PROCEEDINGS REPRINT
SPIE—The International Society for Optical Engineering

Reprinted from

Science of Artificial Neural Networks

21-24 April 1992
Orlando, Florida

SPIE PROCEEDINGS SERIES
Volume 1710

BOLTZMANN DISTRIBUTIONS AND NEURAL NETWORKS: MODELS OF UNBALANCED INTERPRETATIONS OF REVERSIBLE PATTERNS

F. Masulli, M. Riani, E. Simonotto, and F. Vannucci

Department of Physics – University of Genoa
INFM – Unit of Genoa

Via Dodecaneso 33, 16146 Genova - Italy

ABSTRACT

The paper describes a neural network model of the perceptual alternation of ambiguous patterns with unbalanced alternative interpretations. The network is made up by binary "neurons" fully and symmetrically interconnected. An energy function can be introduced; therefore the analogy between the presented model and magnetic systems is exploited to study the statistical properties of the system. On the basis of considerations related to statistical mechanics, the probabilities of "occupation" of the two phase-space regions, associated with the two interpretations of an ambiguous figure, can be determined and analyzed.

1. INTRODUCTION

In designing artificial visual systems or automatic speech translators, many intriguing problems are posed by the ambiguity often arising from the modelling of the processes associated with the interpretation of visual or acoustical inputs. For example, if we analyze the task of an artificial visual system, we can notice that most of the processing of the signal coming from the videocamera, aims to overcome various causes of ambiguity, such as optical image distortion, inappropriate lighting, reconstruction of a tridimensional scene from a twodimensional image, and so on. All such problems are computationally heavy.

On the contrary, the perceptual systems of biological organisms (not only humans or mammals but also insects) are very efficient in accomplishing such functions, and extract easily and without ambiguity the structured and significant parts of a scene (i.e., signals and "figures") from the unstructured and insignificant ones (i.e., noise or "background"), thus achieving, a useful tridimensional interpretation of the scene.

This difference between the skills of artificial and biological systems could be ascribed to their different learning abilities (i.e., capabilities of utilizing previous experiences) and to the massive parallel information processes performed by the nervous systems of living organisms.

However there are peculiar situations in which even human perceptual ability decreases; for instance, in the case of a so called "ambiguous pattern" (see Fig. 1), the same visual input can be associated with and elicit two different interpretations, does giving rise to a cyclic perceptual alternation of the two competitive interpretations.

Figure 1: Two ambiguous figures: the Mach Pyramid and the Necker Cube

This characteristic phenomenon has been widely investigated in the context of visual perception, both experimentally[2,3,4,8,12,23] and theoretically[7,9,10,11,14,22], in order to derive from such unusual situations some hints useful in defining the perceptual processes associated with more usual unambiguous figures.

An efficient tool for studying the connections between the perceptual processes related to ambiguous or unambiguous figures is represented by the analysis of the perceptual behaviour of a subject in observing a set of patterns whose degree of ambiguity can be reduced from one to about zero by means of various components of the set.

If a pattern elicits two alternative interpretations, A and B, with the same probability of occurrence ($p_A = p_B = 1/2$) (i.e., both interpretations are perceived for equivalent periods during the observation time), the degree of ambiguity is equal to one. We can define the degree of ambiguity of a pattern as the ratio between the probability of occurrence of the less probable interpretation and that of the more probable one. Obviously, in the case of a completely unambiguous pattern, the degree of ambiguity is equal to zero, since the probability of the only one interpretation is equal to one, and the probability of alternative (non-existent) interpretations is equal to zero.

An experimental study of the phenomenon of perceptual alternation, using a set of patterns with different degrees of ambiguity, was carried out in our laboratory[16].

In the present work, we report some preliminary results obtained by a model based on an autoassociative neural network with a stochastic behaviour that simulates the interactions between the neural network and other neurons of the central nervous system. It is possible to assume that the interactions between the network and the central nervous system can be summarized by the effect of a thermal bath. As a consequence, we can apply techniques from statistical mechanics to the study of the network behaviour.

In the paper, the equilibrium properties of the system are studied by analyzing the effects of some parameters of the network (e.g., the number of neurons associated with the two different interpretations of a pattern; the values of the learning factors, etc.) on the probabilities of occurrence of the two alternative interpretations.

