, J. Bekkedahl, A. R. Bulsara, and F. Moss, "Signal enhancement in a r.f. SQUID using stochastic resonance," *Il Nuovo Cimento*, vol. 17 D, no. 7–8, pp. 811–817, Luglio–Agosto 1995.
- [37] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*, 5th ed. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [38] N. Hohn and A. N. Burkitt, "Enhanced stochastic resonance in threshold detectors," in *Proc. 2001 IEEE Int. Joint Conf. on Neural Networks (IJCNN '01)*, 2001, pp. 644–647.
- [39] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," in *Proc. National Acad. Sci.*, vol. 79, 1982, pp. 2554–2558.
- [40] —, "Neural networks with graded response have collective computational properties like those of two-state neurons," in *Proc. National Acad. Science*, vol. 81, 1984, pp. 3088–3092.
- [41] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
- [42] M. E. Inchiosa and A. R. Bulsara, "Nonlinear dynamic elements with noisy sinusoidal forcing: Enhancing response via nonlinear coupling," *Phys. Rev. E*, vol. 52, no. 1, pp. 327–339, July 1995.
- [43] M. E. Inchiosa, J. W. C. Robinson, and A. R. Bulsara, "Information-theoretic stochastic resonance in noise-floor limited systems: The case for adding noise," *Phys. Rev. Lett.*, vol. 85, pp. 3369–3372, Oct. 2000.
- [44] P. Jung, "Stochastic resonance and optimal design of threshold detectors," *Phys. Lett. A*, vol. 207, pp. 93–104, Oct. 1995.
- [45] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
- [46] —, *Fuzzy Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [47] B. Kosko and S. Mitaïm, "Robust stochastic resonance: Signal detection and adaptation in impulsive noise," *Phys. Rev. E*, vol. 64, no. 051 110, October 2001.
- [48] J. E. Levin and J. P. Miller, "Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance," *Nature*, vol. 380, pp. 165–168, Mar. 1996.
- [49] R. Linsker, "Self-organization in a perceptual network," *Computer*, vol. 21, no. 3, pp. 105–117, Mar. 1988.

- [50] —, "A local learning rule that enables information maximization for arbitrary input distributions," *Neural Comput.*, vol. 9, no. 8, pp. 1661–1665, Nov. 1997.
- [51] A. Longtin, "Autonomous stochastic resonance in bursting neurons," *Phys. Rev. E*, vol. 55, no. 1, pp. 868–876, Jan. 1997.
- [52] J. Maddox, "Toward the brain-computer's code?," *Nature*, vol. 352, p. 469, Aug. 1991.
- [53] —, "Bringing more order out of noisiness," *Nature*, vol. 369, p. 271, May 1994.
- [54] R. Mannella and V. Palleschi, "Fast and precise algorithm for computer simulation of stochastic differential equations," *Phys. Rev. A*, vol. 40, no. 6, pp. 3381–3386, Sept. 1989.
- [55] R. J. Marks II, B. Thompson, M. A. El-Sharkawi, W. L. J. Fox, and R. T. Miyamoto, "Stochastic resonance of a threshold detector: image visualization and explanation," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS 2002)*, vol. 4, 2002, pp. IV-521–IV-523.
- [56] B. McNamara, K. Wiesenfeld, and R. Roy, "Observation of stochastic resonance in a ring laser," *Phys. Rev. Lett.*, vol. 60, no. 25, pp. 2626–2629, June 1988.
- [57] S. Mitaim and B. Kosko, "Adaptive stochastic resonance," *Proc. IEEE: Special Issue on Intelligent Signal Processing*, vol. 86, pp. 2152–2183, Nov. 1998.
- [58] R. P. Morse and E. F. Evans, "Enhancement of vowel coding for cochlear implants by addition of noise," *Nature Medicine*, vol. 2, no. 8, pp. 928–932, Aug. 1996.
- [59] F. Moss and P. V. E. McClintock, Eds., *Noise in Nonlinear Dynamical Systems*. Cambridge, UK: Cambridge Univ. Press, 1989, vol. I–III.
- [60] D. C. Munson, "A note on Lena," *IEEE Trans. Image Processing*, vol. 5, p. 3, Jan. 1996.
- [61] C. Nicolis, "Stochastic aspects of climatic transitions-response to a periodic forcing," *Tellus*, vol. 34, pp. 1–9, 1982.
- [62] C. L. Nikias and M. Shao, *Signal Processing With Alpha-Stable Distributions and Applications*. New York: Wiley, 1995.
- [63] X. Pei, K. Bachmann, and F. Moss, "The detection threshold, noise and stochastic resonance in the Fitzhugh-Nagumo neuron model," *Phys. Lett. A*, vol. 206, pp. 61–65, Oct. 1995.
- [64] X. Pei, L. Wilkens, and F. Moss, "Light enhances hydrodynamic signaling in the multimodal caudal photoreceptor interneurons of the crayfish," *J. Neurophysiol.*, vol. 76, no. 5, pp. 3002–3011, Nov. 1996.
- [65] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, UK: Cambridge Univ. Press, 1993.
- [66] D. F. Russell, L. A. Willkens, and F. Moss, "Use of behavioral stochastic resonance by paddle fish for feeding," *Nature*, vol. 402, pp. 291–294, Nov. 1999.
- [67] N. G. Stocks, "Information transmission in parallel threshold arrays," *Phys. Rev. E*, vol. 63, no. 041 114, 2001.
- [68] P. Tsakalides and C. L. Nikias, "The Robust Covariation-Based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive noise environments," *IEEE Trans. Signal Processing*, vol. 44, pp. 1623–1633, July 1996.
- [69] M. Usher and M. Feingold, "Stochastic resonance in the speed of memory retrieval," *Biolog. Cybern.*, vol. 83, pp. L11–L16, 2000.
- [70] R. A. Wannamaker, S. P. Lipshitz, and J. Vanderkooy, "Stochastic resonance as dithering," *Phys. Rev. E*, vol. 61, no. 1, pp. 233–236, Jan. 2000.
- [71] K. Wiesenfeld and F. Moss, "Stochastic resonance and the benefits of noise: From ice ages to crayfish and SQUIDS," *Nature*, vol. 373, pp. 33–36, Jan. 1995.

Sanya Mitaim received the B.Eng. degree in control engineering from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. He received the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles.

He is an Assistant Professor with the Department of Electrical Engineering, Faculty of Engineering, Thammasat University, Pathumthani, Thailand. His research interests include neural and fuzzy techniques in nonlinear signal processing and noise processing.

Bart Kosko received the degrees in philosophy, economics, applied mathematics, and electrical engineering.

He is a Professor of electrical engineering with the University of Southern California, Los Angeles. He is the author of the textbooks *Neural Networks and Fuzzy Systems* (Prentice-Hall: Englewood Cliffs, NJ, 1992), and *Fuzzy Engineering* (Prentice-Hall: Englewood Cliffs, NJ, 1997); the novel *Nanotime* (Avon: New York, 1997); the trade books *Fuzzy Thinking* (Hyperion: New York, 1993), and *Heaven in a Chip* (Random House: New York, 2000). He edited *Neural Networks for Signal Processing* (Prentice-Hall: Englewood Cliffs, NJ, 1992) and coedited (with Simon Haykin) *Intelligent Signal Processing* (IEEE/Wiley: New York, 2001).

Dr. Kosko is an elected governor of the International Neural Network Society and has chaired and co-chaired many neural and fuzzy system conferences.

Almost All Noise Types Can Improve the Mutual Information of Threshold Neurons That Detect Subthreshold Signals

Bart Kosko

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

Sanya Mitaim

Department of Electrical Engineering
Faculty of Engineering
Thammasat University
Pathumthani 12121, Thailand

Abstract—Two new theorems show that small amounts of noise can increase the mutual information of threshold neurons that 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. The second theorem shows that this effect holds for all infinite-variance noise types in the broad class 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.

I. STOCHASTIC RESONANCE IN THRESHOLD NEURONS

Small amounts of noise can improve the performance of threshold neurons or of neurons with steep signal functions. Several researchers have found some form of this “stochastic resonance” (SR) effect [5], [7], [8], [14], [16], [18], [19], [22], [23], [33], [36] when either mutual information or input-output correlation 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.

We prove that the mutual-information form of the SR effect holds for *any* finite-variance noise type that obeys a simple mean condition and for *any* infinite-variance noise type from the broad family of stable distributions. All signals are subthreshold. Stable probability densities have finite dispersions but infinite variances and higher-order moments. The dispersion controls the width of the bell curve for symmetric stable densities (see Figure 2). Figure 1 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 stochastic resonance. The second 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. Figure 3 shows the SR effects for four types of non-Cauchy noise with infinite variance. The noise stream itself is a local form of free energy. So the combined results support Linsker’s hypothesis [28], [29] that neurons have evolved to maximize the information content of their local environment.

II. THRESHOLD NEURONS AND SHANNON’S MUTUAL INFORMATION

We use the standard discrete-time threshold neuron model [5], [14], [20], [22], [24], [26]

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

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 $y = 0$. 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

$$\begin{aligned} P_{Y|S}(0|0) &= \Pr\{s + n < \theta\} \Big|_{s=-A} \\ &= \Pr\{n < \theta + A\} = \int_{-\infty}^{\theta+A} p(n)dn \end{aligned} \quad (2)$$

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

$$\begin{aligned} P_{Y|S}(0|1) &= \Pr\{s + n < \theta\} \Big|_{s=A} \\ &= \Pr\{n < \theta - A\} = \int_{-\infty}^{\theta-A} p(n)dn \end{aligned} \quad (4)$$

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

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 [5] and Gaussian inputs [36]. 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 [9] to measure the noise enhancement or stochastic resonance (SR) effect [5], [10], [16], [21], [36]. The discrete Shannon mutual information of the input S and output Y is the difference

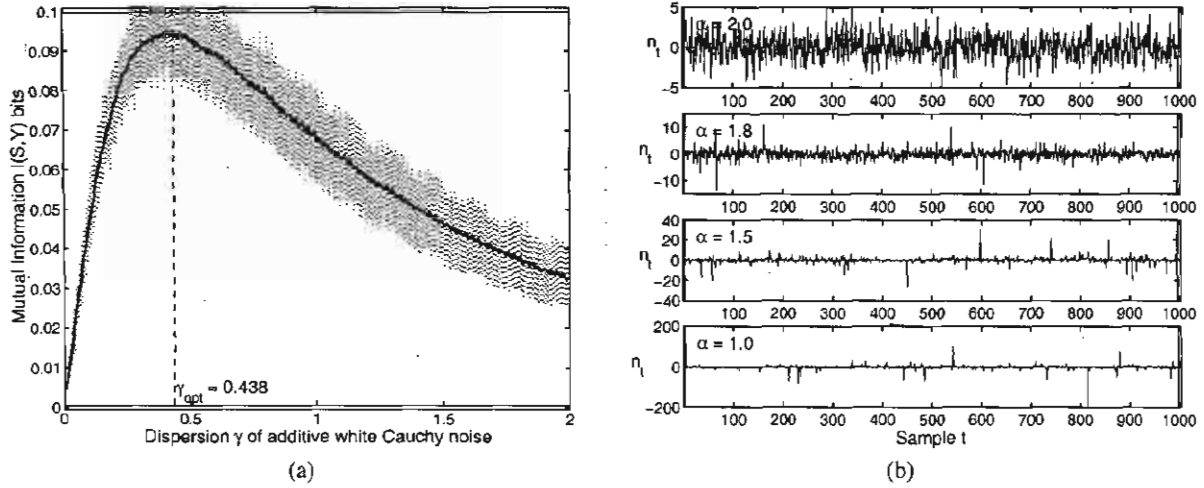


Fig. 1. Stochastic resonance with infinite-variance Cauchy noise. (a) The graph shows the smoothed input-output mutual information of a threshold neuron as a function of the dispersion of additive white Cauchy noise n_t . The dispersion γ controls the width of the Cauchy bell curve. The vertical dashed 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 $\gamma_{opt} \approx 0.438$ and thus shows the SR effect. The noisy signal-forced threshold neuron has the form (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 ($\alpha = 0$) 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 [6], [37] generated these realizations.

between the output unconditional entropy $H(Y)$ and the output conditional entropy $H(Y|X)$:

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

$$= - \sum_y P_Y(y) \log P_Y(y) \quad (8)$$

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

$$= - \sum_y P_Y(y) \log P_Y(y) \quad (9)$$

$$+ \sum_s P(s) \sum_y P(y|s) \log P(y|s) \quad (9)$$

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

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

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

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 [9] 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.

III. 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 sub-threshold 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

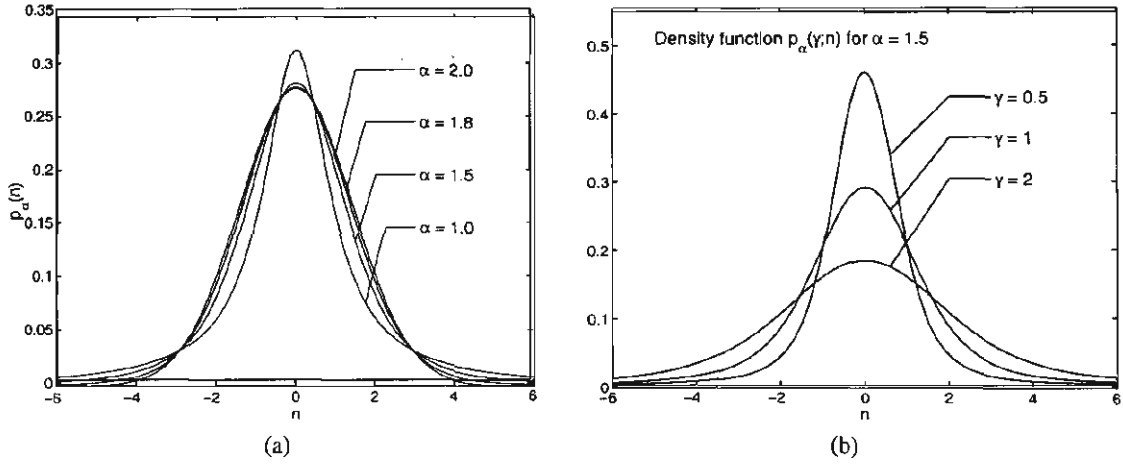


Fig. 2. Samples of standard symmetric ($\beta = 0$) alpha-stable probability densities. (a) Density functions with zero location ($a \approx 0$) and unit dispersion ($\gamma = 1$) for $\alpha = 2, 1.8, 1.5$, and 1 . The densities are bell curves that have thicker tails as α decreases and thus that model 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. (d) Density functions for $\alpha = 1.5$ with dispersions $\gamma = 0.5, 1$, and 2 .

