A Neural Network Model for the Formation of Topographic Maps in the CNS: Development of Receptive Fields

K. Obermayer, H. Ritter and K. Schulten
Beckman Institute and Department of Physics
University of Illinois at Urbana-Champaign, Urbana, IL 61801

Abstract: The "somatotopic map" of the body surface of animals and humans reflects an ordered, neighborhood preserving connectivity between tactile skin receptors and cortical neurons. In a previous study [1] we investigated a large scale neural network model containing 16,384 neurons and 800 receptors for the formation and plasticity of a somatotopic map of the hand-surface, where we showed, that a somatotopic map of a hand-shaped sensory surface emerged during a sequence of randomly applied local stimuli.

In this paper we discuss the properties of the model in the high-dimensional limit. We present results from a Monte-Carlo-simulation indicating a facilitation of the map-ordering process if the dimensionality of the input space is increased, and we provide a mathematical analysis of the development of localized receptive fields via the input-selection mechanism. The analytical results are compared with data from large scale simulations using a Connection-Machine CM-2 and very good agreement is found.

1. Introduction

Topographic representations of sensory surfaces are a widespread architectural feature of the brain of higher animals. They can be found in the cerebellum (fractured maps), in thalamic nuclei and in several areas of the sensory and motor regions of the cortex, e.g. the visual, auditive and somatosensory fields as well as in the motor-cortex [2, 3, 4, 5, 6].

Several models have been designed to explain the formation and plasticity of such maps. One class of models [7, 8] explains the formation of topographic maps by directional axon growth. Although this process is crucial for the formation of a (at least) crudly, topographic spatial map between a sensory surface and the corresponding areas in the brain, it cannot account for the observed representation of more abstract stimulus features like frequency, doppler-shift and movement direction [9, 10].

A second class of models is primarily concerned with the emergence of feature-detectors and their structured distribution over the cortical surface [11, 12, 13, 14]. Within these models, feature-detectors typically develop by input-selection, modifying the synaptic connection strengths between the sensory receptors and the neurons in the network. A structured distribution of feature-detectors over the network is enforced by suitable lateral interactions via intra-layer connections.

A particularly parsimonious neural network model for the formation and structured distribution of feature detectors by input-selection was proposed by Kohonen [15, 16]. In previous work Kohonen's algorithm has been successfully applied to model the formation of maps in the auditive [17] and somatosensory [18] cortex as well as the formation of maps of more abstract "semantic" features pertaining to language [19].

So far these investigations have been restricted to situations with relatively low dimensional input patterns and the dimensionality of the applied stimuli was of the same order as the number of relevant feature variables. In biological systems the dimensionality of the input-patterns is much higher, while the number of parameters used to describe the "optimal" stimuli may still be small, so that the stimuli "occupy" a low dimensional manifold embedded into a high-dimensional input space. It is unclear, how the properties of the model-system changes , when the input-dimension is scaled up to a biological more realistic size.

In a previous study [1] we investigated a large scale version of Kohonen's model, containing 16,384 neurons and 800 receptors for the formation and plasticity of a somatotopic map of the hand-surface. We showed, that a somatotopic map of a hand-shaped sensory surface emerged during randomly applied local stimuli and that the model readapts upon partial deprivation of sensory input much in the same way as is found in experiments [20, 21].

In this paper we shall discuss the convergence properties of the model in the high-dimensional limit and the development of the receptive fields by the input-selection mechanism.

2. Model and Algorithm:

Our model consists of a square-shaped "sensory surface", representing e.g. a small patch of the skin, and a two-dimensional 128 x 128-lattice of 16,384 formal neurons modelling the corresponding somatosensory area within the cortex (Fig.1). The sensory surface comprises up to 800 randomly distributed tactile receptors, each of which is connected to each neuron in the lattice, leading to a total of 13 million adaptive connections, whose initial strengths are choosen at random and which are modified during the simulation. "Tactile stimuli" were modelled by localized Gaussian excitation profiles, which describe the output r_i of receptor i at position $\vec{x}_i$ as a function of its distance from the center $\vec{x}_s$ of the stimulus by:

