4. Kohonen's Network Model

This chapter describes Kohonen's network model. We will discuss how the cells of a neuron layer coordinate their sensitivity to sensory signals in such a way that their response properties to signal features vary in a regular fashion with their position in the layer, an organization observed in many parts of the brain. After some neurophysiological background information, a mathematical formulation of the model will be presented. Simulations will give a first impression of the main features of the model.

4.1. Neurophysiological Background

The model employs a neuron layer A, usually assumed to be a two-dimensional sheet. This layer is innervated by d input fibers (*axons*), which carry the input signal and excite or inhibit the neurons of the layer via synaptic connections, as illustrated schematically in Fig. 4.1. In the following, we consider conditions under which the excitation of the neurons is restricted to a spatially localized region in the layer. The location of this region is then determined by those neurons that respond most intensively to the given stimulus. The neuron layer acts as a *topographic feature map*, if the location of the most strongly excited neurons is correlated in a regular and continuous fashion with a restricted number of signal features of interest. Neighboring excited locations in the layer then correspond to stimuli with similar features. Of course, a single layer can only make a few important features visible in this way. In the simplest case, we may be dealing with the stimulus position on a sensory

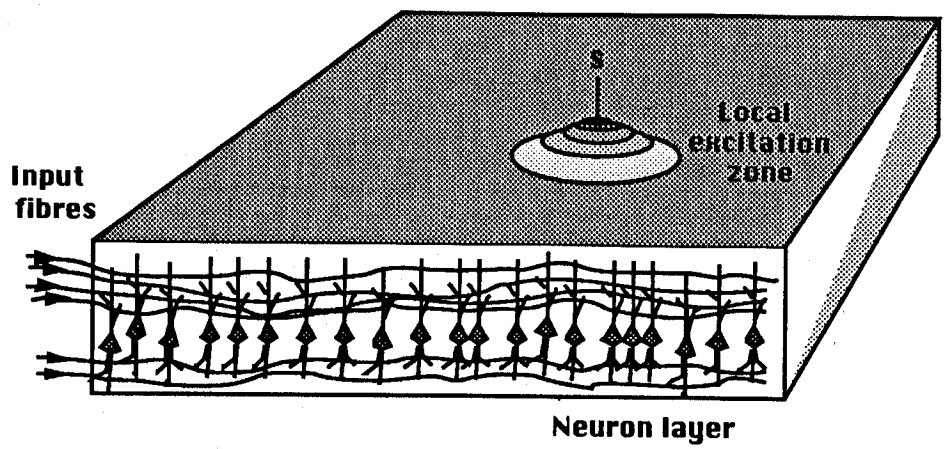

Figure 4.1 Schematic representation of the neuron layer in Kohonen's model. The nerve fibers running horizontally provide the input signal ("stimulus") and excite the layer neurons via synaptic connections. Lateral interactions between the neurons constrain the reaction to a spatially bounded "excitation zone." The layer acts as a "topographical feature map" if the position **s** of the excitation zone varies in a continuous way with the presence of stimulus features of interest.

surface, such as the retina or the body's outer surface; simple examples of more abstract features are pitch and intensity of sound signals.

We now describe the principles which enable the formation of such topographic feature maps in Kohonen's model by means of a self-organizing process. An incoming signal $\mathbf{v}$ is given by the average activities v_l of the individual incoming fibers $l = 1, 2, \ldots$. We identify the neurons of the layer by their two-dimensional position vectors $\mathbf{r} \in A$, with A being a two-dimensional grid. Every neuron $\mathbf{r}$ forms in its dendritic tree a weighted sum $\sum_l w_{\mathbf{r}l} v_l$ of the incoming activities v_l, where $w_{\mathbf{r}l}$ expresses the "strength" of the synapse between axon l and neuron $\mathbf{r}$. Here, $w_{\mathbf{r}l}$ is positive for an excitatory synapse and negative for an inhibitory synapse. The resulting excitation of an isolated neuron $\mathbf{r}$ is described by its average spike frequency $f_{\mathbf{r}}^0$. Usually, a relation

$$f_{\mathbf{r}}^0(\mathbf{v}) = \sigma\left(\sum_l w_{\mathbf{r}l} v_l - \theta\right) \tag{56}$$

is assumed for $f_{\mathbf{r}}^0$. Here, $\sigma(x)$ is a "sigmoid" function, increasing monotonically with x, with a qualitative behavior as shown in Fig. 3.10. In particular, $\sigma(x)$ tends asymptotically to the saturation values 0 or 1 for $x \to \pm\infty$. The quantity θ acts as an excitation threshold, below which the neuron responds weakly.

In addition to the coupling to the input fibers, the neurons are connected to each other via synapses. Thus, the layer has internal feedback. If one designates by $g_{\mathbf{r}\mathbf{r}'}$ the coupling strength from neuron $\mathbf{r}'$ to neuron $\mathbf{r}$, any excitation $f_{\mathbf{r}'}$ of neuron $\mathbf{r}'$ provides a contribution $g_{\mathbf{r}\mathbf{r}'}f_{\mathbf{r}'}$ to the total input signal of neuron $\mathbf{r}$. The contributions of all neurons $\mathbf{r}'$ in the layer are additively superimposed onto the external input signal $\sum_l w_{\mathbf{r}l}v_l$. In the stationary case, the neuron activities $f_{\mathbf{r}}$ are thus the solution of the nonlinear system of equations

$$f_{\mathbf{r}} = \sigma\left(\sum_l w_{\mathbf{r}l}v_l + \sum_{\mathbf{r}'} g_{\mathbf{r}\mathbf{r}'}f_{\mathbf{r}'} - \theta\right). \tag{57}$$

Frequently, the feedback accounted for by $g_{\mathbf{r}\mathbf{r}'}$ is due to excitatory synapses ($g_{\mathbf{r}\mathbf{r}'} > 0$) at small distances $\|\mathbf{r}-\mathbf{r}'\|$ and inhibitory synapses ($g_{\mathbf{r}\mathbf{r}'} < 0$) at larger distances $\|\mathbf{r}-\mathbf{r}'\|$. It can be shown that the effect of such "center-surround" organisation of synaptic interactions on the solutions of (57) consists in the formation of excitatory responses that are confined to a neighborhood around the neuron receiving maximal external excitation. In the following, we will not prove this in general, but we would like to demonstrate it using a simplified version of (57).

To this end, we consider the limiting case when the "sigmoid function" $\sigma(x)$ approximates a step function $\theta(x)$ (as defined in Section 3.1). Further, we restrict ourselves to a one-dimensional system without an external input signal (*i.e.*, $v_l = 0$) and with thresholds $\theta = 0$. We assume for $g_{rr'}$ the function

$$g_{rr'} = \begin{cases} 1 & \text{if } |r-r'| \le a, \\ -g & \text{else.} \end{cases} \tag{58}$$

Here, we assume $g > 2a+1$, *i.e.*, neurons at distances exceeding a act inhibitory, while neurons closer than a act excitatory; the strength of the inhibition is given by the value of g. Defining the quantities

$$M = \sum_r f_r, \tag{59}$$

$$m_s = \sum_{r=s-a}^{s+a} f_r, \tag{60}$$

we see that (57) becomes

$$f_r = \theta\Big([1+g]\sum_{r'=r-a}^{r+a} f_{r'} - g\sum_{r'} f_{r'}\Big), \tag{61}$$

or, by using the property $\theta(gx) = \theta(x)$ which holds for $g > 0$,

$$f_r = \theta\big([1 + g^{-1}]m_r - M\big). \tag{62}$$

Because of the θ-function, every neuron can be in only one of the two states $f_r = 0$ or $f_r = 1$. Equation (62), together with (59) and (60) represents a system of equations for the neuron activities f_r. We now show that, as a consequence of the "center-surround" organization of the lateral interactions, (62) only has solutions in which the total excitation is concentrated within a single, connected "cluster" of $a + 1$ consecutive neurons with $f_r = 1$. All of the neurons outside of this cluster are in the quiescent state ($f_r = 0$). To this end, we first prove the following lemma:

Lemma: If the quantities f_r constitute a solution of (62), and if $g > 2a + 1$, then $f_r = 1$ always implies $f_s = 0$ for all $s > r + a$ and all $s < r - a$.

Proof: From (62) it follows because of $f_r = 1$ that the inequality $m_r + g^{-1}m_r > M$ is satisfied. From the definitions (59) and (60) one also has $m_r \leq M$, and together

$$m_r \leq M < m_r + \frac{m_r}{g} \leq m_r + \frac{2a+1}{g} < m_r + 1.$$

Since M and all the m_r are integers, one has $M = m_r$ and, thus, the lemma is proven.

The lemma implies that two active neurons r, s can never be located more than a positions apart ($|r - s| \leq a$). From this, it follows that $M \leq a + 1$, *i.e.*, at most $a + 1$ neurons can be excited at the same time. If s is the leftmost of these neurons, then it follows for each of the a neurons $r \in [s, s + a]$ adjacent to s on the right

$$\begin{aligned}
[1 + g^{-1}]m_r - M &= [1 + g^{-1}] \sum_{r'=r-a}^{r+a} f_{r'} - M \\
&= [1 + g^{-1}] \sum_{r'=s-a}^{s+a} f_{r'} - M \\
&= [1 + g^{-1}]m_s - M > 0.
\end{aligned} \tag{63}$$

Here, the shift of the limits of summation in the next to last step is based on the vanishing of all the $f_{r'}$ for $r' < s$ and $r' > s + a$. For each of the $a+1$ neurons $r = s, s+1, \ldots, s+a$, (63) yields then $f_r = 1$, and since $M \leq a + 1$ all the remaining neurons satisfy

$f_r = 0$. Every solution of (62) therefore consists of a cluster of $a + 1$ adjacent excited neurons.

Similarly, in higher dimension, a sufficiently strong lateral inhibition also leads to the production of a spatially localized excitatory response. In the case of a continuous sigmoid function $\sigma(.)$, the spatial behavior of the excitation is no longer that of a step function, but rather takes a maximum at a position $\mathbf{r}'$ and from there decreases to zero in all directions. The location $\mathbf{r}'$ of the excitatory center is dependent on the input signal v_l (not taken into account in the above derivation). We pay special attention to this position $\mathbf{r}'$, since by mapping every input signal to a position $\mathbf{r}'$, the layer provides the desired map of the space of input signals. One could obtain $\mathbf{r}'$ by solving the nonlinear system of equations (57). Instead of this tedious step, Kohonen suggests an approximation for $\mathbf{r}'$, replacing it with the position of maximum excitation *on the basis of the external signal* v_l *alone, i.e.,* $\mathbf{r}'$ is determined from

$$\sum_l w_{\mathbf{r}'l} v_l = \max_{\mathbf{r}} \sum_l w_{\mathbf{r}l} v_l. \tag{64}$$

Under the two assumptions that the "total synaptic strength" per neuron $\sqrt{\sum_l w_{\mathbf{r}l}^2}$, is constant and the same for every neuron, and that all of the input signals $\mathbf{v}$ have the same "intensity" $\|\mathbf{v}\| = 1$, the condition

$$\|\mathbf{w}_{\mathbf{r}'} - \mathbf{v}\| = \min_{\mathbf{r}} \|\mathbf{w}_{\mathbf{r}} - \mathbf{v}\|, \tag{65}$$

which often is more convenient from a mathematical point of view, yields the same result for $\mathbf{r}'$. Here, $\|\mathbf{x}\|$ indicates the Euclidean vector norm $\sqrt{\sum_l x_l^2}$, and vector $\mathbf{w}_{\mathbf{r}} \equiv (w_{\mathbf{r}1}, \ldots, w_{\mathbf{r}d})^T$ is a compact notation for the synaptic strengths of neuron $\mathbf{r}$.

