Noise-enhanced categorization in a recurrently reconnected neural network

Christopher Monterola* and Martin Zapotocky†
Max-Planck Institut für Physik Komplexer Systeme, Nöthnitzerstrasse 38, 01187, Dresden, Germany
(Received 5 October 2004; published 23 March 2005)

We investigate the interplay of recurrence and noise in neural networks trained to categorize spatial patterns of neural activity. We develop the following procedure to demonstrate how, in the presence of noise, the introduction of recurrence permits to significantly extend and homogenize the operating range of a feed-forward neural network. We first train a two-level perceptron in the absence of noise. Following training, we identify the input and output units of the feed-forward network, and thus convert it into a two-layer recurrent network. We show that the performance of the reconnected network has features reminiscent of nondynamic stochastic resonance: the addition of noise enables the network to correctly categorize stimuli of subthreshold strength, with optimal noise magnitude significantly exceeding the stimulus strength. We characterize the dynamics leading to this effect and contrast it to the behavior of a more simple associative memory network in which noise-mediated categorization fails.

DOI: 10.1103/PhysRevE.71.036134 PACS number(s): 84.35.+i, 72.70.+m, 89.75.Fb, 02.50.Ey

I. INTRODUCTION

The role of noise in neural processing has received significant attention in recent years, particularly in the context of stochastic resonance (SR) [1–3]. In the SR mechanism, noise is used in a counterintuitive way to permit detection of subthreshold signals of a particular frequency by a thresholding neuron (TN). Recently, such noise-driven frequency detection was demonstrated also in a network of TN's [4].

In the case of SR, information is encoded in the temporal structure of the stimulus. This is in contrast to "spatial" encoding, in which the stimulus is mapped into the pattern of activity across a set of neurons at a specific time. As opposed to the case of frequency-encoding systems, the beneficial role of noise in spatially encoding systems has not been modeled extensively. The previous studies [5] have concentrated on the case of recurrent networks and of weak stimuli that are periodic or constant in time.

In this article, we contrast the cases of feed-forward and recurrent networks performing categorization of transient spatially encoded stimuli. We first train a feed-forward network to recognize a set of Gaussian-distributed profiles of unit strength. We then evaluate the performance of the network for stimuli of arbitrary strength and in the presence of noise. For stimuli below a certain strength, the feed-forward network fails to classify correctly, independently of the magnitude of noise. To obtain enhancement of performance by noise, we reconnect the feed-forward network into a recurrent network (Fig. 1). We present the stimulus at the initial time step and let the network evolve to its attractor state before the performance is evaluated. The recurrently reconnected network (RNN) has a high success rate of classification of subthreshold stimuli provided that noise of suitable magnitude is added during the network dynamics. The optimal noise magnitude is several times higher than the stimulus magnitude, in a manner reminiscent of SR or noise dithering [6,7]. The performance of the network thus provides an example of noise-mediated signal processing in a spatially extended system. To investigate how general our findings are, we also discuss a Hopfield network that performs a similar classification task, but fails to benefit from noise. We give an intuitive understanding of the difference between the two cases based on the attractor structure of the networks.

II. MODEL OF CATEGORIZING NETWORK

We use a standard three-layer feedforward neural network (NN) to model the task of spatial categorization. The mapping of the N inputs $\{x_i; i=1,2,\ldots,N\}$ to the N outputs $\{\Phi_k; k=1,2,\ldots,N\}$ as mediated by J hidden nodes per input is defined by the expression

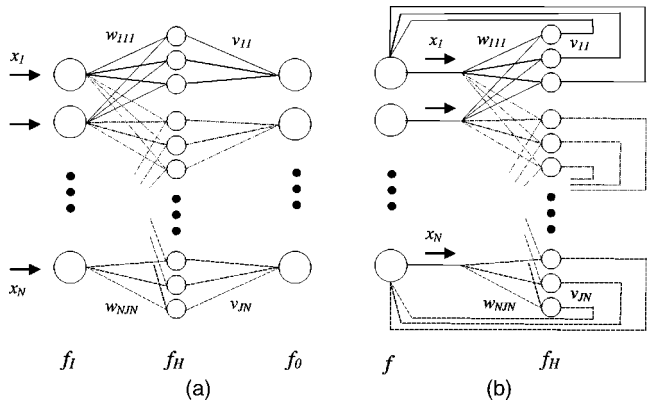

FIG. 1. *Network architecture.* (a) Three-layer fully connected feed-forward network where $f_I(x)=x$, $f_H(x)=1.7159\tanh(2x/3)$, and $f_0(x)=1/[1+\exp(-x)]$. (b) The collapsed two-layer recurent network, now with each node acting as a thresholding neuron. The thresholding function is chosen to be $f(x)=f_I(f_0(x))=1/[1+\exp(-x)]$ for the input layer and remains $f_H(x)=1.7159\tanh(2x/3)$ for the hidden layer.

*Electronic address: chris@mpipks-dresden.mpg.de
†Electronic address: martinz@mpipks-dresden.mpg.de

$$\Phi_k = f_0\left(\sum_{j=1}^{J} Y_{jk} v_{jk}\right), \quad Y_{jk} = f_H\left(\sum_{i=1}^{N} f_I(x_i) w_{ijk}\right), \qquad (1)$$

where $\{w_{ijk}\}$ and $\{v_{jk}\}$ are the weights that are fixed using the standard gradient-descent method (see Appendix A). The hidden (f_H) and output (f_0) nodes have thresholded activation functions while the input (f_I) nodes have linear activation functions.

During training, we feed inputs of form

$$x_k = S \exp\left(\frac{-[k-X]^2}{2\sigma}\right) \Big/ (\pi\sigma)^{1/2},$$

i.e., Gaussian profiles with strength (S), mean (X), and variance (σ). Learning was achieved by updating the weights so that these inputs are mapped into one of the three desired memory states ($\mathbf{\Psi}^1, \mathbf{\Psi}^2, \mathbf{\Psi}^3$). These were chosen to be Gaussian profiles with identical strength and variance ($S=1, \sigma=2$) but varying mean ($X^1=7, X^2=13, X^3=19$).

Spatial information is assumed to be contained in the mean of the input state $\boldsymbol{x}^{t=0}$. The state to evolve to among the $\mathbf{\Psi}^i$ is therefore chosen to be the one with the closest mean or the $\mathbf{\Psi}^i$ with smaller X^i in the special case of equal distance.

The set of training states was defined by the following 12×10^6 possible parameter combinations: $X^i-2 \leqslant X^{t=0} \leqslant X^i+2$ (with step size $\Delta X^{t=0}=2.0\times10^{-3}$), $0.5 \leqslant \sigma^{t=0} \leqslant 3.5$ ($\Delta\sigma^{t=0}=1.5\times10^{-3}$), and $S^{t=0}=1.0$. The input states with $X^{t=0}$ equal to one of the desired means X^i will be referred to as states with "correct" mean.