In the following sections, we shall describe the main aspects of the perceptual alternation phenomenon, the existing theoretical models of the phenomenon based on artificial neural networks, and, finally, the properties of our model.

2. GENERAL ASPECTS OF THE PERCEPTUAL ALTERNATION PHENOMENON

2.1 Phenomenological Aspects

Let us summarize some properties of perceptual alternation that have been deduced from experiments carried out in our laboratory, or from previous works, and that should be taken into account by a model of this phenomenon.

- Once an observer has been informed that a given stimulus can elicit two or more alternative interpretations, the actually perceived configuration is spontaneously replaced by the alternative one. This fact gives rise to a sequence of *cyclic fluctuations* of the two competitive interpretations, which, however, are never elicited simultaneously[2,5]. On the contrary, it often occurs that the alternation process does not start[12] if the observer is unaware of the two possible interpretations of the pattern used as a stimulus; in this case, the perceived configuration remains stable.

- The rate of perceptual fluctuation increases in the first 2 or 3 minutes, then it reaches a *stable value*. During this phase of the phenomenon, called the *stationary phase*, experimental measures are reproducible [3,5].

- In the stationary phase, the durations of the two alternative interpretations are stochastically distributed around their mean values according to a two-parameter Gamma distribution[3,9], whose analytic form can be expressed as follows:

$$p(t)\,dt = \frac{b^n t^{n-1} \exp(-bt)}{\Gamma(n)}\,dt, \tag{1}$$

  where Γ(n) is the Euler-Gamma function, and the values of the parameters b and n can be derived from experimental data using the method of moments:

$$b = \frac{\bar{t}}{\sigma^2}, \qquad n = \frac{\bar{t}^2}{\sigma^2}, \tag{2}$$

  where $\bar{t}$ is the average and σ^2 is the variance.

  These mean values (usually ranging between few and about ten seconds) and the related Gamma distribution parameters are individual characteristics, which also depend on the stimulus considered.

- As stated in the previous section, the degree of ambiguity of a pattern can be derived from the *a posteriori* probabilities of the alternating percepts, A and B, namely, $p_A = \bar{t}_A/(\bar{t}_A+\bar{t}_B)$ and $p_B = \bar{t}_B/(\bar{t}_A+\bar{t}_B)$. The ambiguity value is given by the p_B/p_A ratio between the two probabilities; this ratio is obviously equal to the $\bar{t}_B/\bar{t}_A$ one.

  Experimental results show an exponential dependence of the so defined ambiguity on the differences between the complexities of the two alternative interpretations[16], measured using Structural Information Theory[6].

2.2. Neural Network Models

In the past few years, different models of visual perceptual alternation have been devised, which are based on artificial neural networks[10,11,14,22]. The first model we shall examine is the one by Feldman[11]; it describes the process of *disambiguation* of the percept associated with a pattern, i.e., the process linked to the choice of the first interpretation of the pattern by the observer. Instead the other models, by Kawamoto and Anderson[14], Ditzinger and Haken[10], Riani, Masulli and Simonotto[22], consider the dynamical aspects of the phenomenon.

In Feldman's model [11] and its modifications [18,24], the units of the network generally represent the different salient features of the two alternative interpretations of an ambiguous pattern. In the particular case of the Necker cube[18], the network contains 16 nodes, each representing one of the 8 vertices of the cube in its two different perspectives. Each unit is positively connected with the other units related to the competitive interpretation. The dynamical evolution of the system will stop when a minimum of the "energy" is reached. This minimum might be associated with one of the two different perspectives of the Necker cube, as well as with an "impossible" interpretation of the cube, that is, a cube with two "front" faces. To reach the energy minima, the model follows the standard gradient descent technique, or a simulated annealing scheduling.

In the model by Kawamoto and Anderson[14], the two different interpretations of an ambiguous figure are associated with two orthogonal vectors, $\vec{f}_1$ and $\vec{f}_2$, containing the coded representations of the two interpretations respectively. This model represents a peculiar application of the more general "Brain-State-in-a-Box" (BSB) model by Anderson and coworkers[1], based on an autoassociative network. The activation values of the network elements are limited in order to be bounded within a set of states contained in a box whose corners represent the stable states of the network. In the case of an ambiguous pattern, the system is forced to "learn" the corners associated with the different interpretations by use of the connection matrix $C = \lambda_1 \vec{f}_1 \vec{f}_1^{\mathrm{T}} + \lambda_2 \vec{f}_2 \vec{f}_2^{\mathrm{T}}$, where λ_1 and λ_2 are non-zero eigenvalues.