Thus, we now see how the map is related to the synaptic strengths $w_{\mathbf{r}l}$. An input signal $\mathbf{v}$ is mapped to the position $\mathbf{r}'$ implicitly defined by (65). For fixed synaptic strengths, (65) defines a nonlinear projection of the space of input signals onto the two-dimensional layer. In the following, we will use the notation

$$\phi_{\mathbf{w}} : \mathbf{v} \mapsto \mathbf{r}' = \phi_{\mathbf{w}}(\mathbf{v}) \tag{66}$$

to refer to this mapping. The index $\mathbf{w}$ shall remind us of the mapping's dependence on the synaptic strengths of all neurons.

This leads to the second important issue, the determination of synaptic strengths $\mathbf{w}$ providing "useful" maps. In the nervous

systems of higher animals, a detailed genetic specification of all synaptic strengths is not possible. This specification would require an exact knowledge of the way input signals are coded, a condition which even for technical applications, for example due to tolerances, is difficult to satisfy. Moreover, a system with fixed values $w_{\mathbf{r}l}$ could not respond to subsequent changes of the coding, *e.g.*, due to drift or aging processes; this obviously would contradict the high capacity for adaptation of biological systems. Apparently, such flexibility requires that the neurons be able to find suitable synaptic strengths, starting from arbitrary or only roughly correct initial settings.

In the present model, the only source of information for this process is assumed to be a sequence of input stimuli entering the layer, occurring randomly according to some statistical probability distribution. Each stimulus causes at synapse $w_{\mathbf{r}l}$ the coincidence of a presynaptic activity v_l and the resulting postsynaptic activity of neuron r. The postsynaptic activity of neuron r is just the value of the excitatory response of the layer at the position r. Its magnitude includes all interaction effects within the layer and should be computed from (57). Kohonen's model now makes the simplifying assumption that this response can be written as a function $h_{\mathbf{r}\mathbf{r}'}$ of two position variables r and r′, whose "shape" (with respect to variation of r) is fixed, but whose position (denoted by the second variable r′) depends on the stimulus. Specifically, the position r′ is taken to be the position of the excitation maximum, *i.e.*, r′ is defined by (64) or (65), and r is the location of the neurons whose response is to be described by $h_{\mathbf{r}\mathbf{r}'}$. The model then prescribes for the change of synaptic strengths $w_{\mathbf{r}l}$ the expression

$$\Delta w_{\mathbf{r}l} = \epsilon(h_{\mathbf{r}\mathbf{r}'}v_l - h_{\mathbf{r}\mathbf{r}'}w_{\mathbf{r}l}). \tag{67}$$

The first term corresponds to the "Hebbian learning rule" mentioned earlier, according to which a synapse is strengthened in the case of correlated pre- and postsynaptic activity. The second term is a decay term for the synaptic strengths, which is proportional to the postsynaptic activity. The relative scaling between the first term and the second (decay) term is normalized to unity by appropriate scaling of v. Here, ϵ determines the size of a single adaptation step ($0 < \epsilon < 1$). If ϵ is chosen to be a function $\epsilon(t)$, decreasing gradually with the number t of learning steps from large initial values to small final values, then at the beginning the system is rapidly able to learn coarsely the

correct synaptic strengths. However, for large ϵ, the fluctuation of the map caused by each learning step is also large. Hence, if the map is to stabilize asymptotically in an equilibrium state, one must let ϵ decrease to zero. On the other hand, a permanent "residual plasticity" can be realized with low fluctuations of the map by means of a small, nonvanishing final value for ϵ.

Based on (67), every synaptic change is limited to a neighborhood zone about the excitation center. In this zone, the synaptic connections are changed such that a subsequent reoccurrence of the same or a similar stimulus will lead to an increased excitation. The shape of the function $h_{\mathbf{rr}'}$ controls the size of the neighborhood zone and, thus, of the number of neurons affected by a single adaptation step.

4.2. Simplification and Mathematical Definition

The precise form of the excitatory response $h_{\mathbf{rr}'}$ appears not to be critical for the qualitative behavior of the system under the learning rule (67) and could only be obtained by numerical solution of (57). Hence, in the present model, the exact solution is only approximated qualitatively by means of a given choice of $h_{\mathbf{rr}'}$. To this end, for $h_{\mathbf{rr}'} \geq 0$ a unimodal function depending only on the distance $\mathbf{r} - \mathbf{r}'$ with its maximum at $\mathbf{r} = \mathbf{r}'$ and approaching zero for large distances is assumed. An appropriate choice is given by the Gaussian

$$h_{\mathbf{rr}'} = \exp(-(\mathbf{r} - \mathbf{r}')^2/2\sigma_E^2). \tag{68}$$

The radius σ_E of this excitatory function determines the length scale on which the input stimuli cause corrections to the map. As a rule, it is better if the coarse structure of the map is allowed to form first, before the fine structure is incorporated into the map. This is made possible by choosing σ to be a function $\sigma(t)$ starting with a rather large initial value $\sigma(0)$ and decreasing slowly with the number of learning steps toward a small final value. This can be interpreted as gradually increasing the "selectivity" of the individual neurons in the course of the learning process.

Each learning step requires the arrival of an input stimulus $\mathbf{v}$. For the model, these input stimuli are treated as independent

random variables from a vector space V, and their occurrence is determined by a probability density $P(\mathbf{v})$.

A final simplification is that the neuron positions $\mathbf{r}$ are taken to be the points of a discrete periodic lattice A.

Thus, Kohonen's model can be described by the following algorithm (Kohonen 1982a, 1984a):

0. *Initialization*: Start with appropriate initial values for the synaptic strengths $w_{\mathbf{r}l}$. In the absence of any a priori information, the $w_{\mathbf{r}l}$ can be chosen at random.
1. *Choice of Stimulus*: Choose, according to the probability density $P(\mathbf{v})$, a random vector $\mathbf{v}$ representing a "sensory signal."
2. *Response*: Determine the corresponding "excitation center" $\mathbf{r}'$ from the condition

$$\|\mathbf{v} - \mathbf{w}_{\mathbf{r}'}\| \leq \|\mathbf{v} - \mathbf{w}_{\mathbf{r}}\| \quad \text{for all } \mathbf{r} \in A. \tag{69}$$

3. *Adaptation Step*: Carry out a "learning step" by changing the synaptic strengths according to

$$\mathbf{w}_{\mathbf{r}}^{new} = \mathbf{w}_{\mathbf{r}}^{old} + \epsilon h_{\mathbf{r}\mathbf{r}'}(\mathbf{v} - \mathbf{w}_{\mathbf{r}}^{old}) \tag{70}$$

and continue with step 1.

The mapping

$$\phi_{\mathbf{w}} : V \mapsto A, \quad \mathbf{v} \in V \mapsto \phi_{\mathbf{w}}(\mathbf{v}) \in A, \tag{71}$$

where $\phi_{\mathbf{w}}(\mathbf{v})$ is defined through the condition

$$\|\mathbf{w}_{\phi_{\mathbf{w}}(\mathbf{v})} - \mathbf{v}\| = \min_{\mathbf{r} \in A} \|\mathbf{w}_{\mathbf{r}} - \mathbf{v}\| \tag{72}$$

which constitutes the *neural map* of the input signal space V onto the lattice A which is formed as a consequence of iterating steps 1.–3.

To illustrate this algorithm, the relationships are schematically shown again in Fig. 4.2. The ensemble of all possible input values forms the shaded manifold V, from which a point $\mathbf{v}$ is chosen as "stimulus" for the network in step 1. This leads to a selection (step 2) of an excitation center $\mathbf{s}$ among the neurons (lattice A). All neurons in the neighborhood of this center (highlighted) participate in the subsequent adaptation (step 3). It consists in a "shift" of the vectors $\mathbf{w}_{\mathbf{r}}$ towards $\mathbf{v}$. The magnitude

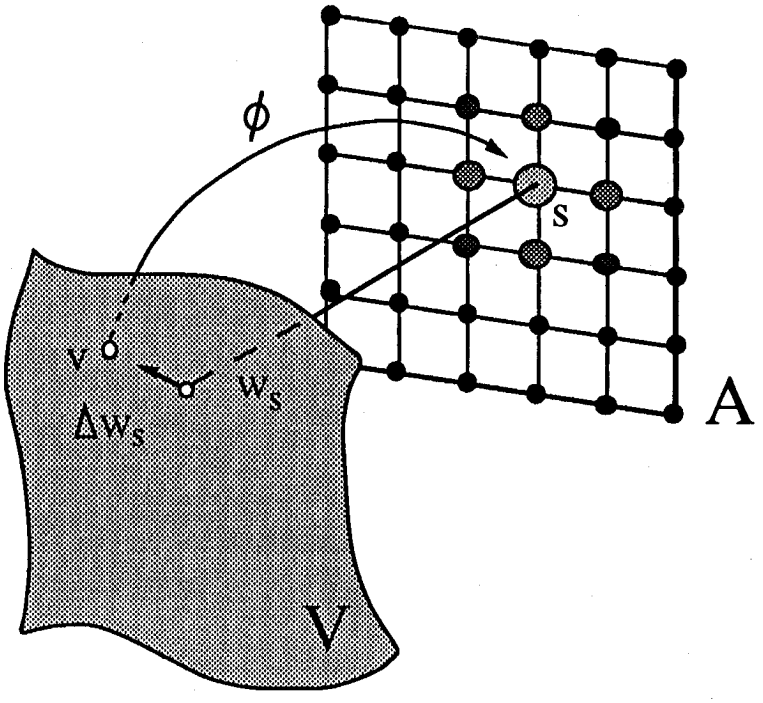

Figure 4.2 The adaptation step in Kohonen's model. The input value $\mathbf{v}$ selects a center $\mathbf{s}$ in whose neighborhood all neurons shift their weight vectors $\mathbf{w_s}$ towards the input $\mathbf{v}$. The magnitude of the shift decreases as the distance of a unit from the center $\mathbf{s}$ increases. In the figure, this magnitude is indicated by different sizes and gray values. The shift of weights is only depicted, though, for unit $\mathbf{s}$.

of this shift is fixed by the learning step size ϵ and by the function $h_{\mathbf{rs}}$.

Mathematically, the algorithm represents a so-called *Markov process*. A Markov process is defined by a set of states and a set of transition probabilities between states. These transition probabilities determine a stochastic process that, given some initial state, produces a sequence of states. This sequence is obtained by using the transition probabilities from the current state to choose a successor, which then becomes the current state for the next step (for a thorough discussion of Markov processes see for example Gardiner 1985 or van Kampen 1981).

In the present model, each possible state is given by a set of values for all the synaptic strengths $\mathbf{w} \equiv (\mathbf{w_{r_1}}, \mathbf{w_{r_2}}, \ldots, \mathbf{w_{r_N}})$ in the system (N denotes the number of neurons). The function $\phi_{\mathbf{w}}$ associates with each such state a mapping that, as we have discussed, has the interpretation of a "neural map" of some feature space. The update of a state $\mathbf{w}$ is obtained as a result of applying (70), *i.e.*, the decision for the update is caused by the

input stimulus $\mathbf{v} \in V$. Each update represents a "learning step" and can be thought of as a local "distortion" of the associated "neural map." Beginning with an initial state that corresponds to a completely disordered map, the goal of the algorithm is to arrive at a state (more precisely, the system shall enter a subset of its state space comprising states differing only by small "statistical fluctuations", see Chapter 14) that corresponds to an ordered, "topology-conserving map" of the stimulus space V, in which some relevant features of input stimuli are two-dimensionally (in the case of a neural sheet) represented. In order to reach such state and make it stationary asymptotically, the learning step length ϵ must slowly tend to zero.

The training process is qualitatively in good agreement with observed features of the formation of certain neural projections in the brain. The resulting maps predominantly represent those directions of the stimulus space V along which the input stimuli change most strongly. These directions, which often correspond to stimulus features of particular interest, may vary locally within V. Therefore, a good projection requires a nonlinear mapping. Usually, the map tries to maintain the neighborhood relationships between the input stimuli under this mapping process. Therefore, Kohonen named the resulting maps "topology-conserving feature maps." Furthermore, the map automatically takes into account the statistical weight $P(\mathbf{v})$ of the input stimuli. Regions of V from which many input stimuli occur become "magnified" and are thus projected with better resolution than regions of less frequently occurring signals. An appropriate choice for the rate of decrease of ϵ and σ with the number of learning steps is important for good results and rapid convergence. If the decrease is too rapid, the synaptic strengths "freeze" before the map has reached an equilibrium state. If the decrease is too slow, the process takes longer than necessary.

