Complex Systems **10** (1996) 55–71

Information-theoretic Aspects of Neural Stochastic Resonance

Joseph C. Park
Department of Ocean Engineering,
Florida Atlantic University,
Boca Raton, FL, USA

P. S. Neelakanta
Department of Electrical Engineering,
Florida Atlantic University,
Boca Raton, FL, USA

Abstract. Information flow through a neural network learning to recognize state-transition statistics produced from a nonlinear, bistable detector under conditions of stochastic resonance (SR) is investigated. The information flow dynamics are examined in terms of an information-theoretic cost-function defined by the relative informational entropy associated with an ensemble of training sets averaged over the temporal evolution of training cycles. The network architecture consists of a multilayer perceptron evolving under the guidance of the back-propagation algorithm. For the purpose of emulating SR, a Schmitt-trigger logic is utilized as the nonlinear detector, and generates state-transitions exhibiting SR in response to a sine-wave signal superimposed with gaussian noise. The output statistics of the Schmitt trigger are used to train the multilayer perceptron towards recognizing the extent of SR present in the detector state-transition dynamics. It is demonstrated that information flow dynamics under conditions of SR are inherently more informatic (or less negentropic) than cases wherein the state-transition statistics are dominated by nonSR conditions, that is, under higher or lower signal-to-threshold ratios. Some details concerning SR in relation to biological neurons are also discussed.

1. Introduction

In sensory detection and classification applications, the inevitable presence of noise typically degrades performance of the system. However, in certain nonlinear systems concerned with the detection and classification of weak, periodic signals, an optimum signal-to-noise ratio may be encountered at a nonzero value of the additive noise level. This phenomenon has been termed as stochastic resonance (SR).

SR was originally postulated as a mechanism to explain the 100000 year periodicity of glacial epochs in the climate of Earth [1]. It purported that a weak periodicity in solar radiation caused by variations in the orbital dynamics of Earth could result in periodic transitions of the climatic behavior of Earth between glacial and interglacial states. The initial experimental detection of SR was obtained from measurements of a signal-plus-noise driven bistable electronic circuit, namely the Schmitt trigger [2]. (A Schmitt trigger is a bistable, dual-threshold device that changes its state when an input stimulus crosses the upper threshold from below, or when the stimulus crosses the lower threshold from above.) SR has also been verified in a bidirectional ring laser, in which a periodic signal not coupled with noise, would not result in correlated laser modulations [3]. Further, SR has also been demonstrated in many other dynamical systems such as the single-well potential system, and in the "integrate and fire" dynamics employed commonly to model the response of sensory neurons [4]. More recently, SR has been verified in a number of important biological processes such as information-transfer from mechanoreceptor hair cells situated on the tail of the red swamp crayfish *Procambarus clarkii* [5], cercal sensory neurons of crickets [6], signal-transduction across the voltage-dependent ion-channels of the cell membrane [7], in neuronal networks prepared from temporal lobe hippocampal sections of the mammalian brain [8], and even in the exteroreceptive somatic nervous system [9] as well as muscle-spindle afferents [10] of *Homo sapiens*. Biological neurons operate in an inherently noisy environment, and the fact that SR has been observed in them indicates that such noise is an integral part of neurocellular activity. From a signal processing perspective, SR results in an optimum input/output signal-to-noise ratio at nonzero values of noise intensity. Viewed from the information theory standpoint, this means that noise contributes to a significant extent in the development of a maximum rate of information transfer through the system.

Contemporary studies have applied information measures (entropies or mutual information) to quantify the noise-induced, maximum information transfer resulting from SR [11–13]. The relevant results demonstrate that models of a single neuron do exhibit an information rate maximization at nonzero noise levels. They do not, however, illustrate whether or not cooperative populations of a neural assembly make use of SR in their information transfer endeavors. The present study, therefore, is intended to elucidate the influence of SR on neural learning in an artificial neural network (ANN).

The detection and transfer of information from neural sensory input to the mammalian brain is known to be encoded in the time-interval sequence between "firings" of neurons which constitute a train of action potentials. A statistical examination of these time-interval sequences (or interspike intervals) reveals that the coding follows a stochastic process. Even though the statistics of action potential trains have been investigated comprehensively [14], it is not established precisely how the associated sensory information is encoded. It has been assumed historically that appropriate mathematical models which represent the statistics of time-interval sequences of neurosensory action potentials refer to a homogeneous Poisson point process; and, they belong to the family of modified Poisson point processes with the inclusion of temporal refractory effects. These models are inherently renewal processes, indicating that they predict independent, identically-distributed, interevent intervals. Such intervals are implicitly uncorrelated; and therefore can be described completely (in a statistical sense) by their interevent-interval histograms (IIH). However, it has also been shown [15] that sequences of interevent intervals may exhibit positive correlations over long time periods, and renewal process based statistical models are rather inadequate for a complete description of such action potential statistics having correlatory (markovian) attributes. In fact, if one were to measure the action potential rate in mammalian auditory nerves, it can be observed that fluctuations in the rate do not often subside to any significant extent even for very long averaging periods; that is, the rate exhibits sustained fluctuations on all time scales [15]. Therefore, it has been suggested [16] that a complete statistical model for the action potentials should be based on a fractal-stochastic-point process.