On the presentation of the ambiguous figure, the system starts evolving to reach the stable state with the greater eigenvalue. To allow the network to shift to the stable state associated with the second interpretation, the model, once the corner has been reached, is forced to use an *antilearning* process, that decreases the associated eigenvalue until the corner is left. In this way, given the alternative prevalence of each of the two eigenvalues, the system can shift between the two alternative interpretations of the pattern, thus giving rise to a cyclic process, which has however a deterministic behaviour, without the stochastic aspects of the experimental phenomenon.

The model by Ditzinger and Hacken[10] is based on an associative memory implemented by a synergetic system. In the learning phase, an energy function is derived by using state vectors encoding prototype patterns so that, after the learning process, such vectors are associated with local energy minima. In a general case, the dynamical evolution of the state vector $\vec{q}$ of the system stops when it reaches the energy minimum corresponding to the stored pattern $\vec{v}^{(k)}$ with the maximum overlap with the initial state vector. In the case of an ambiguous pattern with the two alternative interpretations stored as prototype vectors, $\vec{v}^{(1)}$ and $\vec{v}^{(2)}$, the evolution of the system can oscillate between the two corresponding minima, if the so-called "attention parameters", λ_1 and λ_2, of the model are let to saturate, i.e., to have dynamical evolutions governed by differential equations[10]. With the addition of two fluctuating forces, $\vec{F}_1$ and $\vec{F}_2$, with zero mean values, the model shows a qualitative agreement with experimental Gamma distributions.

To reproduce the stochastic aspects of the perceptual alternation phenomenon, in particular, the statistical Gamma distributions of the durations of the two percepts, we proposed a network, called Multi-Layer Network model (MLN)[17,22], that is made by elementary blocks, or cell assemblies, working in parallel and independently of one another, and organised into two layers. The properties of each cell assembly (which is basically an autoassociative network with continue-valued units), are described in details in [20,21]; some of these properties are included in the model presented in the paper, and reported in paragraph 3.1.

In the MLN, the input stimulus is stochastically mapped into a cell assembly of the lower layer, with such a probability that it will be present, on average, on one or two units. Moreover, the input to a neuron of the upper layer is equal to the sum of the activities of the corresponding neurons of the lower layer.

With the addition of synaptic noise, the activity of the upper layer of the MLN can reach a stationary phase in which the durations of the two alternative interpretations of an ambiguous pattern are stochastically distributed around their respective mean values, and the simulated distributions are well fitted by a Gamma distribution.

Another important phenomenological aspect related to the perception of ambiguous patterns is the dependence of perceptual reversals on the complexities of the two alternative interpretations. From this point of view, the MLN model yields results that are in agreement with experimental ones, but only at qualitative level.

To overcome this limitation of the MLN, in section 3 we propose an alternative Boltzmannian model that describes this peculiar aspect in a more accurate way.

2.3. Interpretation complexities and occurrence probabilities

In this paragraph, we analyze in depth the effect of the complexities of the two interpretations of an ambiguous pattern on the probabilities of the interpretations themselves.

According to Structural Information Theory[15], the complexity of an interpretation is linked to the minimum number of rules required to generate the interpretation itself. Experiments on ambiguous patterns that elicit interpretations of identical complexity (e.g., the Necker cube[3,4]) give nearly equal mean duration times for the two interpretations; by contrast, when the two interpretations are different in complexity, the simplest interpretation becomes dominant.

Such results were obtained by studying a series of eight patterns based on the Mach pyramid[16]. The series started with the "classic" Mach pyramid (Fig. 2, left), the two interpretations (roof and room) of which differ by 2 information units, and finished with an unbalance of 20 information units in the last drawing (Fig. 2, right). All the patterns of the series can be perceived as a truncated pyramid (percept A) or a room (percept B). The figures were obtained by adding to the original pattern some elements favouring, step by step, the interpretation "room".

From the point of view of Structural Information Theory, the elements added to the central square of each figure increase to the same extent the complexities of the two interpretations, while the elements added to the side walls increase the complexity of the room interpretation less than the pyramid complexity.