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 $I(S, Y) \rightarrow 0$ as $\sigma \rightarrow 0$.

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: $I(\sigma) \rightarrow 0$ as $\sigma \rightarrow 0$. Recall that $I(S, Y) = 0$ if and only if S and Y are statistically independent [9]. 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 $\sigma \rightarrow 0$ for some signal symbols $s \in S$ and $y \in Y$. The two-symbol alphabet set 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 $\sigma \rightarrow 0$. 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 (2) and (4) imply that

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) \quad (17)$$

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

$$= \int_{\theta-A}^{\theta+A} p(n)dn \quad (19)$$

Similarly for $y = '1'$:

$$P_{Y|S}(1|0) = \int_{\theta+A}^{\infty} p(n)dn \quad (20)$$

$$P_{Y|S}(1|1) = \int_{\theta-A}^{\infty} p(n)dn \quad (21)$$

Then

$$P_{Y|S}(1|0) - P_{Y|S}(1|1) = - \int_{\theta-A}^{\theta+A} p(n)dn \quad (22)$$

The result now follows if

$$\int_{\theta-A}^{\theta+A} p(n)dn \rightarrow 0 \quad \text{as } \sigma \rightarrow 0 \quad (23)$$

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 $\varepsilon = \frac{1}{2}d(\theta - A, m) = \frac{1}{2}(\theta - A - m) > 0$. So $\theta - A - \varepsilon = \theta - A - \varepsilon + m - m = m + (\theta - A - m) - \varepsilon = m + 2\varepsilon - \varepsilon = m + \varepsilon$. Then

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) = \int_{\theta-A}^{\theta+A} p(n)dn \quad (24)$$

$$\leq \int_{\theta-A}^{\infty} p(n)dn \quad (25)$$

$$\leq \int_{\theta-A-\varepsilon}^{\infty} p(n)dn \quad (26)$$

$$= \int_{m+\varepsilon}^{\infty} p(n)dn \quad (27)$$

$$= \Pr\{n \geq m + \varepsilon\} \quad (28)$$

$$= \Pr\{n - m \geq \varepsilon\} \quad (29)$$

$$\leq \Pr\{|n - m| \geq \varepsilon\} \quad (30)$$

$$\begin{aligned} &\leq \frac{\sigma^2}{\varepsilon^2} && \text{by Chebyshev's inequality} && (31) \\ &\rightarrow 0 && \text{as } \sigma \rightarrow 0 && (32) \end{aligned}$$

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

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) \leq \frac{\sigma^2}{\varepsilon^2} \rightarrow 0 \quad \text{as } \sigma \rightarrow 0 \quad (33)$$

Q.E.D.

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 [26], [32]. 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 [4], [13], [17], [34]. Figure 2 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 $0 < \alpha \leq 2$. 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 [4] 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. 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 [25], [34]. 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 [1], [2], [17], [34]:

$$\varphi(\omega) = \exp\left\{i a \omega - \gamma |\omega|^\alpha \left(1 + i \beta \text{sign}(\omega) \tan \frac{\alpha \pi}{2}\right)\right\} \quad \text{for } \alpha \neq 1 \quad (34)$$

and

$$\varphi(\omega) = \exp\{i a \omega - \gamma |\omega| (1 - 2i \beta \ln |\omega| \text{sign}(\omega) / \pi)\} \quad \text{for } \alpha = 1 \quad (35)$$

where

$$\text{sign}(\omega) = \begin{cases} 1 & \text{if } \omega > 0 \\ 0 & \text{if } \omega = 0 \\ -1 & \text{if } \omega < 0 \end{cases} \quad (36)$$

and $i = \sqrt{-1}$, $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, and $\gamma > 0$. The parameter α is the characteristic exponent. Again the variance of an alpha-stable density distribution 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 $\beta = 0$. The theorem below still holds even when $\beta \neq 0$. 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 (34)–(35). 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 (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

$$\int_{\theta-A}^{\theta+A} p(n) dn \rightarrow 0 \quad \text{as } \gamma \rightarrow 0 \quad (37)$$

The characteristic function $\varphi(\omega)$ of alpha-stable noise density $p(n)$ has the exponential form (34)–(35). This reduces to a simple complex exponential in the zero-dispersion limit:

$$\lim_{\gamma \rightarrow 0} \varphi(\omega) = \exp\{i a \omega\} \quad (38)$$

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

$$\lim_{\gamma \rightarrow 0} p(n) = \delta(n - a) \quad (39)$$

Then

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) = \int_{\theta-A}^{\theta+A} p(n) dn \quad (40)$$

$$= \int_{\theta-A}^{\theta+A} \delta(n - a) dn \quad (41)$$

$$= 0 \quad (42)$$

because $a \notin [\theta - A, \theta + A]$. Q.E.D.

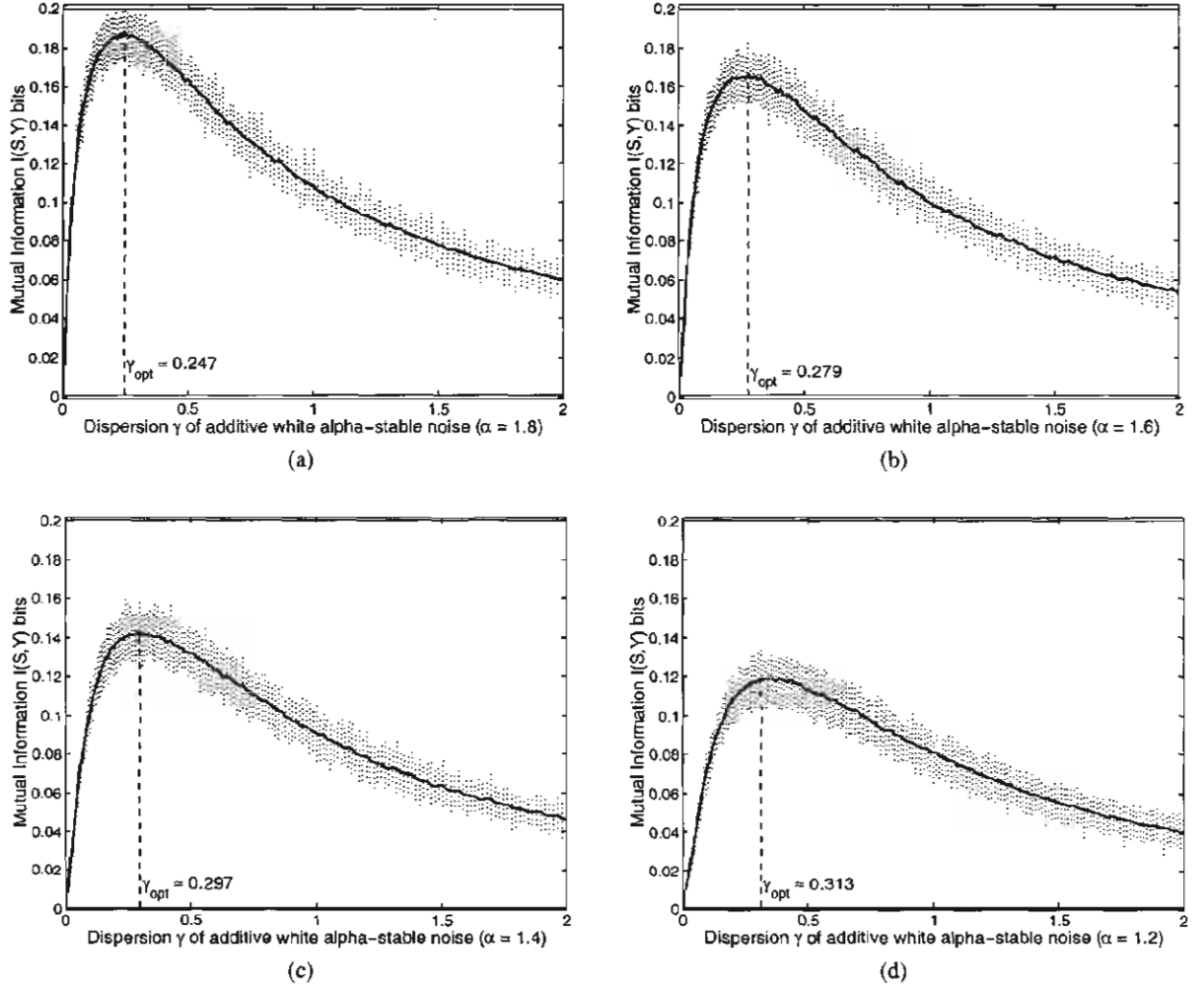


Fig. 3. Stochastic resonance with non-Cauchy infinite-variance symmetric alpha-stable noise. The graphs show the smoothed input-output mutual information of a threshold neuron as a function of the dispersion of additive white alpha-stable noise n_t . The vertical dashed lines show the absolute deviation between the smallest and largest outliers in each sample average of 100 outcomes. The noisy signal-forced threshold neuron has the form (1). The symmetric alpha-stable 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 noise bell curves have thicker tails for smaller α values: (a) $\alpha = 1.8$, (b) $\alpha = 1.6$, (c) $\alpha = 1.4$, and (d) $\alpha = 1.2$. Note that the maximum bit count $I(\gamma_{opt})$ decreases for more impulsive noise (lower α values).

Note that the above theorems hold for any success probability $0 < p < 1$ of the Bernoulli signal.

Figures 1 and 3 instantiate Theorem 2 for three types of infinite-variance symmetric alpha-stable noise. Figure 1 shows the SR effect for the highly impulsive Cauchy case of $\alpha = 1$. Figure 3 shows the SR effect for the comparatively less impulsive cases of $\alpha \approx 1.8$, $\alpha = 1.6$, $\alpha \approx 1.4$, and $\alpha = 1.2$ noise bell curves. Frequent and violent noise spikes interfere with the signal in all three cases.

IV. CONCLUSIONS

We have shown that almost all noise types produce stochastic resonance in threshold neurons that use sub-threshold signals. These results help explain the widespread occurrence of the SR effect in mechanical and biological threshold systems [3], [11], [12], [15], [27], [30], [31], [35]. The broad generality of the results suggests that SR should occur in any nonlinear system whose input-output structure approximates a threshold system. The infinite-variance result further implies that such widespread SR effects should be robust against violent noise impulses. The combined results

support the hypothesis [28], [29] that neurons have evolved to maximize their local information—at least if they process subthreshold signals in the presence of noise.

ACKNOWLEDGMENTS

National Science Foundation Grant No. ECS-0070284 and Thailand Research Fund Grants No. PDF/29/2543 and No. RSA4680001 supported this research.