After training, $\{v_{jk}\}$ are then rewired to the input nodes as depicted in Fig. 1(b). The resulting recurrently RNN is now composed of $N(1+J)$ purely thresholding TN's. The RNN is initiated with an input profile $\boldsymbol{x}^{t=0}$ and then evolves recurrently, with output $\mathbf{\Phi}^t$ at step t becoming the input $\boldsymbol{x}^{t+1}$ at step $t+1$. Note that the state of the RNN after one iteration, $\boldsymbol{x}^{t=1}$, is equivalent to the output of the corresponding nonrecurrent network [non-RNN, Fig. 1(a)]. Throughout the discussion, we use a network with $N=25$ input neurons and $3N$ hidden neurons [8].

The results discussed below (i.e., noise-mediated categorization and uniformity of success rate) are obtained also in the network in which the thresholding functions f and f_H are both chosen to be $1/[1+\exp(-x)]$ or $1.7159\tanh(2x/3)$ [10]. The "mixed" choice $f=1/[1+\exp(-x)]$ and $f_H=1.7159\tanh(2x/3)$ results in fast convergence during training and directly corresponds to the standard implementation of the two-layer perceptron [10,11].

III. RESULTS AND DISCUSSION

The similarity of $\boldsymbol{x}^t$ to the target state $\mathbf{\Psi}^i$ was evaluated using Linfoot's measures of structural content (C), fidelity (F), and correlation quality (Q) [12]. C provides a measure of relative sharpness of the two states, F measures their general similarity and is just the mean square error subtracted from 1, and Q quantifies the alignment of their peaks. The two states are identical when $C=F=Q=1$. In the following, we declare that the current state of the network matches a given target state when C, F, and Q are within 5% of 1 (under such a condition $\mathbf{\Psi}^i$ is visually indistinguishable from $\boldsymbol{x}^t$).

To evaluate convergence to a target state in the case of the RNN, C, F, and Q are measured in 80 successive time steps ($T_{CFQ}=80$, from $t=921$ to $t=1000$). We consider convergence to be complete only if the matching criterion is satisfied at each time step in this range. The extension to $T_{CFQ}>1$ assures that the recovery measure is robust against transients (temporary trappings in the target state) in the presence of noise. For the non-RNN case, $T_{CFQ}=1$ as convergence is evaluated at $\boldsymbol{x}^{t=1}$.

A. Network in the absence of noise

Following 10^5 epochs of training [14], the nonrecurrent network can categorize perfectly the 12×10^6 possible inputs. Generalization by interpolation of resolutions smaller than the training resolutions is achieved perfectly within the training range. Furthermore, extrapolation to include the ranges $0.6<S^{t=0}<2$ and $\sigma^{t=0}<60$ is also attained [Fig. 2(a)].

By adding recurrence, the dynamic range is further improved by two orders of magnitude in strength ($0.12<S^{t=0}<100$) and by a factor of 2 in variance ($\sigma^{t=0}<120$) as shown in Fig. 2(a). Shown in Figs. 2(b)–2(f) are trajectories for representative stimuli from the regions in Fig. 2(a) in which the non-RNN fails, but the recurrently reconnected network achieves perfect categorization after time t_r. Again, note that $t=1$ corresponds to the non-RNN case.

Figure 2(b) shows that the state with high variance is correctly classified—i.e., converges to the state $\mathbf{\Psi}^1$ with the correct mean after $t_r=13$. The non-RNN fails to classify correctly as its output $\boldsymbol{x}^{t=1}$ is dissimilar to $\mathbf{\Psi}^1$. Likewise, an initially strong signal [$S=50$, Fig. 2(c)] or weak signal [$S=0.5$, Fig. 2(d)] achieves recovery at $t_r=3$ and $t_r=5$.

When zero stimulus is presented to the RNN [Fig. 2(e)], the network reaches a limit cycle in which it alternates between $\boldsymbol{x}^t \cong 0$ and a uniform state $\boldsymbol{x}^t \cong 0.5$ with the constant 0.5 traceable from Eq. (1) [15]. This is also the case for the evolution of very weak signals $S^{t=0}<0.12$ [Fig. 2(f)], resulting in classification failure. Hence, $S^{t=0}<0.12$ defines our set of subthreshold stimuli. It is in this region that we demonstrate the "stochastic-resonance-like" behavior of our system in the presence of noise.

B. Interplay of recurrence and noise

We first look at the dynamics of the RNN in the presence of pure noise (stimulus strength $S^{t=0}=0$). Uniform white noise (UWN) of width α and zero mean was added to the state $\boldsymbol{x}^t$ at each time step, starting at $t=0$ [16]. In a real biological network such noise can correspond, e.g., to the spontaneous activity in the input neurons or to synaptic noise. The most probable response [17] of the network to pure noise can be classified as follows. For weak noise [$\alpha \leqslant 0.16$, Fig. 3(a)], the RNN evolution resembles the limit cycle dynamics obtained for zero noise and subthreshold stimulus. For moderate noise [$0.16<\alpha\leqslant0.29$, Fig. 3(b)], the system has a tendency to be permanently trapped in one of the memory states [$\mathbf{\Psi}^2$ in case of Fig. 3(b)]. For strong noise