The experimental durations of the two percepts were well fitted by the exponential functions:

$$\bar{t}_A = ke^{-ac_A+bc_B}, \quad \bar{t}_B = je^{-ac_B+bc_A} \tag{1}$$

which imply:

$$p_A = \frac{\bar{t}_A}{\bar{t}_A + \bar{t}_B} = \frac{1}{1 + se^{r(c_A - c_B)}} \tag{2}$$

where $s = j/k$ and $r = a + b$. The probabilities p_A and, obviously, p_B are sigmoidal functions of $c_A - c_B$. The fit of the parameters r, s of a typical subject (mean subject) gives the values $r = 0.0209$ and $s = 1.50$.

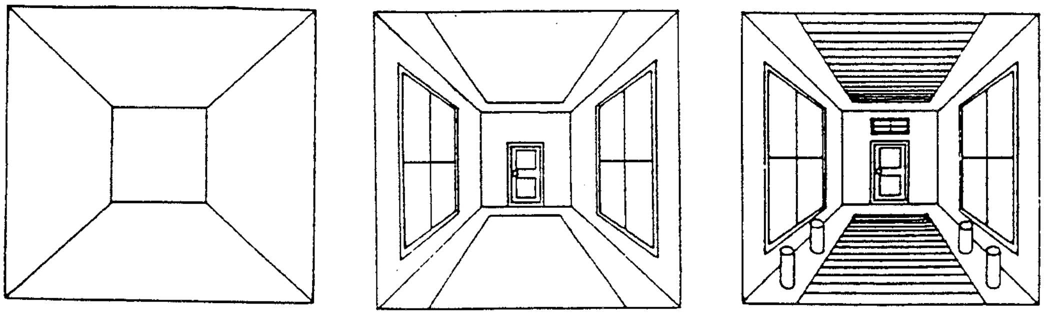

Figure 2: Three of the eight patterns used in the experiment quoted in the text

3. THE NEURAL AUTOASSOCIATOR WITH STABLE TEMPERATURE

3.1 Basic assumptions

The model we propose is based on a network of binary neurons (activation levels 0 and 1), fully interconnected with a symmetric connection matrix. An energy function can be introduced in such a model. Therefore the formal analogy between this neural network model and magnetic systems (spin glass) can be exploited to study the statistical properties of the system. After training the network, by using a Hebbian-like rule, to store two configurations representing the two interpretations of an ambiguous stimulus, we can characterize the network configurations by studying the properties of energy.

The perceptual features characterizing both alternative interpretations must be coded in the activities of the network "neurons". So the network can seen as being composed of two populations of neurons associated with the features of the interpretations. As stated in the phenomenological section, the two interpretations exclude each other; hence one population must exert an inhibiting influence on the other, and vice versa.

The connection matrix W can be obtained by learning, through some trials, the two competitive interpretations[22]. This can be represented by two learning vectors: the first has positive values of the components corresponding to one population and negative values of the components corresponding to the other population; the second has positive and negative components, inverted with respect to the first learning vector (the positive and negative values of these components can be different from the ones of the first vector).

Using a Hebbian-like rule with these two learning vectors gives a connection matrix containing two square blocks (representing the positive autoconnections of each subvector to itself) and two rectangular blocks (which are the inhibitory connections of one population to the other).

All connections in a block share the same value, except for the autoconnections of each neuron to itself, which are set to zero. Due to the symmetry of the matrix, which forces the two rectangular blocks to share the same value, the learning matrix of our model is fully described by three parameters, two for the positive autoconnections and one for the inhibitory one, as follows:

$$W = \begin{pmatrix} 0 & p & p & \cdots & -r & -r \\ p & 0 & p & \cdots & -r & -r \\ p & p & 0 & \cdots & -r & -r \\ & \cdot & & & \cdot & \\ & \cdot & & & \cdot & \\ -r & -r & -r & \cdots & 0 & q \\ -r & -r & -r & \cdots & q & 0 \end{pmatrix} \tag{3}$$

3.2 Energy

As shown by Hopfield[13], we can introduce an energy function for a neural network with a symmetric connection matrix. The analytic expression for energy is a function of network configuration:

$$E = -\sum_{i=1}^{N}\sum_{j=1}^{N} S_i W_{ij} S_j \tag{4}$$

The peculiar form of the connection matrix allows us to simplify the energy function.