To illustrate the basic properties of this approach, we now consider a few simulation examples of the process.

4.3. Simulation Examples

In the first example, a neural network creates a map or image of an unknown region G with curved boundary. Only indirect sensory signals are available to the network. These come from

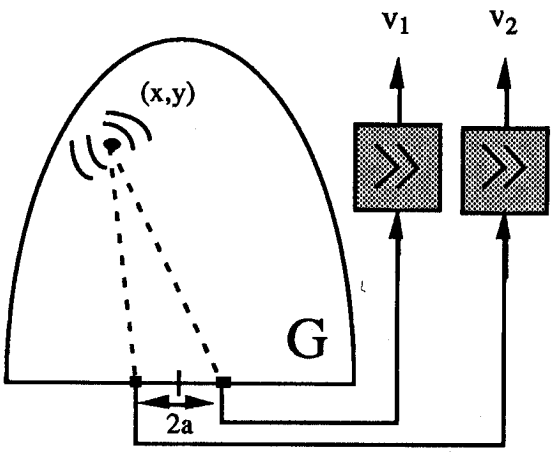

Figure 4.3 Region G containing the sound source. The two microphone positions are marked at the lower boundary of G. The microphone signals are fed into two logarithmic amplifiers, whose output signals v_1, v_2 serve as input for the network.

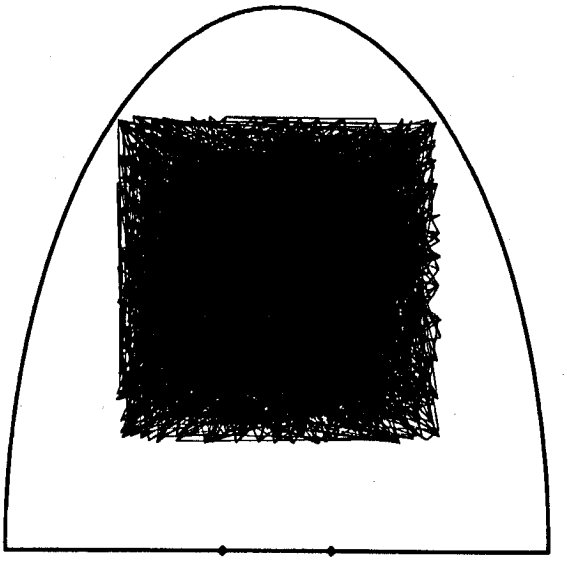

Figure 4.4 Initial relation between neurons and points in G. Initially, each neuron is assigned to a point of G chosen randomly from the filled quadrant. This assignment ignores any neighborhood relations. This is evident from the completely irregular "embedding" of the lattice in the quadrant.

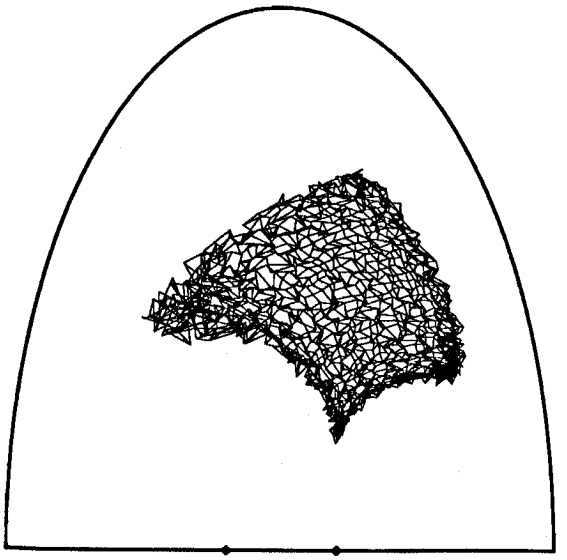

Figure 4.5 After 100 learning steps, an assignment has already formed which roughly reproduces the neighborhood relations of points of G in the lattice. However, the distribution of "responsibilities" of neurons for the region G is still very inhomogeneous.

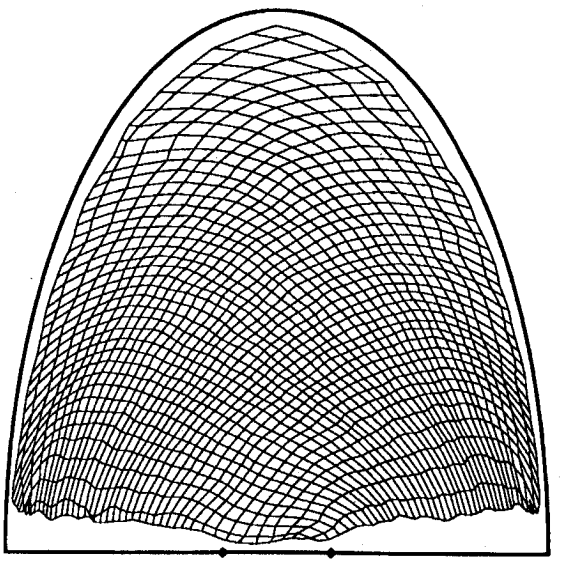

Figure 4.6 After 40,000 learning steps, a good correspondence between lattice neurons and points of G has formed. This corresponds to the choice of curvilinear coordinates, mapping the region G onto the square neuron lattice.

a source of sound moving around in G. From time to time, the sound source emits a sound signal (of constant intensity), and the position in G of each sound emission is random. The sound signal is received by two microphones, each connected to an amplifier with logarithmic characteristics (Fig. 4.3). The two amplifier output signals v_1, v_2 are the "sensory signals," and they are fed via two "axons" to the 1600 "neurons" of a model network.† The "neurons" are arranged in a planar 40×40 lattice. Every single model neuron $\mathbf{r}$ is characterized by a two-component vector $\mathbf{w_r} = (w_{\mathbf{r}1}, w_{\mathbf{r}2}) \in G$ of "synaptic strengths." Each neuron is to adjust its vector $\mathbf{w_r}$ gradually in such a way as to become sensitive for a small subset of input signals $\mathbf{v} = (v_1, v_2)^T$. This subset corresponds to a small subarea of G within which the moving source may be located. This subarea constitutes the "receptive field" of the particular neuron in the "environment" G. The neurons are to coordinate the formation of their receptive fields in such a way that — in the manner of a topographic map — the arrangement of neurons in the lattice reflects the arrangement of their respective receptive fields in the environment. This is achieved if each point of the region G corresponds to a point in the neural lattice such that the neighborhood relation between points is preserved under the correspondence, *i.e.*, the network becomes associated with a "continuous" image of G. This correspondence gives a simple example of a sensory map or sensory image of an environment, here the region in front of the two microphones. Similar "auditive maps" occur in the brain. However, this simulation example is only intended to serve as an illustration of the algorithm and makes no claim of corresponding to any brain map.

In Figs. 4.4–4.6, the evolution of the assignment of neurons to positions is shown in detail. For each neuron $\mathbf{r} \in A$, the location (x, y) of its receptive field in G has been marked, as assigned by the map. Marked locations are connected by a line if their corresponding neurons are adjacent on the lattice. (Thus, in place of the image itself, the embedding of the lattice A in G is shown,

† In the computer simulation, sound source, microphone, and amplifier are represented as follows: if the sound source is at the position (x, y), the output signals v_1 and v_2 of the two amplifiers are given by

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} -\log[(x-a)^2 + y^2] \\ -\log[(x+a)^2 + y^2] \end{pmatrix},$$

where $2a$ is the separation of the microphones.

from which the map can be obtained as its inverse.) Initially, the assignment is completely random, and there is no agreement between the arrangement of neurons and the corresponding locations (Fig. 4.4). After only a few signals, the coarse structure of the assignment has been found (Fig. 4.5), until finally after 40,000 sound signals a good assignment is achieved (Fig. 4.6). In this case, the algorithm has automatically found a nonlinear coordinate transformation mapping the region G with curved boundary onto a square lattice A. The resulting coordinate transformation takes the frequency distribution of the arriving signals into account, as illustrated in the simulation result shown in Fig. 4.7. Instead of a homogeneous distribution of source locations, the signals from the indicated circular region in G were now emitted with a three times higher probability than in the remaining part of G. Within both regions the probability density was constant. In all other respects the simulation was identical to that presented in Fig. 4.4–4.6. As a consequence of the inhomogeneous stimulus distribution, substantially more neurons are assigned to positions in the circular region. This corresponds to a higher resolution of the map for this part of G, which is a desirable result, since a concentration of assignments within regions where signals frequently occur leads to a more efficient use of the network.

However, the frequency with which a signal occurs is not always an indication of its importance. Varying importance of signals can also be taken into account by regulating the plasticity of the network. For example, one can adjust the size of a learning step according to an a priori importance attributed to the signals. This increases the "attentiveness" of the network for signals deemed more important and has the same effect as correspondingly more frequent occurrence. This is illustrated in Fig. 4.8, which shows the result of a simulation with sound emission probability again uniform throughout all of G. However, in contrast to Fig. 4.4–4.6, the network reacted to every sound event from within the circle with an adaptation step that was three times larger than for a sound event from the remaining part of G. The result thus obtained is practically identical to that of Fig. 4.7.

In the example presented, the space of stimuli G is mapped onto a lattice A of the same dimensionality. If the space of stimuli possesses a higher dimensionality, the map tries to project the higher-dimensional space as faithfully as possible by means

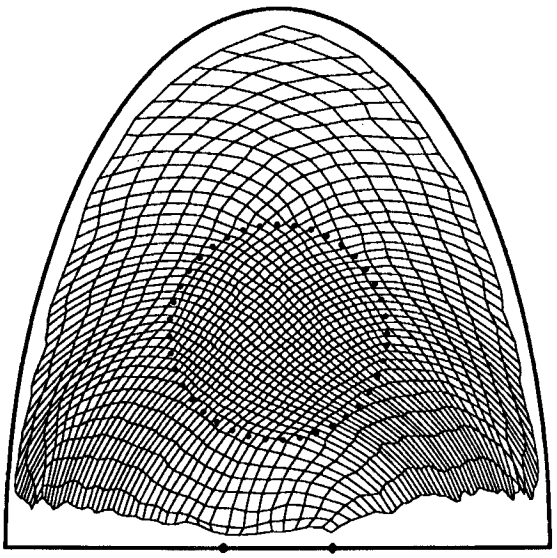

Figure 4.7 Result of the same simulation as in Fig. 4.6, except that within the circular region marked by dots signals were emitted with a three times higher probability than in the remaining region of G. In this case, more neurons code positions in the circular region. This corresponds to a higher resolution of the map created for this region.