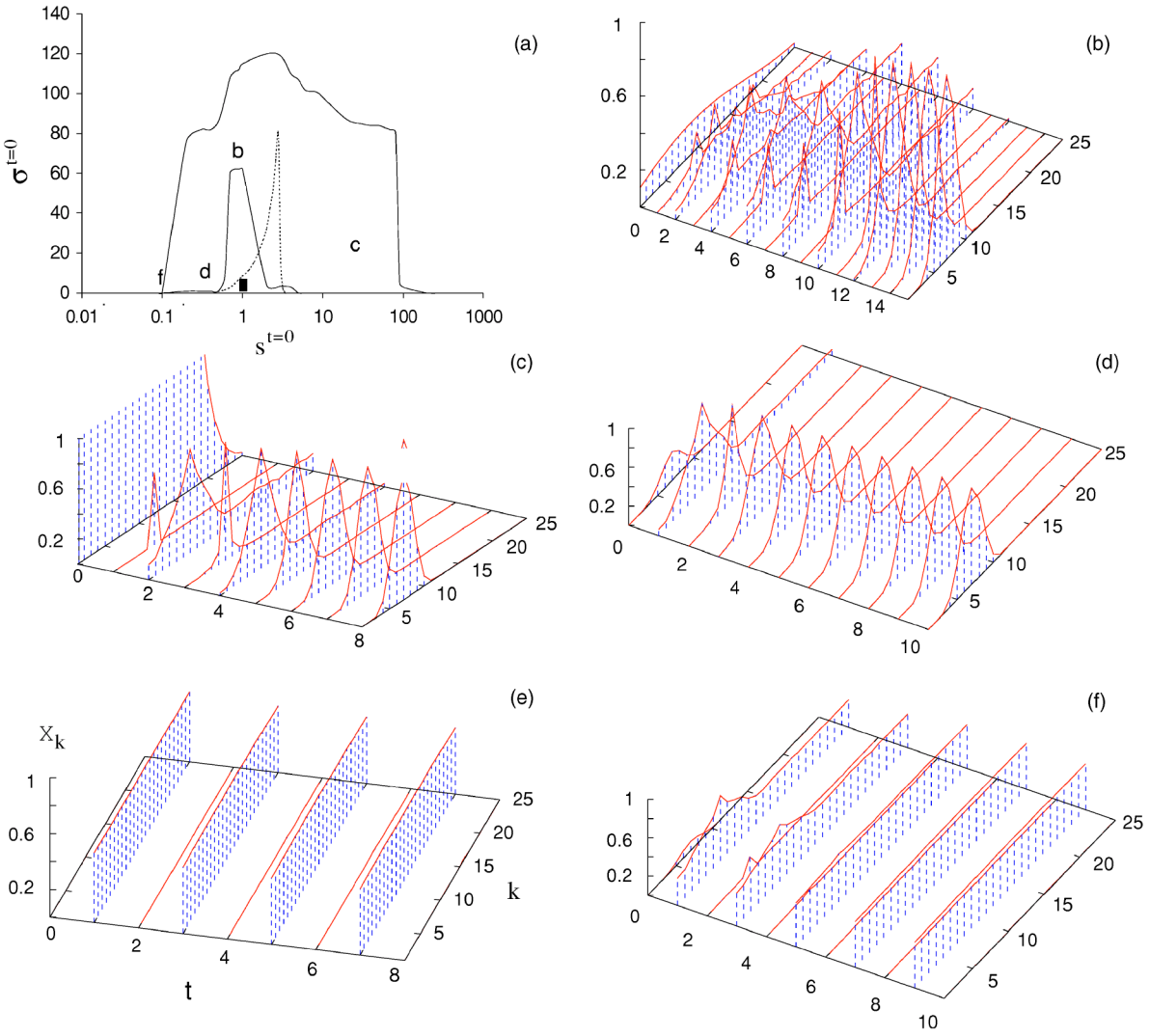

FIG. 2. *Response of the recurrently reconnected network to stimulus in the absence of noise.* (a) The region under the solid curve represents the dynamic range where the RNN (outer solid line) and non-RNN (inner solid line) can operate perfectly for $|X^{t=0}-X^i| \leqslant 2$. The non-RNN network was trained to classify in the smaller region represented by the solid rectangle. Also shown is the dynamic range of the Hopfield network (dotted curve; see Sec. III D). Note that the scale for stimulus strength $S^{t=0}$ is logarithmic. (b)–(f) Dynamics of the RNN starting with input states that are representative of the regions defined in (a): (b) Highly dispersed stimulus: $\sigma^{t=0}=65.0$, $X^{t=0}=9.0$, $S^{t=0}=1.0$. (c) Strong stimulus: $\sigma^{t=0}=20.0$, $X^{t=0}=9.0$, $S^{t=0}=50.0$. (d) Weak stimulus: $\sigma^{t=0}=2.0$, $X^{t=0}=7.0$, $S^{t=0}=0.5$. (e) No stimulus $S^{t=0}=0.0$. (f) Very weak stimulus: $\sigma^{t=0}=2.0$, $X^{t=0}=7.0$, $S^{t=0}=0.1$. Note that $t=1$ corresponds to the non-RNN case.

[$\alpha > 0.29$, Fig. 3(c)], $\boldsymbol{x}^t$ is trapped in the memory states $\boldsymbol{\Psi}^i$ transiently and visits all of them at random.

We now discuss three phenomena exhibited by the network categorizing in the presence of noise when recurrence is imposed: (1) enhanced robustness to noise, (2) uniformity of success rate throughout the dynamic range, and (3) noise-mediated categorization.

1. Enhanced robustness to noise

To evaluate the performance of the network in the presence of noise, we define the success rate S_{rate} as the fraction of runs starting with a given stimulus that resulted in complete convergence to the correct memory state. (In the presence of noise, the Linfoot criteria are evaluated on the output $\boldsymbol{x}^t$ of the network, before noise is added.) Figure 4 shows S_{rate} for the non-RNN [Fig. 4(a)] and RNN [Fig. 4(b)] as a function of noise strength α. At signal strength $0.2 \leqslant S^{t=0} \leqslant 50$, the near-perfect performance of the RNN is robust against noise up to $\alpha \simeq 0.4$ [Fig. 4(b)]. In contrast, for the non-RNN, S_{rate} decays with increasing α and reaches nearly 0 at $\alpha=0.4$ for $S^{t=0}<0.5$ [Fig. 4(a)].

For $\alpha > 0.4$ and $0.5 < S^{t=0} < 2$, it appears at first that the non-RNN is more robust against noise than the RNN. This is no longer the case, however, if we require the non-RNN to satisfy the Linfoot criteria for 80 successive runs with identical stimulus (i.e., a criterion comparable to the one used for the RNN). At $\alpha \approx 0.11$, S_{rate} then abruptly becomes zero (figure not shown) for the non-RNN.

Without recurrence, even in the absence of noise, the non-RNN fails when $S^{t=0}<0.6$ or $S^{t=0}>2.0$. The non-RNN therefore has a dynamic range (DR) of less than one decade, with