$$r_i = A \cdot \exp\left[-(\vec{x}_i - \vec{x}_s)^2/\sigma_r^2\right] \tag{1}$$

The width σ_r and the "intensity" A of the stimuli were held constant throughout the simulation. For each adaptation step the stimulus center $\vec{x}_s$ was choosen at random within the sensory surface. Each neuron (k,l) at position $\vec{y}_{kl}$ of the neuronal sheet computes a weighted sum:

$$o_{kl} = \sum_i w_{kli} r_i \tag{2}$$

over all receptor outputs, where w_{kli} denotes the connection strength from receptor i to neuron (k,l). The input for each neuron is, therefore, described by an 800-dimensional vector $\vec{r} = (r_1, r_2 ... r_{800})^T$. We now make an important notational convention: For the rest of this paper we will reserve the index pair (r,s) for the neuron whose sum o_{kl} is largest for the presented stimulus $\vec{r}$. Therefore, (r,s) is a function of the stimulus $\vec{r}$. Following Kohonen, the output o_{kl} of neuron (r,s) is replaced by a Gaussian output function $h_{rs;kl}$ centered at its position $\vec{y}_{rs}$ in the lattice:

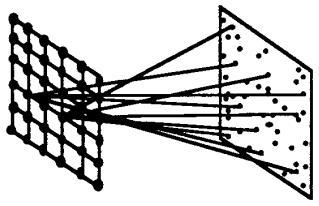

Fig.1: Neuron layer (left) and sensory surface (right) with randomly distributed receptors

$$h_{rs;kl}(t) = \exp\left[-(\vec{y}_{rs} - \vec{y}_{kl})^2/\sigma_h^2(t)\right] \tag{3}$$

The width $\sigma_h(t)$ decreases during the simulation from an initial value σ_i to a final value σ_f to allow the neurons to become specific for a certain part of the input space. The introduction of the output-function $h_{rs;kl}$ is an algorithmic "shortcut" to account for the effect of lateral connections between the neurons [22]. For each stimulus presentation, the connection strengths are changed according to a Hebb-type learning rule:

$$w_{kli}(t+1) = (w_{kli}(t) + \varepsilon(t) h_{rs;kl}(t) \cdot r_i) / \sqrt{\Sigma_i (w_{kli}(t) + \varepsilon(t) h_{rs;kl(t)} \cdot r_i)^2} \tag{4}$$

To accelerate the convergence of the simulation the learning step width $\epsilon(t)$ starts from a relatively large initial value ε_i and decreases linearly to a final value ε_f.

3. Convergence properties of the algorithm:

Fig. 2a shows a completely ordered map of the receptor surface obtained from an initially random connectivity after 10,000 adaptive steps. The network is projected onto the receptor surface such, that each mesh-point (k,l) coincides with the centroid $\vec{s}_{kl} = \sum_i w_{kli}\vec{x}_i / \sum_i w_{kli}$ of the receptive field of neuron (k,l). The rate of convergence is surprisingly fast: after approx. 0.5 adaptive steps per neuron the algorithm has not only detected the two-dimensional manifold within the 800-dimensional input-space, but has also mapped it onto the two-dimensional neuronal sheet in a topology-preserving manner. In a previous simulation [18] with a small network of 900 neurons and only two input-dimensions 20,000 adaptive steps were needed to form a completely ordered map, giving a ratio of 22.0 adaptive steps per neuron. The improved convergence in the large-scale system may stem partially from the fact, that the high-dimensional input-space facilitates the ordering process and makes the formation of topological defects, such as shown Fig. 2b, less likely.