Interestingly, regardless of the model employed to describe the temporal attributes of the neural spike occurrences, the presence of interevent correlations do not alter the outcome of the associated IIH. The IIH can, therefore, serve as a tool to demarcate the presence or absence of SR and is adopted as a simplified statistical tool to model the sensory neuron response.

In this paper, the information flow through an ANN adapting to the IIH produced by a simulated neuron in response to input conditions of SR is investigated. ANNs are a convenient tool for modeling the flow of neural information. They are loosely based on the information-processing structure of biologic neurons, and derive their ability to learn complex mappings from their massively parallel interconnection structure and the inherent nonlinearity of individual neurons. ANNs, in general, operate in two distinct phases: The first is the training phase, where the network adjusts its internal parameters in response to a set of training data or rule-bases. Second, is the predictive phase, where the trained network responds to input data and produces a functional mapping. ANNs can be classified generally either as supervised, or unsupervised architectures. Supervised paradigms require a "teacher" that produces an error-output in response to the training data, and minimization of the error (or cost-function) directs the learning process, which is essentially an adjustment of internal parameters of the network. An unsupervised network does not require an external teacher to generate an error, but relies on a rule-base to adjust its internal parameters in response to the network output during training. ANNs can be classified further as either feedforward, or feedback types. This classification refers to the flow of information during the predictive phase of operation.

As pointed out by the authors in [17], the flow of information during the training phase in an ANN is a competitive endeavor between the positive and negative information available. Positive information relates to the actual learning of the network, namely, its ability to predict the desired outcome with an increasing degree of confidence. Negative information refers to confusing the learning process by information arising from extraneous parameters or due to the inevitable presence of noise. It is therefore indicated in [17] that it is more appropriate to model the information transfer during the learning process in terms of the classical definition of information based on entropy concepts [18]. Further, it is shown in [17] that the dynamics of information flow during ANN learning correspond to that of a stable, chaotic system with a dominant (well-defined) attractor basin.

The essence of this paper is to synthesize the dynamical behavior inherent in SR with the information rate dynamics of neural learning in order to assess the information transfer characteristics of neural systems learning to optimally exploit SR. It is demonstrated that the flow of information during the training phase of an ANN is more informatic when the ANN models a dynamical system exhibiting SR, than in the case of a similar nonSR system. That is, a nonSR dynamic results in a larger information-theoretic entropy (network uncertainty) during learning than the SR case.

The paper is organized as follows. First, the statistical characteristics of state-transitions exhibited by biological neurons in response to SR are reviewed. Next, the Schmitt trigger as a bistable nonlinear detector is described, and a general threshold parameter is identified to quantify the signal-plus-noise input in relation to the thresholds of the detector. Results of computer simulations which generate state-transition statistics of the Schmitt trigger in response to the signal-plus-noise are then presented and contrasted to relevant biological results. Comparison of the biological response statistics with those produced from the computer simulations are used as a criterion to establish the presence or absence of SR. Pertinent computer generated response statistics and corresponding detector threshold parameters are subsequently used as training sets for the ANN. The ANN is then described, along with details of its implementation that provide a mechanism for quantifying the flow of neural information during learning. This is achieved through the use of an information-theoretic cost function in the backpropagation algorithm. The results of the ANN learning evolutions are then quantified and discussed in terms of the average uncertainty perceived by the ANN during training. Lastly, concluding remarks regarding the flow of neural information through systems attempting to maximize the extraction of information from weak periodic signals buried in noise by means of the mechanism of SR presented in this paper are discussed.

2. Interevent-interval histograms and stochastic resonance

2.1 Biological characteristics of interevent-interval histograms

In [19] measured IIH data obtained 23 years apart from the auditory nerve fibers of a squirrel monkey, and from the primary visual cortex of a cat, subjected to periodic stimulus are analyzed. The following substantiative features of biologic IIH in response to weak periodic stimuli are observed.

1. Response modes are located at integer multiples of the stimulus period.

2. Response mode amplitudes decay approximately exponentially, as indicated by the linear slope of mode amplitudes on a logarithmic scale.

The second feature suggests that the spike rates are governed by rate processes.

Noise is required to produce biologically justifiable IIH for at least the following two reasons.

1. The firing threshold of the system is above that of the periodic stimulus amplitude.

2. Noise is required to produce response modes at intervals other than the stimulus period.

In [19] the sensory neuron information transmission on the basis of bistable Schmitt trigger dynamics is modeled. By comparing the resulting state-transition statistics of the bistable model with the experimentally obtained physiological results, it was concluded that noise is a requisite component in order to produce experimentally justifiable IIH when neurons are stimulated by periodic inputs.

