Kohonen's Self-Organizing Maps: Exploring their Computational Capabilities

Helge Ritter and Klaus Schulten
Physik-Department
Technische Universität München
D-8046 Garching
Federal Republic of Germany

Abstract: It is demonstrated that the computational capabilities of Kohonen's self-organizing mapping algorithm can be applied to problems from such diverse fields as sensory mappings, combinatorial optimization and motor control. In addition we present some recent mathematical results characterizing important properties of the algorithm in these situations.

1. Introduction

Connections in a biological neural network of a higher animal cannot be genetically prespecified in a detailed manner. Consequently much of the "neural wiring diagram" can only evolve after birth and requires sensory and motor experiences for proper maturation (Held and Hein 1963). At present, very little is known about higher levels of this extensive process of organization. On a lower level, the adaptive formation of topology-conserving neural projections can be observed in many different parts of the brain (Kaas et al. 1983, King et al. 1988, Harris 1986). Models successfully accounting for the formation of such projections from simple principles have been proposed and investigated by several authors (Grossberg 1976ab, v.d.Malsburg and Willshaw 1977, Willshaw and v.d. Malsburg 1979, Overton and Arbib 1982). Kohonen demonstrated the capability of similar principles for the adaptive formation of more abstract feature maps (Kohonen 1982abc, 1984a, Kohonen et al. 1984b). We believe that his model constitutes a most valuable step towards a better understanding of the question, how the organization of higher levels of synaptic connections can occur on the basis of simple and fairly low level synaptic modification rules without the need for extensive instruction beyond exploratory sensory and motor experience. In this contribution, we want to demonstrate that the computational capabilities of Kohonen's algorithm provide an unified approach to such diverse fields as sensory mappings, combinatorial optimization and learning in motor control. To this end we give in Sec. 2 a brief account of the algorithm. In Sec. 3 we show several applications in the aforementioned fields and in Sec. 4 we present some recent analytical results for the model.

2. The Algorithm

The aim of Kohonen's algorithm (Kohonen 1982a, 1984a) is to generate a mapping of a higher dimensional space V of input signals onto an, usually two-dimensional, discrete lattice A of formal neurons. The map is generated by establishing a correspondence between inputs from V and neurons in the lattice such, that the topological (neighborhood) relationships among the inputs are reflected as faithfully as possible in the arrangement of the corresponding neurons in the lattice. This renders a "non-linearly flattenend" two-dimensional version of the input space which for many tasks constitutes a very useful data structure.

The correspondence is obtained iteratively by a sequence of training steps, which can be formulated in terms of synaptic modification laws (Kohonen 1984a) for the neurons. However, for the purposes of this paper, we present the algorithm in an abstract form without explicit reference to neurons.

Each input signal is represented by a vector $\mathbf{v} \in V$. For each training step an input $\mathbf{v} \in V$ is chosen randomly according to some probability of occurrence $P(\mathbf{v})$. Each location $\mathbf{r}$ of the lattice A carries a vector $\mathbf{w_r} \in V$. The vectors $\mathbf{w_r}$ map lattice locations $\mathbf{r}$ to points in V and for each training step this mapping is adjusted by the following two steps:

1. Determine lattice location $\mathbf{s}$ for which

$$||\mathbf{w_s} - \mathbf{v}|| = \min_{\mathbf{r} \in A} ||\mathbf{w_r} - \mathbf{v}||$$

where $\mathbf{v}$ is the randomly chosen input signal for the current step.

2. For all sites $\mathbf{r}$ in the neighborhood of $\mathbf{s}$ (with $\mathbf{s}$ included) adjust

$$\mathbf{w}_\mathbf{r}^{(new)} = \mathbf{w}_\mathbf{r}^{(old)} + \epsilon h_{\mathbf{rs}}(\mathbf{v} - \mathbf{w}_\mathbf{r}^{(old)}), \tag{1}$$

where $0 \leq h_{\mathbf{rs}} \leq 1$ is a prespecified adjustment function of the distance $||\mathbf{r} - \mathbf{s}||$ and ϵ is a learning step size. $h_{\mathbf{rs}}$ has its maximum at $\mathbf{r} = \mathbf{s}$ and usually decays to zero, as $||\mathbf{r} - \mathbf{s}||$ increases.