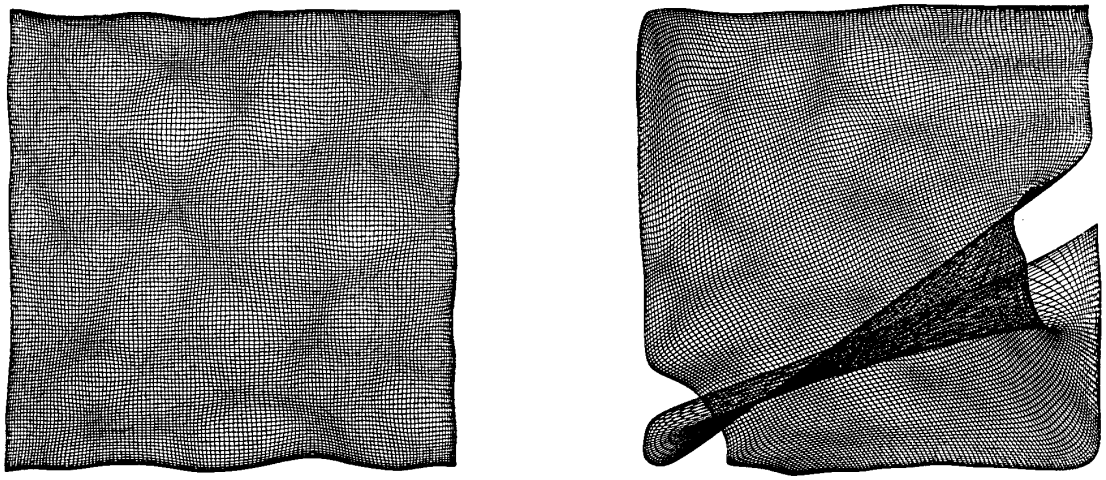

Fig.2: (a) (left) completely ordered map, (b) (right) map showing a "topological defect". The parameter of the simulation were: $\sigma_i = 55$, $\sigma_f = 5$, $\varepsilon = 0.05$, $\sigma_p = 0.15$ and 10,000 steps.

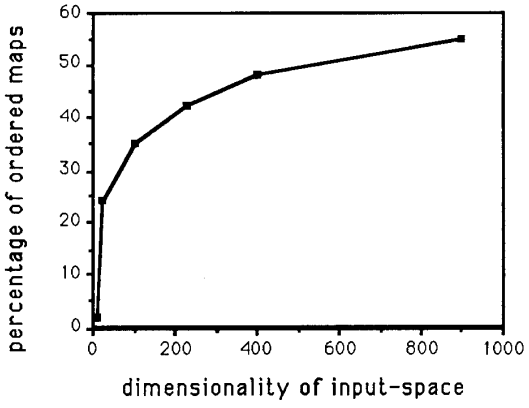

Fig. 3

To investigate this effect we performed Monte-Carlo-Simulations for six ensembles of 50 networks with 16,384 neurons each, but with different dimensionalities (9, 25, 100, 225, 400 and 900) of the input space. The parameters of the simulation ($\sigma_i = 80$, $\sigma_f = 0$, $\varepsilon = 0.08$,$\sigma_p = 0.12$, 5,000 steps) were the same for each of the 300 networks except for the initial connection strengths and for the sequence of applied stimuli, which both were chosen at random. Fig. 3 shows the percentage of completely ordered maps obtained within each ensemble. The increase of this percentage for higher dimensionalities shows, that the scaling of this neural network to larger numbers of neurons and a high-dimensional input space does not necessarily lead to a steep increase in learning time.

Instead it indicates that the convergence properties of the algorithm, which uses cooperativity between neurons, may be even facilitated in a larger system. This is a behavior that is very desirable for neural network algorithms in general.

4. Development of localized receptive fields

An essential aspect of the map-formation process is the gradual formation of localized receptive fields for each neuron. From the algorithm it is immediately clear, that each receptive field must correspond to some small volume in the high-dimensional input-space, which for each neuron (k,l) is given by the set of all inputs that are closer to the vector $\vec{w}_{kl} = (w_{kl1}, ..., w_{kld})^T$, $d \in [1,900]$ than to any other vector $\vec{w}_{k'l'}$. However, this does not make any statement as to whether these sets correspond to receptive fields, that are also *spatially localized in the two-dimensional receptor surface itself.*

In the following, we will give an analytic derivation that this is indeed the case, and compare the results

of this derivation with the data obtained from the simulation. First we will compute the shift of the centroid of a receptive field under an adaptation step. Substituting (4) into the expression for the centroid $\vec{s}_{kl}$ given in the previous section, expanding, and keeping only the leading order terms in the learning step size ε, we obtain

$$\vec{s}_{kl(t+1)} = \vec{s}_{kl}(t) + \varepsilon(t) h_{rs;kl}(t) \frac{\sum_i r_i}{\sum_i w_{kli(t)}} (\vec{x}_s - \vec{s}_{kl}(t)) + O(\varepsilon^2) \tag{5}$$