2.2 Computer simulations of stochastic resonance

As indicated above, the basic statistical description of interevent intervals which result from the transition dynamics of nonlinear systems exhibiting SR have been verified in both biological and electronic systems. In this paper, the approach due to [19] is pursued and a Schmitt-trigger logic is employed to simulate the interevent transitions under conditions of SR. Statistics of the interevent transitions are then extracted as depicted in Figure 1 for subsequent application in the ANN. The Schmitt trigger is a dual-state, dual-threshold device. Let HI and LO represent the two stable states, and denote the corresponding thresholds as α_{U} and α_{L} respectively, where $\alpha_{\mathrm{U}} > \alpha_{\mathrm{L}}$. The state-point logic of the Schmitt trigger is that LO$\rightarrow$HI if the input crosses α_{U} from below and the state is not already HI; and, HI$\rightarrow$LO if the input crosses α_{L} from above and the state is not already LO.

Let the noise considered be zero-mean gaussian-distributed with a variance σ^2. It is added to a periodic process with a peak-to-peak amplitude of $2A$. Relevant to the application of this signal-plus-noise to the bistable Schmitt trigger, it is assumed that $\alpha_{\mathrm{U}} > A$. The lower threshold α_{L} is assumed to take a value near A but below a_{U}. An appropriate system parameter that encompasses the relevant features of signal-plus-noise applied to a threshold detector is the ratio of threshold excess power, namely $(\alpha_{\mathrm{U}} - A)^2$

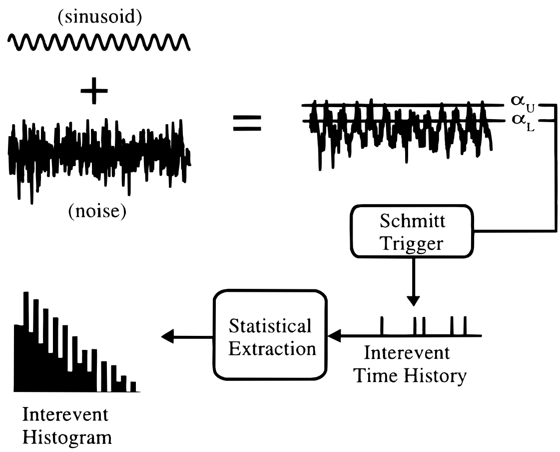

Figure 1: Extraction of interevent statistics from a Schmitt trigger subjected to a small amplitude sinusoid, embedded in noise, producing conditions of SR.

to the variance (σ^2) of the noise. Hence, the system threshold parameter γ can be defined as follows:

$$\gamma^2 = \frac{(\alpha_U - A)^2}{\sigma^2}. \tag{1}$$

This single parameter can be used to characterize the power of the excess signal-plus-noise required to initiate a state-transition of the detector relative to that of the noise. In the case when $\sigma \to \infty$, then $\gamma \to 0$ and it is pointless to attempt extracting useful information from the statistics of the detector state transitions. If the noise variance is greater than the required threshold excess power, that is, when $(\alpha_U - A) < \sigma$, it follows that $0 < \gamma < 1$. This parameter regime encompasses the low signal-to-noise conditions and detector threshold levels which constitute a hallmark of SR. In the event of the unlikely occurrence that the noise variance and excess power are precisely matched, then $(\alpha_U - A) = \sigma$, with $\gamma = 1$; and lastly, if the transition threshold is set far above the level of signal amplitude in relation to the noise, then $(\alpha_U - A) > \sigma$, resulting in $\gamma > 1$.

A computer program was implemented with the Schmitt-trigger logic to produce IIH data for any value of γ using a sine wave plus zero-mean gaussian, or uniform noise. The number of points of signal-plus-noise data which were processed in each simulation was 50000. Figure 2 is a typical IIH that results from threshold and noise conditions specified by $\gamma = 0.8$. It may be noted that the primary response modes are located at integral multiples of interevent intervals normalized by the period of the sine wave, and the amplitude of the modes decay approximately exponentially, in agreement with the statistics of biological IIH as indicated in [19]. (The optimum range of γ which resulted in biologically plausible IIH ranged from 0.7 to 0.9.)

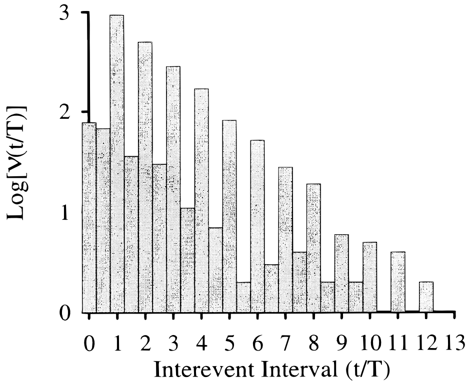

Figure 2: IIH produced from a Schmitt trigger subjected to a sine wave plus gaussian noise with a threshold parameter $\gamma = 0.8$. The interevent intervals (t) are normalized with respect to the sine wave period (T). ν is the number count of Schmitt trigger state-transitions that occurred.