By decreasing the step size ϵ and the lateral width of $h_{\mathbf{rs}}$ slowly with increasing number of training steps, the algorithm can be shown to gradually yield values for the vectors $\mathbf{w_r}$ which define a (discretized) neighborhood conserving mapping between lattice sites $\mathbf{r}$ and points of the input space V (Kohonen, op.cit.).

3. Applications

In this Section we shall demonstrate the computational capabilities of the mapping algorithm by examples in three different fields: sensory mappings, combinatorial optimization and learning in motor control. Applications have been explored also in pattern recognition (Bertsch and Dengler 1987, Kohonen op.cit.).

Throughout the simulations we chose $\epsilon(t) = \epsilon_i \cdot (\epsilon_f/\epsilon_i)^{t/t_{max}}$, $h_{\mathbf{rs}} = \exp(-||\mathbf{r} - \mathbf{s}||^2/2\sigma(t)^2)$ and $\sigma(t)$ as $\epsilon(t)$ with $\epsilon_{f,i}$ replaced by $\sigma_{f,i}$. The variable t counts the learning steps.

The first simulation shows the formation of a somatotopic map between the tactile receptors of a hand surface and a model cortex of 30×30 formal neurons arranged as a regular square lattice (Ritter and Schulten 1986). In this case the space V consists of the activity patterns of the set of tactile receptors covering the hand surface and each

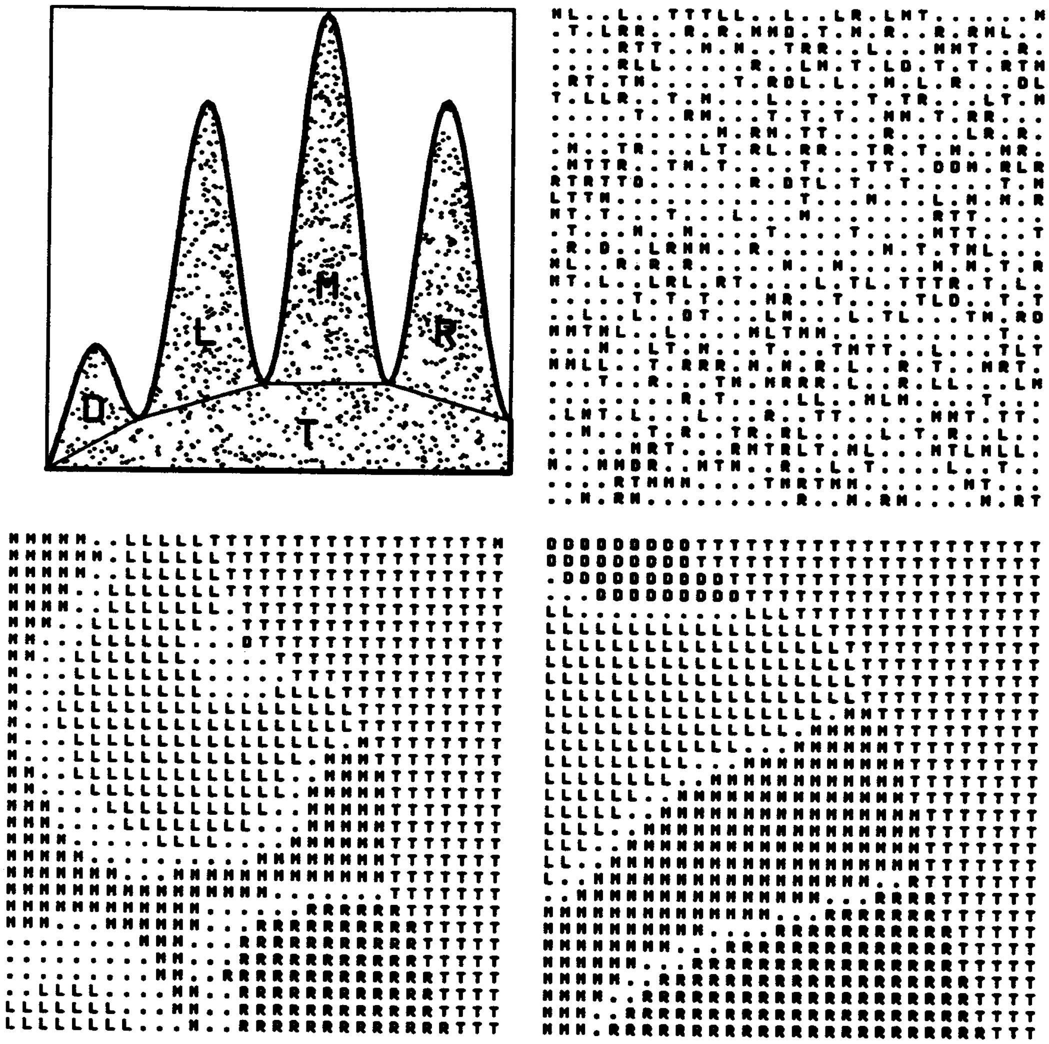

Figure 1a-d

component of a vector $\mathbf{w_r}$ specifies the degree of connectivity between neuron $\mathbf{r}$ and a receptor. If the probability distribution $P(\mathbf{v})$ is restricted to inputs $\mathbf{v}$ corresponding to spatially localized activity distributions in the receptor sheet, it can be shown that an ordered somatotopic map, which preserves the neigborhood relationships in the receptor sheet, emerges even from random initial connectivities. In this situation, the process itself can be mapped to an equivalent situation for which the pattern space V is replaced by a two-dimensional space and Figs. 1b-d show different stages of a simulation of this equivalent process, starting from a random connectivity. Each picture shows the correspondence of the 900 cortex sites and one of the five hand areas denoted by letters D, L, M, R and T in Figure 1a (neurons on sites labelled by a dot have not yet developed a preference to either of the five areas).

The second application is inspired by Durbin and Willshaw's solution of the travelling salesman problem by an elastic net method (Durbin and Willshaw 1987). We show that this type of problem can be solved in a very similar fashion by applying Kohonen's original algorithm, which differs in the way the adjustments are produced. To this end input vectors $\mathbf{v}$ and weight vectors $\mathbf{w_r}$ are chosen two-dimensional and the probability distribution $P(\mathbf{v})$ is concentrated to a set of $L = 30$ randomly chosen locations of "towns" in the unit square. Instead of a two-dimensional lattice of neurons a linear, closed chain of 100 "neurons" is chosen for A. Figures 2a,b,c show the resulting mapping process for $\epsilon_i = \epsilon_f = 0.8$, $\sigma_i = 50$, $\sigma_f = 1$ and $t_{max} = 10\,000$. Each stage of the process is visualized by showing the square together with the image of the chain A under the mapping $\mathbf{r} \in A \mapsto \mathbf{w_r} \in [0,1]^2$. A regular 100-gon was chosen as initial configuration (Fig.2a). During the training sequence it gradually deforms into a path connecting the 30 points to a closed tour (Fig.2b: after 7 000 steps, Fig.2c: after 10 000 steps). The tendency of the algorithm to preserve neighborhood structure results in attempting especially short tours as final configurations. In fact, for the simulation shown the algorithm detected the shortest possible tour, although this is not always guaranteed and slightly longer tours than minimal may be obtained.