Under the assumption of a homogeneous map, i.e. that $\sum_i w_{kli} = const_{(kl)}$ (which holds for maps without topological defects), we can absorb $\sum_i r_i / \sum_i w_{kli}(t)$ into a redefinition of $\varepsilon(t)$, which yields

$$\Delta\vec{s}_{kl}(t+1) = \varepsilon(t)' h_{rs;kl}(t)(\vec{x}_s - \vec{s}_{kl}(t)) \tag{6}$$

if ε is sufficiently small that higher order terms again may be dropped. Therefore, as long as all stimuli have the same shape, the evolution of the centroids $\vec{s}_{kl}$ also follows a Kohonen-type learning rule.

For a further analysis of the development of the receptive fields, we now consider the mean square radius

$$G_{kl} = \sum_i (\vec{x}_i - \vec{s}_{kl})^2 w_{kli} / \sum_i w_{kli} \tag{7}$$

which is a measure of the spatial extent of the receptive field of neuron (k,l). The new value $G_{kl}(t+1)$ after presentation of a tactile stimulus at a location $\vec{x}_s$ can be computed by inserting (4) and (6) into (8). With some calculation, the resulting expression can be simplified to

$$G_{kl}(t+1) = G_{kl}(t) + \frac{\varepsilon(t) h_{rs;kl}(t)}{\sum_i w_{kli}(t)} \sum_i (\vec{x}_i - \vec{s}_{kl})^2 \left(r_i - w_{kli}(t) \frac{\sum_j r_j}{\sum_j w_{klj}(t)} \right). \tag{8}$$

If we introduce the definition

$$\Gamma(\vec{x}_s) = \frac{1}{\sum_i r_i} \sum_i (\vec{x}_i - \vec{x}_s)^2 r_i \tag{9}$$

we obtain from (8) for the change $\Delta G_{kl}(t) = G_{kl}(t+1) - G_{kl}(t)$:

$$\Delta G_{kl} = -\varepsilon h_{rs;kl} \frac{\sum_i r_i}{\sum_i w_{kli}} \{ G_{kl} - \Gamma(\vec{x}_s) - (\vec{x}_s - \vec{s}_{kl})^2 \} + O(\varepsilon^2) \tag{10}$$

The new quantity $\Gamma(\vec{x}_s)$ can be interpreted as the mean square radius of the tactile stimulus at position $\vec{x}_s$. The expectation value of the change ΔG_{kl} in the receptive field radius is then given by the average of (10) over the stimulus probability distribution:

$$E(\Delta G_{kl} | G_{kl})(t) = \int \langle \Delta G_{kl}(t) \rangle P(\vec{x}_s) d^2\vec{x}_s \tag{11}$$

where $\langle \ldots \rangle$ denotes the average over all stimulus shapes and $P(\vec{x}_s)$ is the probability distribution of the positions of the stimulus center. Inserting (10) into (11) we get:

$$\begin{aligned} E(\Delta G_{kl} | G_{kl}) \;=\; & -\frac{\varepsilon}{\sum_i w_{kli}} \int h_{rs;kl} \{ \langle \sum_i r_i \rangle G_{kl} - \langle \Gamma(\vec{x}_s) \sum_i r_i \rangle \\ & - \langle \sum_i r_i \rangle (\vec{x}_s - \vec{s}_{kl})^2 \} P(\vec{x}_s) d^2\vec{x}_s + O(\varepsilon^2) \end{aligned} \tag{12}$$

The expectation value $E(\Delta G_{kl} | G_{kl})$ vanishes (in the limit of small ε), if the receptive field size assumes its equlibrium value. In this case G_{kl} is given by

$$G_{kl} = \frac{\int h_{rs;kl} \langle ([\Gamma(\vec{x}_s) + (\vec{x}_s - \vec{s}_{kl})^2] \cdot \sum_i r_i \rangle P(\vec{x}_s) d^2\vec{x}_s}{\int h_{rs;kl} \langle \sum_i r_i \rangle P(\vec{x}_s) d^2\vec{x}_s} \tag{13}$$