REFERENCES

- [1] V. Akgiray and C. G. Lamoureux, "Estimation of Stable-Law Parameters: A Comparative Study," *Journal of Business and Economic Statistics*, vol. 7, pp. 85–93, January 1989.
- [2] H. Bergstrom, "On Some Expansions of Stable Distribution Functions," *Arkiv för Matematik*, vol. 2, pp. 375–378, 1952.
- [3] H. A. Braun, H. Wissing, K. Schäfer, and M. C. Hirsch, "Oscillation and Noise Determine Signal Transduction in Shark Multimodal Sensory Cells," *Nature*, vol. 367, pp. 270–273, January 1994.
- [4] L. Breiman, *Probability*, Addison-Wesley, 1968.
- [5] A. R. Bulsara and A. Zador, "Threshold Detection of Wideband Signals: A Noise-Induced Maximum in the Mutual Information," *Physical Review E*, vol. 54, no. 3, pp. R2185–R2188, September 1996.
- [6] J. M. Chambers, C. L. Mallows, and B. W. Stuck, "A Method for Simulating Stable Random Variables," *Journal of the American Statistical Association*, vol. 71, no. 354, pp. 340–344, 1976.
- [7] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, "Aperiodic Stochastic Resonance," *Physical Review E*, vol. 54, no. 5, pp. 5575–5584, November 1996.
- [8] J. J. Collins, C. C. Chow, and T. T. Imhoff, "Stochastic Resonance without Tuning," *Nature*, vol. 376, pp. 236–238, July 1995.
- [9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.
- [10] G. Deco and B. Schürmann, "Stochastic Resonance in the Mutual Information between Input and Output Spike Trains of Noisy Central Neurons," *Physica D*, vol. 117, pp. 276–282, 1998.
- [11] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, "Noise Enhancement of Information Transfer in Crayfish Mechanoreceptors by Stochastic Resonance," *Nature*, vol. 365, pp. 337–340, September 1993.
- [12] S. Fauve and F. Heslot, "Stochastic Resonance in a Bistable System," *Physics Letters A*, vol. 97, no. 1,2, pp. 5–7, August 1983.
- [13] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley & Sons, 1966.
- [14] L. Gamaitoni, "Stochastic Resonance in Multi-Threshold Systems," *Physics Letters A*, vol. 208, pp. 315–322, December 1995.
- [15] L. Gamaitoni, P. Hänggi, P. Jung, and F. Marchesoni, "Stochastic Resonance," *Reviews of Modern Physics*, vol. 70, no. 1, pp. 223–287, January 1998.
- [16] X. Godivier and F. Chapeau-Blondeau, "Stochastic Resonance in the Information Capacity of a Nonlinear Dynamic System," *International Journal of Bifurcation and Chaos*, vol. 8, no. 3, pp. 581–589, 1998.
- [17] M. Grigoriu, *Applied Non-Gaussian Processes*, Prentice Hall, 1995.
- [18] S. M. Hess and A. M. Albano, "Minimum Requirements for Stochastic Resonance in Threshold Systems," *International Journal of Bifurcation and Chaos*, vol. 8, no. 2, pp. 395–400, 1998.
- [19] N. Hohn and A. N. Burkitt, "Enhanced Stochastic Resonance in Threshold Detectors," in *Proceedings of the 2001 IEEE International Joint Conference on Neural Networks (IJCNN '01)*, 2001, pp. 644–647.
- [20] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proceedings of the National Academy of Science*, vol. 79, pp. 2554–2558, 1982.
- [21] M. E. Inchiosa, J. W. C. Robinson, and A. R. Bulsara, "Information-Theoretic Stochastic Resonance in Noise-Floor Limited Systems: The Case for Adding Noise," *Physical Review Letters*, vol. 85, pp. 3369–3372, October 2000.
- [22] P. Jung, "Stochastic Resonance and Optimal Design of Threshold Detectors," *Physics Letters A*, vol. 207, pp. 93–104, October 1995.
- [23] P. Jung and G. Mayer-Kress, "Stochastic Resonance in Threshold Devices," *Il Nuovo Cimento*, vol. 17 D, no. 7-8, pp. 827–834, Luglio-Agosto 1995.
- [24] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice Hall, 1992.
- [25] B. Kosko, *Fuzzy Engineering*, Prentice Hall, 1996.
- [26] B. Kosko and S. Mitaim, "Robust Stochastic Resonance: Signal Detection and Adaptation in Impulsive Noise," *Physical Review E*, vol. 64, no. 051110, October 2001.
- [27] J. E. Levin and J. P. Miller, "Broadband Neural Encoding in the Cricket Cercal Sensory System Enhanced by Stochastic Resonance," *Nature*, vol. 380, pp. 165–168, March 1996.
- [28] R. Linsker, "Self-Organization in a Perceptual Network," *Computer*, vol. 21, no. 3, pp. 105–117, March 1988.
- [29] R. Linsker, "A Local Learning Rule That Enables Information Maximization for Arbitrary Input Distributions," *Neural Computation*, vol. 9, no. 8, pp. 1661–1665, November 1997.
- [30] A. Longtin, "Stochastic Resonance in Neuron Models," *Journal of Statistical Physics*, vol. 70, no. 1/2, pp. 309–327, January 1993.
- [31] V. I. Melnikov, "Schmitt Trigger: A Solvable Model of Stochastic Resonance," *Physical Review E*, vol. 48, no. 4, pp. 2481–2489, October 1993.
- [32] S. Mitaim and B. Kosko, "Adaptive Stochastic Resonance," *Proceedings of the IEEE: Special Issue on Intelligent Signal Processing*, vol. 86, no. 11, pp. 2152–2183, November 1998.
- [33] S. Mitaim and B. Kosko, "Adaptive Stochastic Resonance for Noisy Threshold Neurons Based on Mutual Information," in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '02)*, Honolulu, May 2002, pp. 1980–1985.
- [34] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, John Wiley & Sons, 1995.
- [35] D. F. Russell, L. A. Willkens, and F. Moss, "Use of Behavioural Stochastic Resonance by Paddle Fish for Feeding," *Nature*, vol. 402, pp. 291–294, November 1999.
- [36] N. C. Stocks, "Information Transmission in Parallel Threshold Arrays," *Physical Review E*, vol. 63, no. 041114, 2001.
- [37] P. Tsakalides and C. L. Nikias, "The Robust Covariation-Based MUSIC (ROC-MUSIC) Algorithm for Bearing Estimation in Impulsive Noise Environments," *IEEE Transactions on Signal Processing*, vol. 44, no. 7, pp. 1623–1633, July 1996.

Stochastic Resonance in Threshold Neurons: Noise Enhances Mutual Information

Sanya Mitaim[†] and Bart Kosko^{*}

[†]Department of Electrical Engineering, Faculty of Engineering
Thammasat University, Pathumthani 12121, Thailand

^{*}Department of Electrical Engineering, Signal and Image Processing Institute
University of Southern California, Los Angeles, California 90089-2564, USA

Abstract

This paper presents theoretical and simulation results that show the stochastic-resonance (SR) effect in threshold neurons. Small amounts of additive noise can enhance input-output mutual information that measures the performance of threshold systems that process subthreshold input signals. The SR result holds for all possible noise probability density functions with finite variance as well as the entire uncountably infinite class of alpha-stable probability density functions. The SR result for alpha-stable noise densities shows that the SR effect in threshold and threshold-like systems is robust against occasional or even frequent violent fluctuations in noise. Regression analysis reveals (1) an exponential relationship for the optimal noise dispersion as a function of the alpha bell-curve tail thickness and (2) an approximate linear relationship for the SR-maximal mutual information as a function of the alpha bell-curve tail thickness.

Keywords: stochastic resonance, impulsive noise, alpha-stable noise, infinite-variance statistics, threshold systems, mutual information, dithering.

I. Stochastic Resonance in Threshold Neurons

Noise can sometimes help neural or other nonlinear systems. Figure 1 shows that a small amount of alpha-stable impulsive pixel noise improves the ‘baboon’ image while too much noise degrades the image.

Several researchers have found that threshold neurons and other threshold systems exhibit stochastic resonance [7], [17], [19], [21], [22], [25], [26], [38]: Small amounts of noise improve the threshold neuron’s input-output correlation measure [9], [10] or mutual information [7], [30], [38]. All of these simulations and analyses assume a noise probability density function that has finite variance. Most further assume that the noise is simply Gaussian or uniform. Yet the statistics of real-world noise can differ substantially from these simple and finite-variance

probability descriptions. The noise can be impulsive and irregular and have infinite variance and infinite higher-order moments.

The paper shows that finite-variance noise as well as infinite-variance (impulsive) noise can enhance mutual information in threshold neurons that process subthreshold input signals. Two theorems [30] confirm the existence of this SR effect. The first theorem shows that threshold neurons exhibit the SR effect for all finite-variance noise densities if the system performance measure is Shannon’s mutual information and if the mean or location parameter falls outside an interval that one can often pick in advance. The second theorem shows that this also holds for all infinite-variance densities that belong to the large class of stable distributions. Both theorems assume that all signals are subthreshold signals. The paper also presents statistical findings on the relationship of the SR effects and the bell-curve tail thickness parameter from simulation experiments. The regression analysis confirms and extends the exponential relationship between the optimal noise dispersion and the alpha bell-curve tail thickness [29]. This exponential relationship corresponds to a similar one for infinite-variance SR systems that use a signal-to-noise ratio or a cross correlation for the system performance measure [29]. Regression also shows that the SR-maximal mutual information in noisy threshold neurons depend approximately linearly on the bell-curve tail thickness for symmetric alpha-stable noise.

Figure 2 shows some typical symmetrical noise densities whose bell curves have thick tails that produce infinite variance and often highly impulsive noise spikes. Figure 3 shows a simulation instance of both Theorem 2 and the empirical trends in Figure 4. Infinite-variance alpha-stable noise produces the SR effect when plotted against the Shannon mutual information of the threshold system. This also holds for the impulsive Cauchy noise that belongs to this alpha-stable family for $\alpha = 1$. The linear regression results in Table 1 and Figure 4 reveal the exponential relationship between the optimal noise dispersion and alpha bell-curve tail thickness. The linear

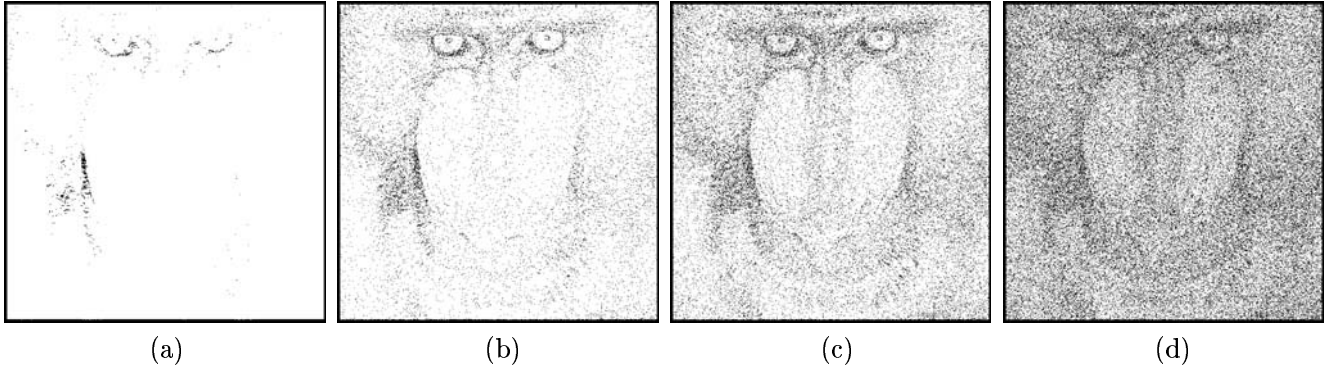


Figure 1: Alpha-stable ($\alpha = 1.5$) pixel noise can improve the quality of an image through a stochastic-resonance or dithering process [16], [40]. The noise produces a nonmonotonic response: A small level of noise sharpens the image features while too much noise degrades them. These noisy images result when we apply a pixel threshold to the ‘baboon’ image. The system quantizes the noisy 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 θ : $y = g((x + n) - \theta)$ 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 $\theta = 0.1$. Thresholding the original ‘baboon’ image gives the faint image in (a). The alpha-stable noise n has zero “mean” for images (b)-(d). The noise scale κ grows from (b)-(d): $\kappa = 0.1$ in (b), $\kappa = 0.2$ in (c), and $\kappa = 0.5$ in (d). Note that the noise dispersion is $\gamma = \kappa^\alpha$ where $\alpha = 1.5$ in this case.

dependence of the log-transformed optimal noise dispersion on the bell-curve thickness becomes quadratic when the signal amplitude is too small or too close to the neuron’s threshold. They also shows a similar pattern: The linear dependence of the SR-maximal mutual information on the bell-curve thickness also becomes quadratic when the signal amplitude is too small or too close to the neuron’s threshold.

II. Threshold Neurons and Shannon’s Mutual Information

We use the standard discrete-time threshold neuron model [7], [17], [23], [25], [27], [29]

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

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 system uses subthreshold binary signals: $A < \theta$. The symbol ‘0’ denotes the input signal $s = -A$ and output signal $y = 0$. The symbol ‘1’ denotes the input signal $s = A$ and output signal $y = 1$. Then the conditional probabilities $P_{Y|S}(y|s)$ are

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

$$= \Pr\{n < \theta + A\} = \int_{-\infty}^{\theta+A} p(n)dn \quad (2)$$

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

$$P_{Y|S}(0|1) = \Pr\{s + n < \theta\} \Big|_{s=A} \\ = \Pr\{n < \theta - A\} = \int_{-\infty}^{\theta-A} p(n)dn \quad (4)$$

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

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 [7] and Gaussian inputs [38]. Our theorems and proofs in [30] 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 [11] to measure the noise enhancement or “stochastic resonance” (SR) effect [7], [12], [19], [24], [38]. 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 $H(Y|X)$:

$$I(S, Y) = H(Y) - H(Y|S) \\ = - \sum_y P_Y(y) \log P_Y(y) \quad (7)$$

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

$$= \sum_{s,y} P_{SY}(s, y) \log \frac{P_{SY}(s, y)}{P_S(s)P_Y(y)} \quad (9)$$

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

$$I(S, Y) = E \left[\log \frac{P_{SY}(s, y)}{P_S(s)P_Y(y)} \right] \quad (10)$$

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 [11] pseudo-distance measure $I(S, Y) = \sum_s \sum_y P_{SY}(s, y) \log \frac{P_{SY}(s, y)}{P_S(s)P_Y(y)}$. 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.

III. SR for Threshold Neurons: Theoretical and Simulation Results

Two theorems state that almost all finite-variance and infinite-variance noise densities produce the SR effect in threshold neurons with subthreshold signals. This holds for all probability distributions on a two-symbol input alphabet. The proofs in [30] show that if $I(S, Y) > 0$ then eventually the mutual information $I(S, Y)$ tends toward zero as the noise variance σ or noise dispersion γ 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 $m = E[n]$ or the noise location a of the alpha-stable distributions does not lie in the signal-threshold interval $[\theta - A, \theta + A]$.

Theorem 1. Suppose that the threshold signal system (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 system (1) exhibits the

nonmonotone SR effect in the sense that $I(S, Y) \rightarrow 0$ as $\sigma \rightarrow 0$.

Theorem 2 applies to *any* alpha-stable noise model. The density need not be symmetric. We use the alpha-stable bell-curve probability density functions to model many types of impulsive noise. A general alpha-stable probability density function f has characteristic function φ [1], [3], [20], [34]:

$$\varphi(\omega) = \exp \left\{ i a \omega - \gamma |\omega|^\alpha \left(1 + i \beta \text{sign}(\omega) \tan \frac{\alpha \pi}{2} \right) \right\} \quad \text{for } \alpha \neq 1 \quad (11)$$

and

$$\varphi(\omega) = \exp \{ i a \omega - \gamma |\omega| (1 - 2i \beta \ln |\omega| \text{sign}(\omega) / \pi) \} \quad \text{for } \alpha = 1 \quad (12)$$

where

$$\text{sign}(\omega) = \begin{cases} 1 & \text{if } \omega > 0 \\ 0 & \text{if } \omega = 0 \\ -1 & \text{if } \omega < 0 \end{cases} \quad (13)$$

and $i = \sqrt{-1}$, $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, and $\gamma > 0$. The parameter α is the characteristic exponent. 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 $\beta \neq 0$. The *dispersion* parameter γ acts like a variance because it controls the width of a symmetric alpha-stable bell curve [5], [15], [20], [34]. 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. There are no known closed forms of the α -stable densities for most α 's. Numerical integration of φ produced the simulation results in Figure 2.

Alpha-stable models tend to work well when the noise or signal data contains “outliers” — and all do to some degree. 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 [28], [34]. 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. Figure 2 shows realizations of four symmetric alpha-stable noise random variables.