We have previously noted that all excitatory connections in a population share the same positive value, while the inhibitory ones share the same negative value. Thus, by evaluating the double sum involved in the energy function, we find that all network configurations with the same number of active neurons in each population are isoenergetic, that is, they have the same energy value. In a configuration with a active neurons of the A population and b active neurons of the B population, the energy value is given by:

$$E = -a(a-1)p - b(b-1)q + 2abr \tag{5}$$

We can consider energy both as a function of the network state and as a function of the couple of integer numbers (a, b) indicating the active neurons in the two populations.

Using this set of variables, energy is defined for every couple (a, b) that satisfies: $0 \leq a \leq N_a$; $0 \leq b \leq N_b$, where N_a and N_b are the numbers of neurons in the two population.

A couple (a, b) represents a set of isoenergetic states. The degeneracy of a couple (i.e., the number of states the couple represents) is obtained by a simple combinatorial calculation:

$$\mathcal{N}(a,b) = \begin{pmatrix} N_a \\ a \end{pmatrix} \begin{pmatrix} N_b \\ b \end{pmatrix} \tag{6}$$

The only couples with degeneracy values equal to one are those at the domain vertices.

Let us assume now that the network evolves according to Hopfield's dynamic rule. This rule allows only those transitions which decrease the energy value, and where the activation value of only one neuron can be changed.

Moreover a state is an equilibrium state if and only if its four nearest neighbours in the (a, b) space have larger energy values, that is, if they satisfy:

$$\begin{aligned} E(a,b) &< E(a+1,b) \quad ; \quad & E(a,b) &< E(a-1,b) \\ E(a,b) &< E(a,b+1) \quad ; \quad & E(a,b) &< E(a,b-1) \end{aligned} \tag{7}$$

Obviously, fewer conditions must be satisfied if we consider a couple (a, b) on the boundary of the energy domain. Then a network state is a local minimum if its configuration satisfies:

$$\begin{aligned} br - ap &> 0 \quad ; \quad & p &< ap - br \\ ar - bq &> 0 \quad ; \quad & q &< bq - ar \end{aligned} \tag{8}$$

or fewer inequalities if the state is on the boundary.

In the learning phase we required that only two states should be "stored" in the network, namely, those configurations with one population of neurons fully activated and the competing population fully inhibited.

It is easy to verify that the two vertices $(N_a, 0)$ and $(0, N_b)$, in the non trivial case of $N_a, N_b > 1$, are local minima if $p, q, r > 0$. To prove this assertion, let us first consider the vertex $(N_a, 0)$ and the network states represented by the configurations $(N_a - 1, 0)$ and $(N_a, 1)$, that is the configurations that the state $(N_a, 0)$ can reach by changing the activation value of only one neuron. The vertex is a local minimum if two inequalities are satisfied:

$$\begin{aligned} E(N_a,0) &< E(N_a-1,0) \\ E(N_a,0) &< E(N_a,1) \end{aligned} \tag{9}$$

which implies:

$$\begin{aligned} 2(1-N_a)p &< 0 \\ -2N_a r &< 0 \end{aligned} \tag{10}$$

The two inequalities are verified if $p, r > 0$ and $N_a > 1$; if population A consists of only one neuron, the weight matrix does not depend on the parameter p as we do not allow self-excitation. If we analyze the stability of the second vertex, we find two inequalities which require $q > 0$ and $r > 0$.

To ensure that these will be the only two energy minima for the network, we must check only the other two vertices $(0, 0)$ and (N_a, N_b), as conditions (8) are two by two incompatible for a generic state. Whereas the vertex $(0, 0)$ has the same energy value as its contiguous states, the vertex (N_a, N_b) is a local energy minimum provided that the two following conditions are satisfied:

$$p > \frac{N_b\, r}{N_a - 1} \quad ; \quad q > \frac{N_a\, r}{N_b - 1} \tag{11}$$

As a next step in the procedure, we first check computationally that at least one of the two inequalities has not been satisfied.

Now, if we start the network in any initial configuration and follow its dynamic evolution, we observe that, after a transient period, the network reaches one of the two memories.