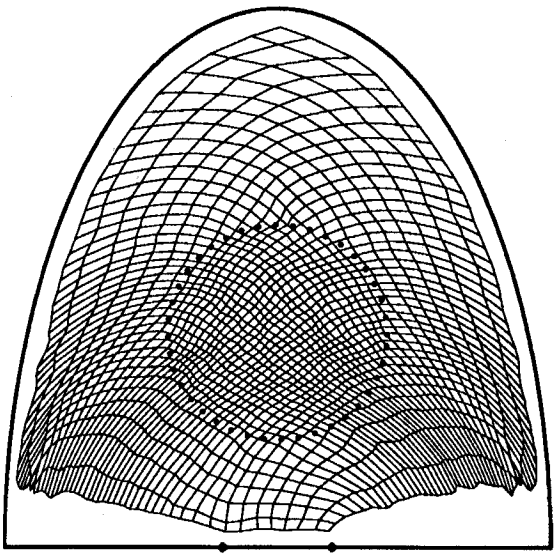

Figure 4.8 The same effect as in Fig. 4.7 can be achieved by a signal-dependent adjustment of the plasticity of the neurons. In this simulation, the sound signals were again emitted as in Fig. 4.4–4.6 with a homogeneous probability everywhere in G, but the learning step size ϵ was increased by a factor of three if the sound source was located in the circular region.

of an appropriate "convolution." To illustrate this behavior, we consider a one-dimensional neural "net," *i.e.*, a neuron chain. For the input signal, we take a random sequence of two-dimensional vectors v, whose values are homogeneously distributed in the unit square. For $h_{\mathbf{rr}'}$, we choose the Gaussian (68) with $\sigma(t) = 100 \cdot (0.01)^{10^{-5}t}$. The correspondence between neurons and points of the square is again represented as an embedding of the neuron chain into the square, as in the previous example. This assignment is initially made at random as shown in Fig. 4.9a. After 200 iterations, the curve has attained a U-shaped configuration (Fig. 4.9b). At this time, the range σ of the function $h_{\mathbf{rr}'}$ is still large and, hence, structure has formed only at this length scale. As σ decreases further, structures gradually form at shorter length scales as well (Fig. 4.9c, 50,000 iterations). Eventually, after 100,000 iteration steps, the hierarchically convoluted graph of Fig. 4.9d has emerged. The network thus tries to fill the two-dimensional region while reproducing the neighborhood relations as well as possible. The degree of success is evident from the similarity of the curve created in this way to the finite approximation of a so-called "Peano curve." This is an

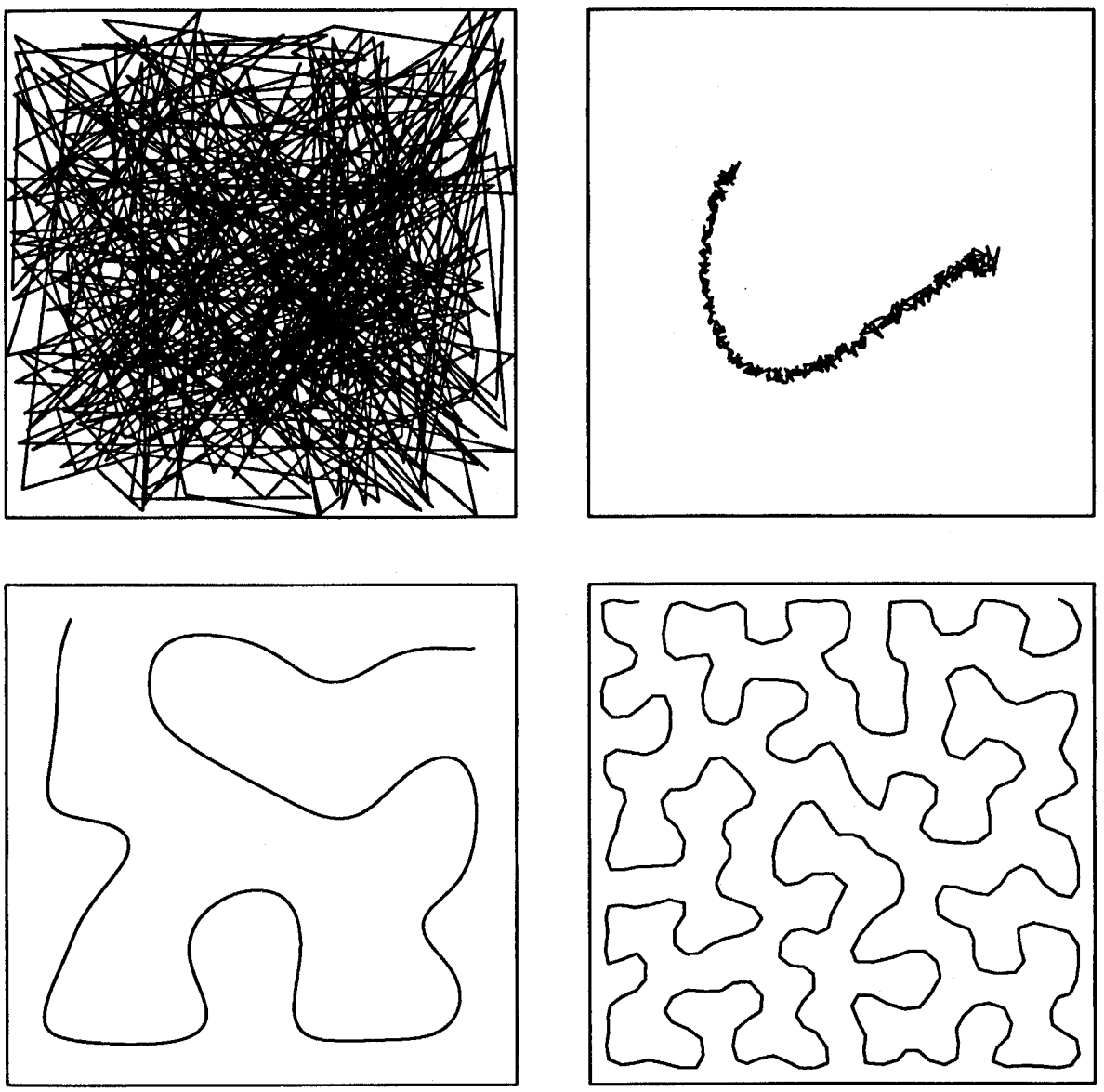

Figure 4.9 Mapping between a "neural" chain and a squared stimulus space. From top left to bottom right: a) randomly chosen initial assignment; b) coarse assignment after 200 Markov steps; c) after 50,000 Markov steps; d) assignment obtained after 100,000 Markov steps resembling a "Peano curve."

infinitely, recursively convoluted fractal curve representing the solution of the problem of mapping a one-dimensional interval continuously onto a two-dimensional surface.

However, as a rule one is interested in mapping of higher-dimensional regions onto a two-dimensional image. Indeed, Kohonen used the procedure successfully to map spectra of different speech sounds (phonemes) to separate map positions. Here, the tonal similarity relations between the individual phonemes are translated into locational relations in the image. This constitutes a very important preprocessing step for the problem of

artificial speech recognition. The subsequent steps require the analysis of transitions between individual phonemes, *i.e.*, of time sequences. The possibility of employing the procedure also for such purposes shall be indicated in the following concluding example. At the same time, this example will clarify how in the course of the formation of a map hierarchic relations can also be represented.

The source of the signal is a Markov process (here used as a simple model of a temporal signal and to be distinguished from the learning algorithm itself) with 10 states. The aim is to create a map of the possible transitions between states of the process. Transitions to the same successor state are to be adjacent in the map. A state i, $i = 0,\ldots,9$, is assumed to have one of the five states $i-3$, $i-2$, $i-1$, $i+1$ or $i+2$ (modulo 10) as a possible successor. A transition from state i to state j is coded by a 20-component vector v with components $v_k = \delta_{k,i} + \delta_{k,j+10}$. A transition occurs at each time step, and all transition probabilities have the same value 0.2. A lattice consisting of 20×20 neurons is used, and the Gaussian (68) is chosen for $h_{\mathbf{rr}'}$. The remaining parameter values of the simulation are $\sigma(t) = 5 \cdot 0.2^{t/t_{max}}$, $\epsilon(t) = 0.9 \cdot (0.05/0.9)^{t/t_{max}}$ and $t_{max} = 5,000$ learning steps. Additionally, for the computation of the distances $\|\mathbf{v} - \mathbf{w}(\mathbf{r})\|$, a "metric" was used which weights the differences in the last 10 components of v twice as strongly as those of the first 10 components. In Fig. 4.10, the 20×20-lattice of neurons is represented. For each lattice site, two numbers $i, j \in \{0,\ldots,9\}$ indicate the initial and final state of the transition assigned to the respective neuron. The initial distribution was again chosen randomly. Figure 4.10 shows the map obtained after 5,000 learning steps. For each of the 50 allowed transitions, an "island" of neurons responding to this transition has formed, and the islands are in turn arranged in such a way that islands corresponding to transitions to the same successor state form a larger cluster. This corresponds to a hierarchical arrangement and is a consequence of the described choice of weight, the successor state obtaining a higher weight than the predecessor state in the choice of the excitation center. This choice dominates the formation of the "large-scale" structure of the map, *i.e.*, the structure on the level of "clusters of islands." This illustrates that, by an appropriate choice of metric (the choice of weight corresponds to a choice of metric), it is possible to arrange for certain features (here successors) to be grouped together hierar-

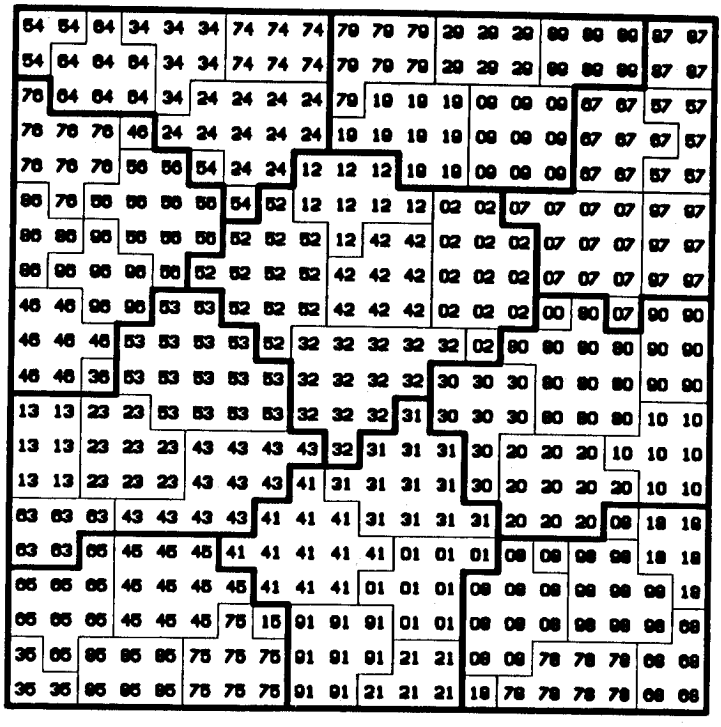

Figure 4.10 Mapping of the transitions $i \rightarrow j$ of a Markov process with states $i, j = 0, ..., 9$ onto a lattice consisting of 20 × 20 neurons. For each lattice location, the transition to which the corresponding neuron best responds is indicated as jk. Neurons with the same transition are adjacent to one another within islands. Islands with the same successor in turn form "clusters." This corresponds to a hierarchical distribution of the neuron specificities over the lattice.

chically in the map. By the inclusion of *contextual information*, such a hierarchical grouping can emerge from the data itself. For example one can create "semantic maps" which arrange words in hierarchies of meaning. This ordering is gradually found by the system itself in the course of a learning phase, where simple English sentences can serve as "training data" (Ritter and Kohonen 1989).

After this initial overview, we consider in the following chapters a series of information processing tasks, for which the choice is motivated by their significance for biological systems. At the same time, we investigate how self-organizing maps can be useful in solving such problems. While viewing biological examples as a guide, we will occasionally consider technical applications when appropriate. This applies particularly to Chapter 6, which gives a solution to the "traveling salesman problem" and Chapters 10–13, which are concerned with applications to robotics.

5. Kohonen's Network for Modeling the Auditory Cortex of a Bat

In this chapter we employ Kohonen's model to simulate the projection of the space of the ultrasound frequencies onto the auditory cortex of a bat (Martinetz, Ritter, and Schulten 1988). The auditory cortex is the area of the cerebrum responsible for sound analysis (Kandel and Schwartz, 1985). We will compare the results of the simulation with available measurements from the cortex of the bat *Pteronotus parnelli rubiginosus*, as well as with an analytic calculation.

For each animal species, the size of an area of neural units responsible for the analysis of a particular sense strongly depends on the importance of that sense for the species. Within each of those areas the extent of the cortical representation of each input stimulus depends on the required resolution. For example, the fine analysis of the visual information of higher mammals is accomplished in the *fovea*. The fovea is a very small area of the retina in the vicinity of the optical axis with a very high density of *rods* and *cones*, the light sensitive receptors in the eye. The especially high density gives rise to a significantly higher resolution in this area than in the regions of the retina responsible for the peripheral part of the visual field. Although the fovea is only a small part of the total retina, the larger part of the visual cortex is dedicated to the processing of signals from the fovea. Similarly nonproportional representations have also been found in the somatosensory system and in the motor cortex. For example, particularly large areas in the somatosensory and the motor cortex are assigned to the hand when compared to the area devoted to the representation of other body surfaces

or limbs (Woolsey 1958).