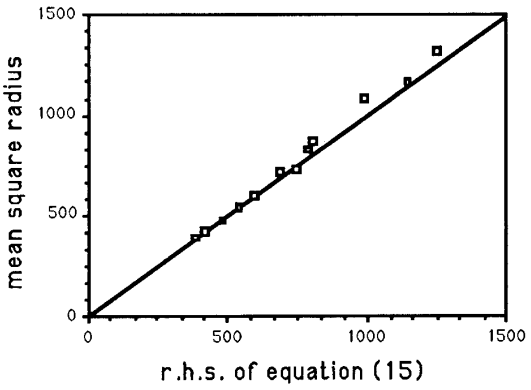

Fig.4: Receptive field radii

If the stimulus distribution varies sufficiently smoothly over the receptor surface, and if each stimulus excites many receptors simultaneously, we can neglect any variations of $\langle \sum_i r_i \rangle$, $\Gamma_{\vec{x}_s}$ and $P_{\vec{x}_s}$ within the integration range where $h_{rs;kl}$ is significant. Then we can simplify (13) to

$$G_{kl} = \langle \Gamma(\vec{x}_s) \rangle + \frac{\int h_{rs;kl}(\vec{x}_s - \vec{s}_{kl})^2 d\vec{x}_s}{\int h_{rs;kl} d\vec{x}_s} \qquad (14)$$

For a simple interpretation of this result, we assume that the mapping from stimulus centroids $\vec{x}_s$ to neuron coordinates (r, s) is "smooth" and conformal (i.e. angle-preserving). In this case we can approximate the integral in equation (14), and we obtain

$$G_{kl} \approx \Gamma + M^{-1}\sigma_h^2 \qquad (15)$$

where

$$\sigma_h^2 = \int h_{rs;mn}\{(r-m)^2 + (s-n)^2\} dmdn / \int h_{rs;nm} dmdn \qquad (16)$$

denotes the mean square radius of the output-function and M is the local magnification factor of the mapping from stimuli centroids $\vec{x}_s$ to neuron coordinates (r, s). Equation (15) essentially shows, that the neurons develop receptive fields whose area (which is proportional to G_{kl}) is the sum of two terms: the first term is essentially the area of a typical stimulus ($\propto \Gamma$) and the second term is essentially the area ($\propto \sigma^2$ of the adjustment zone in the neuron layer, but "projected back" (inverse magnification factor M^{-1}) onto the receptor sheet. Therefore, for predominantly localized tactile stimuli, and a small adjustment zone the neurons will develop localized receptive fields.

Fig.4 compares this theoretical result with data from a simulation (16,384 cells, 784 receptors, $\sigma_r = 0.15$, 6×10^4 steps), where σ_h was slowly varied between $\sigma_h = 100$ and $\sigma_h = 5$. The diagram shows the mean square radius of the receptive field averaged over 2,300 neurons from the center of the neuronal sheet plotted versus the r.h.s of equation (15). The dots represent the results of the simulation and the solid line gives the theoretical relationship. The correspondence is very accurate, except for parameter values leading to large receptive fields, for which edge effects become noticeable.

Fig. 5 illustrates the state of a typical receptive field at the beginning of another simulation (Fig. 5a), after 3,000 iterations (Fig. 5b) and after 10,000 iterations (Fig. 5c) (for this run $\sigma_h = 50 \ldots 5$, $\sigma_p = 0.12$

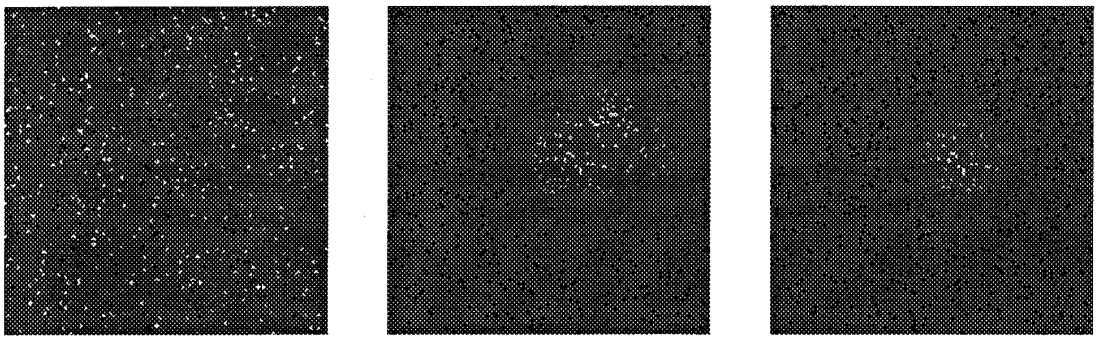

Fig.5: Development of a localized receptive field: (a) (left) initial connectivity, (b) (center) after 3000 and (c) (right) after 10,000 iterations

