Stochastic Resonance in the Perceptual Interpretation of Ambiguous Figures: A Neural Network Model

M. Riani and E. Simonotto
Dipartmento di Fisica, Università di Genova, Istituto Nazionale di Fisica della Materia, Unità di Genova via Dodecaneso 33, 16146 Genova, Italy
(Received 27 December 1993)

We describe the results of computer simulations of the dynamical behavior of an autoassociative network with a two-dimensional energy landscape. Such a network can model some aspects of the phenomenon of perceptual bistability in the presence of ambiguous figures. The network can be operated at either zero or nonzero temperatures which represent an internal system noise. Our results show that, under the influence of a weak periodic external signal, the network exhibits a maximum in the signal-to-noise ratio at an optimum noise level: the characteristic signature of stochastic resonance.

PACS numbers: 87.10.+e, 05.40.+j

Studies of the perception of ambiguous figures have a long history [1,2]. Perception of these kinds of figures (e.g., the Necker cube [2]; see Fig. 1) is characterized by noisy bistable dynamics, that is, the two different interpretations, elicited by the figure, are alternatively perceived by the observer with a stochastic time course. Numerous experiments have shown that the times between such reversals are approximately gamma distributed [3]. Such distributions are common in biology and can be interpreted in terms of the noise driven motion of a state point which randomly crosses a threshold or surmounts an energy barrier. Recently, there has been a revival of interest in these results in connection with the development of dynamical models of brain function during such reversals in perception [4]. There has also been a growing interest in *stochastic resonance* (SR) associated with noisy nonlinear systems [5]. This is a dynamical behavior wherein noise may enhance the transmission of information through certain systems, such that a defined signal-to-noise ratio (SNR) achieves a maximum for an optimum value of the noise intensity. SR has been demonstrated in numerous physical experiments [5,6] and, more recently, in a simple sensory neuron [7]. The theory of SR was first advanced as a possible explanation of the observed periodic recurrences of the Earth's Ice Ages [8], and, stimulated by an interesting experiment with a bistable ring laser [9], has been the object of numerous recent theoretical studies [10]. The possible importance of SR for the processing of information in neural systems seems evident at all levels from the lower *physiological* levels to the higher *cognitive* ones. Indeed, it has long been recognized that noise can improve the performance of certain neural networks [11], and it may be possible that an optimum noise level can achieve the maximum improvement.

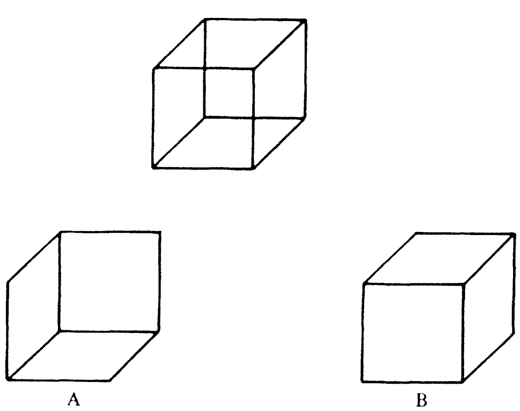

FIG. 1. Necker cube with its two alternative interpretations.

In this Letter, we consider a noisy autoassociative neural network which has previously been shown to be an accurate model of the bistable perceptual process involved in the interpretation of ambiguous figures [12]. Our results indicate that SR, as well as other recently studied features of noisy bistable dynamics, can easily be demonstrated in this system.

Our network is made up of binary neurons (activation levels 0 and 1) which are globally, or all to all, connected [13]. The connection matrix is symmetric with zero diagonal elements, that is, without individual element autoconnections. The network is split into two groups of neurons numbering n_A and n_B, respectively. Each of these groups is associated to a single perceptive state, A and B, respectively, representing the two possible interpretations of the figure. In experiments, the two perceptive interpretations are mutually exclusive, and, in order to reproduce this feature, the excitation of the n_A group must inhibit the n_B group and vice versa. We have thus designed a connection matrix, W, to be obtained through learning the two competitive interpretations by using a Hebbian rule, which is made up of four parts: two square blocks representing the positive (excitatory) connections within n_A and n_B, respectively, and two rectangular blocks representing the negative (inhibitory) interconnections linking the two groups. For simplicity, the connections among neurons within each group are taken to be of equal strength m (n) within the group A (B), whereas the

0031-9007/94/72(19)/3120(4)$06.00

interconnections between the two groups are taken to be negative of strength $-r$. The network described above is a particular case of the general Hopfield one [14], for which an energy function, with its classical expression, can be introduced,

$$E=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}S_iW_{ij}S_j\,, \tag{1}$$

where S_i is the activation of the ith neuron, and W_{ij} is an element of the synaptic connection matrix. The detailed value of E will depend on the network configuration. The particular form of our connection matrix W allows us to simplify this energy function by evaluating the double sum, whence all network configurations with the same number of active neurons in each population have the same energy value. This value is given by

$$E=-\tfrac{1}{2}a(a-1)m-\tfrac{1}{2}b(b-1)n+abr \tag{2}$$

in a configuration of the network with a active neurons in the population n_A and b active neurons in the population n_B. Therefore the energy level of the network is determined by only two independent variables, a and b, with finite integer values, and, obviously, by the three parameters m, n, and r of the connection matrix.

An example of the energy function is shown in Fig. 2 for the parameter values $m=n=0.02$, $r=0.04$. It has two minima located at the vertices $(n_A,0)$ and $(0,n_B)$. These two minima correspond to the two stored configurations in which one population is fully activated and the other fully inhibited. Constraints must be imposed on the connection matrix such that the energy will have only

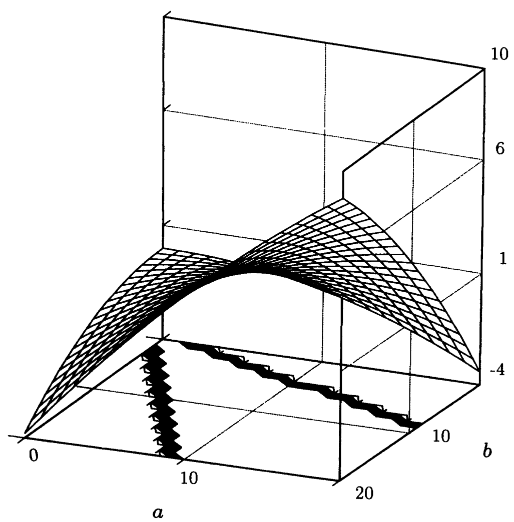

FIG. 2. The energy landscape for the parameter values $m=n=0.02$ and $r=0.04$. Alternate perceptual interpretations of an ambiguous figure are indicated by trajectories lying in the basins of attraction of the points ($a=0$, $b=20$) and ($a=20$, $b=0$). The basins at $T=0$, discussed in the text, are bounded by the dark lines lying in the a-b plane.

these two configurations as minima. In fact, as shown in [12], one can easily derive from Eq. (2) that only the vertices of the bidimensional domain can be minima of the energy function. Whereas the vertex (0,0) is always isoenergetic with its nearest neighbors, the vertex (n_A,n_B) is not a (local) minimum if the condition $r\geq\min[(n_A-1)m/n_B,(n_B-1)n/n_A]$ is satisfied.

We assume that the network evolves according to a modified Hopfield dynamical rule, that is, the fixed point dynamics of the original Hopfield model (which can be interpreted as the transition of a system from an excited state to its fundamental state at the temperature $T=0$ K) is replaced by a dynamics generated by thermal noise. The network state is asynchronously updated at each time step by changing the activation level ($0\rightarrow1$ or $1\rightarrow0$) of one, randomly chosen neuron. After each update, a new value of the energy function is calculated, and the corresponding state is either accepted or rejected according to the following. (a) The Hopfield rule ($T=0$): accept only if $\Delta E\leq0$, otherwise reject; and (b) the modified Hopfield rule ($T>0$), using the Metropolis algorithm [15]: accept with probability one if $\Delta E\leq0$, accept with probability $p=\exp(-\Delta E/k_BT)$ if $\Delta E>0$. In this way the effects of nonzero temperature, or noise, are accounted for by the simulation. For $T=0$, basins of attraction are defined by the jagged lines in the a-b plane. Any trajectory begun on the grid to the lower left (upper right) of the lines will with probability 1 arrive at the vertex $b=0$ ($a=0$). Trajectories beginning on grid locations between the lines may terminate on either vertex.