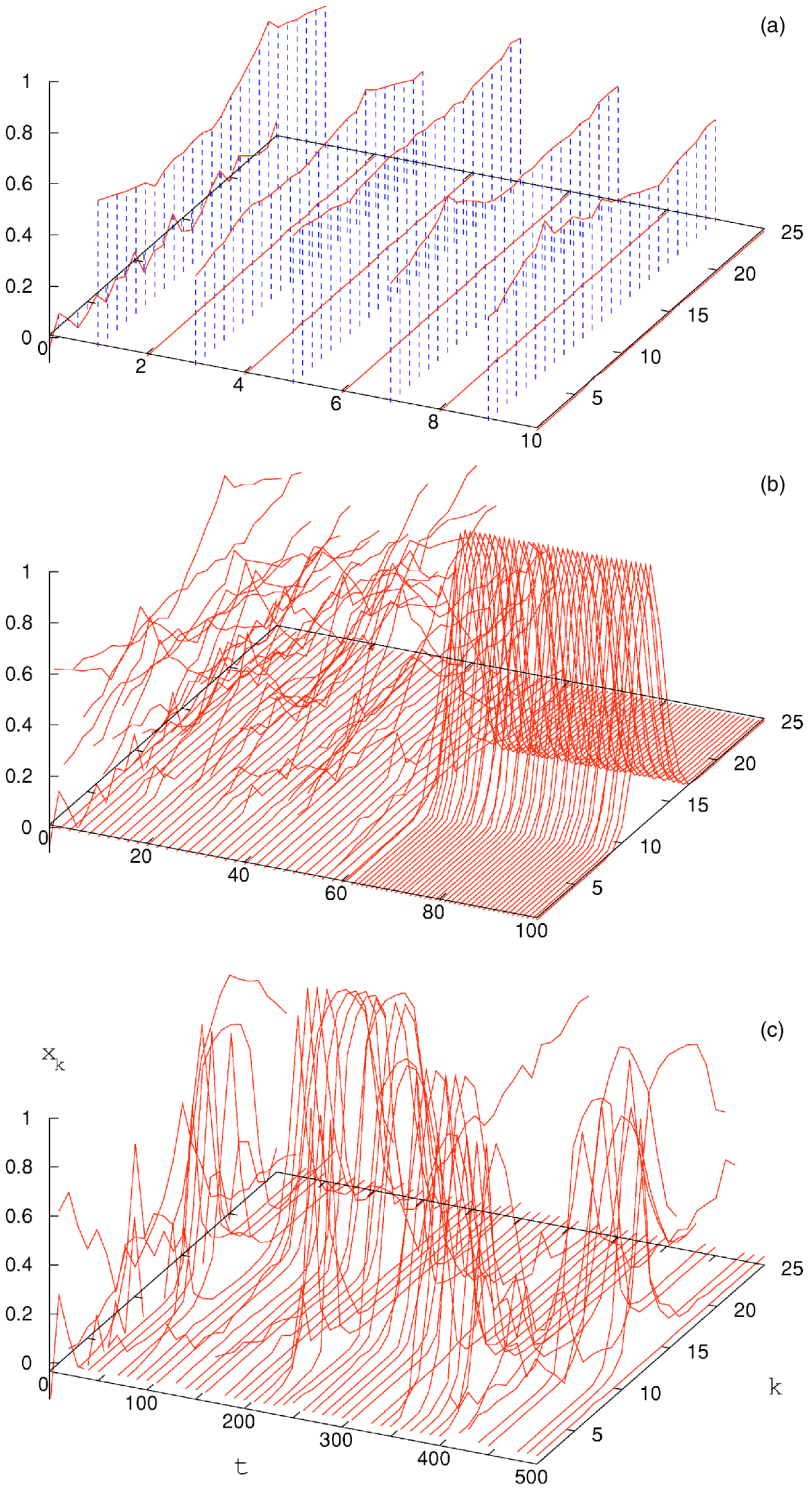

FIG. 3. *Typical recurrently reconnected network dynamics for zero stimulus* and (a) weak noise ($\alpha=0.14$), (b) moderate noise ($\alpha=0.22$), and (c) strong noise ($\alpha=0.50$). The output $\boldsymbol{x}^t$ of the network (before noise is added) is shown for $t\geqslant 1$; for $t=0$, $\boldsymbol{x}^0$ plus UWN is shown.

low robustness against noise. In contrast, RNN possesses a DR of over two decades in $S^{t=0}$ (0.12–85.0) and is robust up to noise strength $\alpha\approx 0.4$ independent of $S^{t=0}$.

2. Uniformity of success rate throughout dynamic range

The success rate of the RNN is approximately uniform throughout its dynamic range: $S_{rate}\simeq 1$ for $\alpha<0.4$ and $S_{rate}\simeq 0$ for $\alpha>0.5$, independently of signal strength $S^{t=0}$ [Fig. 4(b)]. To verify that this result is not a consequence of the strict condition we imposed on S_{rate}, we lowered both T_{CFQ} (down to $T_{CFQ}=1$, only evaluated at $t=1000$) and Linfoot's tolerance (up to 20% variation of C, F, Q from 1.0). We obtained curves that are similar to Fig. 4(b) except that at higher noise strengths ($\alpha>0.5$), S_{rate} saturates at a nonzero value (0.2 for $T_{CFQ}=1$ and 20% tolerance). The nonzero success rate at high noise strength originates from the transient (typically fewer than ten iterations) trapping of the system in one of the incorrect memory states $\boldsymbol{\Psi}^i$ [Fig. 3(e)]—the rea-

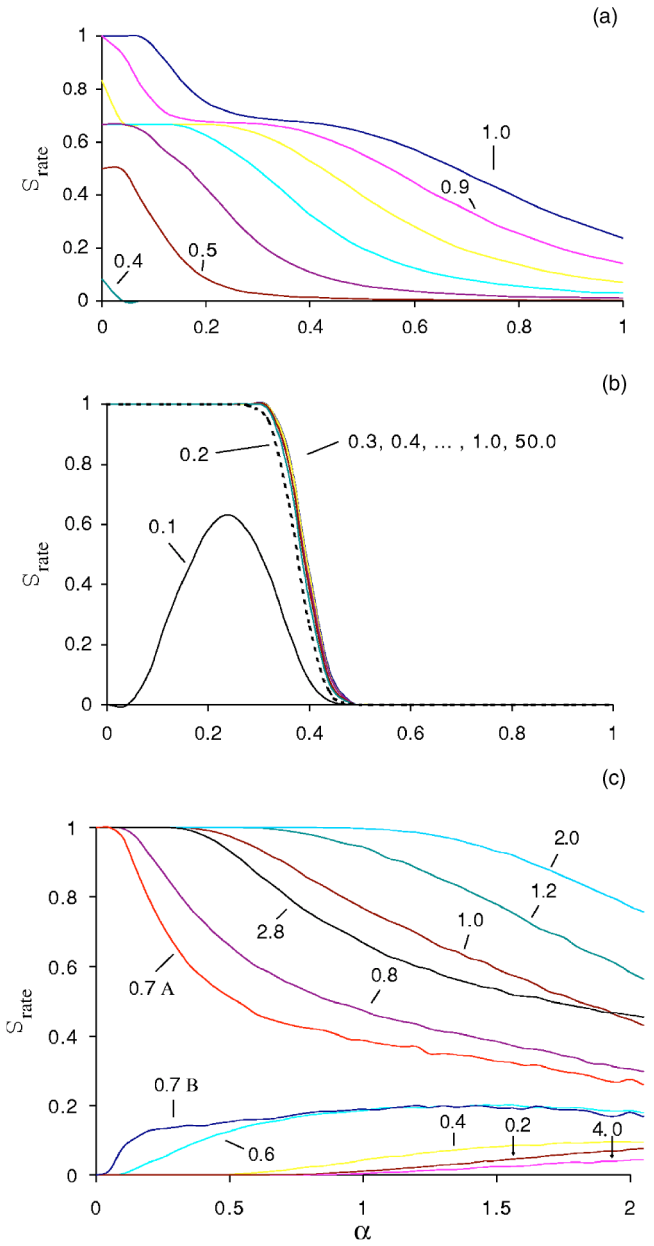

FIG. 4. *Success rate of (a) nonrecurrent, (b) recurrently reconnected, and (c) Hopfield networks in the presence of noise*. Average S_{rate} is shown as a function of noise strength α at the indicated stimuli strengths $S^{t=0}$. In (a) and (b), S_{rate} is averaged over stimuli with $\sigma^{t=0}=2$ and $X^{t=0}$ in the range $X^i-2\leqslant X^{t=0}\leqslant X^i+2$ (resolution of $X^{t=0}=1.0$), with 50 runs performed for each $X^{t=0}$. In (c), all curves except 0.7*A* and 0.7*B* were averaged over stimuli with $\sigma^{t=0}=4$ and with $X^{t=0}$ as in (a) and (b). The curve 0.7*A* shows S_{rate} for stimuli with strength $S=0.7$ and "correct" mean $X^{t=0}=X^i$, while the curve 0.7*B* shows the average over $X^{t=0}=X^i\pm 2$ and $X^i\pm 1$. Here 1000 runs were performed for each stimulus in (c).

son why we set T_{CFQ} to a relatively high value (=80) in Fig. 4(b). The corresponding $S_{rate}\approx 0.2$ is lower than 1/3 as in the presence of noise of high strength the network state is not always close to one of the Ψ^i.

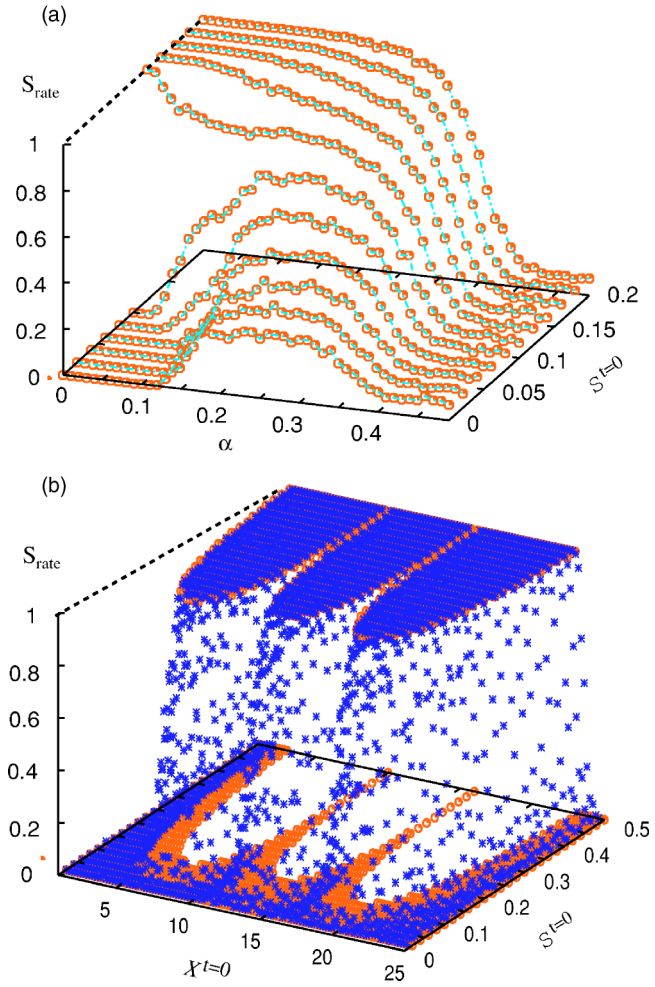

FIG. 5. (Color online) *Noise-mediated categorization of weak stimuli by the recurrently reconnected network*. (a) Success rate of the RNN averaged over stimuli with $X^i-3<X^{t=0}<X^i+3$. (b) Success rate of the RNN in the absence (○) and presence (*) of noise with strength $\alpha=0.24$. In all cases $\sigma^{t=0}=2.0$. Linfoot's criteria are evaluated with a $T_{CFQ}=10$ from $t=991$ to $t=1000$ and averaging is done over 50 runs for each stimulus.

3. Noise-mediated categorization

In the absence of noise, the RNN fails to respond to stimuli of strength $S^{t=0}<0.12$ [i.e., it evolves as if no stimulus is presented, Figs. 2(e) and 2(f)]. When noise is added, however, such subthreshold stimuli can be classified correctly. This is clearly seen for $S^{t=0}=0.1$ in Fig. 4(b), and $S^{t=0}<0.12$ in Fig. 5(a). The shape of the $S_{rate}(\alpha)$ curves is reminiscent of stochastic resonance [1,3,4]: the highest S_{rate} is obtained at a nonzero optimal noise strength α_{opt}. The average S_{rate} for signals of strength $S^{t=0}=0.1$ reaches 60% at $\alpha_{opt}\approx 0.24$. Below $S^{t=0}=0.12$, the optimal S_{rate} decays linearly with $S^{t=0}$, reaching 30% at $S^{t=0}=0$. The 30% $\approx 1/3$ level reflects the long-term trapping of the network in a randomly selected memory state [Fig. 3(b)] when α falls into the moderate noise range. $S_{rate}=1/3$ therefore represents the base line to which the gain in S_{rate} should be compared. Still, adding noise at strength α_{opt} increases S_{rate} by up to 100% (from 30% to 60%).

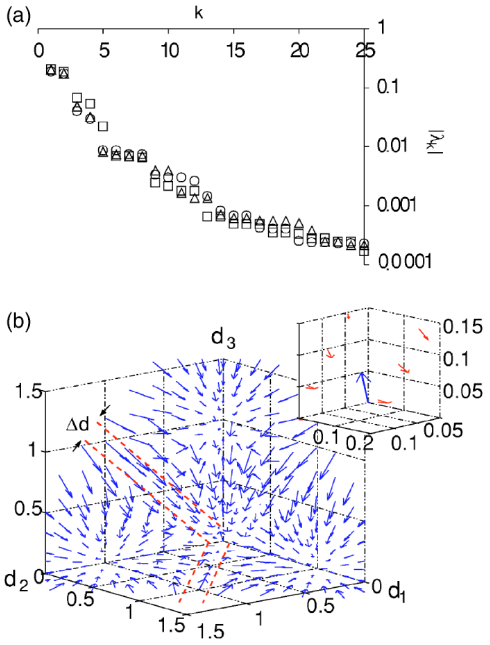

FIG. 6. (Color online) *Attractors of the recurrently reconnected network and their stability*. (a) Eigenvalue spectrum at the three memory states $\boldsymbol{\Psi}^1$ (△), $\boldsymbol{\Psi}^2$ (□), and $\boldsymbol{\Psi}^3$ (◯). (b) Flow diagram of the RNN. Shown are the projections d_i of the state $\boldsymbol{x}^t$ on the memory states $\boldsymbol{\Psi}^i$. Included are trajectories for stimuli from the set $0.0 \leqslant S^{t=0} \leqslant 2.0$, $1.0 \leqslant X^{t=0} \leqslant 25.0$, $1.0 \leqslant \sigma^{t=0} \leqslant 25.0$. Inset: the "velocity vector" $\boldsymbol{d}^{t+1} - \boldsymbol{d}^t$ with vector origin defined by $\boldsymbol{d}^t$. The velocity at $\boldsymbol{x}^{t=0}=0$ has a strong magnitude and moves the system to the flat state $\boldsymbol{x}^t$ with $(d_1=0.68, d_2=0.69, d_3=0.68)$. The magnitude of the velocities is scaled to 0.2 of its original strength.

The addition of noise also improves the classification of signals (of arbitrary strength) that were not included in the training set of the NN ("noise-mediated extrapolation"). This is apparent in Fig. 5(b) for signals with $X^{t=0} \simeq 10$ or $X^{t=0} \simeq 16$ (in the region of strong overlap between two of the states $\boldsymbol{\Psi}^i$) and $X^{t=0} < 4$ or $X^{t=0} > 22$ (near the system boundary).

C. RNN viewed as a dynamical system

The recurrently reconnected network implements a discrete dynamical system that maps $\boldsymbol{x}^t$ into $\boldsymbol{x}^{t+1}$ through: $\boldsymbol{x}^{t+1} = \Gamma(\boldsymbol{x}^t)$, where the mapping function Γ is given by Eq. (1).

By construction, the three target states $\boldsymbol{\Psi}^i$ are fixed points of the map. Note that the state $\boldsymbol{x}=0$ is not a fixed point. To evaluate the stability of $\boldsymbol{\Psi}^i$, we examine the eigenvalues $\boldsymbol{\lambda}$ of the Jacobian matrix Ω of the system [18]. Figure 6(a) shows that the largest modulus of $\boldsymbol{\lambda}$ is ≈ 0.21; as $|\lambda_k| < 1$ for all k, $\boldsymbol{\Psi}^i$ are stable fixed points [19].

The trajectories $\boldsymbol{x}^t$ for the noiseless case are visualized in Fig. 6(b), where the three axes (d_1, d_2, d_3) are the vector projections of $\boldsymbol{x}^t$ on the three target states $\boldsymbol{\Psi}^i$, in the units of $|\boldsymbol{\Psi}^i|$ [i.e., $d_i = (\boldsymbol{\Psi}^i \cdot \boldsymbol{x}^t)/|\boldsymbol{\Psi}^i|^2$]. The trajectories form four clusters, three of which converge to one of $\boldsymbol{\Psi}^i$, and the fourth cluster consists of a thin region (with thickness $\Delta d \approx 0.24$) oriented along the diagonal directions [see Fig. 6(b)]. In the latter region, $\boldsymbol{x}^t$ asymptotically alternates between $(d1, d2, d3) = (0.00, 0.00, 0.00)$ and (0.68, 0.69, 0.68). The attractor of this region, which includes the zero-stimulus trajectory, is thus a limit cycle of period 2.

We are now in a position to understand the results of Sec. III B in terms of the flow represented in Fig. 6(b). The weaker the stimulus strength $S^{t=0}$, the higher is the percentage of stimuli of this strength that become trapped in the limit cycle attractor, rather than converging to one of the desired fixed points $\boldsymbol{\Psi}^i$; with no noise, all states with $S^{t=0} < 0.1$ become trapped. In the presence of noise, the trajectory can switch in between the basins of attraction of the zero-noise system. When the noise strength α becomes comparable to the thickness $\Delta d \approx 0.24$, the trajectories starting in the undesired basin are likely to escape from it. If noise is too strong ($\alpha > 0.29$), the trajectory is unlikely to become permanently trapped in the vicinity of one of the target states, and instead hops randomly between the corresponding basins of attraction [see also Fig. 3(c)]. The existence of the optimal value of noise strength α_{opt} is then seen to be a compromise maximizing the chance of getting out of the undesired basin and subsequently not escaping the desired basin.

D. Comparison to a Hopfield network

The arguments of the previous section are quite general and raise the question of whether the effects discussed by us are a generic feature of recurrent networks. In this section, we describe a simple associative memory network that fails to exhibit the enhancement of performance by noise which we demonstrated in the two-layer recurrently reconnected network. As described in Appendix B, we constructed a Hopfield network with three target states ${}^{\mathrm{H}}\boldsymbol{\Psi}^q$ of means $X^1 = 7$, $X^2 = 13$, and $X^3 = 19$. The states are defined by $\{{}^{\mathrm{H}}\Psi_k^q = 1\}$ for $k = X^q \pm 2, X^q \pm 1, X^q$ and $\{{}^{\mathrm{H}}\Psi_k^q = -1\}$ for all other k (k being the index of each neuron). We then evaluated the performance of the network for the set of input states defined by

$$x_k = \frac{2S \exp\left(\dfrac{-[k-X]^2}{2\sigma}\right)}{(\sigma/4)^{1/2}} - 1. \qquad (2)$$

Note that when the sgn function is applied to the input states with $S=1$, $\sigma=4$ and $X=X^q$, one obtains the target states ${}^{\mathrm{H}}\boldsymbol{\Psi}^q$. Other input states are considered to be correctly classified if the network evolves to the target state with closest mean.

We first discuss the dynamics in the absence of noise. The input state $\{x_k=-1\}$ (corresponding to $S^{t=0}=0$) represents the zero stimulus and does not evolve, as $\boldsymbol{x}^t=-1$ is a fixed point of the map. The set of subthreshold stimuli is defined as those that converge to the fixed point $\{x_k=-1\}$ after one iteration. The dynamic range of the network (within which perfect classification is achieved) is indicated in Fig. 2(a). Note that the dynamic range in $S^{t=0}$ is two orders of magnitude narrower compared to the range of the two-layer RNN.

The success rate S_{rate} of the Hopfield network in the presence of noise is plotted in Fig. 4(c). In contrast to the RNN case [Fig. 4(b)], the success rate is not uniform as a function of signal strength. Within most of its narrow dynamic range,

the Hopfield network is more robust than the RNN.

For large enough noise strength α, the success rate of the Hopfield network does become nonzero [Fig. 4(c)] even for subthreshold stimuli ($S^{t=0} \leqslant 0.6$). However, S_{rate} never exceeds $1/3$—i.e., the success rate that corresponds to classification performed at random. Therefore the implemented Hopfield network does not show noise-mediated categorization.

This illustrates that the conditions necessary for noise-mediated detection of subthreshold stimuli will not always be met in recurrent neural networks. The failure of the Hopfield network to benefit from noise can be understood using arguments similar to those of Sec. III C. Compared to the RNN case, the size of the basin of attraction for the zero-stimulus fixed point $\{x_k=-1\}$ is larger, both in absolute terms and relative to the size of basins of attraction for the memory states $^{\mathbf{H}}\boldsymbol{\Psi}^q$ [20]. In addition, the dynamics is strongly influenced by the spurious fixed points $\mathrm{sgn}(\pm{}^{\mathbf{H}}\boldsymbol{\Psi}^1 \pm {}^{\mathbf{H}}\boldsymbol{\Psi}^2 \pm {}^{\mathbf{H}}\boldsymbol{\Psi}^3)$. When designing a neural network capable of noise-mediated categorization, particular care should be given to the minimization of the basins of attraction of unwanted attractors.

IV. RELATED IDEAS AND CONCLUSION

We first briefly discuss our results in the context of biological neural systems. Consider a recurrent neural network that processes spatially encoded sensory information—for example, the olfactory bulb network [21]. The input to the network is generated by sensory neurons at the periphery (the olfactory epithelium). Even in the absence of sensory stimulus, many sensory neurons have high levels of spontaneous activity (up to 100 spikes/sec in typical vertebrate olfactory sensory neurons [22]). Following the presentation of a brief sensory stimulus, the sensory neurons provide a transient, spatially encoded input to the network [23]. If the interval between successive presentations of stimuli allows sufficient time for the network to evolve recurrently [24], the output from the network will show the noise-mediated features discussed above. Spontaneous activity in the sensory neurons can in this context be viewed as a necessary part of a biological mechanism that optimizes the detection of weak stimuli by the system.

Identification and discrimination of odors can be viewed as a categorization task performed by the olfactory system. Psychophysical studies show that the perceived quality of an odor is typically invariant to odor concentration over two orders of magnitude [25]. Recently, Brody and Hopfield proposed a phenomenological model based on many-are-equal operations, designed in part to explain this invariance [26]. Our results show that in the presence of noise such a feature can emerge also by allowing simple recurrence.

In summary, we presented a detailed case study designed to investigate the interplay of recurrence and noise in neural networks. We showed that in the presence of noise, the introduction of recurrence extends and homogenizes the operating range of a two-layer neural network trained to categorize spatial patterns of neural activity. In particular, we demonstrated that subthreshold signals can be correctly classified by the recurrent network when noise of magnitude exceeding the signal strength is introduced in the network dynamics. We also gave an example of a simpler recurrent network of the Hopfield type where noise-mediated categorization fails. We gave an intuitive explanation of the difference between these two cases based on the fixed point structure of the two systems. A more systematic understanding of how recurrent neural networks should be designed to be capable of efficient noise-mediated signal processing remains an interesting subject for further study.

ACKNOWLEDGMENT

We thank Karsten Kruse for useful comments on the manuscript.

APPENDIX A: TRAINING PROCEDURE FOR THE RECURRENTLY RECONNECTED NETWORK

The synaptic weights $\{w_{ijk}\}$ and $\{v_{jk}\}$ were initialized randomly within $[-1/\chi, 1/\chi]$ where χ is the number of free parameters connected to the weights ($\chi=JN$ and $\chi=J$, respectively, for $\{w_{ijk}\}$ and $\{v_{jk}\}$) [11,27]. The activation functions for the input, hidden, and output nodes are $f_I(x)=x$, $f_H(x)=1.7159\tanh(2.0x/3.0)$, and $f_0(x)=1/[1+\exp(-x)]$, respectively. The threshold function f_H is chosen so as to allow faster convergence and better memory representation [11,27,28]. The linear function f_I in this (training) stage is chosen so that the network can be wrapped in the next (recurrent) stage as depicted in Fig. 1(b), hence eventually producing a network with $N(1+J)$ purely TN's.

Fixation of the desired memory state $\boldsymbol{\Psi}$ for an initial profile $\boldsymbol{x}^{t=0}$ is attained through gradient descent by updating the weights according to

$$w_{ijk}^{t+1} = w_{ijk}^{t} + \eta\frac{\partial E_k}{\partial w_{ijk}} + \gamma\Delta w_{ijk}^{t-1}, \tag{A1}$$

where $E_k=(\Psi_k-\Phi_k)^2/2$ is the cost function and $\gamma\Delta w_{ijk}^{t-1} = \gamma(w_{ijk}^{t-2}-w_{ijk}^{t-1})$ is the momentum term [11,27]. The learning (η) and momentum (γ) rates are adaptively varied within $[10^{-5} \leqslant (\eta,\gamma) \leqslant 10^{-1}]$ to account for the variation of the error surface along different regions of weight dimension [29]. These rules enhance the minimization of E_k, preventing unnecessary oscillations in the search space [11,27–29]. A similar update rule is implemented for v_{ij}.

APPENDIX B: IMPLEMENTATION OF THE HOPFIELD NETWORK

The one-layer, recurrent, associative memory network of the Hopfield type was designed to perform a classification task similar to that for which we trained the network of

Appendix A. The desired memory states ${}^{\mathrm{H}}\boldsymbol{\Psi}^q$ were chosen to be the discretized version of the Gaussian profiles that were used as memory states of the RNN:

$$ {}^{\mathrm{H}}\Psi_k^q = \mathrm{sgn}\left(\frac{2S \exp\left(\frac{-[k-X^q]^2}{2\sigma} \right)}{(\sigma/4)^{1/2}} - 1 \right), \tag{B1}$$

where $\sigma=4$, and $S=1$. The means (X^q) that characterize the three memory states are $X^1=7$, $X^2=13$, and $X^3=19$—in exact correspondence to those in Appendix A.

The synaptic weight (w_{ij}) that connects the ith neuron to the jth neuron is fixed using the Hebbian rule [11,28]

$$ w_{ji} = w_{ji} = \begin{cases} \frac{1}{N}\sum_{q=1}^{3} \left({}^{\mathrm{H}}\Psi_i^q \right)\left({}^{\mathrm{H}}\Psi_j^q \right), & j \neq i, \\ 0, & j = i. \end{cases} \tag{B2}$$

The state of the network $\boldsymbol{x}^t$ evolves recurrently according to the map

$$ x_j^{t+1} = \mathrm{sgn}\left(\sum_{i=1}^{25} w_{ji} x_i^t \right), \quad j = 1,2,\ldots,25. \tag{B3}$$

The update of $\boldsymbol{x}^t$ is done asynchronously; i.e., the element j is chosen randomly and one at a time.

[1] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, Rev. Mod. Phys. **70**, 223 (1998).

[2] P. Dayan and L. F. Abbot, *Theoretical Neuroscience: Computational and mathematical modeling of neural systems* (MIT Press, Cambridge, MA, 2001).

[3] G. Wenning and K. Obermayer, Phys. Rev. Lett. **90**, 120602 (2003); S. Laughlin and T. Sejnowski, Science **301**, 1870 (2003); T. Hoch, G. Wenning and C. Obermayer, Phys. Rev. E **68**, 011911 (2003); B. Kosko and S. Mitaim, Neural Networks **16**, 755 (2003); M. Chacron, A. Longtin, and L. Maler, Network Comput. Neural Syst. **14**, 803 (2003); I. Goychuk and P. Hänggi, Phys. Rev. E **69**, 021104 (2004).