and $t_{max} = 10^4$). The dot locations mark the positions of the tactile receptors on the receptor surface, while their brightness encodes the strength of their connection to the neuron under consideration. Initially, the field is very diffuse (Fig.5a), but contracts rapidly (Fig.5b), until it finally is well localized (Fig.5c).

5. Topological defects and cells with multiple receptive fields

The results of part 4 are no longer valid for maps with topological defects, where neurons with multiple receptive fields evolve (Fig. 6). Fig. 7 shows the result of a simulation leading to an incompletely ordered map with a multiple representation of parts of the sensory surface. For the sake of a better visualization of the map, the sensory surface (Fig. 7a) was divided into 16 squares filled with different gray-shadings. Fig. 7b shows a view onto the model-cortex. Each pixel represents a neuron (k, l) and its gray-value coincides with the gray-value of the location $\vec{s}_{kl}$ of the receptive field center in Fig. 7a. There are three areas where the representation of the sensory surface is topographically correct, separated by bands of discontinuous shifts. Within this bands the position of the center $\vec{s}_{kl}$ of the receptive field changes rapidly. Fig. 7c shows the (brightness-coded) mean square radius of the receptive fields of the model neurons. Dark shaded areas correspond to specific cells with their radius G_{kl} given by (7), while bright areas code for "unspecific" cells with large G_{kl}, i.e. with either large or multiple receptive fields. A comparison between Fig. 7b and Fig. 7c shows, that the regions of rapid position change coincide with the regions containing unspecific neurons. Cells of this type allow to map distant locations on the receptor surface onto neighboring locations of the cortical sheet without introducing a discontinuity in the response properties of the neuron.

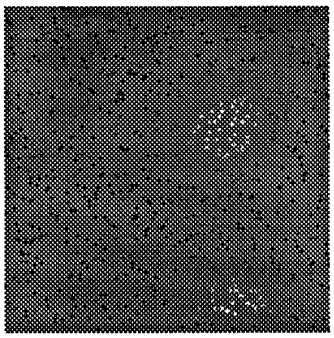

Fig. 6: Double receptive field

Topological defects in maps and cells with multiple receptive fields have both been reported for the somatosensory system [21]. However, we were unable to find reports about neurophysiological experiments adressing the relation of "unspecific" cells to "defects" within the somatotopic representation. Further information about this issue would shed interesting light on the question to what extent a small set of principles, such as those embodied in Kohonen's algorithm, can accurately explain important features of topographic maps within the central nervous system.

Fig. 7: a) Distribution of receptors over the sensory surface, b) distribution of (color-coded) centroids of the neurons, c) color coded mean square radius of the receptive fields

Acknowledgement: This work was supported by the National Center for Supercomputer Applications in Urbana/Champaign. One of the authors (K.Obermayer) was supported by a scholarship from the Boehringer-Ingelheim-Fonds.

6. References:

[1] Obermayer K. et al. (1990), in Proceedings of the ICNC-Conference, Düsseldorf 1990, in press
[2] Lemon R. (1988), Trends in Neur. Sci. (Vol. 11) 11:501
[3] Kaas J.H. et al. (1983), Annu. Rev. Neurosci. 6:325
[4] Sparks D.L., Nelson J.S. (1987), TINS 10:312
[5] Suga N., O'Neill W.E. (1979), Science 206:351
[6] Tunturi A.R. (1952), Am. J. Physiol. 162:489
[7] von der Malsburg C., Willshaw D.J. (1977), PNAS USA 74:5176
[8] Malsburg C. (1973), Kybernetik 14:85
[9] Takeuchi A., Amari S. (1979), Biol. