Noise-Driven Manifestation of Learning in Mature Neural Networks

Christopher Monterola and Caesar Saloma*

National Institute of Physics, University of the Philippines, Diliman, Quezon City 1101, The Philippines

(Received 1 April 2002; published 14 October 2002)

We show that the generalization capability of a mature thresholding neural network to process above-threshold disturbances in a noise-free environment is extended to subthreshold disturbances by ambient noise without retraining. The ability to benefit from noise is intrinsic and does not have to be learned separately. Nonlinear dependence of sensitivity with noise strength is significantly narrower than in individual threshold systems. Noise has a minimal effect on network performance for above-threshold signals. We resolve two seemingly contradictory responses of trained networks to noise — their ability to benefit from its presence and their robustness against noisy strong disturbances.

DOI: 10.1103/PhysRevLett.89.188102 PACS numbers: 84.35.+i, 07.05.Mh, 87.19.La, 89.75.Hc

Noise-enhanced sensitivity has already been observed in several biological neural systems such as (i) periodically stimulated crayfish neurons [1], (ii) information encoding in the primary auditory nerve of the squirrel monkey [2], (iii) prey targeting of paddle fish [3] and crickets [4], (iv) human visual perception [5], and (v) heart rate compensation for blood pressure regulation [6]. However, no one has yet shown that the ability to benefit from noise is intrinsic in trained (mature) neural networks. Recently [7,8], models of single spiking or integrate-and-fire (IF) neurons in noisy environments were analyzed individually, but their collective behavior in trained networks is yet to be studied. Essentially, an IF neuron is a threshold system that provides an operational model for real biological neurons [9].

Here, we show that the generalization capability of a thresholding neural network that already learned to process specific features of above-threshold (strong) disturbances in a noise-free environment is extended to subthreshold (weak) disturbances by ambient noise without retraining. Learning is a collective behavior and generalization is its key manifestation. We therefore demonstrate that learning (and not merely detection) could be manifested in previously inaccessible domains with suitable ambient noise.

The three-layer, single-output feedforward network consists of M thresholding input nodes for sampling an external disturbance, and H hidden nodes that exhibit a sigmoidal response. Each input node integrates all incoming contributions including ambient noise and fires when the threshold condition is met. Each hidden node sums the outputs from M input nodes. The output neuron integrates the inputs from the H hidden neurons and yields the network output in accordance with a sigmoidal response. Hence, the individual nodes are highly simplified versions of an IF neuron.

The network samples and processes a disturbance instantaneously — another simplification that permitted us to analyze relevant network properties about learning. Similar analysis is extremely difficult with networks of full-model IF neurons due to phase synchronization and temporal response considerations.

The mature network has a finite dynamic range and could perform various frequency filtering operations with bandlimited inputs. Filtering, which is an essential aspect of feature recognition [10], permits biological systems to distinguish a predator from a prey and vice versa.

We verify if a mature network could benefit "naturally" from ambient noise — a crucial capability for adaptation and survival in difficult environments. Trained artificial networks perform reliably even with noisy inputs, which is their most important advantage over other statistical techniques in pattern recognition and other related applications [11–13]. We therefore resolve two seemingly contradictory responses of networks to noise: their performance robustness against noisy strong disturbances and its ability to utilize noise to detect weak disturbances.

Description of network.—The network response Ψ_N^q is described by $\Psi_N^q = f_o(\sum_h d_{ho}^q y_h^q)$, where $y_h^q = f_H(\sum_m w_{mh}^q x_m^q)$, $w_{mh}^q(d_{ho}^q)$ represents the interconnection weight from the mth input to the hth-hidden node after the qth iteration, and $m = 1, 2, \ldots, M$; and $h = 1, 2, \ldots, H$. The hidden and output response (activation) functions are sigmoidal: $f_H(z) = f_o(z) = [1 + \exp(-z)]^{-1}$, where $0 \le f_H(z) \le 1$.