An external periodic signal can be introduced in the form of a small additive perturbation to the energy,

$$E=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}S_iW_{ij}S_j+H_0\sin(\omega_0t)\left(\sum_{i\in A}S_i-\sum_{i\in B}S_i\right), \tag{3}$$

where in the second term H_0 and ω_0 are the amplitude and frequency of the signal, and the summations extend over all neutrons in the populations n_A and n_B. Adopting the analogy between Hopfield networks and spin systems, the external signal corresponds to an oscillating weak magnetic field oriented in two opposite directions on the two populations. During alternate half cycles, the field favors the activation of one population and the inhibition of the other. Trajectories generated during one half cycle then preferentially migrate toward one of the vertices and toward the other vertex on the alternate half cycles. Thus the probability to realize one of the interpretations is periodically modulated. The time series of a trajectory for $T>0$ is thus represented by a series of switching events between the attractors. These events occur at more-or-less random times but with some degree of coherence with the signal. For moderate values of the temperature, the signal frequency was small compared to the mean rate of switching between attractors (and always very small compared to the rate at which the net-

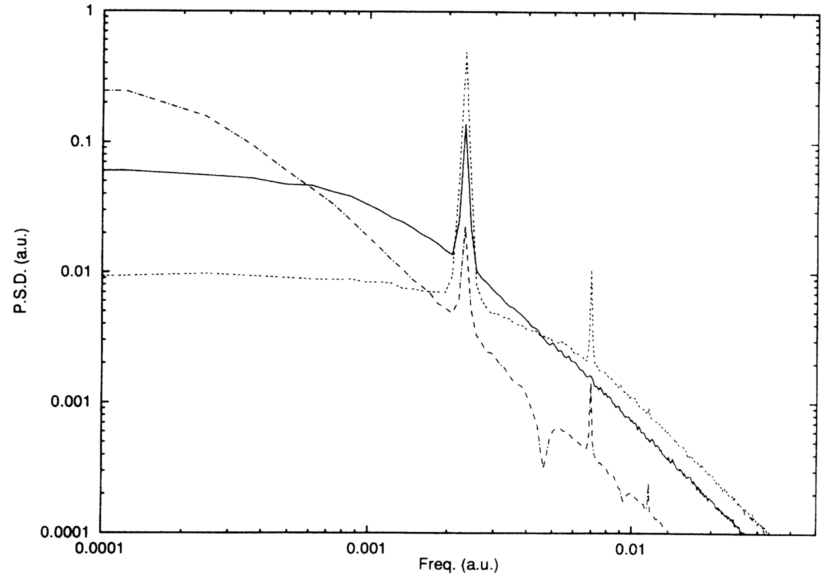

FIG. 3. Power spectra of the time series representing sequential perceptive interpretations for three temperatures: dashed, $T=0.65$; dotted, $T=1.20$; and solid, $T=2.90$ (scaled for $k_B \equiv 1$).

work was updated). This corresponds to the adiabatic approximation adopted in nearly all SR theories [10].

The power spectrum of the time series was computed and averaged with representative results shown in Fig. 3. The details of these spectra are very similar to those previously observed for SR systems, that is, they are characterized by a broad band, Lorentzian noise background, with signal features in the form of sharp peaks at the signal frequency and its odd harmonics. Three spectra, taken at small, intermediate, and large noise intensities, are shown. It is clear that the largest signal features, at both the fundamental and the third harmonic, are evident at the intermediate noise level. Moreover, a dip or "hole" is evident at the first two even harmonics. These holes, which appear only for low noise intensities, have been observed previously [16], were at first unexplained, but have now been accounted for in a recent and elegant theory [17].

SR can be observed by comparing the signal strength, S (the area under the fundamental peak), to the noise level at the fundamental frequency, $N(\omega_0)$. The SNR is thus calculated from the usual formula,

$$\mathrm{SNR} = 10 \log_{10}[S/N(\omega_0)]$$

in dB. The results for three different signal frequencies are shown in Fig. 4. We note that for higher frequencies the maximum moves toward larger values of T (higher noise levels) and its amplitude is decreased. Both the shape of the SNR versus T curves and this behavior of the maximum are characteristic of SR as demonstrated by the theory [10].

We conclude with the observation that, while it has been accepted for some time that noise must play some role in perceptive processes, the nature of that role, and indeed whether the noise is useful or not, remains quite obscure. Our demonstration of SR in a realistic perceptual model, albeit of a very simple process, is suggestive of usefulness and expands the range of application into the area of brain function.

We are grateful to Frank Moss for stimulating suggestions and helpful comments. We also thank Donatella Petracchi, who took part in early discussions on this work. This work was supported by the National Research Council (CNR) of Italy, within the strategic project: "Nonlinear Dynamics in Biological Systems."

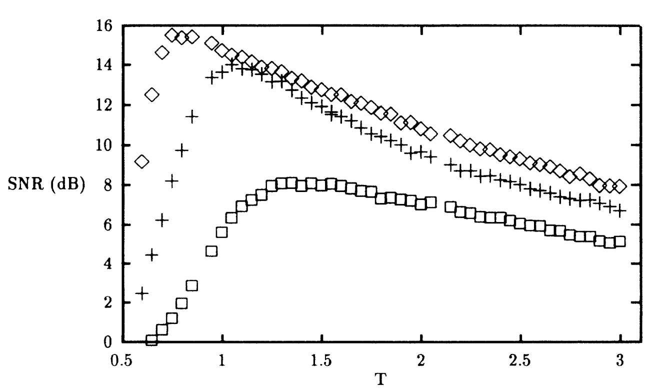

FIG. 4. The SNR versus T for three frequencies: diamonds, $f=0.0009$; plus signs, $f=0.0023$; and squares, $f=0.01$. The characteristic signature of SR, a maximum SNR at an optimum temperature (noise level), is evident.

[1] See, for example, H. Goldhamer, Am. J. Psychol. **46**, 189 (1934), and references therein.
[2] L. A. Necker, Philos. Mag. **3**, 329 (1832).

[3] A. Borsellino, A. De Marco, A. Allazetta, S. Rinesi, and B. Bartolini, Kybernetic **10**, 139 (1972); K. T. Brown, Am. J. Psychol. **68**, 358 (1955); A. Borsellino, F. Carlini, M. Riani, M. T. Tuccio, A. De Marco, P. Penengo, and A. Trabucco, Perception **11**, 263 (1982).

[4] H. Gang, T. Ditzinger, C. Z. Ning, and H. Haken, Phys. Rev. Lett. **71**, 807 (1993); T. Ditzinger and H. Haken, Biol. Cybern. **63**, 453 (1990); **61**, 279 (1989); M. Riani, F. Masulli, and E. Simonotto, Proc. SPIE Int. Soc. Opt. Eng. **1469**, 166 (1991).

[5] See, for example, *Proceedings of the NATO Advanced Research Workshop on Stochastic Resonance in Physics and Biology*, edited by F. Moss, A. Bulsara, and M. F. Shlesinger [J. Stat. Phys. **70**, 1 (1993)].

[6] F. Moss, in "An Introduction to Some Contemporary Problems in Statistical Mechanics," edited by G. H. Weiss (SIAM, Philadelphia, to be published).

[7] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, Nature (London) **365**, 337 (1993).

[8] R. Benzi, S. Sutera, and A. Vulpiani, J. Phys. A **14**, L453 (1981); R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, Tellus **34**, 10 (1982); C. Nicolis, Tellus **34**, 1 (1982).

[9] B. McNamara, K. Wiesenfeld, and R. Roy, Phys. Rev. Lett. **60**, 2626 (1988).

[10] B. McNamara and K. Wiesenfeld, Phys. Rev. A **39**, 4854 (1989); P. Jung and P. Hänggi, Europhys. Lett. **8**, 505 (1989); P. Jung, Z. Phys. B **16**, 521 (1989); L. Gammaitoni, F. Marchesoni, E. Menichella-Saetta, and S. Santucci, Phys. Rev. Lett. **62**, 349 (1989); P. Jung and P. Hänggi, Phys. Rev. A **44**, 8032 (1991).

[11] J. Buhmann and K. Schulten, Biol. Cybern. **56**, 313 (1987).

[12] F. Masulli, M. Riani, and E. Simonotto, Proc. SPIE Int. Soc. Opt. Eng. **1710**, 267 (1992).

[13] S. Watanabe and S. H. Strogatz, Phys. Rev. Lett. **70**, 2391 (1993), and references therein; P. Jung, U. Behn, E. Pantazelou, and F. Moss, Phys. Rev. A **46**, R1709 (1992); A. Bulsara and G. Schmera, Phys. Rev. E **47**, 3734 (1993).

[14] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982); **81**, 3088 (1984).

[15] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[16] T. Zhou and F. Moss, Phys. Rev. A **41**, 4255 (1990); L. Kiss, Z. Gingl, Z. Marton, J. Kertesz, F. Moss, G. Schmera, and A. Bulsara, in *Proceedings of the NATO Advanced Research Workshop on Stochastic Resonance in Physics and Biology* (Ref. [5]), p. 451.

[17] V. A. Shneidman, P. Jung, and P. Hänggi (to be published).