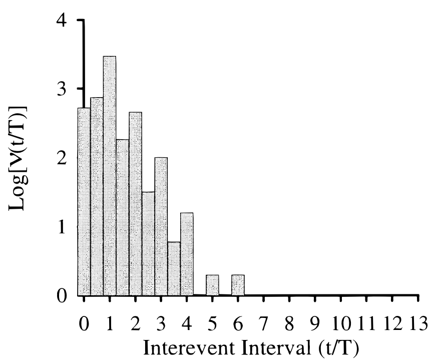

Figure 3: IIH produced from a Schmitt trigger subjected to a sine wave plus gaussian noise with a threshold parameter $\gamma = 0.1$. The interevent intervals (t) are normalized with respect to the sine wave period (T). ν is the number count of Schmitt trigger state-transitions that occurred.

When values of γ are less than 0.7, the detector threshold is set slightly above the level of the periodic signal amplitude, and the resulting state-transitions are dominated by the periodic signal as illustrated in Figure 3. In the case of $\gamma > 1$, the threshold level is excessively distant from the signal amplitude (in relation to the noise power), so that the noise fluctuations control the detector dynamics. The resulting IIH then tends to exhibit amplitudes distributed across the range of interevent-intervals as shown in Figure 4.

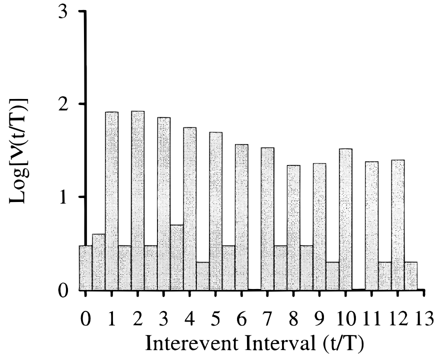

Figure 4: IIH produced from a Schmitt trigger subjected to a sine wave plus gaussian noise with a threshold parameter $\gamma = 1.5$. The interevent intervals (t) are normalized with respect to the sine wave period (T). ν is the number count of Schmitt trigger state-transitions that occurred.

3. Multilayer perceptron neural network

In this section a multilayer perceptron ANN is implemented to evaluate learning dynamics involving SR. The perceptron is a supervised feedforward network (Figure 5). That is, it requires an error-source or cost-function of the network output in response to training data during the learning phase. The network architecture is configured in a multilayer structure. Specifically, the network consists of one input layer, one hidden layer, and one output layer. The input layer contains 26 neurons, the hidden layer has 15 neurons, and there is a single neuron in the output layer as depicted in Figure 5.

The input layer serves as a signal-multiplexer, and routes the value of each input to each neuron in the hidden layer. Connecting each neuron in successive layers is a trainable weight. The weight between the ith and jth units, denoted as w_{ij}, is a numerical value which is multiplied by the output of the ith unit. This weighted value is then presented as one of the multiinputs to the jth unit. Each weight is modified during the training process to produce a minimum error output from the network. Neurons in the hidden and output layers perform computations as follows. Let

$$X_i = \sum w_{ij} x_j$$

represent the weighted sum of the multiinputs x_j. This summed input is processed by the activation function to produce the neurons output signal O_i. The activation function provides each neuron with a nonlinear transfer function, so as to allow the processing of large input values devoid of overload, while simultaneously facilitating sensitive response to low-level input activity. The activation functions used in the hidden layers are sigmoidal Bernoulli functions $L_Q(x)$, with $Q = 1/2$, [20] and a linear function in the output layer. The linear output activation functions allow the network output to converge

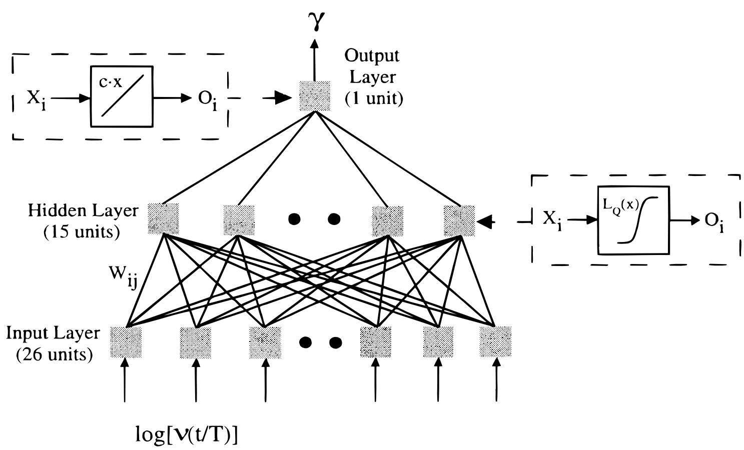

Figure 5: Architecture of the multilayer perceptron implemented to learn the state-transition statistics associated with SR.

to values outside the ± 1 interval set by the Bernoulli function bounds. The input and hidden layer also have an additional bias unit clamped to a fixed output of -1, connected to each unit in the succeeding layer through a trainable weight.