The disturbance is described by $i(x) = (A/2)[\sin(2\pi f x) + 1]$, where $0 \le A$ (7-bit resolution) ≤ 1, f = spatial frequency, $x = m\Delta x$, and Δx is the sampling interval. The sampling of $i(x)$ by the uniformly separated M input nodes (separation $= \Delta x = 1$) satisfies the Nyquist sampling criterion. Element $i(m)$, which is not unique in the sampled sequence $\{i(m)\}$, represents the input to the mth input node which outputs a "1" if $i(m) \ge A_{th} = 0.5$. Otherwise, the node outputs a zero. Only $M/2$ f values are possible: $f = 1/M, 2/M, \ldots, 1/2$ (cutoff frequency), where $T = M$ is the sampling period.

We utilize a network with $M = 50$ and $H = 10$ ($2^6 M = 3200$). Other values within, $6 \leq H \leq 35$, were also highly trainable. Our training set consists of 128 $\{i(m)\}$ corresponding to $f = f_r$ at 2^7 different A values, plus 2432 other $\{i(m)\}$ corresponding to randomly chosen $f(\neq f_r)$ and A values.

The $\{w^q_{mh}\}$ are determined via the gradient descent backpropagation method [14] by minimizing a nonnegative cost function: $E^q = 0.5(\Psi^q_N - \Psi^q_t)^2$. The desired response Ψ_t is equal to 1.0 if (i) $A > A_{th} = 0.5$, and (ii) f value of $i(x)$ is equal to a preset value f_r. The weight-change rule is $w^{q+1}_{mh} = w^q_{mh} - \eta[(\partial E^q)/(\partial w^q_{mh})]$, with the learning rate $\eta = 0.001$. A similar relation holds for $\{d^q_{ho}\}$. The network learned quickly ($E^{500} = 0$) for any f_r except at $f_r = 1/2$ which is prone to aliasing [15]. The mature network generalizes perfectly with a test set that consists of the 640 remaining $\{i(m)\}$. Note that the network could not be trained with a set of weak $\{i(m)\}$ only. It also could not learn the task if the Nyquist criterion is not satisfied due to aliasing.

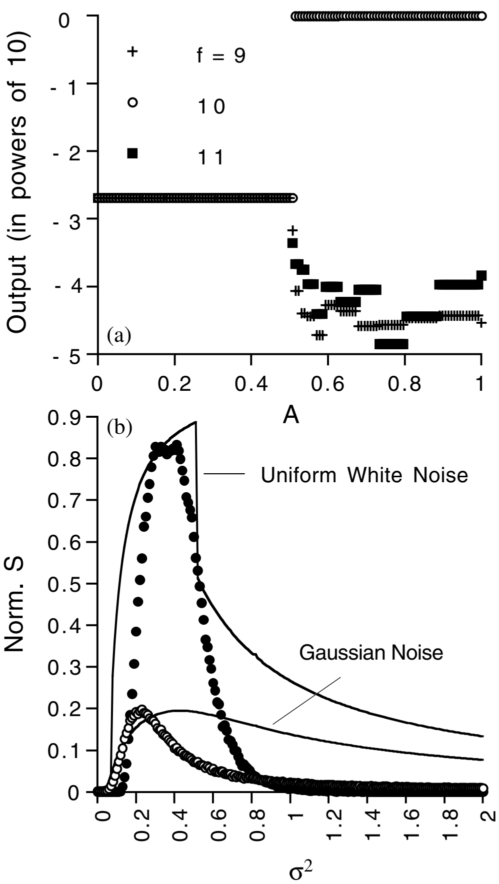

FIG. 1. Performance of a mature network ($f_r = 1/5$, 5000 trials). (a) $\Psi^{500}_N(A)$ for $f = 9/50$ (cross-hair), 1/5 (circle), 11/50 (filled squares), and $\sigma^2 = 0$; and (b) normalized $S_N(\sigma^2)$ for UWN (filled circles) and GN (circles) with $A = 0.4$. Solid curves in (b) are $S_T(\sigma^2)$ (UWN: $k = 2$; GN: $k = 0.28$) for an STS which outputs a "1" if $i(x) > A_{th} = 0.5$, and "0" otherwise. $S_T(\sigma^2)$ behavior is f_r independent.

Performance of trained network for different f_r.—Figure 1(a) plots Ψ^{500}_N vs A for the test set. Without noise, the mature network recognizes all $i(x)$ with $f = f_r = 1/5$, only if $A > 0.5$. However, noise-dithered weak disturbances, $i(x) + n_\sigma(x)$, with $f = f_r$, become recognizable at the right noise strength (variance) σ^2.