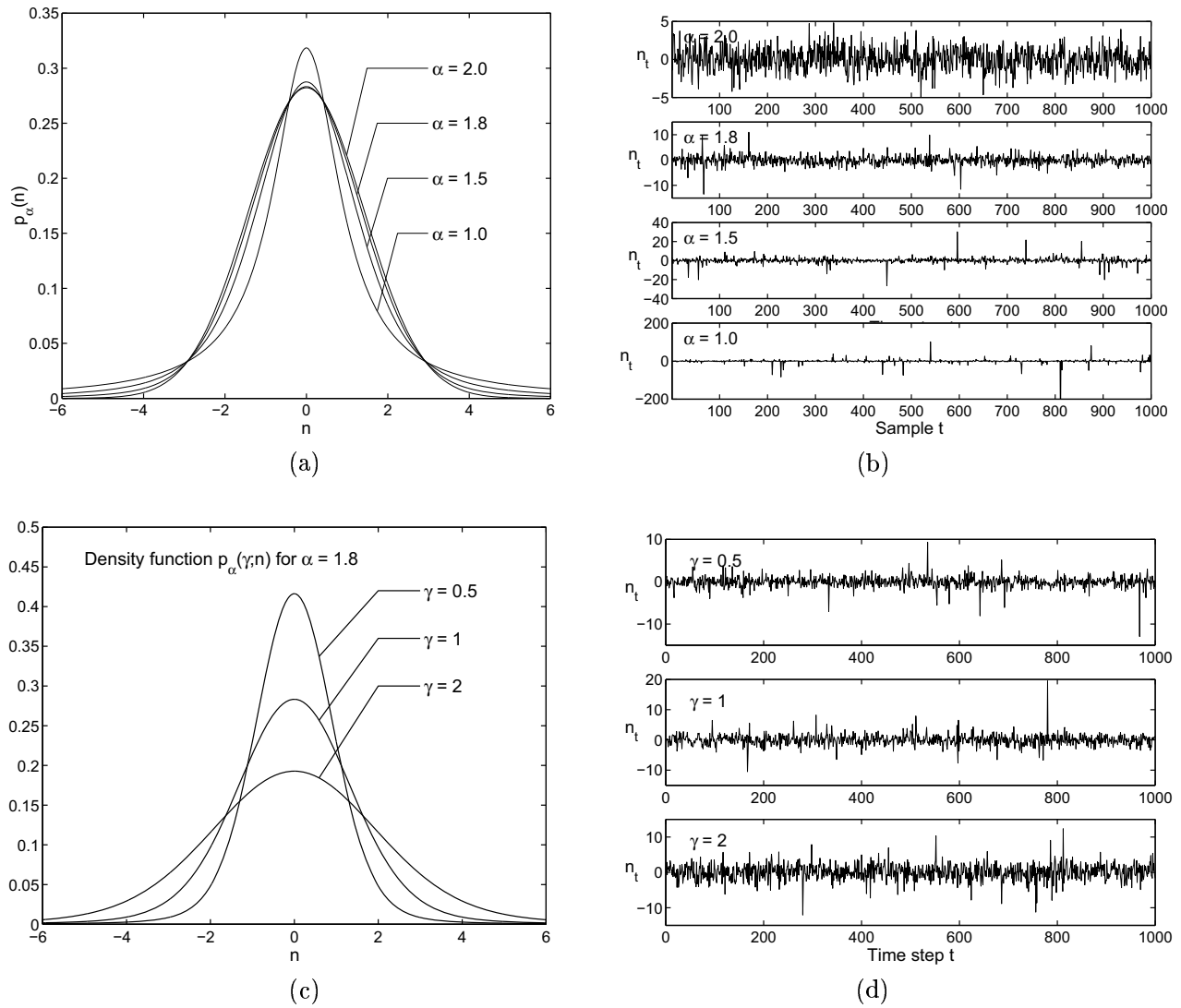


Figure 2: Samples of standard symmetric alpha-stable probability densities and their realizations. (a) Density functions with zero location ($a = 0$) and unit dispersion ($\gamma = 1$) for $\alpha = 2, 1.8, 1.5$, and 1 . The densities are bell curves that have thicker tails as α decreases and thus that 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) Samples of alpha-stable random variables with zero location and unit dispersion. 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 [8], [39] generated these realizations. (c) Density functions for $\alpha = 1.8$ with dispersions $\gamma = 0.5, 1$, and 2 . (d) Samples of alpha-stable noise n for $\alpha = 1.8$ with dispersions $\gamma = 0.5, 1$, and 2 .

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

Figure 3 gives a typical example of the SR effect for finite-variance noise and highly impulsive noise with infinite variance. The alpha-stable noises have $\alpha = 2$ (Gaussian), $\alpha = 1.8$, $\alpha = 1.4$, and $\alpha = 1$ (Cauchy). So frequent and violent noise spikes in-

terfere with the signal. Figure 3 also illustrates the empirical trends in Figure 4: A falling tail-thickness parameter α produces an increasing optimal noise dispersion γ_{opt} but a decreasing SR-maximal mutual information $I_{max}(S, Y)$.

We next show the exponential relationship between the optimal noise dispersion γ_{opt} and the bell-curve tail-thickness parameter α : $\gamma_{opt}(\alpha) = 10^{\beta_{\gamma,0} + \beta_{\gamma,1}\alpha}$ for parameters $\beta_{\gamma,0}$ and $\beta_{\gamma,1}$ that depend on the signal amplitude A . Then the log-transformation of the optimal dispersion gives the linear model $\log \gamma_{opt}(\alpha) =$

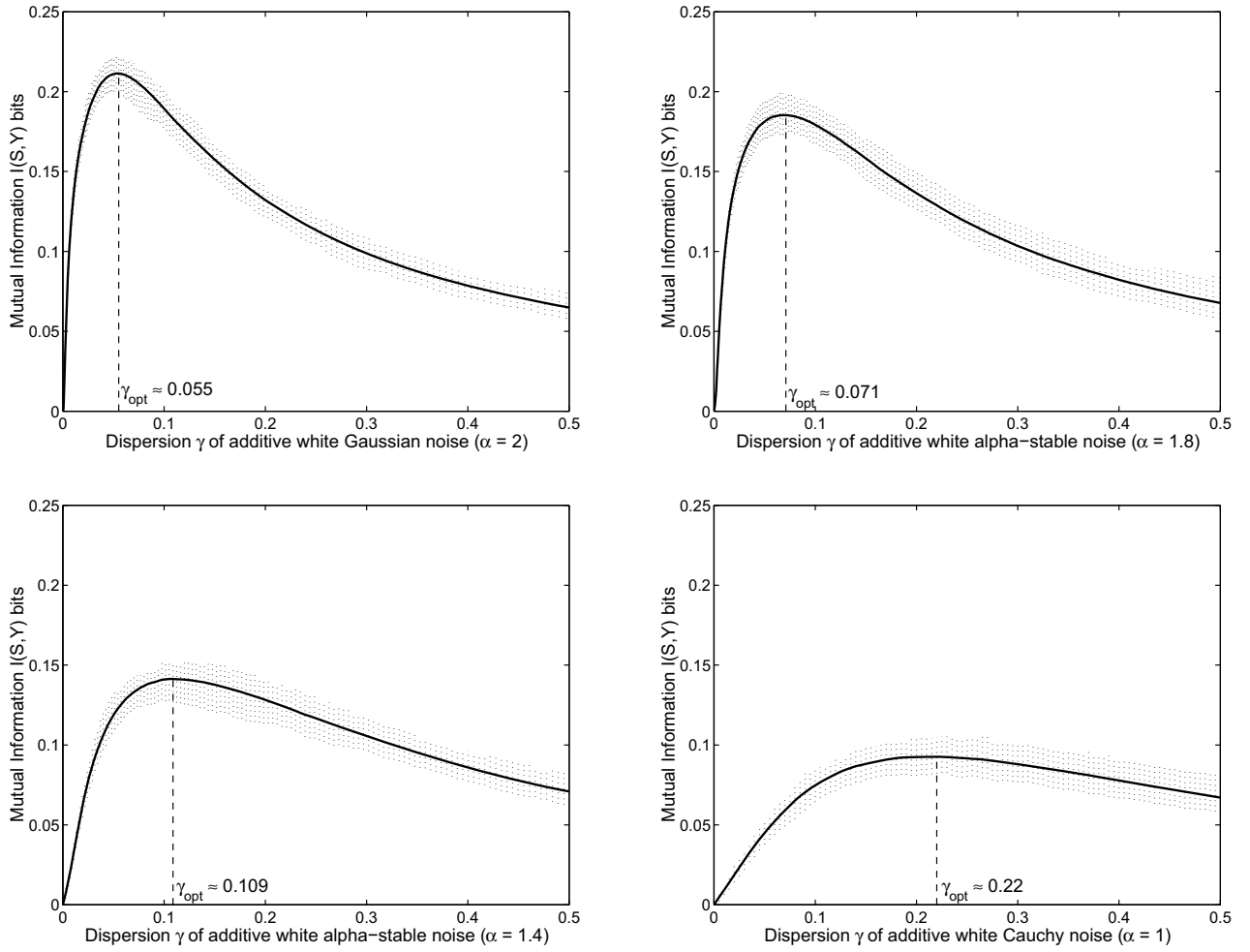


Figure 3: Stochastic resonance with finite-variance and highly impulsive (infinite-variance) alpha-stable noise. The graphs show the smoothed input-output mutual information of a threshold system as a function of the dispersion of additive white alpha-stable noise n_t with $\alpha = 2$ (Gaussian noise), $\alpha = 1.8$, $\alpha = 1.4$, and $\alpha = 1$ (Cauchy noise). The vertical dashed lines show the absolute deviation between the smallest and largest outliers in each sample average of 100 outcomes. The system has a nonzero noise optimum at $\gamma_{opt} \approx 0.055$ for $\alpha = 2$, $\gamma_{opt} \approx 0.071$ for $\alpha = 1.8$, $\gamma_{opt} \approx 0.109$ for $\alpha = 1.4$, and $\gamma_{opt} \approx 0.220$ for $\alpha = 1$ and thus shows the SR effect. The noisy signal-forced threshold system has the form (1). The Cauchy noise n_t adds to the bipolar input Bernoulli signal s_t . The system has threshold $\theta = 0.5$. The input Bernoulli signal has amplitude $A = 0.4$ 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. Note that decreasing the tail-thickness parameter α increases the optimal noise dispersion γ_{opt} and decreases the SR-maximal mutual information $I_{max}(S, Y)$ as in Figure 4.

$\beta_{\gamma,0} + \beta_{\gamma,1}\alpha$. Table 1 shows the estimated parameters $\hat{\beta}_{\gamma,0}$ and $\hat{\beta}_{\gamma,1}$ and the coefficient of determination r_i^2 for different input signal amplitudes in the threshold neuron using SPSS software. All observed significant levels or p -values were less than 10^{-4} . The exponential trend's exponent is linear for most amplitudes but becomes quadratic for very small amplitudes and for amplitudes close to the threshold $\theta = \frac{1}{2}$ (or $\gamma_{opt}(\alpha) = 10^{\beta_{\gamma,0} + \beta_{\gamma,1}\alpha + \beta_{\gamma,2}\alpha^2}$ for a quadratic fit to the data). Figure 4(a) shows samples of the log-linear plots.

Table 1 and Figure 4(b) also show an approximate

linear relationship $I_{max}(S, Y; \alpha) = \beta_{I,0} + \beta_{I,1}\alpha$ for the SR-maximal mutual information $I_{max}(S, Y)$ as a function of the tail-thickness parameter α . There is a clear linear trend for most amplitudes A but the trend becomes quadratic for very small amplitudes and for amplitudes close to the threshold $\theta = \frac{1}{2}$.

IV. Conclusions

This paper shows that almost all noise types produce stochastic resonance in threshold systems that use subthreshold signals. This result helps explain the widespread occurrence of the SR effect in me-

Input signal amplitude A	Linear regression estimates of $\log \gamma_{opt}$				Linear regression estimates of $I_{max}(S, Y)$			
	Linear model			Quadratic model $r^2_{\gamma,q}$	Linear model			Quadratic model $r^2_{I,q}$
	Regression coefficients				Regression coefficients			
	$\hat{\beta}_{\gamma,0}$	$\hat{\beta}_{\gamma,1}$	$r^2_{\gamma,l}$		$\hat{\beta}_{I,0}$	$\hat{\beta}_{I,1}$	$r^2_{I,l}$	
0.050	0.1002	−0.6087	0.9321	0.9723	−0.0008	0.0022	0.9370	0.9972
0.100	0.1180	−0.6261	0.9558	0.9888	−0.0031	0.0086	0.9440	0.9990
0.150	0.1078	−0.6251	0.9679	0.9921	−0.0068	0.0190	0.9521	0.9995
0.200	0.0915	−0.6214	0.9699	0.9942	−0.0113	0.0329	0.9612	0.9998
0.250	0.0694	−0.6172	0.9781	0.9959	−0.0161	0.0500	0.9715	0.9997
0.300	0.0439	−0.6148	0.9869	0.9961	−0.0207	0.0698	0.9816	0.9993
0.350	0.0116	−0.6211	0.9935	0.9961	−0.0236	0.0920	0.9913	0.9987
0.400	−0.0313	−0.6367	0.9947	0.9951	−0.0229	0.1161	0.9976	0.9981
0.450	−0.1107	−0.6688	0.9757	0.9944	−0.0120	0.1408	0.9905	0.9975
0.490	−0.2805	−0.8053	0.8987	0.9863	0.0336	0.1527	0.9145	0.9959

Table 1: Linear regression estimates of the SR-optimal log dispersion γ_{opt} and the SR-maximal mutual information $I_{max}(S, Y)$ as functions of the bell-curve tail-thickness parameter α from a symmetric alpha-stable noise density. The parameters $\beta_{\gamma,0}$ and $\beta_{\gamma,1}$ relate $\log \gamma_{opt}$ and α through a linear relationship: $\log \gamma_{opt}(\alpha) = \beta_{\gamma,0} + \beta_{\gamma,1}\alpha$ for base-10 logarithm. Likewise the parameters $\beta_{I,0}$ and $\beta_{I,1}$ relate $I_{max}(S, Y)$ and α through a linear relationship: $I_{max}(S, Y; \alpha) = \beta_{I,0} + \beta_{I,1}\alpha$. The coefficients of determination $r_{\gamma,l}^2$ and $r_{I,l}^2$ shows how well the linear models fits the data. The Table also shows the coefficients of determination $r_{\gamma,q}^2$ for the quadratic model $\log \gamma_{opt}(\alpha) = \beta_{\gamma,0} + \beta_{\gamma,1}\alpha + \beta_{\gamma,2}\alpha^2$ of the optimal dispersion as well as $r_{I,q}^2$ for the quadratic model of the SR-maximal mutual information. All observed significant levels or p -values were less than 10^{-4} .