3.1 Training phase

The objective of the training process is to allow the network to learn the functional mapping of input data to the desired output vector. This is achieved by repeatedly presenting to the network a known set of input/output pairs (training sets) and adjusting the weights to minimize some measure of error, or cost-function, between the desired output and the computed network output. In the case of the multilayer perceptron, the conventional error minimization approach refers to the so-called backpropagation algorithm [21, 22]. The fundamental entity used in the weight adjustment process is the error ϵ_i of the network output O_i at the ith unit, deviating from a target value T_i. It is used to calculate the effective gradient of the weight modification term in the backpropagation algorithm. The effective gradient δ_j has two distinct definitions depending on whether or not a target value is available for a particular unit. In the case of network output units, for which a target is known, δ_j is defined as the distance (error) of the jth unit times the derivative of the activation function evaluated at the output value of the ith unit. That is, $\delta_j = (\partial O_i / \partial X_i)\epsilon_j$ where X_i represents the ith unit input to the activation function. When the unit resides in a hidden or input layer, a target value is not available for computation of the network error ϵ. In this case, the definition is modified such that the product of cumulative effective gradients from the next layer and the interconnection weights are backpropagated to these units. In other words: $\delta_i = (\partial O_i / \partial X_i) \sum_j \delta_j w_{ij}$. With the appropriate expression for the error gradient in hand, the basic prescription for adjustment of weights at the nth training step is given by the well-known Widrow–Hoff delta rule [21], namely,

$$w_{ij}(n) = w_{ij}(n-1) + \eta \delta_j O_i = w_{ij}(n-1) + \Delta w_{ij}(n)$$

where η is the learning rate.

In regions of the error surface where large gradients exist, the δ terms may become inordinately large. The resulting weight modifications will also be large leading to extensive oscillations of the network output, preventing convergence to the true error minimum. The learning coefficient can be set to an extremely small value to counteract this tendency; however, this would drastically increase the training time. To avoid this situation, the weight modification can be given a "memory" so that it will no longer be subject to abrupt changes. That is, the weight change algorithm is specified by:

$$\Delta w_{ij}(n) = \eta \delta_j O_i + \lambda[\Delta w_{ij}(n-1)]$$

where λ is known as the momentum parameter. If λ is set to a value close to 1, the search in the parameter space will be determined by the gradient accumulated over several epochs instead of a single iteration, improving the stability of the network towards convergence.

In order to train the network as robustly as possible, it is desirable to expose it to an ensemble of training sets during the learning phase. This will avoid incorporation of details specific to a particular training pattern into the memory of the network. Therefore, the network is trained with L distinct training patterns for each realization of signal-plus-noise and threshold conditions. These L training patterns are presented sequentially to the network at each learning step n.

3.2 Cost-function

The cost-function commonly employed in ANN training is the euclidean distance which is specified by $\epsilon_i = (T_i - O_i)$. This error-metric is usually represented in absolute terms as the root of the square error

$$\epsilon_{\mathrm{SE}} = \sqrt{(T_i - O_i)^2}$$

and is referred to as the square error (SE) cost-function. The SE is solely a measure of the network deviation from the desired goal, and it does not quantify implicitly the information flow dynamics of the learning process. Alternatively, the cost-function can be specified in terms of relative informational entropy (ϵ_{RE}) between the present state of the network output and the desired network goal [18]:

$$\epsilon_{\mathrm{RE}} = \frac{1+T_i}{2} \log\left[\frac{1+T_i}{\beta + O_i}\right] + \frac{1-T_i}{2} \log\left[\frac{1-T_i}{\beta + O_i}\right] \tag{2}$$

where

$$\beta = a^2 + b^2 \left[\left(\frac{1}{\tanh(bO_i)} \right)^2 + \left(\frac{1}{\sinh(bO_i)} \right)^2 \right] - \frac{2ab}{\tanh(aO_i)\tanh(bO_i)}$$

and $a = (3Q+1)/2$, $b = (3Q-1)/2$, with Q as the parameter of the Bernoulli activation function $L_Q(x)$. During learning, weight modifications are made proportional to the gradient of the network error. In the information-theoretic domain, this means that learning is taking place concordant with the rate of information flow towards or away from the desired goal as dictated by the instantaneous stochastic state of the network.