The dependence of the network's success rate S_N with σ^2 is nonlinear for both uniform-white (UWN) and Gaussian noise (GN) [see Fig. 1(b)]. For $\sigma^2 < 0.2$, $S_N(\sigma^2)$ increases with σ^2. The detection onsets are at $\sigma^2 \approx 0.1$, and $\sigma^2 \approx 0.17$, for GN and UWN, respectively. UWN dithering yields higher peak S_N values and decreases more rapidly with σ^2 such that $S_N < 0.001$ at $\sigma^2 > 1.4$. With GN, $S_N < 0.07$ at $\sigma^2 > 2$. Similar $S_N(\sigma^2)$ behavior was also observed with other f_r except at $f_r = 1/2$. The benefits of noise dithering were unseen in non-f_r weak $i(x)$.

We compare the $S_N(\sigma^2)$ with the $S_T(\sigma^2)$ of a simple threshold system (STS, threshold $= A_{th}$) that has already been characterized by Gammaitoni [16]. The onsets of $S_T(\sigma^2)$ are found at $\sigma^2 \approx 0.08$ (UWN) and 0.03 (GN), which is expected since $A = 0.4$ and $A_{th} = 0.5$. The tuning parameter k [16] was adjusted to make the peak $S_N(\sigma^2)$ and $S_T(\sigma^2)$ comparable for a given noise type.

The $S_N(\sigma^2)$ is narrower and its peak value occurs at a smaller σ^2 (GN: $\sigma^2 \approx 0.2$; UWN: $\sigma^2 \approx 0.3$) than that of $S_T(\sigma^2)$. It also decreases much faster to zero. With GN, the full width at half maximum of $S_N(\sigma^2)$ is approximately equal to that of $[S_T(\sigma^2)]^7$. With UWN, the width of $S_N(\sigma^2)$ is approximately equal to that of $[S_T(\sigma^2)]^2$.

At this point, we present only the UWN-driven $S_N(\sigma^2)$ dithering because of their higher peak values and faster decay with increasing σ^2.

Performance of trained network for different A values.—Noise-aided generalization is evaluated further at different weak A values. Figure 2 reveals that detection

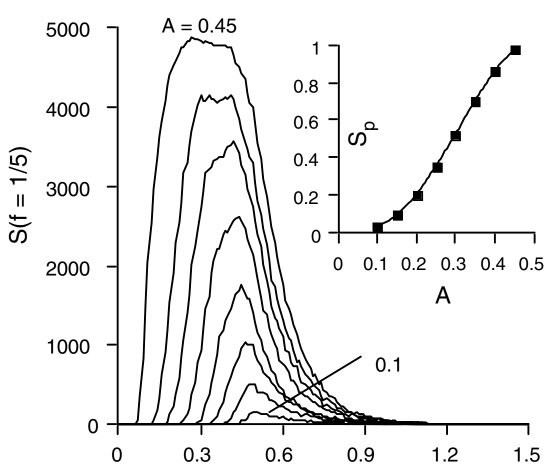

FIG. 2. Performance of mature network ($f_r = 1/5$, 5000 trials). Normalized $S_N(\sigma^2)$ for $f = 1/5$ and $A = 0.45$ (topmost curve), 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, and 0.1 (lowermost curve). Inset plot: Peak S_p vs A.

begins at $\sigma^2 \approx (0.5 - A)$. Within $0.10 \leq A \leq 0.45$, the peak S_p of $S_N(\sigma^2)$ decreases with A according to (best-fit curve) $S_p = -13.727A^4 - 9.817A^3 + 17A^2 - 2.52A + 0.127$, with an average standard deviation of ± 0.0049. Via the Newton-Raphson method, the smallest detectable A is determined to be at 0.08 ± 0.0008, representing a sensitivity enhancement of 625%.

Multifrequency disturbances.—Feature extraction involves the recognition of a specific set of f components in a multifrequency disturbance [10]. The effect of ambient noise is tested on mature networks that could perform low-, band-, or high-pass filtering of strong $i(x)$. Low-pass and high-pass filtering permit the determination of the energy content and fine details of a disturbance, respectively.