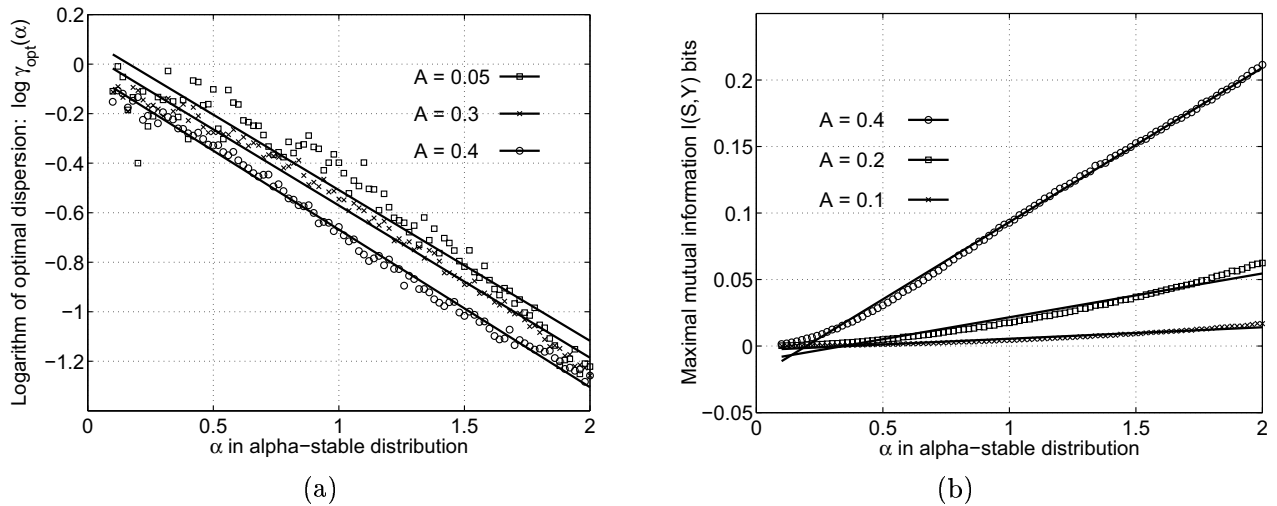


Figure 4: Linear regression estimates. (a) Exponential law for optimal noise dispersion γ_{opt} as a function of bell-curve thickness parameter α for the mutual-information performance measure and for different signal amplitudes A . The optimal noise dispersion γ_{opt} depends on the parameter α through the exponential relation $\gamma_{opt}(\alpha) = 10^{\beta_{\gamma,0} + \beta_{\gamma,1}\alpha}$ for parameters $\beta_{\gamma,0}$ and $\beta_{\gamma,1}$ (or $\gamma_{opt}(\alpha) = 10^{\beta_{\gamma,0} + \beta_{\gamma,1}\alpha + \beta_{\gamma,2}\alpha^2}$ for a quadratic fit to the data). (b) Linear regression for maximal mutual information $I_{max}(S, Y)$ as a function of bell-curve thickness parameter α for different signal amplitudes A . The maximal mutual information $I_{max}(S, Y)$ depends on the parameter α through the linear relationship $I_{max}(S, Y; \alpha) = \beta_{I,0} + \beta_{I,1}\alpha$ for parameters $\beta_{I,0}$ and $\beta_{I,1}$ (or $I_{max}(\alpha) = \beta_{I,0} + \beta_{I,1}\alpha + \beta_{I,2}\alpha^2$ for a quadratic fit to the data). Table 1 shows the estimated parameters for different input Bernoulli signal amplitudes A .

chanical and biological threshold systems [4], [13], [14], [18], [31], [33], [37]. The broad generality of the results suggests that SR should occur in any nonlinear system whose input-output structure approximates a threshold system as in the many models of continuous neurons [6], [32], [36]. This applies to a broad class of physical and biological systems. The infinite-variance result further implies that such widespread SR effects should be robust against violent noise impulses. The result further suggests that scientists and engineers should consider the practical use of noise when they design their information and signal processing systems [2], [35], [41].

Acknowledgments

National Science Foundation Grant No. ECS-0070284 and Thailand Research Fund Grant Nos. PDF/29/2543 and RSA4680001 supported this research.

References

- [1] V. Akgiray and C. G. Lamoureux, "Estimation of Stable-Law Parameters: A Comparative Study," *Journal of Business and Economic Statistics*, vol. 7, pp. 85–93, January 1989.
- [2] T. R. Albert, A. R. Bulsara, G. Schnera, and M. Inghiosa, "An Evaluation of the Stochastic Resonance Phenomenon as a Potential Tool for Signal Processing," in *Conference Record of the Twenty-Seventh Asilomar Conference on Signals, Systems & Computers*, November 1993, vol. 1, pp. 583–587.
- [3] H. Bergstrom, "On Some Expansions of Stable Distribution Functions," *Arkiv för Matematik*, vol. 2, pp. 375–378, 1952.
- [4] H. A. Braun, H. Wissing, K. Schäfer, and M. C. Hirsch, "Oscillation and Noise Determine Signal Transduction in Shark Multimodal Sensory Cells," *Nature*, vol. 367, pp. 270–273, January 1994.
- [5] L. Breiman, *Probability*, Addison-Wesley, 1968.
- [6] A. R. Bulsara, A. J. Maren, and G. Schnera, "Single Effective Neuron: Dendritic Coupling Effects and Stochastic Resonance," *Biological Cybernetics*, vol. 70, pp. 145–156, 1993.
- [7] A. R. Bulsara and A. Zador, "Threshold Detection of Wideband Signals: A Noise-Induced Maximum in the Mutual Information," *Physical Review E*, vol. 54, no. 3, pp. R2185–R2188, September 1996.
- [8] J. M. Chambers, C. L. Mallows, and B. W. Stuck, "A Method for Simulating Stable Random Variables," *Journal of the American Statistical Association*, vol. 71, no. 354, pp. 340–344, 1976.
- [9] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, "Aperiodic Stochastic Resonance," *Physical Review E*, vol. 54, no. 5, pp. 5575–5584, November 1996.
- [10] J. J. Collins, C. C. Chow, and T. T. Imhoff, "Stochastic Resonance without Tuning," *Nature*, vol. 376, pp. 236–238, July 1995.
- [11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.
- [12] G. Deco and B. Schürmann, "Stochastic Resonance in the Mutual Information between Input and Output Spike Trains of Noisy Central Neurons," *Physica D*, vol. 117, pp. 276–282, 1998.
- [13] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, "Noise Enhancement of Information Transfer in Crayfish Mechanoreceptors by Stochastic Resonance," *Nature*, vol. 365, pp. 337–340, September 1993.
- [14] S. Fauve and F. Heslot, "Stochastic Resonance in a Bistable System," *Physics Letters A*, vol. 97, no. 1,2, pp. 5–7, August 1983.
- [15] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley & Sons, 1966.
- [16] L. Gammaitoni, "Stochastic Resonance and the Dithering Effect in Threshold Physical Systems," *Physical Review E*, vol. 52, no. 5, pp. 4691–4698, November 1995.
- [17] L. Gammaitoni, "Stochastic Resonance in Multi-Threshold Systems," *Physics Letters A*, vol. 208, pp. 315–322, December 1995.
- [18] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, "Stochastic Resonance," *Reviews of Modern Physics*, vol. 70, no. 1, pp. 223–287, January 1998.
- [19] X. Godivier and F. Chapeau-Blondeau, "Stochastic Resonance in the Information Capacity of a Nonlinear Dynamic System," *International Journal of Bifurcation and Chaos*, vol. 8, no. 3, pp. 581–589, 1998.
- [20] M. Grigoriu, *Applied Non-Gaussian Processes*, Prentice Hall, 1995.
- [21] S. M. Hess and A. M. Albano, "Minimum Requirements for Stochastic Resonance in Threshold Systems," *International Journal of Bifurcation and Chaos*, vol. 8, no. 2, pp. 395–400, 1998.
- [22] N. Hohn and A. N. Burkitt, "Enhanced Stochastic Resonance in Threshold Detectors," in *Proceedings of the 2001 IEEE International Joint Conference on Neural Networks (IJCNN '01)*, 2001, pp. 644–647.
- [23] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proceedings of the National*

- Academy of Science*, vol. 79, pp. 2554–2558, 1982.
- [24] M. E. Inchiosa, J. W. C. Robinson, and A. R. Bulsara, “Information-Theoretic Stochastic Resonance in Noise-Floor Limited Systems: The Case for Adding Noise,” *Physical Review Letters*, vol. 85, pp. 3369–3372, October 2000.
 - [25] P. Jung, “Stochastic Resonance and Optimal Design of Threshold Detectors,” *Physics Letters A*, vol. 207, pp. 93–104, October 1995.
 - [26] P. Jung and G. Mayer-Kress, “Stochastic Resonance in Threshold Devices,” *Il Nuovo Cimento*, vol. 17 D, no. 7-8, pp. 827–834, Luglio-Agosto 1995.
 - [27] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice Hall, 1992.
 - [28] B. Kosko, *Fuzzy Engineering*, Prentice Hall, 1996.
 - [29] B. Kosko and S. Mitaim, “Robust Stochastic Resonance: Signal Detection and Adaptation in Impulsive Noise,” *Physical Review E*, vol. 64, no. 051110, October 2001.
 - [30] B. Kosko and S. Mitaim, “Stochastic Resonance in Noisy Threshold Neurons,” *Neural Networks*, vol. 16, no. 5-6, pp. 755–761, June-July 2003.
 - [31] J. E. Levin and J. P. Miller, “Broadband Neural Encoding in the Cricket Cercal Sensory System Enhanced by Stochastic Resonance,” *Nature*, vol. 380, pp. 165–168, March 1996.
 - [32] A. Longtin, “Stochastic Resonance in Neuron Models,” *Journal of Statistical Physics*, vol. 70, no. 1/2, pp. 309–327, January 1993.
 - [33] V. I. Melnikov, “Schmitt Trigger: A Solvable Model of Stochastic Resonance,” *Physical Review E*, vol. 48, no. 4, pp. 2481–2489, October 1993.
 - [34] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, John Wiley & Sons, 1995.
 - [35] H. C. Papadopoulos and G. W. Wornell, “A Class of Stochastic Resonance Systems for Signal Processing Applications,” in *Proceedings of the 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-96)*, May 1996, pp. 1617–1620.
 - [36] H. E. Plesser and S. Tanaka, “Stochastic Resonance in a Model Neuron with Reset,” *Physics Letters A*, vol. 225, pp. 228–234, February 1997.
 - [37] D. F. Russell, L. A. Willkens, and F. Moss, “Use of Behavioural Stochastic Resonance by Paddle Fish for Feeding,” *Nature*, vol. 402, pp. 291–294, November 1999.
 - [38] N. G. Stocks, “Information Transmission in Parallel Threshold Arrays,” *Physical Review E*, vol. 63, no. 041114, 2001.
 - [39] P. Tsakalides and C. L. Nikias, “The Robust Covariation-Based MUSIC (ROC-MUSIC) Algorithm for Bearing Estimation in Impulsive Noise Environments,” *IEEE Transactions on Signal Processing*, vol. 44, no. 7, pp. 1623–1633, July 1996.
 - [40] R. A. Wannamaker, S. P. Lipshitz, and J. Vanderkooy, “Stochastic Resonance as Dithering,” *Physical Review E*, vol. 61, no. 1, pp. 233–236, January 2000.
 - [41] S. Zozor and P.-O. Amblard, “Stochastic Resonance in a Discrete Time Nonlinear AR(1) Models,” *IEEE Transactions on Signal Processing*, vol. 47, no. 1, pp. 108–122, January 1999.

Sanya Mitaim received the B.Eng. degree in control engineering from King Mongkut's Institute of Technology Ladkrabang, the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California. He is Assistant Professor of Electrical Engineering at Thammasat University. His research interests include neural and fuzzy techniques in nonlinear signal processing and noise processing.

Bart Kosko received the B.S. degrees in philosophy and economics from the University of Southern California (USC), the M.S. degree in applied mathematics from the University of California, San Diego, and the Ph.D. degree in electrical engineering from the University of California, Irvine.

He is Professor of Electrical Engineering at USC. He is author of the books *Fuzzy Engineering* (Prentice-Hall, 1996), *Neural Networks and Fuzzy Systems* (Prentice-Hall, 1992), *Fuzzy Thinking* (Hyperion, 1993), *Heaven in a Chip* (Random House, 2000), and the novel *Nanotime* (Avon, 1997). He is an elected Governor of the International Neural Network Society and has chaired and co-chaired many neural and fuzzy system conferences.

Weighted Fuzzy C-Means Algorithm for Room Equalization at Multiple Locations

Sirichai Turmchokkasam

Department of Electronic and Telecommunication
Engineering, Faculty of Engineering
Bangkok University, Pathumthani, 12120, Thailand
E-mail: sirichai.t@bu.ac.th

Sanya Mitaim

Department of Electrical Engineering
Faculty of Engineering, Thammasat University
Pathumthani, 12120, Thailand
E-mail: msanya@engr.tu.ac.th

Abstract

This paper studies the problem of simultaneous room response equalization for multiple listeners. The classical fuzzy c-means (FCM) algorithm already proves useful to deriving an equalizing filter for multiple listeners. The FCM algorithm gives fewer prototypes for several room responses. But the FCM algorithm uses equal weights for all parts of sound reverberation. The paper proposes an algorithm to obtain a room acoustic equalization filter based on weighted fuzzy c-mean (WFCM) algorithm. The use of WFCM can give clustering more flexibility in obtaining a prototype for each cluster. Then combining the prototypes to obtain a representative room response that derives the inverse filter will be more effective. We use spectral deviation to measure how well each equalizing filter works. Experiments show that an equalizing filter obtaining from the prototypes using WFCM gives better performance for most listener locations.