In contrast no nonproportional projections have been found so far in the auditory cortex of higher mammals. The reason for this is perhaps that the acoustic signals perceived by most mammals contain a wide spectrum of frequencies; the signal energy is usually not concentrated in a narrow range of frequencies. The meow of a cat, for example, is made up of many harmonics of the base tone, and no region of the frequency spectrum plays any particular function in the cat's survival. The auditory cortex of cats was thoroughly examined, and the result was that frequencies, as expected, are mapped onto the cortex in a linearly increasing arrangement without any regard for particular frequencies. The high-frequency units lie in the *anterior* and the low-frequency units lie in the *posterior* region of the cortex. According to available experimental evidence, the auditory cortex of dogs and monkeys is structured very similarly (Merzenich et al. 1975).

5.1. The Auditory Cortex of a Bat

In bats, nonproportional projections have been detected in the auditory cortex. Due to the use of sonar by these animals, the acoustic frequency spectrum contains certain intervals which are more important. Bats utilize a whole range of frequencies for orientation purposes. They can measure the distances to objects in their surroundings by the time delay of the echo of their sonar signals, and they obtain information about the size of the detected objects by the amplitude of the echo.

In addition, bats are able to determine their flight velocity relative to other objects by the Doppler shift of the sonar signal that they transmit. This ability to determine the Doppler shift has been intensively studied in *Pteronotus parnelli rubiginosus*, a bat species which is native to Panama (Suga and Jen 1976). This species has developed this ability to the extent that it is able to resolve relative velocities up to 3 cm/s, enabling it to detect even the beating of the wings of insects, its major source of nutrition. The transmitted sonar signal consists of a pulse that lasts about 30 ms at a frequency of 61 kHz. For the analysis of the Doppler-shifted echoes, this bat employs a special part of its auditory cortex (Suga and Jen 1976).

The Doppler shift Δf of the sonar frequency by an object moving in the same line with the bat is determined by

$$\frac{\Delta f}{f_e} = \frac{2v_{bat}}{c} - \frac{2v_{obj}}{c}. \tag{73}$$

Here f_e is the bat's sonar frequency, *i.e.*, 61 kHz, v_{bat} is the bat's velocity, v_{obj} is the object's flight velocity, and c is the velocity of sound. The factor of two is due to the fact that both the transmitted signal and the echo are Doppler shifted. If the bat knows its own velocity, it can determine v_{obj} from the Doppler shift Δf.

Excellent sonar capabilities are certainly indispensable for the bat's survival. To be able to detect a frequency shift of 0.02% which corresponds to the stated relative velocity of 3 cm/s, assuming a sound velocity of 300 m/s, a particularly high resolution of frequencies around the sonar frequency is necessary. Therefore, it would not be surprising if the interval around 61 kHz of the frequency spectrum were disproportionately represented in the part of the auditory cortex responsible for the Doppler analysis. Investigations on *Pteronotus parnelli rubiginosus* indeed support this expectation (Suga and Jen 1976).

Figure 5.1 shows the results of observations by Suga and Jen (1976). In part B of Fig. 5.1 one can clearly see that the one-dimensional frequency spectrum essentially extends continuously and monotonically from the posterior to the anterior region of the auditory cortex. In addition, one recognizes a region around the sonar frequency of 61 kHz with a very high resolution. To emphasize this anomaly, the region shaded in part A of Fig 5.1 has been displayed separately in part C. This region corresponds to the frequency interval which is especially important for the bat and extends monotonically from a minimum frequency of about 20 kHz up to a maximum frequency of about 100 kHz. The position and *best frequency* for each measurement in the shaded region of A is also shown in part C of Fig. 5.1. As "best frequency" for a neuron, one picks the frequency that causes the highest excitation of that neuron. One clearly sees that the majority of the measured values are clustered around the sonar frequency, as is expected. Almost half of the anterior-posterior region is used for the analysis of the Doppler-shifted signals. This provides the particularly high resolution of 0.02% which gives the bat its fine navigational and insect hunting abilities.

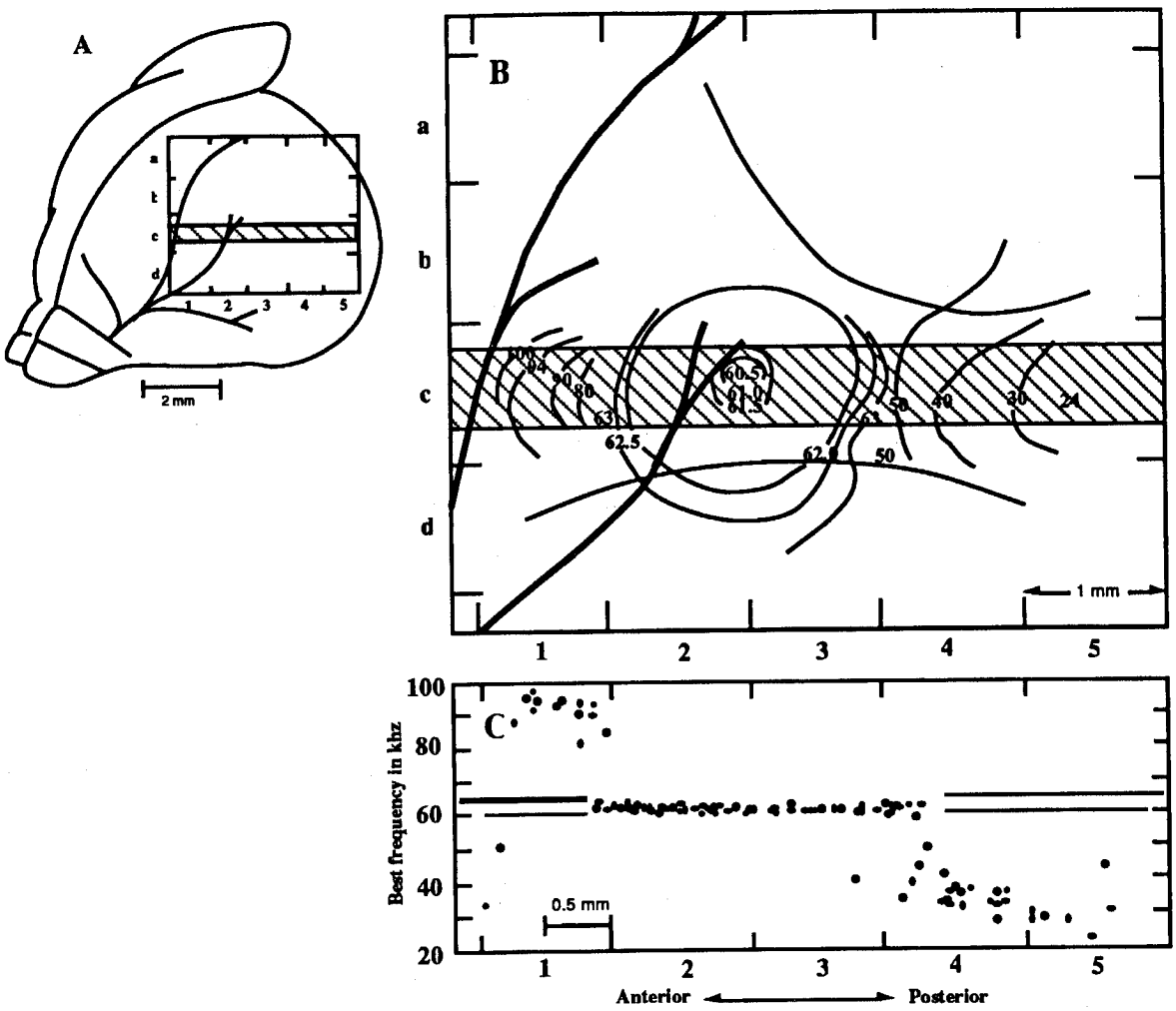

Figure 5.1 (A) Dorsolateral view of the bat's cerebrum. The auditory cortex lies within the inserted rectangle. (B) Distribution of "best frequencies" on the auditory cortex, the rectangle in (A). (C) Distribution of "best frequencies" along the region shaded in (A) and (B). The distribution of measured values around 61 kHz has been enlarged (after Suga and Jen 1976).

5.2. A Model of the Bat's Auditory Cortex

The development of the projection of the one-dimensional frequency space onto the auditory cortex, with special weighting of the frequencies around 61 kHz, will now be simulated by Kohonen's model of self-organizing maps. For this purpose we will model the auditory cortex by an array of 5×25 neural units.

The space of input stimuli is the one-dimensional ultrasound spectrum of the bat's hearing. In our model this spectrum will be simulated by a Gaussian distribution of Doppler-shifted

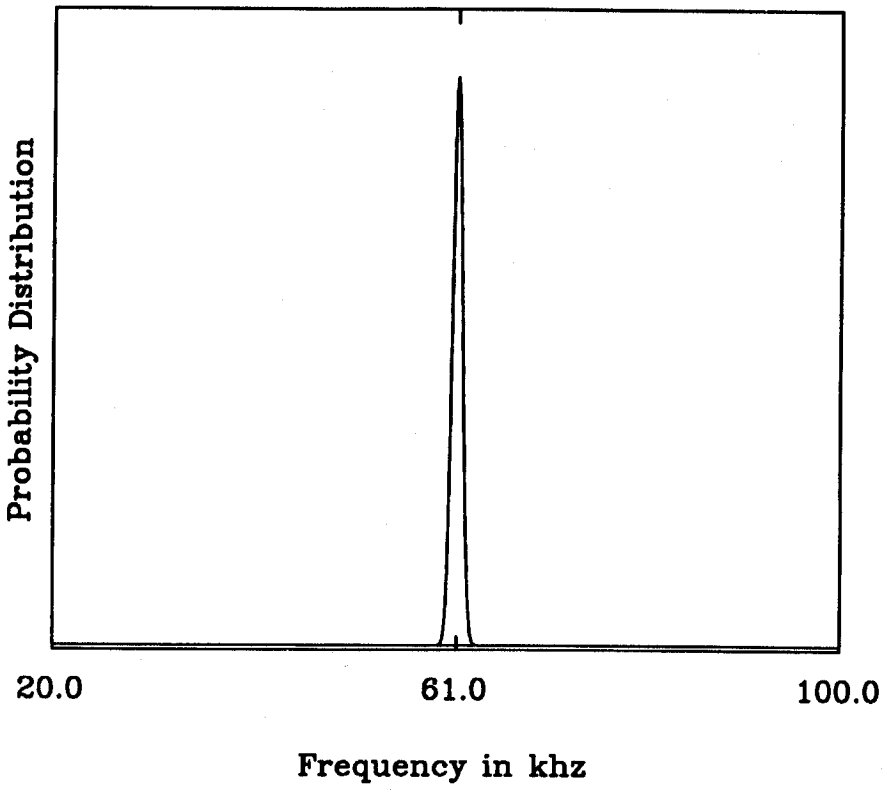

Figure 5.2 The relative probability density of the input signals versus frequency. Doppler-shifted echoes occur exactly three times as often as signals from the white background noise.

sonar echoes on top of a white background noise. The background noise in the range from 20 to 100 kHz depicts signals from external ultrasound sources. In addition, there is a peak near 61 kHz which consists of the echoes from objects moving relative to the bat. We describe this peak of Doppler-shifted sonar signals by a Gaussian distribution centered at 61 kHz with a width of σ_r=0.5 kHz. This corresponds to a root mean square speed difference of the sonar-detected objects of about 2 m/s. Doppler-shifted sonar signals occur in our model three times as often as signals from the white background noise. Figure 5.2 shows the weighted probability distribution.