A network ($M = 50$, $H = 10$) was trained to identify a set of equiamplitude $i(x)$ with preselected f values. Four different sets are considered ($A = 0.4$): Set I ($f_r = 1/50, 1/5, 21/50, 23/50$), Set II ($1/50 \leq f_r \leq 1/10$), Set III ($1/5 \leq f_r \leq 7/25$), and Set IV ($21/50 \leq f_r < 1/2$). The network training set consists of 4×128 $\{i(m)\}$ corresponding to the four f_r at 2^7 different A values each, plus 2048 other $\{i(m)\}$ corresponding to randomly chosen $f(\neq f_r)$ and A values. Training was rapid and perfect results were achieved in less than 500 iterations in any of the sets.

In Fig. 3(a) are the resulting $S_N(\sigma^2)$ for Set I, which have onsets at $\sigma^2 \approx 0.1$. Their peaks are within $0.3 \leq \sigma^2 \leq 0.4$, with an average value of $\langle S_p \rangle = 4211 \pm 290$. Weak $i(x)$ with non-Set I f values are hard to detect (see inset S plots). The standard deviation of $\langle S_p \rangle$ for Set I signals is within the $\langle S_p \rangle$ yielded by non-Set I signals. In Fig. 3(b) are the $S_N(\sigma^2)$ for Set II, with $\langle S_p \rangle = 3983.2 \pm 538$. In Fig. 4(a) are the $S_N(\sigma^2)$ for Set III, where $\langle S_p \rangle = 4201 \pm 209$. In Fig. 4(b) are the corresponding $S_N(\sigma^2)$ for Set IV, where $\langle S_p \rangle = 4303 \pm 397$.

Figures 3 and 4 illustrate that ambient noise is beneficial only for $i(x)$ with f values that belong to a preselected set. For the four sets of f_r, the $S_N(\sigma^2)$ all have onsets at

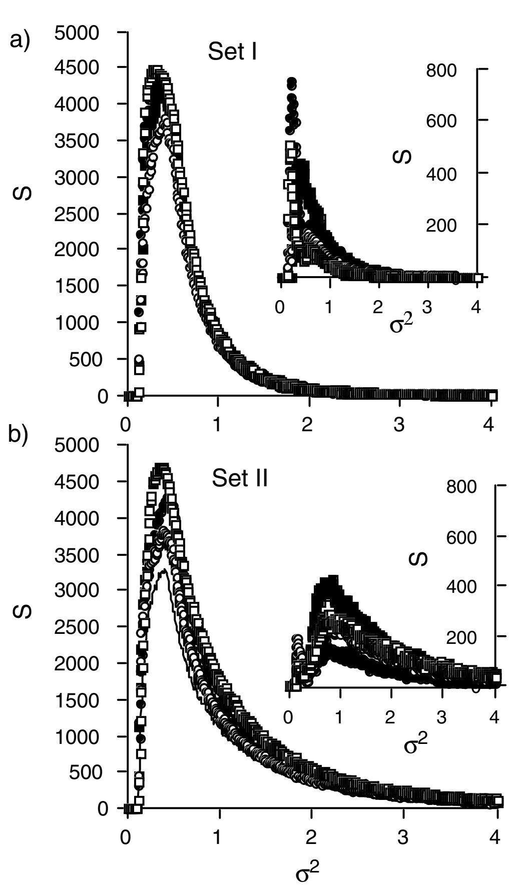

FIG. 3. $S_N(\sigma^2)$ for $A = 0.4$ (5000 trials). (a) Selected $f_r = 1/50$ (squares), $1/5$ (circles), $21/50$ (filled squares), and $23/50$ (filled circles); (b) low-f band, $f_r = 1/50$ (squares), $1/25$ (circles), $3/50$ (filled squares), $2/25$ (filled circles), and $1/10$ (solid line). Inset plots in (a) are for $f = 3/50$ (squares), $11/50$ (circles), $2/5$ (filled squares), and $12/25$ (filled circles). Inset plots in (b) are for $f = 11/50$ (squares), $6/25$ (circles), $13/50$ (filled squares), and $7/25$ (filled circles).