3.3 Gross-features of network uncertainty

Since the relative entropy error measure is an implicit measure of the relative informational entropy, an examination of the neural learning curve under the direction of relative entropy quantifies the flow of neural information during the learning endeavor. Specifically, the value of relative entropy (ϵ_{RE}) at any discrete learning step n, averaged over the L input training patterns, will quantify the average relative uncertainty of the current network organization:

$$H_L(n) = \frac{1}{L} \sum_{l=1}^{L} \epsilon_{\mathrm{RE}}(n). \tag{3}$$

Further, it is necessary to ensure that the network training is truly representing the average information available for processing during learning. Therefore, we seek to minimize the possibility that a single network learning cycle, consisting of $n = 1, 2, \ldots, N$ sequential presentations of the L input patterns, would result in an unrepresentative network evolution due to a particular choice of random numbers that were selected to initialize the interconnection weights. To avoid this, the network is trained $m = 1, 2 \ldots, M$ times, with the L training sets presented N times, using a different random number seed to initialize the network weights for each of the M runs. The average network uncertainty at a particular learning step n for L training sets after M training cycles, is then given by:

$$H_M(n) = \frac{1}{M} \sum_{m=1}^{M} H_{Lm}(n). \tag{4}$$

This quantity represents the average informational relative entropy of the network goal-directed organization as the network seeks to minimize the divergence of its output from that of the desired goal. It will be used to examine the discrete-step temporal evolution of neural learning as the multi-layer perceptron learns to recognize the statistics of SR.

4. Simulation results

The perceptron is trained to learn the functional mapping from IIH input, to threshold parameter output, for the cases of SR and nonSR discussed earlier in section 2.2. The number of training sets for each threshold parameter realization is $L = 10$. The network was trained $M = 10$ times for each threshold parameter specification with a distinct random number seed in each case. The inputs to the ANN are the amplitude values of the IIH computed from the Schmitt-trigger logic, corresponding to a fixed value of γ (the ordinate of Figures 2, 3, or 4). The inputs are specified at 26 values covering the range from 0 to 12.5 in increments of 0.5, corresponding to the interevent intervals normalized by the forcing period (the abscissa of Figures 2, 3, or 4). The network output is defined by the respective value of the threshold parameter γ. The ANN was trained *via* the backpropagation algorithm with the relative entropy error-metric. A learning coefficient of $\eta = 0.0015$ was adopted along with a momentum term of $\lambda = 0.9$. In order to provide a valid comparison of learning dynamics for different threshold parameter conditions, the random number generator which sets the initial interconnection weights is initialized with the same seed for each of the successive M training cycles at the start of each training run. For example, to compare the learning dynamics for $\gamma = 0.8$ and $\gamma = 0.1$, each of the M training cycles for the distinct values of γ are started with the same random number seed. This ensures that the evolution of network dynamics for each value of γ are initiated identically.

Figure 6 compares H_M as a function of training iterations for the case of SR ($\gamma = 0.8$) with that of the nonSR statistics where the detector dynamics are dominated by the periodic signal ($\gamma = 0.1$). It can be seen that when SR is present, the informational entropy during the early learning stage is significantly lower than when it is absent. It is also observed that the SR network configuration completes learning (cessation of oscillatory behavior) at $n = 60$, while for the nonSR case, network convergence is achieved at $n = 70$. There is not a significant deviation from these values up to $n = 120$ learning steps, the maximum extent of the simulations.

A comparison of the network information dynamics of the SR case ($\gamma = 0.8$) with the noise-dominated statistics, $\gamma = 1.5$, is presented in Figure 7. It is observed that the early learning phase is not significantly different, except for the large uncertainty of nonSR dynamics at the initiation of training. Both systems have converged to a stable network configuration at $n = 60$, however, the nonSR system has migrated to a state of larger entropy. The nonSR system is, therefore, less certain about the proximity of its output to the goal than in the SR case. Figures 6 and 7 demonstrate that the nonlinear detector statistics inherent under conditions of SR result in a more efficient organization of information flow during neural learning than when SR is absent. Therefore, the learning of an ANN under conditions of SR are robustly more informatic (less negentropic) than under similar conditions when SR is not present.

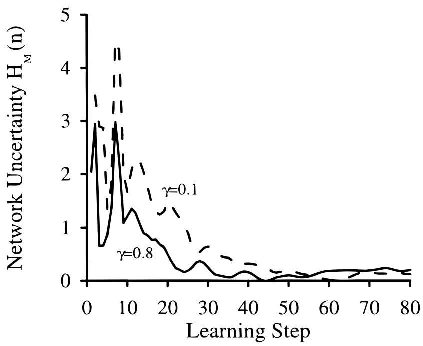

Figure 6: Evolution of ANN average information flow for threshold parameters of $\gamma = 0.1$ and $\gamma = 0.8$.