[4] C. Monterola and C. Saloma, Phys. Rev. Lett. **89**, 188102 (2002).

[5] S. K. Han, W. S. Kim, and H. Kook, Phys. Rev. E **58**, 2325 (1998); T. Kanamaru and Y. Okabe, *ibid.* **62**, 2629 (2000).

[6] Z. Gingl, L. B. Kiss, and F. Moss, Europhys. Lett. **23**, 191 (1995).

[7] L. Gammaitoni, Phys. Rev. E **52**, 4691 (1995).

[8] Results analogous to those discussed for $N=25$ were obtained for $N \leqslant 100$ $(N=4,9,16,50,100)$ and an arbitrary number of memory states below $N/2$. We chose $J=3$ to obtain a network of high capacity per neuron. (The capacity of the three-layer feedforward network is $C=[16/(\pi-2)][\ln(J)]^{1/2}N$ [9]. The capacity per neuron, $C/N(J+1)$, reaches the maximum value of 3.8 N between $J=2$ and $J=3$.)

[9] R. Urbanczik, J. Phys. A **30**, L387 (1997).

[10] Results for the network with $f=1/[1+\exp(-x)]$ and $f_H=1/[1+\exp(-x)]$ are given at http://www.mpipks-dresden.mpg.de/~chris. In this case, the period-two limit cycle of the RNN (Sec. III A.) is replaced by a fixed point $\boldsymbol{x} \simeq 0$.

[11] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. (Prentice-Hall, New York, 1999).

[12] The expressions for Linfoot's criterion are $C=\langle \boldsymbol{x}^t \rangle^2 / \langle \boldsymbol{\Psi}^i \rangle^2$, $F=1-\langle \boldsymbol{\Psi}^i - \boldsymbol{x}^t \rangle^2 / \langle \boldsymbol{\Psi}^i \rangle^2$, and $Q=\langle |\boldsymbol{\Psi}^i||\boldsymbol{x}^t| \rangle / \langle \boldsymbol{\Psi}^i \rangle^2$ where $\langle \cdots \rangle$ designates averaging over the state (note that $F=2Q-C$) [13].

[13] F. Huck *et al.*, J. Opt. Soc. Am. A **2**, 1644(1985).

[14] During each training epoch, the full set of 12×10^6 stimuli was presented in random order. The NN was implemented in c++ and compiled using gcc; the run time for training was 40 min using a DEC ALPHA computer with 1 GHz processor speed. We note that the results are independent of the initial weights. Even though the set of the final values of the weights are different for each weight initialization, results consistent with those described in Sec. III (i.e., oscillation between 0 and 0.5, noise-enhanced categorization, etc.) are obtained for each of the 50 initializations we tested. A representative set of the obtained weights $\{w_{ijk}\}$ and $\{v_{jk}\}$ is listed at http://www.mpipks-dresden.mpg.de/~chris.

[15] The transition $0 \rightarrow 0.5$ follows immediately from Eq. (1): $Y_{jk}=f_H(0)=1.7159\tanh(0)=0$, $\Phi_k=f_0(0)=1/[1+\exp(-0)]=0.5$. The transition $0.5 \rightarrow 0$ occurs as the sum of the inputs to each unit in the hidden layer is a (large) negative number when the state $\boldsymbol{x}^t$ at the previous time step was uniform. This is a property of the evolved weights.

[16] Noise was generated using the MT19937 generator by Matsumoto and Nishimura (prime period $\approx 10^{6000}$) taken from the GNU scientific library with seed randomized by the CPU clock. Results analogous to those reported for UWN noise were also obtained for Gaussian-distributed noise.

[17] The behavior described in the main text was obtained in 93% of our runs for weak noise $[\alpha < 0.16$ $(\Delta\alpha=0.01)]$, 81% of runs for moderate noise $[0.16 \leqslant \alpha < 0.29$ $(\Delta\alpha=0.01)]$, and 87% of runs for strong noise $[0.29 \leqslant \alpha \leqslant 1.0$ $(\Delta\alpha=0.05)]$. Ten runs were performed at each noise strength.

[18] The explicit expression for λ_k and the corresponding eigenvectors (in a Mathematica 5.0 notepad) can be viewed at http://www.mpipks-dresden.mpg.de/~chris.

[19] H. Tong, *Non-linear Time Series: A dynamical system approach* (Clarendon Press, Oxford, 1990).

[20] We traced the evolution of all the 2^{25} possible combinations for the 25 neurons each with a state of either -1 or 1. The fixed points where the network will eventually settle were distributed as follows: ${}^{\mathbf{H}}\boldsymbol{\Psi}^1=11\%$, ${}^{\mathbf{H}}\boldsymbol{\Psi}^2=13\%$, ${}^{\mathbf{H}}\boldsymbol{\Psi}^3=11\%$, $\{x_k=-1\}=16\%$, $\{x_k=1\}=13\%$, $-{}^{\mathbf{H}}\boldsymbol{\Psi}^1=9\%$, $-{}^{\mathbf{H}}\boldsymbol{\Psi}^2=11\%$, $-{}^{\mathbf{H}}\boldsymbol{\Psi}^3=9\%$, and $\mathrm{sgn}(\pm{}^{\mathbf{H}}\boldsymbol{\Psi}^1 \pm {}^{\mathbf{H}}\boldsymbol{\Psi}^2 \pm {}^{\mathbf{H}}\boldsymbol{\Psi}^3)=5\%$. In the last case we exclude the fixed points when the three signs are either all $(+)$ or all $(-)$ because such are identical to $\{x_k=-1\}$ and $\{x_k=1\}$, respectively. Only about 2% of 2^{25} have settled to fixed points other than the states listed above.

[21] K. Mori, H. Nagao, and Y. F. Sasaki, Network **9**, R79 (1998).

[22] J. P. Rospars *et al.*, Brain Res. **662**, 31 (1994).

[23] This will also be the case when the sensory stimulus is extended in time and the sensory neurons rapidly adapt.

[24] While we evaluated the success rate of recovery of weak signals in Fig. 4(b) after 1000 iterations, in most cases fewer than 20 iterations are required to converge to the correct memory state.
[25] D. Krone *et al.*, Phytother Res. **15**, 135 (2001).
[26] C. Brody and J. J. Hopfield, Neuron **37**, 843 (2003).
[27] T. Watkin and A. Rau, Rev. Mod. Phys. **65**, 499 (1993).
[28] J. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Cambridge, MA, 1991).
[29] Z. Lou, Neural Comput. **3**, 226 (1991); C. Monterola *et al.*, J. Forecast. **21**, 435 (2002).