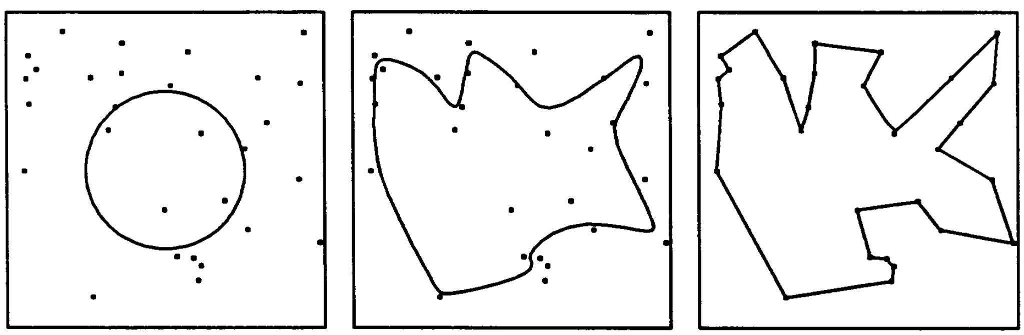

Figure 2a-c

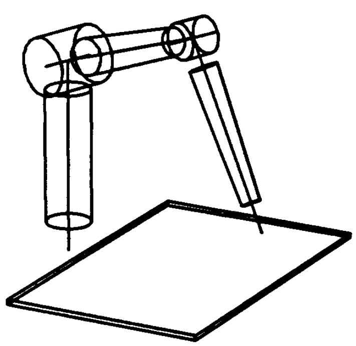

Figure 3

The last example deals with a problem in motor control. We consider a three-jointed robot arm over a table (Fig.3). Its joints are actuated by torques $\vec{\tau} = (\tau_1, \tau_2, \tau_3)^T$ and its joint angles are $\vec{\theta} = (\theta_1, \theta_2, \theta_3)^T$. The arm shall be actuated by short torque pulses and swing freely for the remaining periods ("ballistic movements"). A torque pulse $\vec{\tau}(t) = \vec{\tau} \cdot \delta(t)$ accelerates the end effector to a velocity given by $\mathbf{v} = \mathbf{A}(\vec{\theta})\vec{\tau}$ (we assume the end effector to be initially at rest and we assume absence of gravity). Here $\mathbf{A}$ is a configuration dependent linear mapping, relating a desired velocity $\mathbf{v}$ for the end effector with the required torque pulse amplitude $\vec{\tau}$.