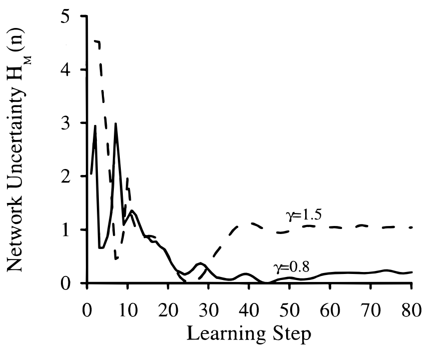

Figure 7: Evolution of ANN average information flow for threshold parameters of $\gamma = 0.8$ and $\gamma = 1.5$.

After training the ANN, it is of interest to assay the total extent of relative entropy experienced by the network during the learning process. This cumulative average uncertainty perceived by the neural complex during learning can be quantified in terms of the relative entropy error measure for the M distinct ensemble presentations of the L input patterns, over N discrete learning steps as:

$$H_{M,N} = \frac{1}{N} \sum_{n=1}^{N} H_M(n). \tag{5}$$

This average informational entropy was computed for the neural learning curves shown in Figures 5 and 6 and is presented in Table 1 for two values of the terminal learning step N. The cumulative uncertainty of equation (5) was computed for the two terminal values of N, since the question of "when has the network converged?" is always open for interpretation. Examination of the early learning stage, $N = 50$, reveals that the total relative uncertainty

Table 1: Total extent of average network uncertainty $H_{M,N}$ perceived during the learning phase.

Learning Steps	Threshold Parameter γ		
N	0.1	0.8	1.5
50	1.148	0.586	1.009
70	0.844	0.462	1.019

for the SR conditions ($\gamma = 0.8$) are roughly one-half of those when SR is absent. This condition remains valid for a terminal training step of $N = 70$, at which point it is clear that the network has fully organized in all cases. Therefore, the informational uncertainty experienced by the neural complex during the learning of a SR process has about one-half of the uncertainty associated with it than the nonSR process.

5. Conclusion

Stochastic resonance arises in a wide variety of nonlinear detectors exposed to the coupling of a weak periodic signal with noise. In the case of a bistable system, one of its attributes is the quasiperiodic modulation of the transition probabilities of dichotomous states of the bistable potential. This indicates that for a neuron modeled as a nonlinear bistable detector, a pulse train constituted of quasiperiodically modulated state-transitions can serve as a robust representation of proliferating neural events along a time axis. In a biological sense, this model refers to the firing events as triggered by threshold crossings of the membrane potential undergoing a biased random walk, driven by excitatory and inhibitory synaptic potentials in the presence of intra- and/or extraneural disturbances. It also reflects the robustness of competitive negentropic and posentropic synergism leading to a stable attractor point for the neural dynamics as observed recently by the authors [17, 23] and others [24]. This robustness refers to achieving an isomorphism between the mapping of information flow at the input and output sets of the neural network that preserve certain properties of the input domain set.

The relative information in the domain space ($\mathcal{D}$) and the range space ($\mathcal{R}$) can be considered in terms of an information density function defined as the difference between the maximum entropy and the observed entropy. In the case of the network output range: $H_{\mathcal{R}} = (H_{\mathcal{R}}^{\max} - H_{\mathcal{R}}^{\text{obs}})$. Similarly, the information density over the input domain space can be defined as $H_{\mathcal{D}} = (H_{\mathcal{D}}^{\max} - H_{\mathcal{D}}^{\text{obs}})$. Within these bounded values of entropy associated with the input domain and the output range of the network, the fidelity of narrowing the output range of information (being close to a target value) can be improved (without sacrificing the variety associated with it), if the input domain of information is manipulated within a narrow range of additive noise corresponding to established signal amplitude and detector threshold conditions. The added noise at the input domain increases the maximum value of disorder so that the network training robustly self-imposes certain constraints and ordering. As a result, the Shannon redundancy is introduced implicitly to reduce the range of uncertainty at the output. The domain-to-range concept of stochastic resonance isomorphism when viewed in Shannon's perspective, indicates that the more negentropic (less informatic) aspect of the input domains, and less negentropic (more informatic) certainty of the output range settling in the basin of convergence, is a synergistic endeavor which utilizes both the randomness (introduced *via* noise) and the orderliness achieved through self-organization of the network. This endeavor is specific to set detection thresholds, facilitating the maximization of net information transfer. The added noise robustly allows the search-process of self-organization in the network to find a more compact set (of less cardinality) which can be accommodated in the global minimal basin of attraction.

Stochastic resonance in both real and artificial neural networks refer implicitly to an adaptive scheme of information processing. As indicated by simulation studies in [25], neural networks require a "parametric tuning" in order to force them to display a robustness in converging towards a steady-state. Otherwise, initial behavior of the network would typically either degenerate rapidly to zero activity, or would tend to an unstable dynamic with wide oscillatory fluctuations of neural firing. Considering stochastic resonance, the parameter tuning can be introduced *via* a narrow range of additive noise with respect to the set detection thresholds, which influences the spatiotemporal neural activity and maximizes the information transfer through the network. The result is an efficient convergence towards network stability.