Initially, a random frequency is assigned to each model neuron of our model cortex. This corresponds to Step 0 of Kohonen's model as described in the last chapter. Due to the one-dimensionality of the space of input stimuli, the synaptic strengths w_r of the model neurons r have only a single component.† An input signal according to a probability distribution

† This is only an idealization that is caused by the explicit use of frequency values. In a more

$P(\mathbf{v})$ causes that model neuron whose momentarily assigned frequency (the so-called "best frequency" of that neuron) lies closest to the input frequency to determine the center of the "activity peak" within which the neurons become significantly excited (Step 2). Next, the "best frequencies" of all neurons of the cortex are modified according to Step 3 of Kohonen's algorithm. After a sufficient number of steps this modification should result in an arrangement of "best frequencies" on the model cortex that is continuous and is adapted to the particular probability distribution of the input signals.

5.3. Simulation Results

In Fig. 5.1B it can be seen that the region of the auditory cortex of *Pteronotus parnelli rubiginosus* responsible for the resolution of the echo is greatly elongated, it being much more extended along the anterior-posterior axis than it is along the perpendicular direction. A similar length-width ratio for the model cortex was chosen in the simulation we will describe. There, the anterior-posterior length contains 25 model neurons and is five times longer than the width of the array.

Figure 5.3 shows the model cortex at different stages of the learning process. Each model neuron is represented by a box containing (the integer part of) the assigned frequency. Figure 5.3a presents the initial state. Each neuron was assigned randomly a frequency value in the range 20 to 100 kHz. As we see in Fig. 5.3.b, after 500 learning steps a continuous mapping between the space of input frequencies and the model cortex has already emerged. The final state, achieved after 5000 learning steps, is depicted in Fig. 5.3.c. One can see the special feature of Kohonen's model that represents the input stimuli on the net of neural units according to the probability with which stimuli occur. The strong maximum of the probability density in our model causes a wide-ranging occupation of the "cortex" with frequencies in the narrow interval around the sonar frequency of 61 kHz.

In this simulation the time dependence of the excitation zone

realistic model one could, for example, code the frequency by different output amplitudes of a set of overlapping filters as they are actually realized in the inner ear. The ordering process demonstrated in the simulation would, however, not be affected by this.

78	45	63	43	75
51	77	28	86	91
52	80	24	34	45
52	85	91	85	84
23	97	26	27	37
96	28	98	86	82
44	99	50	81	42
52	86	74	92	75
66	39	56	41	34
99	23	77	69	61
42	47	81	38	100
93	86	56	45	23
36	70	46	59	22
62	59	80	97	25
64	92	94	23	84
83	59	31	89	47
76	71	41	41	89
72	27	85	27	67
51	23	93	100	27
89	52	32	42	65
48	44	30	65	45
21	34	83	77	31
25	38	82	38	79
47	75	43	86	64
20	73	43	51	64

28	28	30	32	35
28	29	31	33	36
30	31	32	35	37
33	33	35	37	39
36	37	38	39	41
40	41	41	43	44
44	44	45	46	47
48	48	49	49	50
51	52	52	52	53
54	54	55	55	55
57	57	57	57	57
58	58	58	58	58
59	59	59	59	59
60	60	60	60	60
60	60	60	60	60
60	60	60	60	60
60	60	60	60	60
60	60	60	60	60
60	60	60	60	60
61	61	61	61	61
62	61	61	61	61
62	62	62	62	62
63	63	63	63	62
64	64	63	63	63
64	64	64	64	63
65	64	64	64	64

36	33	29	26	25
38	35	32	30	27
40	39	37	35	33
44	42	41	39	39
47	46	44	44	44
50	49	48	48	49
54	53	53	54	54
58	58	58	58	58
59	59	59	59	59
60	60	60	60	60
60	60	60	60	60
61	61	61	61	61
61	61	61	61	61
61	61	61	61	61
61	61	61	61	61
62	62	62	62	62
62	62	62	62	63
62	62	63	63	63
64	64	64	64	65
68	68	68	69	68
72	73	73	74	73
75	77	79	81	81
78	80	84	87	88
81	83	87	91	93
82	85	90	94	96

Figure 5.3 (a)*(left)* The initial state with random frequencies assigned to the neural units. The length-to-width ratio of the array of model neurons is five (anterior-posterior) to one (dorsolateral). Each box represents a neuron and contains the integer part of the current "best frequency" assigned to that neuron. (b)*(middle)* The state of the "auditory cortex" after 500 learning steps. The field has evolved into a state where neighboring neurons have similar "best frequencies;" *i.e.*, the space of input stimuli is represented continuously on the array. (c)*(right)* The "auditory cortex" in the final state, after 5000 learning steps. The region of "best frequencies" around the sonar frequency, which represents the Doppler-shifted input signals, occupies almost half of the model cortex.

σ and of the adaptation step widths ϵ were chosen as follows: $\sigma(t) = \sigma_i[1 + \exp(-5\,(t/t_{max})^2)]$ and $\epsilon(t) = \epsilon_i \exp(-5\,(t/t_{max})^2)$ with $\sigma_i = 5$ and $\epsilon_i = 1$, where t denotes the number of performed learning steps. The final number of learning steps at the end of the simulation was $t_{max} = 5000$.

In accordance with the experimental results from the auditory cortex of *Pteronotus parnelli rubiginosus*, the representation of the input frequencies on our model cortex increases monotonically along the "anterior-posterior" axis. In order to compare the results of our simulation with the measurements, we have presented the distribution of "best frequencies" as in

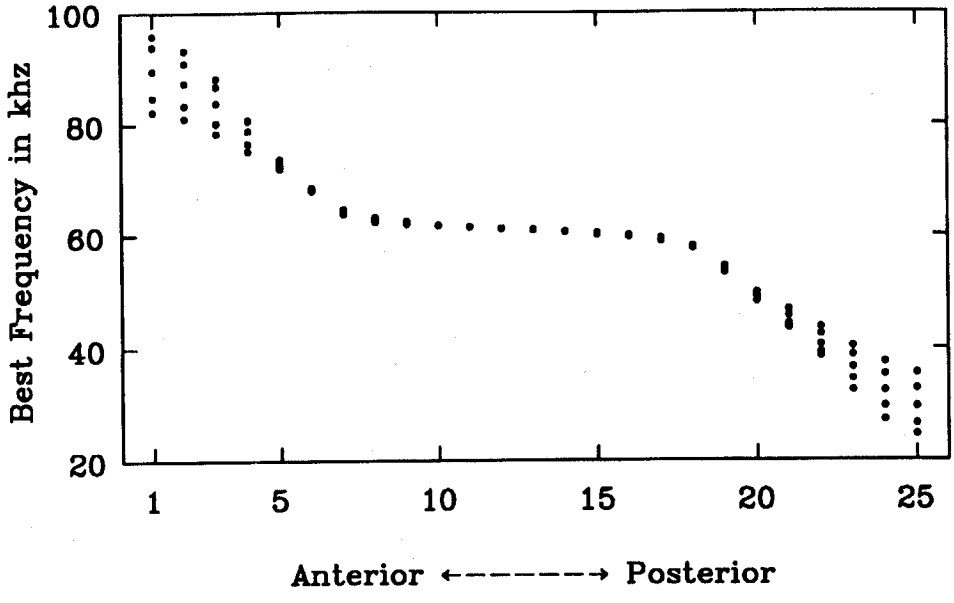

Figure 5.4 The simulation results presented as in Fig. 5.1C. Along the abscissa are the positions 1 through 25 of the model neurons along the "anterior-posterior" axis. The ordinate shows the corresponding "best frequencies." For every value between 1 and 25 five frequency values are represented, one for each of the five neural units along the "dorso-lateral" direction.

Fig. 5.1C. Figure 5.4 depicts the simulation results of Fig. 5.3 in the same way as Fig. 5.1C represents the data of Fig. 5.1A-B. Each model neuron has been described by its position 1 to 25 on the "anterior-posterior" axis as well as by its "best frequency." This representation of the results of the simulation produces a picture very similar to that of the experimental measurements (Fig. 5.1). In both cases a plateau arises that occupies almost half of the cortex and contains the neural units specialized in the analysis of the Doppler-shifted echoes. The size of this plateau is determined by the shape of the probability distribution of the input stimuli. In Section 5.4 we will look more closely at the relation between the shape of the probability distribution and the final cortical representation in Kohonen's model.

5.4. Mathematical Description of the "Cortical Representation"

We want to investigate what mappings between a neural lattice and an input signal space result asymptotically for Kohonen's model. For "maximally ordered" states we will demonstrate a quantitative relation between the "neural-occupation density" in the space of input stimuli which corresponds to the local enlargement factor of the map, and the functional form of the probability density $P(\mathbf{v})$ of the input signals (Ritter and Schulten 1986a). The result will enable us to derive an analytical expression for the shape of the curve shown in Fig. 5.4, including the size of the plateau. Unfortunately, such analytical expressions will be limited to the special case of one-dimensional networks and one-dimensional input spaces. The following derivation is mainly directed at the mathematically inclined reader; it can be skipped without loss of continuity.

To begin, we consider a lattice A of N formal neurons $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$. A map $\phi_{\mathbf{w}} : V \mapsto A$ of the space V onto A, which assigns to each element $\mathbf{v} \in V$ an element $\phi_{\mathbf{w}}(\mathbf{v}) \in A$, is defined by the synaptic strengths $\mathbf{w} = (\mathbf{w}_{\mathbf{r}_1}, \mathbf{w}_{\mathbf{r}_2}, \ldots, \mathbf{w}_{\mathbf{r}_N})$, $\mathbf{w}_{\mathbf{r}_j} \in V$. The image $\phi_{\mathbf{w}}(\mathbf{v}) \in A$ that belongs to $\mathbf{v} \in V$ is specified by the condition

$$\|\mathbf{w}_{\phi_{\mathbf{w}}(\mathbf{v})} - \mathbf{v}\| = \min_{\mathbf{r} \in A} \|\mathbf{w}_{\mathbf{r}} - \mathbf{v}\|, \tag{74}$$

i.e., an element $\mathbf{v} \in V$ is mapped onto that neuron $\mathbf{r} \in A$ for which $\|\mathbf{w}_{\mathbf{r}} - \mathbf{v}\|$ becomes minimal.

As described in Chapter 4, $\phi_{\mathbf{w}}$ emerges in a learning process that consists of iterated changes of the synaptic strengths $\mathbf{w} = (\mathbf{w}_{\mathbf{r}_1}, \mathbf{w}_{\mathbf{r}_2}, \ldots, \mathbf{w}_{\mathbf{r}_N})$. A learning step that causes a change from $\mathbf{w}'$ to $\mathbf{w}$ can formally be described by the transformation

$$\mathbf{w} = \mathbf{T}(\mathbf{w}', \mathbf{v}, \epsilon). \tag{75}$$

Here $\mathbf{v} \in V$ represents the input vector invoked at a particular instance, and ϵ is a measure of the plasticity of the synaptic strengths (see Eq. (70)).

The learning process is driven by a sequence of randomly and independently chosen vectors $\mathbf{v}$ whose distribution obeys

a probability density $P(\mathbf{v})$. The transformation (75) then defines a Markov process in the space of synaptic strengths $\mathbf{w} \in V \otimes V \otimes \ldots \otimes V$ that describes the evolution of the map $\phi_{\mathbf{w}}(\mathbf{v})$. We will now show that the stationary state of the map which evolves asymptotically by this process can be described by a partial differential equation for the stationary distribution of the synaptic strengths.

Since the elements $\mathbf{v}$ occur with the probability $P(\mathbf{v})$, the probability $Q(\mathbf{w}, \mathbf{w}')$ for the transition of a state $\mathbf{w}'$ to a state $\mathbf{w}$, via adaptation step (75), is given by

$$Q(\mathbf{w}, \mathbf{w}') = \int \delta(\mathbf{w} - \mathbf{T}(\mathbf{w}', \mathbf{v}, \epsilon)) P(\mathbf{v})\, d\mathbf{v}. \tag{76}$$

$\delta(\mathbf{x})$ denotes the so-called delta-function which is zero for all $\mathbf{x} \neq 0$ and for which $\int \delta(\mathbf{x}) d\mathbf{x} = 1$. More explicitly, Eq. (75) can be written

$$\mathbf{w}_{\mathbf{r}} = \mathbf{w}'_{\mathbf{r}} + \epsilon\, h_{\mathbf{rs}}(\mathbf{v} - \mathbf{w}'_{\mathbf{r}}) \quad \text{for all } \mathbf{r} \in A. \tag{77}$$

Here $\mathbf{s} = \phi_{\mathbf{w}'}(\mathbf{v})$ is the formal neuron to which $\mathbf{v}$ is assigned in the old map $\phi_{\mathbf{w}'}$.

In the following we take exclusive interest in those states $\phi_{\mathbf{w}}$ that correspond to "maximally ordered maps," and we want to investigate their dependence on the probability density $P(\mathbf{v})$. We assume that the space V and the lattice A have the same dimensionality d. A "maximally ordered map" can then be characterized by the condition that lines in V which connect the $\mathbf{w}_{\mathbf{r}}$ of $\mathbf{r}$ adjacent in the network are not allowed to cross. Figure 5.5 demonstrates this fact with an example of a two-dimensional Kohonen lattice on a two-dimensional space V of input stimuli with a homogeneous probability distribution $P(\mathbf{v})$. The square frame represents the space V. The synaptic strengths $\mathbf{w}_{\mathbf{r}} \in V$ determine the locations on the square which are assigned to the formal neurons $\mathbf{r} \in A$. Each mesh point of the lattice A corresponds to a formal neuron and, in our representation, is drawn at the location that has been assigned to that neuron through $\mathbf{w}_{\mathbf{r}}$. Two locations $\mathbf{w}_{\mathbf{r}}$ are connected by a line if the two corresponding formal neurons $\mathbf{r}$ are neighbors in the lattice A. Figure 5.5a shows a map that has reached a state of "maximal order" as seen by the lack of line crossings between lattice points. In contrast Fig. 5.5b presents a map for which even in the final stage some connections still cross. Such a map is not "maximally ordered."

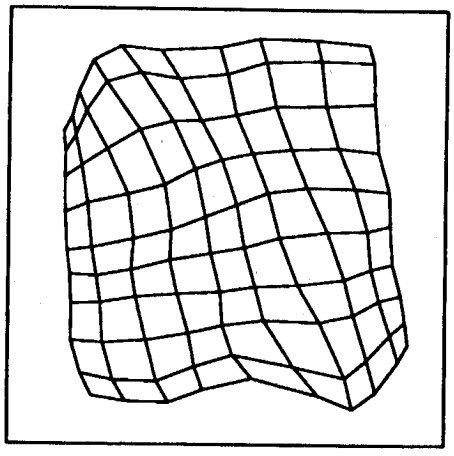

Figure 5.5a An example for a "maximally ordered" state of the network. Network and input signals are both two-dimensional. All input signals originate from the limiting square. In the continuum limit the network nodes are infinitely dense and specify a one-to-one mapping between the network and the square.

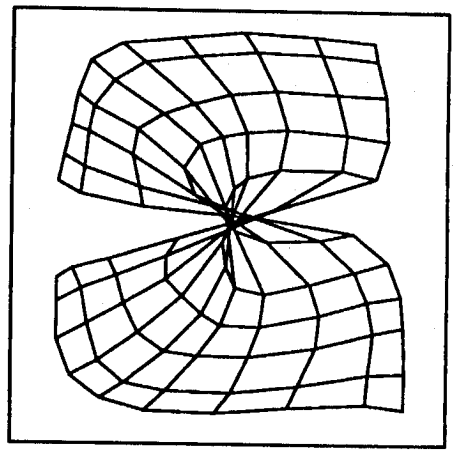

Figure 5.5b An example of an incompletely ordered state of the network, evolved as a consequence of the range $\sigma(t)$ of $h_{\mathbf{rs}}$ to be too short initially (see Eq. (68)). In this case a topological defect develops and the connections between neighboring lattice points cross. In the continuum limit a one-to-one mapping cannot be obtained.

In the following calculation we will make a transition from discrete values of $\mathbf{r}$ to continuous ones. This is possible because in the following we restrict ourselves to "maximally ordered" states where in the transition to a continuum $\mathbf{w_r}$ becomes a smooth function of the spatial coordinate $\mathbf{r}$ in the network.

We consider an ensemble of maps that, after t learning steps, are all in the vicinity of the same asymptotic state and whose distribution is given by a distribution function $S(\mathbf{w}, t)$. In the limit $t \to \infty$, $S(\mathbf{w}, t)$ converges towards a stationary distribution $S(\mathbf{w})$ with a mean value $\bar{\mathbf{w}}$. In Chapter 14 we will show that the variance of $S(\mathbf{w})$ under the given conditions will be of the order of ϵ. Therefore, for an ϵ that is sufficiently slowly approaching zero, all members of the ensemble will result in the same map characterized by its value $\bar{\mathbf{w}}$.

We want to calculate $\bar{\mathbf{w}}$ in the limit $\epsilon \to 0$. In the stationary state, the condition $S(\mathbf{w}) = \int Q(\mathbf{w}, \mathbf{w}')S(\mathbf{w}')\, d\mathbf{w}'$ holds, and, therefore, it also holds that

$$\bar{\mathbf{w}} = \int \mathbf{w}S(\mathbf{w})\, d\mathbf{w} = \int\int \mathbf{w}Q(\mathbf{w}, \mathbf{w}')S(\mathbf{w}')\, d\mathbf{w}d\mathbf{w}'. \qquad (78)$$

In the limit $\epsilon \to 0$ it follows $S(\mathbf{w}) \to \delta(\mathbf{w} - \bar{\mathbf{w}})$ and, therefore,

$$\bar{\mathbf{w}} = \int \mathbf{w}Q(\mathbf{w}, \bar{\mathbf{w}})\, d\mathbf{w}$$

$$= \int \mathbf{T}(\bar{\mathbf{w}}, \mathbf{v}, \epsilon) P(\mathbf{v})\, d\mathbf{v}. \tag{79}$$

Applying Eq. (77) we obtain

$$0 = \epsilon \int h_{\mathbf{rs}}(\mathbf{v} - \bar{\mathbf{w}}_{\mathbf{r}}) P(\mathbf{v})\, d\mathbf{v} \quad \text{for all } \mathbf{r} \in A. \tag{80}$$

We formulate the restriction of maximally ordered maps by two approximating assumptions:

(*i*) We assume that for sufficiently large systems $\bar{\mathbf{w}}_{\mathbf{r}}$ is a function that varies slowly from lattice point to lattice point so that its replacement by a function $\bar{\mathbf{w}}(\mathbf{r})$ on a continuum of r-values is justified.

(*ii*) We assume that $\bar{\mathbf{w}}(\mathbf{r})$ is one-to-one.

We demand also that $h_{\mathbf{rs}}$ at $\mathbf{r} = \mathbf{s}$ has a steep maximum and satisfies

$$\begin{aligned} \int h_{\mathbf{rs}}(\mathbf{r} - \mathbf{s})\, d\mathbf{r} &= 0, \\ \int h(\mathbf{r} - \mathbf{s})(r_i - s_i)(r_j - s_j)\, d\mathbf{r} &= \delta_{ij}\sigma^2, \quad i, j = 1, \ldots, d \end{aligned} \tag{81}$$

where d is the dimension of V and r_j, s_j describe the d Cartesian components of $\mathbf{r}$, $\mathbf{s}$. The constant σ is the range of $h_{\mathbf{rs}}$ which coincides with σ in (68) in case of a Gaussian $h_{\mathbf{rs}}$.

From the above we will derive a differential equation for $\bar{\mathbf{w}}$. Due to the continuum approximation (*i*), the quantity $\min_{\mathbf{r} \in A} \|\mathbf{w}_{\mathbf{r}} - \mathbf{v}\|$ in Eq. (74) vanishes because now for each $\mathbf{v}$ there exists exactly one $\mathbf{r}$ for which $\mathbf{w}_{\mathbf{r}} = \mathbf{v}$ holds. Therefore, we can replace $\mathbf{v}$ in Eq. (80) by $\bar{\mathbf{w}}(\mathbf{s})$. Here $\mathbf{s} := \phi_{\bar{\mathbf{w}}}(\mathbf{v})$ is the image of $\mathbf{v}$ under the map that belongs to $\bar{\mathbf{w}}$. This provides the condition

$$\int h_{\mathbf{rs}}\big(\bar{\mathbf{w}}(\mathbf{s}) - \bar{\mathbf{w}}(\mathbf{r})\big) P(\bar{\mathbf{w}}(\mathbf{s})) J(\mathbf{s})\, d\mathbf{s} = 0. \tag{82}$$

Here

$$J(\mathbf{s}) := \left| \frac{d\mathbf{v}}{d\mathbf{s}} \right| \tag{83}$$

is the absolute value of the Jacobian of the map $\phi_{\bar{\mathbf{w}}}$. With $\mathbf{q} := \mathbf{s} - \mathbf{r}$ as a new integration variable and $\bar{P}(\mathbf{r}) := P(\bar{\mathbf{w}}(\mathbf{r}))$ the expansion of Eq. (82) in powers of $\mathbf{q}$ yields (with implicit summation over repeated indices; *e.g.*, $q_i \partial_i$ is to be summed over

all values of i)

$$
\begin{aligned}
0 &= \int h_{\mathbf{q}0}(q_i\partial_i\bar{\mathbf{w}} + \frac{1}{2}q_iq_j\partial_i\partial_j\bar{\mathbf{w}} + \ldots)\cdot \\
&\qquad \cdot (\bar{P} + q_k\partial_k\bar{P} + \ldots)\cdot(J + q_l\partial_l J + \ldots)\, d\mathbf{q} \\
&= \int h_{\mathbf{q}0}q_iq_j\, d\mathbf{q}\cdot\left((\partial_i\bar{\mathbf{w}})\partial_j(\bar{P}J) + \frac{1}{2}\bar{P}J\cdot\partial_i\partial_j\bar{\mathbf{w}}\right)(\mathbf{r}) + O(\sigma^4) \\
&= \sigma^2\cdot\left[(\partial_i\bar{\mathbf{w}})(\partial_i(\bar{P}J) + \frac{1}{2}\bar{P}J\cdot\partial_i^2\bar{\mathbf{w}}\right](\mathbf{r}) + O(\sigma^4),
\end{aligned}
\tag{84}
$$

where we made use of (81). In order for the expansion (84) to hold it is necessary and sufficient for small σ that condition

$$
\sum_i \partial_i\bar{\mathbf{w}}\left(\frac{\partial_i\bar{P}}{\bar{P}} + \frac{\partial_i J}{J}\right) = -\frac{1}{2}\sum_i\partial_i^2\bar{\mathbf{w}} \tag{85}
$$

or, with the Jacobi matrix $J_{ij} = \partial_j\bar{w}_i(\mathbf{r})$ and $\Delta = \sum_i \partial_i^2$, condition

$$
\mathbf{J}\cdot\nabla\ln(\bar{P}\cdot J) = -\frac{1}{2}\Delta\bar{\mathbf{w}} \tag{86}
$$

is satisfied. For the one-dimensional case we obtain $\mathbf{J} = J = d\bar{w}/dr$ and $\Delta\bar{\mathbf{w}} = d^2\bar{w}/dr^2$ with $\bar{w}$ and r as scalars. In this case the differential equation (86) can be solved. For this purpose we rewrite (86) and obtain

$$
\frac{d\bar{w}}{dr}\left(\frac{1}{P}\frac{d\bar{P}}{dr} + \left(\frac{d\bar{w}}{dr}\right)^{-1}\frac{d^2\bar{w}}{dr^2}\right) = -\frac{1}{2}\frac{d^2\bar{w}}{dr^2} \tag{87}
$$

from which we can conclude

$$
\frac{d}{dr}\ln\bar{P} = -\frac{3}{2}\frac{d}{dr}\ln\left(\frac{d\bar{w}}{dr}\right). \tag{88}
$$

This result allows us to determine the local enlargement factor of the map in terms of the generating probability distribution $P(\mathbf{v})$.

Since $\phi_{\bar{\mathbf{w}}}(\bar{\mathbf{w}}(\mathbf{r})) = \mathbf{r}$ holds, the local enlargement factor M of $\phi_{\bar{\mathbf{w}}}$ can be defined by $M = 1/J$ (compare Eq. (83)). For the one-dimensional case $M = (d\bar{w}/dr)^{-1}$ and we obtain as a relation between input stimulus distribution and cortical representation

$$
M(v) = J^{-1} = \frac{dr}{d\bar{w}} \propto P(v)^{2/3}. \tag{89}
$$

The local enlargement factor $M(v)$ depends on the probability density $P(v)$ according to a power law. It can be shown that the

exponent 2/3 that we found in the continuum approximation undergoes a correction for a discrete one-dimensional system and is then given by $\frac{2}{3} - [3(1+n^2)(1+[n+1]^2)]^{-1}$, where n is the number of neighbors that are taken into account on each side of the excitation center, (*i.e.*, $h_{\mathbf{rs}} = 1$ for $\|\mathbf{r}-\mathbf{s}\| \leq n$ and zero elsewhere) (Ritter 1989). The continuum corresponds to the limit of infinite density of neighbors. Then $n = \infty$ for each finite σ and we obtain the previous result of 2/3.

5.5. "Cortical Representation" in the Model of the Bat's Auditory Cortex

We now apply the mathematical derivation of Section 5.4 to the particular input stimulus distribution that we assumed for our model of the bat's auditory cortex and compare the result with a simulation.

The input stimulus distribution that we assume can be written in the range $v_1 \leq v \leq v_2$ as

$$P(v) = \frac{P_0}{v_2 - v_1} + (1 - P_0)\frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(v - v_e)^2}{2\sigma_r^2}\right) \tag{90}$$

with the parameters σ_r=0.5 kHz, v_e=61.0 kHz, v_1=20 kHz, v_2=100 kHz and P_0=1/4. The width of the distribution of the Doppler-shifted echoes is given by σ_r, and P_0 is the probability for the occurrence of an input stimulus from the white background noise. v_1 and v_2 are the limits of the ultrasound spectrum that we assume the bat can hear.

The integral $I = \int_{v_1}^{v_2} P(v)dv$ is not exactly unity because of the finite integration limits. Since, due to the small σ_r of 0.5 kHz, nearly all the Doppler-shifted echo signals lie within the interval $[20, 100]$ and the deviation of I from unity is negligible. With the choice $P_0 = 1/4$, the Doppler-shifted signals occur three times as often as signals due to the background noise (see also Fig. 5.2). From Eqs. (89) and (90) we find

$$\frac{dr}{d\bar{w}} = C \cdot \left(\frac{P_0}{v_2 - v_1} + (1 - P_0)\frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(v - v_e)^2}{2\sigma_r^2}\right)\right)^{2/3} \tag{91}$$

where C is a proportionality constant. In integral form one has

$$r(\bar{w}) - r_1 = C \cdot \int_{\bar{w}_1}^{\bar{w}} \left(\frac{P_0}{v_2 - v_1} + \frac{1 - P_0}{\sqrt{2\pi}\sigma_r} \times \exp\left(-\frac{(v - v_e)^2}{2\sigma_r^2} \right) \right)^{2/3} dv. \tag{92}$$

We will solve this integral numerically and then compare the resulting $\bar{w}(r)$ with the corresponding values from a simulation.

Since these considerations apply only to the case where the dimensionality of the net and the dimensionality of the space of input stimuli is identical, we stretch the "auditory cortex" and, instead of a 5×25 net as in Figs. 5.3 and 5.4, assume a one-dimensional chain with 50 elements for the present simulation. Starting from a linear, second-order differential equation, we need two boundary conditions, *e.g.*, $\bar{w}_1(r_1)$ and $\bar{w}_2(r_2)$, from our simulation data to be able to adjust the function $r(\bar{w})$ of Eq. (92) uniquely. Since boundary effects at the beginning and the end of the chain were not taken into account in our analytic calculation, the end points can in some cases deviate slightly from our calculated curve. To adjust the curve to the simulation data, we take values for w_1 and w_2 that do not lie too close to the end points; in this case we have chosen $\bar{w}$ at the third and forty-eighth link of the chain, *i.e.*, at $r_1 = 3$ and $r_2 = 48$. The solid curve in Fig. 5.6 depicts the function $\bar{w}(r)$ calculated numerically from Eq. (92) and adjusted to the simulation data. The dots show the values $\bar{w}_r$ that were obtained by simulating the Markov process (75). The representation corresponds to the one in Fig. 5.4. The time dependence of the excitation zone σ and of the adaptation step width ϵ for the simulation were chosen as follows: $\sigma(t) = \sigma_i[1 + \exp(-(5t/t_{max})^2)]$ with $\sigma_i = 10$, $\epsilon(t) = \epsilon_i \exp(-(5t/t_{max})^2)$ with $\epsilon_i = 1$. For the maximal number of learning steps $t_{max} = 20000$ was chosen.

Clearly, the function $\bar{w}(r)$ resulting from Eq. (92) is in close agreement with the simulation results, and even the deviations at the end points are small. One may have expected intuitively that for the magnification holds $M(v) \propto P(v)$, *i.e.*, a magnification proportional to the stimulus density. The corresponding result is presented in Fig. 5.6 as well to demonstrate that this expectation is, in fact, incorrect.

For the present input stimulus distribution, it is possible to estimate the size of the region relevant for the analysis of

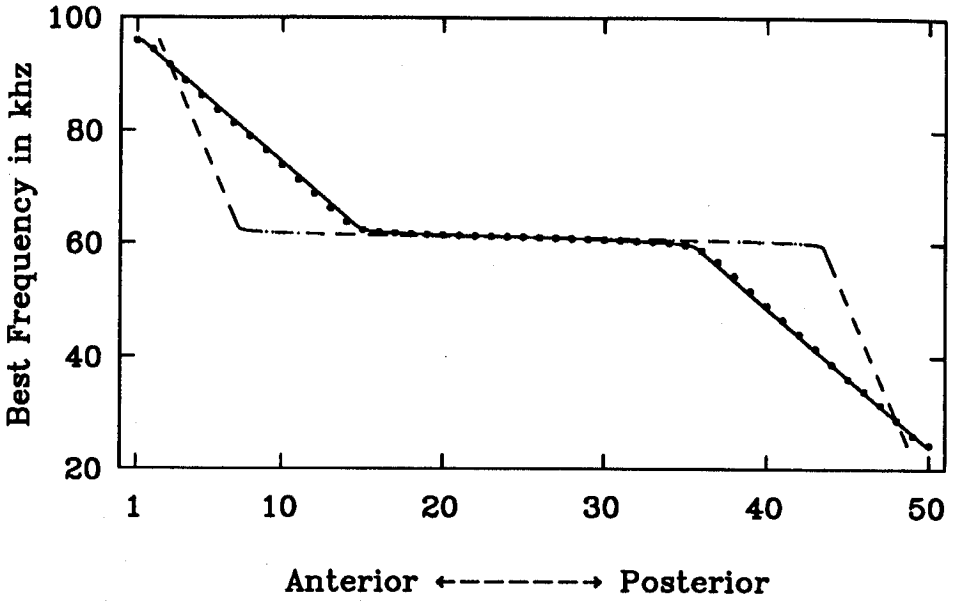

Figure 5.6 A bat's sensitivity to accoustic and sonar signals (cf. Fig. 5.4). The solid curve represents the function $\bar{w}(r)$ calculated from Eq. (92). The dots show the values obtained from simulating the Markov process (75). For comparison we show the result for $M(v) \propto P(v)$ with a dashed line. This result strongly deviates from the simulation data.

the Doppler-shifted signal, *i.e.*, the extension of the 61 kHz plateau in Fig. 5.6. In Eq. (92) we integrate over $P(v)^{2/3}$ and, therefore, the function $r(\bar{w})$ increases sharply for large values of $P(v)$. Hence, the plateau starts where the Gaussian distribution of the Doppler-shifted echoes increases strongly relative to the background. This is approximately the case for $v = v_e - 2\sigma_r$. Accordingly, the plateau ends where the Gaussian peak recedes back into the homogeneous background, *i.e.*, at $v = v_e + 2\sigma_r$. Therefore, the relation

$$\Delta r_{plateau} = C \cdot \int_{v_e-2\sigma_r}^{v_e+2\sigma_r} \left(\frac{P_0}{v_2 - v_1} + (1 - P_0) \frac{1}{\sqrt{2\pi}\sigma_r} \times \exp\left(-\frac{(v - v_e)^2}{2\sigma_r^2} \right) \right)^{2/3} dv. \tag{93}$$

for the size of the plateau holds. Within these integration limits the background portion in the integrand is negligible compared to the values of the Gaussian. Furthermore, we can extend the integration of the integrand that results without the background towards infinity without significant error. The integral can then

be evaluated, yielding the approximation

$$\Delta r_{plateau} \approx C \cdot (1 - P_0)^{2/3} \int_{-\infty}^{\infty} \frac{1}{(\sqrt{2\pi}\sigma_r)^{2/3}} \exp\left(-\frac{2}{3}\frac{v^2}{2\sigma_r^2}\right) dv$$

$$\approx C \cdot \sqrt{\frac{3}{2}} \left(\sqrt{2\pi}\sigma_r (1 - P_0)^2\right)^{1/3}. \tag{94}$$

In order to determine the part of the plateau relative to the overall "auditory cortex," we also need an estimate of the integral in Eq. (92), where we have to integrate over the full band width of input frequencies. To obtain this we split the integration range from v_1=20 kHz to v_2=100 kHz into three regions as follows

$$\Delta r_{total} \propto \int_{v_1}^{v_e - 2\sigma_r} (P(v))^{2/3}\, dv + \int_{v_e - 2\sigma_r}^{v_e + 2\sigma_r} (P(v))^{2/3}\, dv + \int_{v_e - 2\sigma_r}^{v_2} (P(v))^{2/3}\, dv. \tag{95}$$

We have already estimated the second integral in the sum by Eq. (94). Within the integration limits of the other two integrals the contribution of the Gaussian distribution is so small that it can be neglected relative to the background. In addition, $\sigma_r \ll (v_2 - v_1)$, enabling us to write

$$\Delta r_{total} \approx \Delta r_{plateau} + C \cdot (v_2 - v_1) \left(\frac{P_0}{v_2 - v_1}\right)^{2/3}$$

$$\approx \Delta r_{plateau} + C \cdot {P_0}^{2/3} (v_2 - v_1)^{1/3}. \tag{96}$$

If we insert the parameters of our above model of the input stimulus distribution of the bat into the two estimates (94) and (96), we obtain for the size of the 61 kHz region, relative to the size of the total "cortex," the value

$$\frac{\Delta r_{plateau}}{\Delta r_{total}} \approx 39\%.$$

This implies that for our case of a 50-unit chain, the plateau should consist of 19 to 20 neurons. This value agrees very well with the simulation results presented in Fig. 5.6.

By now we have extensively described the basics of Kohonen's model—the self-organization of a topology-conserving map between an input stimulus space and a network of neural

units. We have compared the simulation results of Kohonen's model to experimental data as well as to a mathematical description valid for certain limiting cases. The simulation data have agreed at least qualitatively with the experimental findings. More than a qualitative agreement should not have been expected, considering the many simplifications of Kohonen's model. In contrast to that, the mathematical result for the representation of the input signals relative to their probability corresponds, even quantitatively, very well to the results obtained from simulations.

In Chapter 6 we will become acquainted with a completely different application of Kohonen's model. Instead of a mapping onto a continuum, we will generate a mapping that projects a linear chain onto a discrete set of points. Such a mapping can be interpreted as a choice of a connection path between the points. The feature of the algorithm to preserve topology as much as possible manifests itself in a tendency to minimize the path-length. In this way, very good approximate solutions for the well-known travelling salesman problem can be achieved.