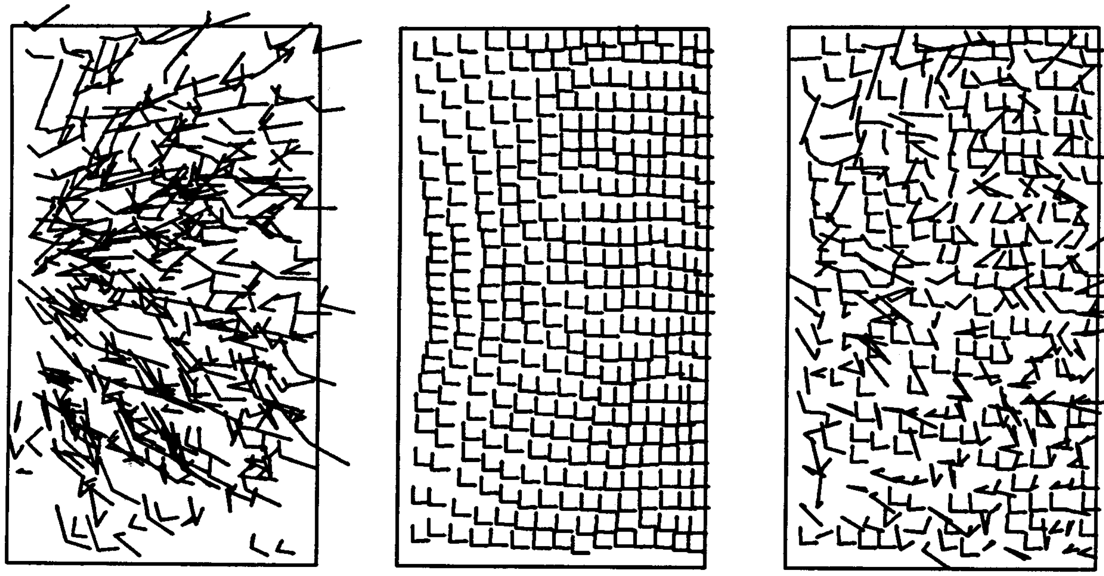

Figure 4a-c

A generally depends in a very complicated manner on configuration and dynamical properties of the arm and, therefore, often is not known explicitly. Here we shall employ an extension of Kohonen's algorithm to learn **A** by trial movements from the velocities **v** actually attained. During training each site **r** shall become responsible for a small subregion of configuration space, located at $\mathbf{w_r}$, and learn the appropriate transformation $\mathbf{A_r}$ valid there. To this end we again employ a lattice, but now we attach to each site **r** in addition to $\mathbf{w_r}$ a linear mapping $\mathbf{A_r}$ transforming desired velocity into required torque amplitude. Training consists of a sequence of random trial movements. Each movement starts from a configuration $\vec{\theta}$, for which the end effector is at a random position on the table. For the simulation the desired velocity **u** of the movement is chosen of unit magnitude and along a random direction. Configuration $\vec{\theta}$ selects the site **s** for which $||\vec{\theta} - \mathbf{w_s}||$ is minimal, and on the basis of the associated matrix $\mathbf{A_s}$ a torque pulse of amplitude $\vec{\tau} = \mathbf{A_s u}$ is applied at the joints. A learning rule of error-correction-type (Widrow and Hoff, 1960) is used to obtain from the velocity **v** actually resulting from $\vec{\tau}$ an improved estimate

$$\mathbf{A}^* = \mathbf{A_s} + ||\mathbf{v}||^{-1}\left(\vec{\tau} - \mathbf{A_s v}\right)\mathbf{v}^T$$

for the correct value of $\mathbf{A_s}$. Then the pair $(\vec{\theta}, \mathbf{A}^*)$ is used to perform an adjustment step on the lattice variables $\mathbf{w_r}$ and $\mathbf{A_r}$, which is analogous to step 2. in Sec.2.

In this case the resulting data structure is a planar map of linear mappings connecting sensory input to motor output. This is reminiscent of a qualitatively similar situation in the Superior Colliculus, where a topologically organized motor map of eye saccade movements is found (Sparks and Nelson 1987).

Figures 4a,b show the resulting maps from a simulation of the system described above. Figure 4a shows the reactions of the end effector prior to learning, when the initial values of the matrices $\mathbf{A_r}$ entail large random errors. The initial values $\mathbf{w_r}$ have been chosen to correspond to a set of random end effector positions on the table, and for each of these positions the resulting actual velocities from two test movements of equal desired velocity parallel to the table sides are shown. Fig. 4b shows the same test after 10 000 trial movements. Now the reactions are in good agreement with the desired ones. A much worse result is obtained, if the lateral interactions provided by the finite width of $h_{\mathbf{rs}}$ in the learning algorithm are absent (Fig. 4c). In this case only part of the units manages to converge to the correct values (Ritter and Schulten 1987).

4. Mathematical Results

In view of the wide applicability of Kohonen's mapping algorithm, a closer mathematical analysis of its basic properties is highly desireable. Some results for this and a closely related algorithm have already been presented in (Kohonen 1982b, 1984a, Cottrell and Fort 1986, Ritter and Schulten 1986, 1988). Here we want to report some more recent results.