References

[1] Benzi, R., Parisi, G., Sutera, A., and Vulpiani, A., "Stochastic Resonance in Climate Change," *Tellus*, **34** (1982) 10–16.

[2] Fauve, S. and Heslot, F., "Stochastic Resonance in a Bistable System," *Physics Letters* **97**A (1983) 5–7.

[3] McNamara, B., Wiesenfeld, K., and Roy, R., "Observation of Stochastic Resonance in a Ring Laser," *Physical Review Letters*, **60** (25) (1988) 2626–2629.

[4] Weisenfeld, K. and Moss, F., "Stochastic Resonance and the Benefits of Noise: From Ice Ages to Crayfish and SQUIDs," *Nature*, **373** (1995) 33–36.

[5] Douglass, J. K., Wilkens, L., Pantazelou, E., and Moss, F., "Noise Enhancement of Information Transfer in Crayfish Mechanoreceptors by Stochastic Resonance," *Nature*, **365** (1993) 337–340.

[6] Miller, J. P. and Levin, J. E., "Broadband Neural Encoding in the Cricket Cercal Sensory System Enhanced by Stochastic Resonance," *Nature*, **380** (1996) 165–168.

[7] Bezrukov, Sergey M., and Vodyanoy, Igor, "Noise-induced Enhancement of Signal Transduction across Voltage-dependent Ion Channels," *Nature*, **378** (1995) 362–364.

[8] Gluckman, B. J., *et. al*, "Stochastic Resonance in Neuronal Network from Mammalian Brain," *Physical Review Letters*, **77** (19) (1996) 4098–4101.

[9] J. J. Collins, T. T. Imhoff, and P. Grigg, "Noise-enhanced Tactile Sensation," *Nature*, **383** (1996) 770.

[10] P. Cordo, *et. al*, "Noise in Human Muscle Spindles," *Nature*, **383** (1996) 769.

[11] Heneghan, C. *et. al*, "Information Measures Quantifying Aperiodic Stochastic Resonance," *Physical Review E*, **54**(3) (1996) R2228–R2231.

[12] Bulsara, A. R., Zador, A., "Threshold Detection of Wideband Signals: A Noise Induced Maximum in the Mutual Information," *Physical Review E*, **54** (3) (1996) R2185–R2188.

[13] Neiman, A. *et. al*, "Dynamical Entropies Applied to Stochastic Resonance," *Physical Review Letters*, **76** (23) (1996) 4299–4302.

[14] Srinivasan, S. K. and Sampath G., "Stochastic Models for Spike Trains of Single Neurons," in *Lecture Notes on Biomathematics*, volume 16 (Springer-Verlag, Berlin, 1972).

[15] Teich, M. C., Johnson, D. H., Kumar A. R., and Turcott, R. G., "Rate Fluctuations and Fractional Power-law Noise Recorded from Cells in the Lower Auditory Pathway of the Cat," *Hearing Research*, **46** (1990) 41–52.

[16] Teich, M. C., "Fractal Character of the Auditory Neural Spike Train," *IEEE Transactions on Biomedical Engineering*, **36** (1989) 150–160.

[17] Neelakanta, P. S., *et al*, "Dynamic Properties of Neural Learning in the Information-theoretic Plane," *Complex Systems*, **9** (1995) 349–374.

[18] Park, J. C., *et. al*, "Information-theoretics Based Error-metrics for Gradient Descent Learning in Neural Networks", *Complex Systems*, 9 (1995) 287–304.

[19] Longtin, A., Bulsara, A., and Moss, M., "Time-interval Sequences in Bistable Systems and the Noise-induced Transmission of Information by Sensory Neurons," *Physical Review Letters*, **67** (5) (1991) 656–659.

[20] Neelakanta, P. S., *et al*, "Langevin Machine: A Neural Network Based on a Stochastically Justifiable Sigmoidal Function," *Biological Cybernetics*, **65** (1991) 331–338.

[21] Wasserman, P. D., *Neural Computing* (Van Nostrand Reinhold, 1989).

[22] Müller, B. and J. Reinhardt, *Neural Networks* (Springer-Verlag, 1990).

[23] P.S. Neelakanta, *et. al*, "Csiszar's Generalized Error Measures for Gradient-descent-based Optimizations in Neural Networks Using the Backpropagation Algorithm," *Connection Science*, **8** (1) (1996) 79–114.

[24] C. vanVreeswijk and H. Sompolinsky, "Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity," *Science*, **274** (5923) (1996) 1724–1726.

[25] Finley, M., "An Experimental Study of the Formation and Development of Hebbian Cell-assemblies by Means of Neural Network Simulation," Technical Report 08333-1-T (Department of Computer Sciences, University of Michigan, Ann Arbor, 1967).