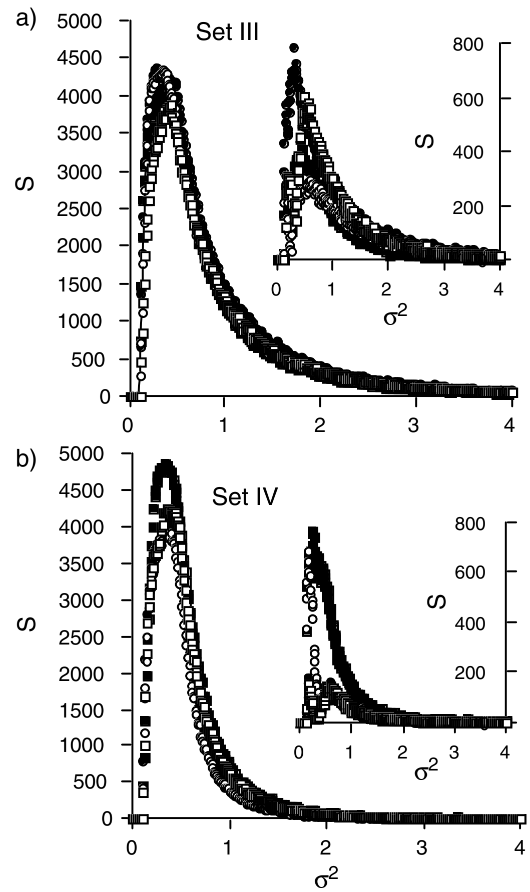

FIG. 4. $S_N(\sigma^2)$ for $A = 0.4$ (5000 trials). (a) Mid-f band, $f_r = 1/5$ (squares), $11/50$ (circles), $6/25$ (filled squares), $13/50$ (filled circles), and $7/25$ (solid line); (b) high-f band, $f_r = 21/50$ (squares), $11/25$ (circles), $23/50$ (filled squares), and $12/25$ (filled circles). Inset of (a) is plots for $f = 8/50$ (squares), $9/50$ (circles), $15/50$ (filled squares), and $16/50$ (filled circles). Inset plots in (b) are for $f = 11/50$ (squares), $6/25$ (circles), $13/50$ (filled squares), and $7/25$ (filled circles).

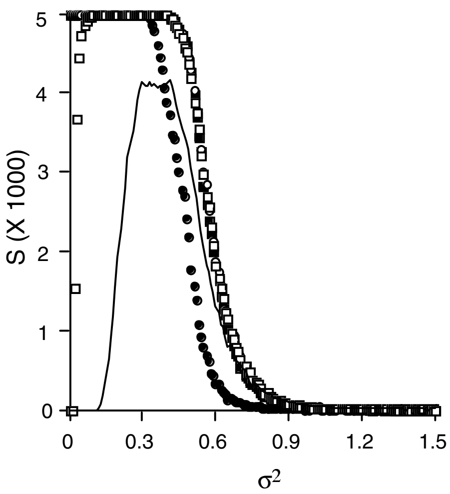

FIG. 5. $S_N(\sigma^2)$ for $A = 0.5$, and $f_r = 10$: $A = 0.5$ (squares), 0.55 (circles), 0.6 (filled squares), and 0.95 (filled circles). Also shown is the $S_N(\sigma^2)$ for $A = 0.4$ (solid line).

$\sigma^2 \approx 0.1$, and peak range ($0.3 \leq \sigma^2 \leq 0.4$). The benefits of noise are therefore realized uniformly in the four filtering tasks. Network response variations among the preselected f_r are within the network's residual performance with other non-f_r values.

Above-threshold disturbances.—Figures 1–4 reveal that ambient noise of the right strength enables a mature network to process weak $i(x)$ without retraining. The influence of noise for strong $i(x)$ is presented in Fig. 5 for the network presented in Figs. 1 and 2 ($f_r = 1/5$). For $A > 0.5$ (squares), S_N increases rapidly to $S_N = 4900$ from $S_N = 0$ within $0.01 \leq \sigma^2 \leq 0.07$. It is well described by a fourth-order polynomial in σ^2. For $A > 0.5$, perfect recognition ($S_N = 5000$) remains possible within $0 \leq \sigma^2 \leq 0.35$. In the same σ^2 range ($A > 0.5$), the network ignores other $i(x)$ with $f \neq f_r$. Both are manifestations of the tolerance of trained networks to noisy strong $i(x)$. For $A = 0.5$, network performance eventually deteriorates with increasing σ^2—an exponential decrease is noted for $S_N(\sigma^2)$ for $\sigma^2 > 0.35$.

The $S_N(\sigma^2)$ for $A = 0.4$ decreases to zero (after peaking) within the (exponential decay) bounds set by the other $S_N(\sigma^2)$ for $A > 0.5$. Hence, the σ^2 range that is useful for weak-signal detection does not compromise network performance for strong $i(x)$. The results in Fig. 5 reconcile two seemingly opposing responses of a "mature" network to ambient noise.

Discussion.—Networks ($M = 50$) within $5 < H \leq 35$ could be trained quickly and yield a nonlinear $S_N(\sigma^2)$ with the same onset and peak location range. However, their ability to benefit from noise differs because S_p decreases with H [UWN: $S_p(H = 7) = 4588$, $S_p(35) = 3933$; GN: $S_p(7) = 1094$, $S_p(35) = 340$]. Hence, the most desirable H value may be established from the network's response to noise-dithered weak $i(x)$—an issue that is difficult to resolve with strong $i(x)$ alone. The network with $H = 5 = M/10$ performed poorly [$S_p(5) = 0$].

We have compared the performance of a mature network with that of an STS. The $S_N(\sigma^2)$ is sharper and has a faster after-peak falloff than $S_T(\sigma^2)$. Ambient noise could improve the network sensitivity significantly—a 625% improvement was observed. For a given A, the network yields a higher onset for $S(\sigma^2)$ and is more robust against accidental triggering by inherent fluctuations in real physical systems. With the right noise strength ($0.1 \leq \sigma^2 \leq 0.5$ for $A = 0.4$), the generalization capability of a mature network is enhanced to accommodate weak $i(x)$ without jeopardizing its performance with strong ones. The same behavior was not observed in networks of (nonfiring) nodes with sinusoidal or linear responses.

We have shown that, with ambient noise, a mature thresholding network could generalize to weak $i(x)$ without retraining. The ability of a mature network to benefit from noise is intrinsic and does not have to be learned separately—no previous experience with noise is necessary.

While we compare the network behavior with that of an STS, it should be emphasized that an STS could not be trained to do a task. A network of STSs is also not trainable at least by the gradient descent method, because the STS response does not have a well-defined derivative.

*Email address: csaloma@nip.upd.edu.ph

[1] A. Longtin, A. Bulsara, and F. Moss, Phys. Rev. Lett. **67**, 656 (1991).
[2] M. C. Teich, S. M. Khanna, and P. Guiney, J. Stat. Phys. **70**, 257 (1993).
[3] D. Russel, L. Wilkens, and F. Moss, Nature (London) **402**, 291 (1997).
[4] J. E. Levin and J. P. Miller, Nature (London) **380**, 165 (1996).
[5] E. Simonotto *et al.*, Phys. Rev. Lett. **78**, 1186 (1997).
[6] I. Hidaka, D. Nozaki, and Y. Yamamoto, Phys. Rev. Lett. **85**, 3740 (2000).
[7] J. Feng and B. Tirozzi, Phys. Rev. E **61**, 4207 (2000).
[8] N. Brunel *et al.*, Phys. Rev. Lett. **86**, 2186 (2001).
[9] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT, Cambridge, Massachusetts, 1997).
[10] N. Mohanty, *Signal Processing: Signals, Filtering, and Detection* (Van Nostrand Reinhold, New York, 1987).
[11] M. Soriano and C. Saloma, Appl. Opt. **37**, 3628 (1998).
[12] T. Watkin and A. Rau, Rev. Mod. Phys. **65**, 499 (1993).
[13] M. Quito, C. Monterola, and C. Saloma, Phys. Rev. Lett. **86**, 4741 (2001).
[14] S. Haykin, *Neural Networks: A Comprehensive Foundation* (Prentice-Hall, New York, 1999), 2nd ed.
[15] J. Proakis and D. Manolakis, *Digital Signal Processing: Principles, Algorithms and Applications* (Prentice-Hall, New York, 1995).
[16] L. Gammaitoni, Phys. Rev. E **52**, 4691 (1995).