Mathematically, the algorithm is a Markov process with transition probabilities

$$Q(\mathbf{w},\mathbf{w}') = \int \delta(\mathbf{w} - \mathbf{T}(\mathbf{w}',\mathbf{v},\epsilon))P(\mathbf{v})\, d\mathbf{v}. \quad (2)$$

where $\mathbf{w} := (\mathbf{w}_{\mathbf{r}_1}, \ldots, \mathbf{w}_{\mathbf{r}_N})$ is a multi-vector comprising all vectors $\mathbf{w_r}$ of A and $\mathbf{T}(\mathbf{w}',\mathbf{v},\epsilon)$ is the new multi-vector $\mathbf{w}$ obtained from $\mathbf{w}'$ under one adaptation step (1) with input $\mathbf{v}$. For a discrete probability density ($p_i > 0$, $\sum_i p_i = 1$)

$$P(\mathbf{v}) = \sum_{i=1}^{L} p_i \delta(\mathbf{v} - \mathbf{q}_i), \quad \mathbf{q}_i \in V. \quad (3)$$

there exists a potential $V(\mathbf{w})$, whose expected change for state $\mathbf{w}$ under a single learning step is given by

$$E(\Delta V|\mathbf{w}) = -\epsilon \sum_{\mathbf{r}} ||\nabla_{\mathbf{w_r}} V||^2 \quad (4)$$

if the learning step size ϵ is small. V itself is given by

$$V(\mathbf{w}) = \frac{1}{2} \sum_{\mathbf{rs}} h_{\mathbf{rs}} \sum_{\mathbf{q}_i \in F_{\mathbf{s}}(\mathbf{w})} p_i (\mathbf{q}_i - \mathbf{w_r})^2, \quad (5)$$

with

$$F_{\mathbf{s}}(\mathbf{w}) = \{\mathbf{v} \in V \,\big|\, ||\mathbf{v} - \mathbf{w_s}|| < ||\mathbf{v} - \mathbf{w_r}||\ \forall \mathbf{r} \in A\}. \quad (6)$$

V is continuous, but only piecewise differentiable. All local extrema, where V is differentiable, are minima. $-V(\mathbf{w})$ can be interpreted as a measure of the degree of topological ordering of the mapping defined by state $\mathbf{w}$, if $h_{\mathbf{rs}}$ is taken as measuring the degree of neighborhood of $\mathbf{r}$ and $\mathbf{s}$†. V usually exhibits numerous local minima, which correspond to mappings with different types of topological defects. (4) shows, that on the average V decreases and the learning process tries to find states minimizing V. Any individual learning step however, can lead to a temporary increase of V. Therefore, similar to Monte-Carlo-annealing, Kohonen's algorithm has to some extent the capability of escaping from bad local minima with ϵ in loose qualitative analogy to a temperature. The number of minima decreases with increasing range of the lateral adjustment function $h_{\mathbf{rs}}$. In the limiting case $h_{\mathbf{rs}} = const.$ (i.e. infinite range), only a single minimum is left, corresponding to $\mathbf{w_r} = const.$ as the only stable state. As the range of $h_{\mathbf{rs}}$ decreases, more and more minima appear, and correspondingly many different stable equilibrium maps become possible. Therefore, starting with long-ranged $h_{\mathbf{rs}}$ and slowly changing $h_{\mathbf{rs}}$ to the desired values adiabatically transforms V from an "easy" to a "difficult" potential. This greatly facilitates achieving low minima for the learning algorithm and, therefore, good mappings without topological defects. Passing to the limit of a continuous probability distribution, V looses its differentiability and cannot any longer be regarded as a potential function. However, we expect that the presence of many different possible configurations is preserved in this case, too. This is in accordance with experience from simulations, which here likewise produce better maps if the range of $h_{\mathbf{rs}}$ is gradually lowered from large values.

5. Conclusion

We have explored Kohonen's algorithm in three problem domains: sensory mappings, combinatorial optimization and motor learning. For a discrete probability distribution of the training inputs, the formation of the mapping can be described as a probabilistic descent in a potential. In view of their wide applicability, the principles of the algorithm may also be inherent to the maturation of biological brains and may help to achieve a better understanding of these processes from a more unified point of view.

References:

Bertsch H, Dengler J (1987) Klassifizierung und Segmentierung medizinischer Bilder mit Hilfe der selbstlernenden topologischen Karte. In Paulus E (ed), 9.DAGM-Symposium Mustererkennung pp 166-170, Springer Informatik Fachberichte 149, Berlin , Heidelberg.