Keyword: room acoustic equalization, inverse filtering, weighted fuzzy c-means algorithm.

1. Introduction

Room acoustic equalization has been a classic inverse filtering problem [1,2]. The design of equalizing filter using simple inverse technique works well for a single-location or a single-listener. But in real-world listening environments we must consider equalization for multiple listeners. Examples of such environments are typical classrooms, meeting rooms, and theatres. Then the problem of designing an equalizing filter that best suits all locations becomes difficult. This also holds for small rooms in which standing waves at low frequencies cause significant variations in the frequency responses at the listening positions [1,3].

Researchers model a room acoustic response at each particular listening position as a linear system [1,3]. So an impulse response or a room impulse response $h(n), n = 0, 1, 2, \dots$ (with frequency response $H(e^{j\omega})$) can model the effect the sound undergoes when it travels from a source to a receiver (from microphone to listener) [3]. The equalizing filter $g(n)$ is to eliminate the effect of the room impulse response. So for a single listener the convolution of the designed equalizing filter and the room impulse response at that particular location should result to the delayed delta function: $g(n) \otimes h(n) = \delta(n - T)$.

However, designing an equalizing filter that simultaneously flattens the overall frequency responses at all locations is not as trivial. Fig. 1 shows a diagram of how the equalizing filter should work for multiple listeners.

Researchers have worked on finding techniques to design the best possible equalizing filter for all listeners [1,3,7]. One technique is to find a representative room response for all locations [1,3]. Then a simple inverse filtering algorithm will give the desired equalizing filter. The fuzzy c-means (FCM) algorithm can perform the task of deriving room response prototypes and then a fuzzy system can combine each prototype to result in a representative response [1,3].

The FCM algorithm considers distances between room responses and put all responses into smaller clusters [5]. The algorithm does not partition the space of room responses: clusters can overlap and a room response can belong to the cluster with degree in the interval $[0,1]$. But the FCM treats each component in the room impulse response with equal weight. This restricts the condition that clusters can form. Some part of room impulse response may have more effect on the reverberation than the other parts. So we propose the use of a weighted fuzzy c-means (WFCM) algorithm to identify the representative response in each cluster. Then inverse filtering the representative response gives an equalizing filter.

Section 2 reviews the FCM and WFCM algorithms and presents a way to derive an equalizing filter. Section 3 shows the experimental results of equalization in a room. The results show that equalizing filter design with the use of WFCM performs better. Section 4 gives the conclusion and discusses future research.

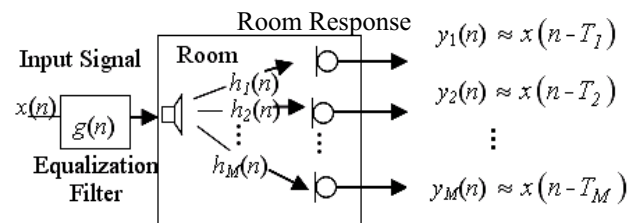


Fig. 1. An equalizing filter $g(n)$ is designed to simultaneously equalize sound reverberation for several locations in a room. Here, $x(n)$ is input signal, $y(n)$ is output signal and $h_i(n)$ is the room response.

2. Equalizing Filter Design and Performance Measure

2.1 Fuzzy c-means algorithm

The fuzzy c-mean (FCM) and its generalized version of weighted fuzzy c-mean (WFCM) clustering algorithms have found many applications in signal processing and communications, system modeling, pattern recognition, computer vision, and data mining [5]. In general a clustering procedure forms a new structure from

data in terms of a smaller number of clusters where a centroid or prototype represents each cluster. Clustering algorithm uses a measure of similarity to perform such grouping.

In our case the data are a set of N room impulse responses in terms of d -dimensional vectors $\underline{h}_i \in \mathbb{R}^d$ or a sequence $\{h_i(n)\}_{n=1}^d$ for $i = 1, 2, \dots, N$. Then a clustering algorithm partitions the data set into c nonempty subclasses or clusters $\Omega_1, \dots, \Omega_c$ (usually we let $1 < c < N$): for each sample $\underline{h}_i$ and cluster Ω_j either $\underline{h}_i \in \Omega_j$ or $\underline{h}_i \notin \Omega_j$ and if $\underline{h}_i \in \Omega_j$ then $\underline{h}_i \notin \Omega_k$ for $k \neq j$. This implies that $\underline{h}_i \in \Omega_j$ with degree 1 or 0. We can use the discrete membership function $\mu_j : \mathbb{R}^d \rightarrow \{0, 1\}$ to define the degree to which a sample $\underline{h}_i$ belongs to the cluster Ω_j . So $\mu_j(\underline{h}_i) = 1$ if $\underline{h}_i \in \Omega_j$ and $\mu_j(\underline{h}_i) = 0$ if $\underline{h}_i \notin \Omega_j$. Then room responses with strong similarity belong to the same cluster and prototype (or centroid) $\hat{\underline{h}}_j$ represents the cluster Ω_j . Since $c < N$ then a set of all prototypes $\{\hat{\underline{h}}_j\}_{j=1}^c$ is a compact representation of the data set.

Fuzzy c -means algorithm (FCM) does not restrict that a sample $\underline{h}_i$ from the data set must belong to only one cluster Ω_j for some j . FCM provides more flexibility to clustering by allowing that a sample can belong to several clusters simultaneously and with degrees as any real numbers between 0 and 1. The membership function is a fuzzy membership function $\mu_j : \mathbb{R}^d \rightarrow [0, 1]$. So FCM does not partition the set of all room responses: clusters can overlap and a room response may belong to each cluster with a degree between 0 and 1.

One of the commonly used similarity measures in FCM is the Euclidean distance d_{ik} between a sample $\underline{h}_k$ and a prototype $\hat{\underline{h}}_i$ [2]:

$$d_{ik}^2 = \|\underline{h}_k - \hat{\underline{h}}_i\|^2 \quad (1)$$

The centroid or prototype of the cluster i is

$$\hat{\underline{h}}_i = \frac{\sum_{k=1}^N \left(\mu_i(\underline{h}_k) \right)^2 \underline{h}_k}{\sum_{k=1}^N \left(\mu_i(\underline{h}_k) \right)^2} \quad (2)$$

$$\text{and } \mu_i(\underline{h}_k) = \left[\sum_{j=1}^c \frac{d_{ik}^2}{d_{jk}^2} \right]^{-1} = \frac{\frac{1}{d_{ik}^2}}{\sum_{j=1}^c \frac{1}{d_{jk}^2}} \quad (3)$$

for $i=1, 2, \dots, c$ and $k=1, 2, \dots, N$. Then an iterative procedure of the FCM algorithm as in [2] determines the cluster prototypes.

2.2 Proposed weighted fuzzy c-means algorithm for generating acoustical room response prototypes

Clustering with the usual unweighted distance in equation (1) may not be the best choice to determine the prototypes of several room responses since the distance does not give more weights to the region that has more effects on reverberation than the others. We propose a weighted distance FCM algorithm [5] for room response clustering:

$$d_{ik}^2 = \left(\underline{h}_k - \hat{\underline{h}}_i \right)^T W \left(\underline{h}_k - \hat{\underline{h}}_i \right) \quad (4)$$

where the weight matrix W is positive definite. We use a diagonal matrix W in our study: $W = \text{diag}(w_{11}, w_{22}, w_{33}, \dots, w_{NN})$. Other forms of positive definite matrix W may give better results. The weights that we used in our experiments were obtained from a linear function of the tap i .

$$w_{ij} = \begin{cases} 10 - \frac{i}{400} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad \text{for } i, j = 1, \dots, 4000 \quad (5)$$

2.3 Obtaining a representative response from room response prototypes

We combine the prototypes $\hat{\underline{h}}_i$ to obtain a representative room impulse response $\hat{\underline{h}}_{rep}$ using the standard additive model (SAM) of fuzzy system [6,8]:

$$\hat{\underline{h}}_{rep} = \frac{\sum_{j=1}^c \left(\sum_{k=1}^c \mu_j(\underline{h}_k) \right)^2 \hat{\underline{h}}_j}{\sum_{j=1}^c \left(\sum_{k=1}^c \mu_j(\underline{h}_k) \right)^2} \quad (6)$$

The SAM allows combining the throughputs of fuzzy systems before defuzzification. The advantage of SAM (as any additive fuzzy model) lies in its ability to approximate any continuous function on a compact (closed and bounded) domain [6].

After obtaining the representative of the room responses we designed and enhanced equalizing filter by inverting the minimum phase component of the representative room response $\hat{\underline{h}}_{rep}$ as in (10).

2.4 Inverse filters

Fig. 1 implies that the magnitude of the equalized room response $|E(e^{j\omega})|$ equals the magnitude of the product of the room response $H_i(e^{j\omega})$ and the equalizing filter $G(e^{j\omega})$

$$|E(e^{j\omega})| = |H_i(e^{j\omega})G(e^{j\omega})| \quad \forall i \quad (7)$$

We want $|E(e^{j\omega})| = 1$. In practice we cannot obtain an equalizing filter $G(e^{j\omega})$ that simultaneously satisfies (7)

for all locations i . So we use the representative response $H_{rep}(e^{j\omega})$ to derive an equalizing filter $G(e^{j\omega})$. So we set

$$|E(e^{j\omega})| = |H_{rep}(e^{j\omega})G(e^{j\omega})| = 1 \quad (8)$$

and obtain the inverse filter by

$$G(e^{j\omega}) = 1/H_{rep}(e^{j\omega}) \quad (9)$$

where $H_{rep}(e^{j\omega})$ is the representative room response and $G(e^{j\omega})$ is the desired equalization filter [4]. Note also that in our approach the corresponding equalization filter is obtained by inverting the minimum phase component:

$$G(e^{j\omega}) = 1/H_{rep,min}(e^{j\omega}) \quad (10)$$

Here $H_{rep,min}(e^{j\omega})$ is the minimum-phase component of

the representative room response $H_{rep}(e^{j\omega})$ and

$H_{rep,ap}(e^{j\omega})$ is the all-pass component:

$$H_{rep}(e^{j\omega}) = H_{rep,min}(e^{j\omega})H_{rep,ap}(e^{j\omega}) \quad (11)$$

Inverse filters can also be obtained from other methods such as inverse system identification, Least Mean Square (LMS) algorithm, etc.

2.5 Spectral Deviation Measure

This section shows the performances using a spectral deviation measure [1,3,7] that has the form

$$\sigma_E = \sqrt{\frac{1}{P} \sum_{i=0}^{P-1} \left[10 \log_{10} |E(e^{j\omega_i})| - AVG \right]^2} \quad (12)$$

where $AVG = \left[\frac{1}{P} \sum_{i=0}^{P-1} 10 \log_{10} |E(e^{j\omega_i})| \right]$. This measure

provides a measure of residual spectral distortion from a constant level. Flatter room responses will have lower values of σ_E .

3. Experimental Results

This section shows the results of room equalization using methods in [1,3] combining with the WFCM algorithm. Fig. 2 shows a room configuration as a reverberant enclosure to test our algorithm. The microphone locations were arranged in a rectangular grid at $N = 9$ locations. The spacing between the microphones in both directions of the grids was roughly the same at about 1 m. The loudspeaker was placed about 0 degree to a vertical axis at a distance of about 1 m from the right bottom of the grid. The axis passed through the right side location of the grid. Both the microphone and the loudspeaker were positioned approximately one meter from the ground.

The number of clusters determined was $c = 3$. Room responses obtained from each of the nine microphone locations were clustered using (2) with the distance (4) and weights (5) to obtain the prototypes $\hat{h}_c$. Then the prototypes are combined to obtain $\hat{h}_{rep}$ as in (6).

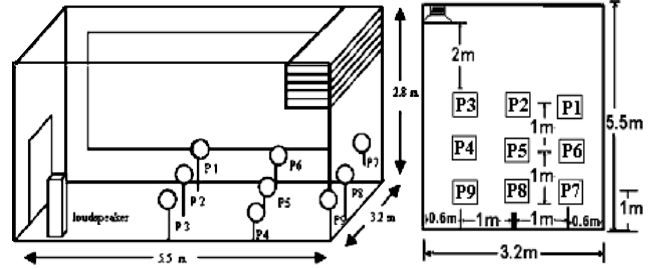


Fig. 2. Nine locations in a room.

The magnitude responses measured at the nine receiver locations are shown in Fig. 3 with the corresponding time-domain in Fig. 4. Magnitude responses of the representative room responses clustered by FCM and WFCM algorithms are depicted in Fig. 5. The results of equalized room magnitude responses using FCM and WFCM algorithms are depicted in Fig. 6 and Fig. 7. The corresponding results of spectral deviation measure σ_E are tabulated in Table 1. The results show that the equalized room magnitude responses using WFCM are flatter than the results of FCM algorithm.

The proposed WFCM algorithm that we used to obtain room response prototypes yielded better results than FCM algorithm for most listener locations. The only exception is that the proposed WFCM performed slightly worse than the FCM algorithm in location P9. The linear weight tended to give more emphasis on direct paths of room responses than others and thus tended to give better prototypes than the non-weighted FCM.

Table 1 Experimental Results: Spectral deviation measure (σ_E)

Location	Original Room Response	Equalized Room Response	
		FCM Algorithm	Linear Weighted FCM Algorithm
P1	14.45	5.43	3.34
P2	14.69	4.76	3.82
P3	14.53	4.43	3.54
P4	14.96	5.56	3.42
P5	14.89	4.87	3.53
P6	14.78	4.86	4.22
P7	14.95	5.64	5.53
P8	15.25	6.23	4.98
P9	14.66	6.15	6.17

4. Conclusion

This paper shows the use of WFCM clustering technique to calculate response prototypes from multiple acoustical room responses. We combined these prototypes using fuzzy SAM to obtain a representative response. Then we obtained the equalization filter from the representative response using inverse filtering technique. We compare the proposed WFCM algorithm with the conventional FCM algorithm. The spectral deviation measure at several room locations show that the WFCM can give better results. This shows that different weights of the distance metric in the WFCM algorithm can give better room response prototypes. Future research will

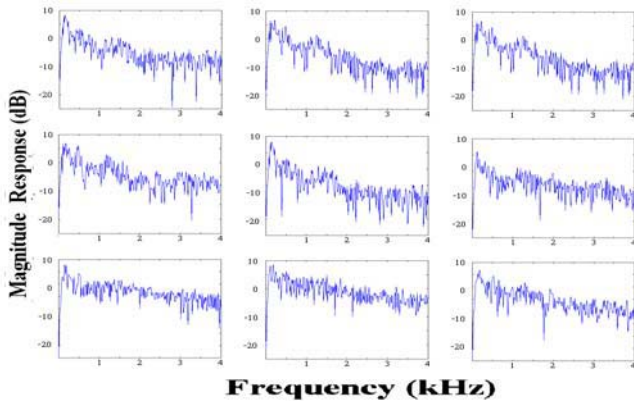


Fig. 3. Magnitude responses at the nine locations

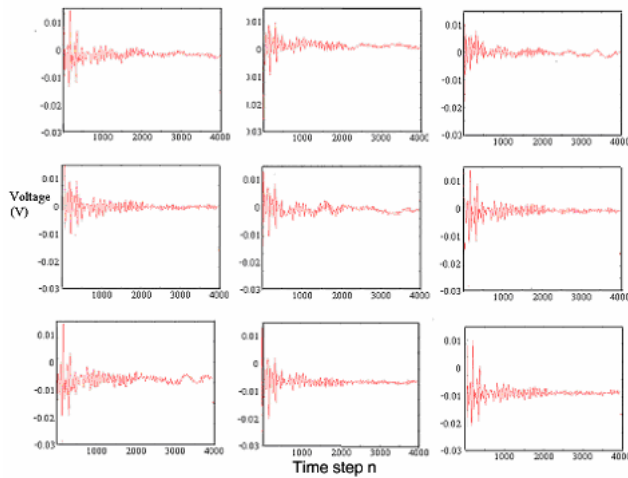


Fig. 4. The room responses in time domain $h_i(n)$.

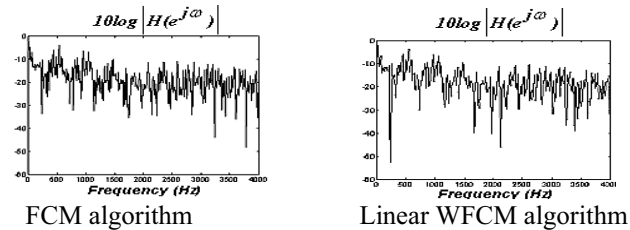


Fig. 5. Magnitude responses of representative room responses clustered by FCM and WFCM algorithms.

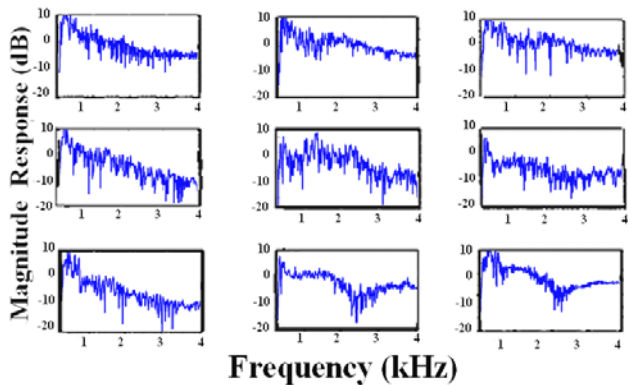


Fig. 6. Equalized room responses by FCM algorithm. The equalizing filter $g(n)$ has 500 taps at 8 kHz sampling frequency.

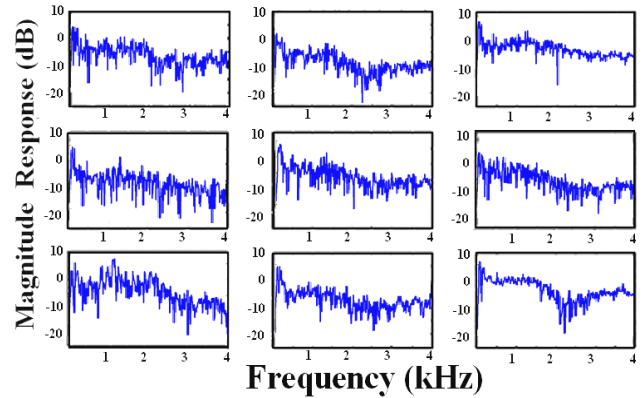


Fig. 7. Equalized room responses by WFCM algorithm. The equalizing filter $g(n)$ has 500 taps at 8 kHz sampling frequency.

focus on how to obtain the best weights for the WFCM to obtain the prototypes. Another possibility is to optimally tune the fuzzy SAM combiner used to obtain the representative room responses from prototypes.

Acknowledgement

Thailand Research Fund Grant RSA4680001 supported this research. S. Turmchokkasam was also supported by a graduate scholarship from Bangkok University to study at Thammasat University.

References

- [1] S. Bharitkar, C. Kyriakakis, "A Cluster Centroid Method for Room Response Equalization at Multiple Locations," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp.55-58, October 2001.
- [2] J. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms," *Plenum Press New York*, 1981.
- [3] S. Bharitkar, C. Kyriakakis, "New Factors in Room Equalization Using a Fuzzy Logic Approach," *Convention Paper Audio Engineering Society*, pp.1-13, September 2001.
- [4] J. Mourjopoulos, "On the variation and invertibility of room impulses response function," *Journal of Sound and Vibration*, Vol. 102(2), pp.217-228, 1985.
- [5] J. M. Leski, "Generalized Weighted Conditional Fuzzy Clustering," *IEEE Transactions on Fuzzy systems*, Vol. 11(6), December 2003.
- [6] B.Kosko, *Fuzzy Logic and Neural Network Handbook*, (Ed.:C.Chen), McGraw-Hill. New York.1996
- [7] Y. Haneda S. Makino, Y. Kaneda, "Multiple-Point Equalization of Room Transfer Function by Using Common Acoustical Poles," *IEEE Transactions on Speech and Audio Processing*, Vol. 5(4), pp. 325-333, July 1997.
- [8] B. Kosko, "Fuzzy System as Universal Approximators," *IEEE Transactions on Neural Networks*, Vol. 43(11), pp. 1329-1333, Nov. 1994.

A Study of Stochastic Resonance Effect in Object Segmentation with Color Thresholding

Sittichote Janpaiboon and Sanya Mitaim

Department of Electrical Engineering, Faculty of Engineering
Thammasat University, Pathumthani, 12120, Thailand
E-mail: toring99@hotmail.com, msanya@engr.tu.ac.th

Abstract

Object segmentation is one of the most important tasks in image analysis and computer vision. Color thresholding provides a fast and simple scheme for such task. But it is sensitive to lighting conditions and other noise effect in images obtained from the real world applications. This paper shows that addition of a small amount of noise can improve the accuracy of such color object segmentation. It shows that this “stochastic resonance” or SR effect does occur for various performance indices that measure how well an object is segmented from the background. We use mutual information, error pixels count, and object position error as performance measures to compare the segmented images obtained from the original thresholding algorithm with the proposed SR-extended algorithm. The study confirms by examples that addition of noise can robustify the color thresholding algorithm and thus provides an alternative for engineers when they need to detect color objects in noisy input images.

Keywords: stochastic resonance, noise processing, image segmentation, color thresholding, mutual information, error pixels count, object position error.

1. Introduction to SR

Stochastic resonance or SR is noise benefit phenomenon. SR occurs when noise enhances a faint signal in a nonlinear system. The system’s performance such as signal-to-noise ratio, cross-correlation, or mutual information increases when the small amount of noise is added to the system and so the system has nonzero-noise optimality. The system’s nonlinearity is often as simple as a memoryless threshold. SR occurs in physical systems such as ring lasers [12], threshold hysteretic Schmitt triggers [6], superconducting quantum interference devices (SQUIDs) [8], flash A/D [15], chemical systems [5]. SR also occurs in biological systems such as rat [2], crayfish [4], cricket [11], river paddlefish [14], and in types of model neurons [13].

Color image segmentation applies to several image analysis and computer vision. Object detection, positioning, classification, and other tasks employ segmentation. Color thresholding is a simple algorithm that can effectively segment objects in controlled lighting condition [1],[14],[18]. The algorithm uses a set of thresholds for each color dimension. But in many applications the conditions of acquired images do not match with the preset thresholds. The conditions of the acquired images can be the difference in lighting conditions derived from physical environments as well as pixel noise occurred in

optical sensors of a digital camera [10]. These deviated conditions can greatly worsen the performance of the color thresholding algorithm to segment objects if the thresholds do not match with the image’s pixel colors. We can adapt thresholds to each image’s lighting conditions but the algorithm is complex [14].

This paper shows that noise can enhance the color threshold algorithm. The proposed noise-added algorithm extends the single-stage color thresholding to a multiple-stage scheme. The experiments tested the SR-extended algorithm with synthetic images of a plain-color circle as our object on a plain background with different color. We modified the synthetic images to be noisy and blurred versions of the original image.

We used several performance indices such as mutual information, error pixel count, and object’s position error to measure how well the algorithm works. The results show that the addition of noise in our extended SR-algorithm can improve these measures when the images already contain noise or other distortion. This improvement still holds for images of actual objects taken with a digital camera. The extended SR-threshold algorithm can improve the performances when the images are not from good environment settings.

2. Color Image Thresholding

The thresholding method described here can be used with general multidimensional, color spaces that have discrete component color levels [1],[7] such as RGB, HIS. This paper uses the RGB color space. A color object is segmented with a set of six threshold values, two for each color dimension in the RGB color space. A pixel which has color component value in the interval of the two thresholds, lower and upper thresholds, is represented as one or otherwise is zero:

$$g(y) = \begin{cases} 1 & \theta_{\min} \leq y \leq \theta_{\max} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where y is the image pixel value, $\theta_{\min}$ is the lower threshold and $\theta_{\max}$ is upper threshold.

Red, Green and Blue color components of the input image are compared and classified to three binary images of each color spaces by the set of RGB thresholds. The AND operation of the pixels of these three binary images gives a target object region.

3. Performance Measures

3.1 Mutual Information Measure

We use Shannon mutual information [3] to measure the SR effect. The discrete Shannon mutual information

of the random variable S and Y has the form

$$I(S, Y) = H(Y) - H(Y|S) \quad (2)$$

$$= -\sum_y P_Y(y) \log P_Y(y) + \sum_s \sum_y P_{S,Y}(s, y) \log P_{Y|S}(y|s) \quad (3)$$

$$= \sum_{s,y} P_{S,Y}(s, y) \log \frac{P_{S,Y}(s, y)}{P_S(s)P_Y(y)} \quad (4)$$

Here $P_S(s)$ is the probability density of the random variable S , $P_Y(y)$ is the probability density of the random variable Y , $P_{Y|S}(y|s)$ is the conditional density of the random variable Y given the random variable S , and $P_{S,Y}(s, y)$ is the joint density of the random variable S and Y .

3.2 Error Pixels Count

This measure directly describes mistaken pixels between two binary images. The bit-wise XOR operation shows the pixels in which their binary values do not match the pixels in another binary image at the same locations. So we can obtain the amount of the error pixels by counting the results of 1 of the bit-wise XOR. The error pixels count C_e between an original binary image S and an output binary image Y having $m \times n$ of dimension has the form

$$C_e = \sum_{i=1}^m \sum_{j=1}^n S_{ij} \otimes Y_{ij} \quad (5)$$

$$\text{where } S_{ij} \otimes Y_{ij} = \begin{cases} 0 & \text{if } S_{ij} = Y_{ij} \\ 1 & \text{if } S_{ij} \neq Y_{ij} \end{cases} \quad (6)$$

3.3 Position Error

In many applications of image processing [1],[18], an object positioning has been a useful function. But noise sensitivity in the image segmentation process causes error in estimating the object region and so error of the estimated position. Thus the position error is a significant measure for showing the SR effect of our approach. We apply this measure to the binary images of classified target objects using the color image thresholding. We determine the position of an object by calculating the centroid of the classified region (region classified as an object):

$$\begin{bmatrix} c_x \\ c_y \end{bmatrix} = \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} x_i \\ y_i \end{bmatrix} \quad (7)$$

where c_x, c_y are row and column coordinates of the centroid. Then the position error P_e is

$$P_e = d(c_{in}, c_{out}) = \sqrt{(c_{xin} - c_{xout})^2 + (c_{yin} - c_{yout})^2} \quad (8)$$

where $\begin{bmatrix} x_i & y_i \end{bmatrix}^T$ is the coordinate vector of a pixel which would be a target object, N is the number of

pixels of an object region, c_{in} and c_{out} are centroids of the target object of original image and noisy image:

$$c_{in} = \begin{bmatrix} c_{xin} & c_{yin} \end{bmatrix}^T \text{ and } c_{out} = \begin{bmatrix} c_{xout} & c_{yout} \end{bmatrix}^T$$

4. Proposed SR Segmentation System and Experimental Results

4.1 SR Segmentation System

We propose a new image segmentation technique using N stages of noisy RGB color thresholding system. Each stage simply adds independent white Gaussian noise to a noisy input image before performing the usual color thresholding. Binary output images of all stages are combined with an OR operation to obtain a binary output image Y of the SR segmentation system. We measure the performance of the SR segmentation system using mutual information $I(S, Y)$, error pixels count C_e , and position error P_e to determine how the binary output image Y match the original binary image S as shown in Fig 1.

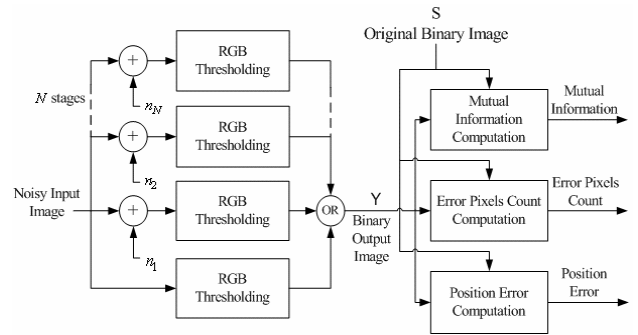


Fig 1. The SR segmentation system consists of the multiple-stage RGB color thresholdings where N is the number of stages and n_i is independent noise for stage i . S and Y are original and output binary images.

4.2 Experimental Results: Synthetic Images

We first tested the SR segmentation system with synthetic images. The original image consisted of an orange circle (as an object) on the green background. Then we added Gaussian noise to the original image and performed the blurring and shadowing operations on it to produce noisy test images shown in Fig. 2. By using this synthetic image we could precisely determine the actual pixels that belong to the object.

The SR segmentation system in Fig. 1 has $N = 1$ stage of RGB color thresholding system. The additive noise is Gaussian. The segmentation system will separate the target object (the orange circle) from the noisy image. In this experiment the RGB color thresholding algorithm uses the following thresholds (based on the 0-255 levels of intensity):

$$\text{Red: } \theta_{R\min} = 120, \quad \theta_{R\max} = 255$$

$$\text{Green: } \theta_{G\min} = 50, \quad \theta_{G\max} = 120$$

$$\text{Blue: } \theta_{B\min} = 0, \quad \theta_{B\max} = 40$$

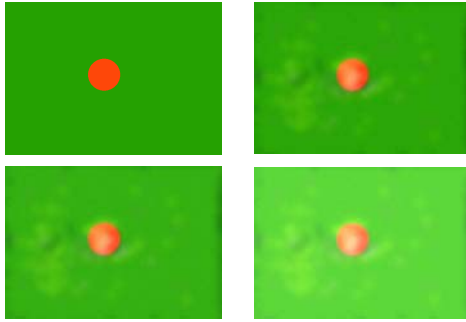


Fig 2. Synthetic images of an orange object on a green background. The top-left image is the original image and the top-right is a noisy, blurred, and shadowed version of the original one. The images at the bottom are brightness-increase versions of the noisy image of the top-right position. We label the top-right image as “Brightness 0%,” the bottom-left as “Brightness 10%,” and the bottom-right as “Brightness 50%”. The SR results of these images are in Fig. 3 (a)-(c). We do not include the Brightness 20% image here.

The images that we tested are the noisy images (with 0%, 10%, 20% and 50% brightness) shown in Fig. 2. We did not show the test image with 20% brightness here. Fig. 3 (a)-(c) show the performances when the noise standard deviation increases. The system has nonzero optimal noise level and thus shows the SR effect for the

synthetic test images. The effect is more pronounced for very noisy images.

Object positioning error can be a good indicator as well. Fig. 4 (a)-(d) graphically show the positions obtained from SR and non-SR segmentation systems comparing to the actual position in all four test images in Fig. 2. The Figure shows that the positions obtained from the SR-system (with *optimal* amount of noise) are more accurate than the ones obtained from the original (non-SR) segmentation system. A specific case of Fig. 4 (d) shows that conventional color thresholding cannot detect any pixels as an object. So we do not have an estimate of the position in this case (and so there is no white circle shown). But the SR segmentation algorithm can find some pixels that belong to the object and gives an estimate of the object position.

4.3 Experimental Results: Real Images

Here we tested the SR segmentation system on images taken from the real world using a digital camera. The images show an orange golf ball on the green carpet as a background. We measured the position of the golf ball against the known mark on the carpet. Three images shown in Fig. 3 (d)-(f) are taken in three different illuminations (different light settings).

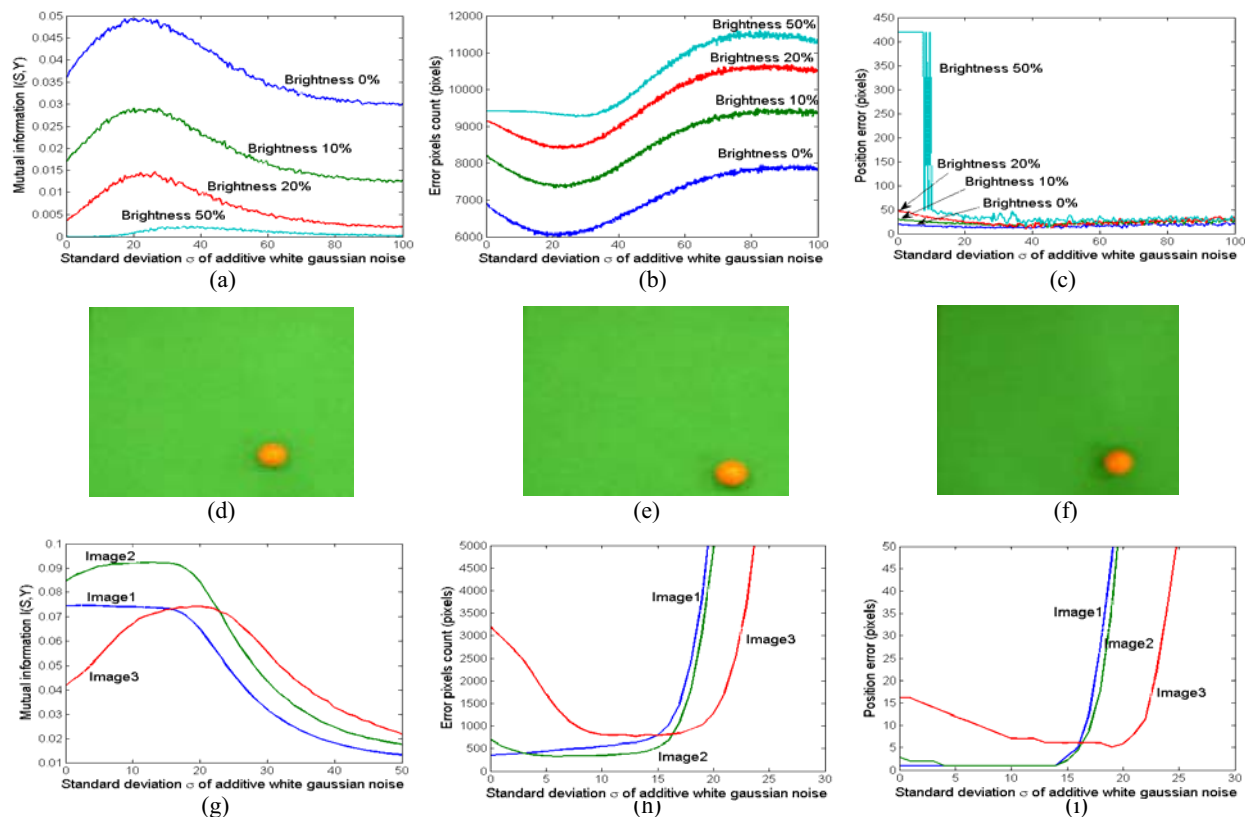


Fig 3. SR effect in image segmentation for both synthetic images real images using SR segmentation system in Fig. 1. The performance measures are mutual information in (a) and (g), error pixels count in (b) and (h), and position error in (c) and (i). The additive noise in the SR segmentation system is Gaussian. The graphs (a)-(c) show the results of the SR segmentation system with $N = 1$ for the four synthetic images. The graphs (g)-(i) show the results of SR segmentation system with $N = 50$ for the actual images in (d)-(f) (Image1 – Image3).

The SR segmentation used in this system is the same as in Section 4.2 (the system in Fig. 1) but now with $N = 50$ stages. The noise in each stage is independent Gaussian noise. The thresholds used for the RGB color thresholding scheme in each stage are (based on the 0-255 levels of intensity):

$$\text{Red: } \theta_{R_{\min}} = 150, \quad \theta_{R_{\max}} = 255$$

$$\text{Green: } \theta_{G_{\min}} = 110, \quad \theta_{G_{\max}} = 205$$

$$\text{Blue: } \theta_{B_{\min}} = 0, \quad \theta_{B_{\max}} = 50$$

These set of thresholds are manually set for optimal segmentation of Image1. So they are different from the optimal thresholds for the synthetic image “Brightness 0%” in Fig 2. Fig. 3 (g)-(i) and Fig. 4 (e)-(g) show how white Gaussian pixel noise can improve our image information and segmentation in terms of mutual information, error pixels count, and position error. The results show that noise can improve the segmentation performances when the preset RGB thresholds do not match the lighting conditions. The *perfect* threshold case of “Image1” also shows that a small amount of noise does not ruin the accuracy of segmentation while it can increase the accuracy of the noisy images (“Image2” and “Image3”).

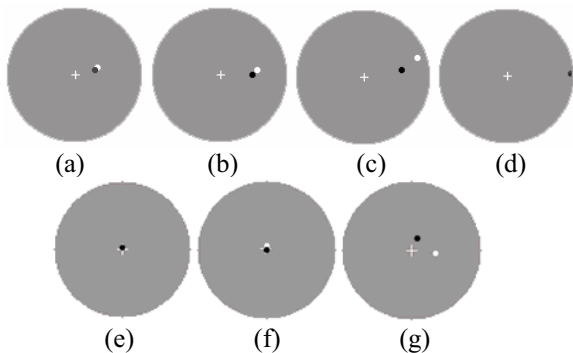


Fig 4. Actual and estimated positions of the objects. The white ‘+’ denotes the actual positions of the objects (at the center). The white circles are positions obtained from non-SR segmentation systems (original RGB color thresholding). The dark circles represent positions from the optimal SR effect in segmentation. The SR system gives estimated position with less error. The top four panels (a)-(d) are results of the four synthetic images where the SR optimal noise has standard deviation (SD) of 22, 24, 24, and 30. The bottom three panels (e)-(g) are results of the three real images with SR optimal noise at SD of 13, 13, and 19. These positions refer to the SR optimal results from Fig. 3.

5. Conclusions

This paper studies the effect of noise addition in object segmentation using RGB color thresholding scheme. The paper shows that the stochastic resonance or SR effect occurs when we use this algorithm to segment a color object from a plane background in an image that is not perfectly captured or that the preset thresholds do not perfectly match the color distribution.

The experimental results suggest that addition of white Gaussian noise can enhance the mutual information between the correct image and the image acquired from the actual environment. The number of error pixels

and the object’s position error also decrease when we add the right amount of noise to the noisy input images. These results confirm that noise can robustify segmentation using RGB thresholding algorithm. For an image with good lighting condition (so the RGB thresholds perfectly match its histograms) and with no other distortion, a little amount of noise only worsens the performance of the segmentation a little white noise can significantly improve the performance of the noisy ones. So for real word applications in which captured images contain a lot of interferences, engineers might consider using optimal noise to design a more robust image segmentation algorithm.

Acknowledgment

Thailand Research Fund Grant RSA 4680001 supported this research.

References

- [1] J. Bruce, T. Balch and M. Veloso, “Fast and Inexpensive Color Image Segmentation for Interactive Robots”, *IEEE/RSJ Conf. Intel. Robot and Systems*, pp. 2061-2066, 2000.
- [2] J. J. Collins, T. T. Imhoff, and P. Grigg, “Noise-enhanced information transmission in Rat SA1 cutaneous mechanoreceptors via aperiodic stochastic resonance,” *J. Neurophysiol.*, vol. 76, no. 1, pp. 642-645, Jul. 1996.
- [3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. NewYork: Wiley, 1991.
- [4] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, “Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance,” *Nature*, vol. 365, pp. 337-340, Sep. 1993.
- [5] M. I. Dykman, T. Horita, and J. Ross, “Statistical distribution and stochastic resonance in a periodically driven chemical system,” *J. Chem. Phys.*, vol. 103, no. 3, pp. 966-972, Jul. 1995.
- [6] S. Fauve and F. Heslot, “Stochastic resonance in a bistable system,” *Phys. Lett. A*, vol. 97, no. 1,2, pp. 5-7, Aug. 1983.
- [7] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice-Hall, 2002.
- [8] A. Hibbs, E. W. Jacobs, J. Bekkedahl, A. R. Bulsara, and F. Moss, “Signal enhancement in a r.f. SQUID using stochastic resonance,” *IlNuovo Cimento*, vol. 17D, pp. 811-817, 1995.
- [9] B. Kosko, and S. Mitaim, “Stochastic resonance in noisy threshold neurons,” *Neural Networks*, vol. 16, pp. 755-761, 2003.
- [10] R. Lenz and D. Fritsch, “Accuracy of videometry with CCD sensors”, *J. Int. Soc. Photogramm. Remote Sens.*, vol. 45, pp. 90-100, Jun. 1990.
- [11] J. E. Levin and J. P. Miller, “Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance,” *Nature*, vol.380, pp. 165-168, Mar. 1988.
- [12] B. McNamara, K. Wiesenfeld, and R. Roy, “Observation of stochastic resonance in a ring laser,” *Phys. Lett. A*, vol. 60, no. 25, pp. 2626-2629, Jun. 1988.
- [13] S. Mitaim, and B. Kosko, “Adaptive stochastic resonance in noisy neurons based on mutual information,” *IEEE Trans. Neural Networks*, vol. 15, no. 6, pp. 1526-1540, Nov. 2004.
- [14] W. N. Lie, “Automatic Target Segmentation by Locally Adaptive Image Thresholding”, *IEEE Trans. Image Processing*, vol. 4, pp. 1036-1046, Jul. 1995.
- [15] D. Rousseau and F. Chapeau-Blondeau, “Suprathreshold Stochastic Resonance and Signal-to-Noise ratio Improvement in Arrays of Comparators”, *Phys. Lett. A*, pp. 280-290, 2004.
- [16] D. F. Russell, L. A. Wilkens, and F. Moss, “Use of behavioral stochastic resonance by paddle fish for feeding,” *Nature*, vol. 402, pp. 291-294, Nov. 1999.
- [17] C. M. Tsai and H. J. Lee, “Binarization of Color Document Images via Luminance and Saturation Color Features”, *IEEE Trans. Image Processing*, vol. 11, pp. 434-451, Apr. 2002.
- [18] Z. Wasik and A. Saffiotti, “Robust Color Segmentation for the RoboCup Domain”, *Int. Conf. Pattern Recognition*, Aug. 2002.