So we can divide the configurations into three regions: the first is made up of configurations that evolve only toward the vertex $(N_a, 0)$, the second one is made up of configurations that evolve only toward the other vertex $(0, N_b)$, and the latter is made up of configurations that choose randomly the vertex to be reached. This classification is easily obtained by computer simulations.

3.3 Perception probabilities

We can use the three classes of configurations previously defined to evaluate the probabilities of perception of the two interpretations. We can say that the network perceives interpretation A of the stimulus when the network configuration belongs to the class of configurations evolving toward the $(N_a, 0)$ vertex (at zero temperature), while interpretation B is represented by the configurations evolving only toward the $(0, N_b)$ vertex.

According to this definition, we can evaluate the occupation probabilities of the two classes of configurations representing the interpretations at a given pseudotemperature T. To this end, we exploit the formal analogy between neural networks fully interconnected with symmetric connection matrix and magnetic systems such as spin glasses.

Standard statistical mechanics suggests that we can express the probability of a state with a given energy E at the temperature T is as:

$$P = Ce^{-\beta E} \tag{12}$$

where C is the normalization factor and $\beta = 1/T$.

Now, to evaluate the probabilities of the two interpretations, we can sum over the two previously defined classes of configurations. This can be done by using the (a, b) representation and by multiplying the probability of a configuration by its degeneracy:

$$\begin{aligned} P(A) &= \frac{\sum_{(a,b)\in A} \mathcal{N}(a,b) e^{-\beta E(a,b)}}{\sum_{all\ states} e^{-\beta E(a,b)}} \\ P(B) &= \frac{\sum_{(a,b)\in B} \mathcal{N}(a,b) e^{-\beta E(a,b)}}{\sum_{all\ states} e^{-\beta E(a,b)}} \end{aligned} \tag{13}$$

It is now possible to analyze graphically the probabilities of the two interpretations as functions of the network parameters. We are mainly interested in studying the ratio $\frac{P(A)}{P(A)+P(B)}$ as a function of one network parameter that acts as an unbalancing parameter modelling the bias in the two interpretations of the pattern.

In Fig. 3, we give the probability ratio as a function of an unbalancing parameter, which, in the first part of the figure is the autoconnection of a population and in the second part, is the number of neurons in a population. As one can see, in both cases, the ratio has a sigmoidal behaviour, resulting in a good agreement with the experimental dependence on the biasing parameter. One can also note that the pseudotemperature alters the slope of the sigmoid.

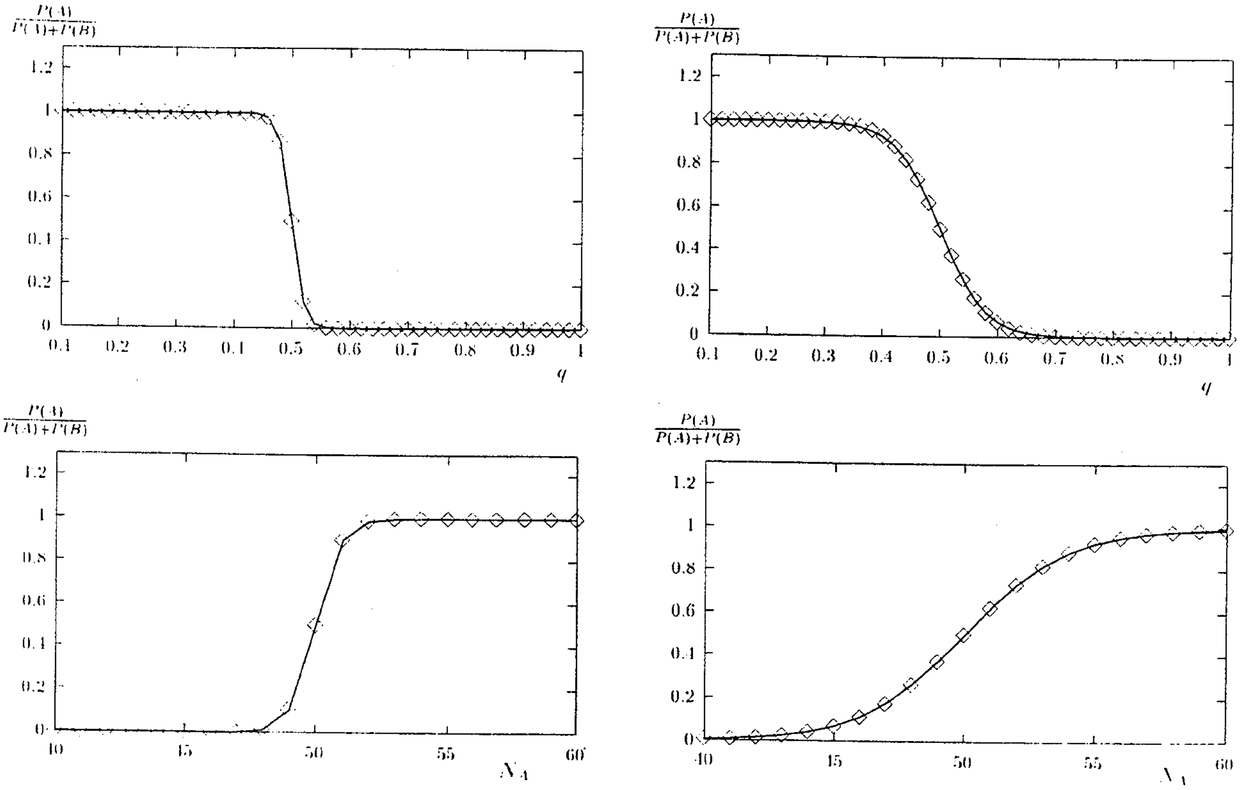

Figure 3: Four examples illustrating the behaviour of the simulation results of $\frac{P(A)}{P(A)+P(B)}$ ($\diamond$) versus q (upper plots) and versus N_a (lower plots). The left figures refer to a high β value ($\beta = .05$), while the right ones refer to a low β value ($\beta = .001$). The solid lines indicate best-fits with sigmoid functions. The values of the other model parameters are: $N_b = 50$, $p = .5$ and $q = .8$.

3.4 Dynamics

In this paragraph, we present some preliminary results on the simulation of the dynamics of the network at a given temperature T.

When we characterized the energy function, we used the Hopfield dynamic rule [13] to classify the network states. This rule can be regarded as the stochastic dynamic rule that governs the evolution of a spin glass at zero temperature. Now we shall extend this dynamic rule to non-zero temperature.

Some algorithms are known to simulate the stochastic dynamics of a system at non-zero temperature. We have chosen to use the Metropolis algorithm [19], which, at low temperatures, converges to the Hopfield dynamics.

At low temperatures (high β values) the network evolves towards one of the two minima of the energy function, at which the network evolution ends; while, at higher temperatures, the network leaves such minima and oscillates between them.

This can be used to model the stochastic multistability of perception of ambiguous stimuli. Through a simple extension of the definition of perception given in the previous section, we can say that the network perceives interpretation A when its configuration belongs to the class evolving toward the $(N_a, 0)$ vertex (class A), or to the class of configurations that can evolve towards both minima, but coming from class A. The same for interpretation B. Then, to state that the network perceives one interpretation, we must consider the network configuration and the network trajectory.

A peculiarity of the perception of ambiguous stimuli lies in the fact that the distribution of the durations of perception follows a Gamma distribution.

Preliminary results of computer simulations show that the perception durations follow a Gamma distribution (for some values of the network parameters), with good results of the χ^2 test. Such results compare favourably with experimental data.

4. ACKNOWLEDGMENTS

This work was supported by grants from MURST, INFM and "Progetto Strategico Reti Neuronali" CNR. Part of this work was carried out while F. Masulli was a Visiting Scientist at the International Computer Science Institute in Berkeley.

5. REFERENCES

1. Anderson, J.A., Silverstein, J.W., Ritz, S.A. and Jones, R.S., "Distinctive Features, Categorical Perception and Probability Learning", Psychol. Rev., 84, 413–451 (1977).
2. Attneave, F.,"Multistability in perception", Scientific American, 225,63-71 (1971).
3. Borsellino, A., De Marco, A., Allazetta, A., Rinesi, S. and Bartolini, B., "Reversal Time Distribution in the Perception of Visual Ambiguous Stimuli," Kybernetik, 10, 139-144 (1972).
4. Borsellino, A., Carlini, F., Riani, M., Tuccio, M.T., De Marco, A., Penengo, P., Trabucco, A., "Effect of Visual Angle on Perspective Reversal for Ambiguous Patterns", Perception, 11, 263-273 (1982).
5. Brown, K.T., "Complete Interocular Transfer of an Adaption Process Responsible for Perceptual Fluctuations with an Ambiguous Visual Figure", Vision Res., 2, 469-475 (1962).
6. Buffart, H. and Leeuwenberg, E., "Structural Information Theory", in *Modern Issues in Perception*, Geisler, H.G., Buffart, H., Leeuwenberg, E. and Sarris, V., Eds., Amsterdam: North-Holland (1983).
7. Caglioti, G., Caianiello, E.R., "A Model for Non-Resolvable Ambiguities", Biol. Cybern., 31, 205-208 (1978).
8. Cohen, L., "Perception of Reversible Figures After Brain Injury", Arch. Neurol. Psych., 81, 119-129 (1959).

9. De Marco, A., Penengo, P., Trabucco, A., Borsellino, A., Carlini, F., Riani, M. and Tuccio, M.T., "Stochastic Models and Fluctuations in Reversal Time of Ambiguous Figures", Perception, 6, 645-656 (1977).
10. Ditzinger, T. and Haken, H., "Oscillations in the Perception of Ambiguous Patterns", Biol. Cybern., 45, 279-287 (1989).
11. Feldman, J.A., "A Connectionist Model of Visual Memory" in *Parallel Models of Associative Memories*, Hinton, G.E. and Anderson, J.A., Eds., 49-81, Erlbaum (1981).
12. Girgus, J.J., Rock, I. and Egatz, R., "The Effect of Knowledge of Reversibility on the Reversibility of Ambiguous Figures", Percept. & Psychoph., 22, 550-556 (1977).
13. Hopfield J.J., "Neural networks as physical systems with emergent collective computational abilities", Proc. Nat. Acad. Sci., USA 79, 2554-2558 (1982).
14. Kawamoto, A.H. and Anderson, J.A., "A Neural Network Model of Multistable Perception", Acta Psychol., 59, 35-65 (1985).
15. Leeuwenberg, E. and Buffart, H., "An Outline of Coding Theory", in *Modern Issues in Perception*, Geisler, H.G., Buffart, H., Leeuwenberg, E. and Sarris, V., Eds., North-Holland (1983).
16. Masulli, F. and Riani, M., "Ambiguity and Structural Information in the Perception of Reversible Figures", Percept. & Psychoph., 45, 501-513 (1989).
17 Masulli, F., Riani, M. and Simonotto E., "A Multilayer Neural Network modelling the Perceptual Reversal of Ambiguous Patterns", in *International Joint Conference on Neural Networks - Washington*, Caudill, M., Ed., Vol. 1, 185-188, Erlbaum (1990).
18. McClelland, J.L., and Rumelhart, D.E., "Explorations in Parallel Distributed Processing", MIT Press (1988).
19. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E., "Equation of state calculations by fast computing machines", Journal of Chemical Physics, 21, 1087-1092 (1953).
20. Riani, M. and Masulli, F., "Toward a neural model of perceptual alternation", in *Advances in System Research and Cybernetics*, G. E. Lasker, Ed., 158-164, The Int. Inst. for Advanced Studies in System Research and Cybernetics, Windsor, Ontario (1988).
21., Riani, M. and Masulli, F., "Modelling perceptual alternation by using ANN's", in *Second Italian Workshop on Parallel Architectures and Neural Networks*, Caianiello, E. R., Ed., 43–50, World Scientific (1990).
22. Riani, M., Masulli, F., Simonotto, E., "Perceptual Alternation of Ambiguous Patterns: a Model based on an Artificial Neural Network", in *Applications of Artificial Neural Networks II*, Rogers, S. K., Ed., Proc. SPIE 1469, 166–177 (1991).
23. Riani, M., Oliva, G.A., Selis, G., Ciurlo, G., & Rossi, P., "Effect of Luminance on Perceptual Alternation of Ambiguous Patterns", Percept. & Mot. Skills, 58, 267-274 (1984).
24. Rumelhart, D.E., Smolensky, P., McClelland, J.L., and Hinton, G.E., " Schemata and Sequential Thought Processes in PDP Models ", in *Parallel Distributed Processing*, McClelland, J.L., Rumelhart, D.E., and the PDP Research Group, Eds., Vol 2., 7-57, MIT Press (1986).