Cotrell M, Fort JC (1986) A Stochastic Model of Retinotopy: A Self-Organizing Process. Biol Cybern 53:405-411

Durbin R, Willshaw D (1987) An analogue approach to the travelling salesman problem using an elastic net method. Nature 326:689-691

Grossberg S (1976a) On the Development of Feature Detectors in the Visual Cortex

† If $h_{\mathbf{rs}}$ connects only nearest neighbors on a chain, and the state $\mathbf{w}$ is some permutation of a fixed set of locations, V represents the sum of quadratic distances for the tour defined by $\mathbf{w}$. This fact was exploited above to solve the travelling-salesman problem by this method.

with Applications to Learning and Reaction-Diffusion Systems. Biol Cybern 21:145-159

Grossberg S (1976b) Adaptive Pattern Classification and Universal Recoding: I. Parallel Development and Coding of Neural Feature Detectors. Biol Cybern 23:121-134

Harris WA (1986) Learned topography: the eye instructs the ear. TINS, March 1986, 97-99

Held R, Hein A., J. Comp. Physiol. Psychol. 56, 872 (1963)

Kaas JH, Merzenich MM, Killackey HP (1983) The Reorganization of Somatosensory Cortex Following Peripheral Nerve Damage in Adult and Developing Mammals. Annual Rev Neurosci 6:325-56

King AJ, Hutchings ME, Moore DR, Blakemore C (1988) Developmental Plasticity in the Visual and Auditory Representations in the Mammalian Superior Colliculus. Nature 332:73-76

Kohonen T (1982a) Self-organized Formation of Topologically Correct Feature Maps. Biol Cybern 43:59-69

Kohonen T (1982b) Analysis of a Simple Self-organizing Process. Biol Cybern 44:135-140

Kohonen T (1982c) Clustering, Taxonomy and Topological Maps of Patterns. Proc 6th Int Conf on Pattern Recognition, Munich pp 114-128

Kohonen T (1984a) Self-Organization and Associative Memory. Springer Series in Information Sciences 8, Heidelberg

Kohonen T, Măkisara K, Saramăki T (1984b) Phonotopic maps, — insightful representation of phonological features for speech recognition. Proc 7th Int Conf on Pattern Recognition, Montreal pp 182-185

von der Malsburg C, Willshaw DJ, (1977) How to label nerve cells so that they can interconnect in an ordered fashion. Proc Natl Acad Sci USA 74:5176-5178

Overton KJ, Arbib MA (1982) The Branch Arrow Model of the Formation of Retino-Tectal Connections. Biol Cybern 45:157-175

Ritter H, Schulten K (1986) On the Stationary State of Kohonen's Self-Organizing Sensory Mapping. Biol Cybern 54:99-106

Ritter H, Schulten K (1986) Topology Conserving Mappings for Learning Motor Tasks. In Denker JS (ed), Neural Networks for Computing, AIP Conf Proceedings 151, Snowbird, Utah, pp 376-380

Ritter H, Schulten K (1987) Extending Kohonen's Self-Organizing Mapping Algorithm to Learn Ballistic Movements. In Eckmiller R and von der Malsburg C (eds), Neural Computers, Springer, Heidelberg, pp 393-406

Ritter H, Schulten K (1988) Convergency Properties of Kohonen's Topology Conserving Maps: Fluctuations, Stability and Dimension Selection. Submitted to Biological Cybernetics.

Sparks DL, Nelson JS (1987) Sensory and motor maps in the mammalian superior colliculus. TINS 10:312-317

Widrow B, Hoff ME (1960) Adaptive switching circuits, WESCON Conv Record, part IV pp 96-104

Willshaw DJ, von der Malsburg C (1979) A Marker Induction Mechanism for the Establishment of Ordered Neural Mappings: Its Application to the Retinotectal Problem. Proc R Soc London B 287:203-243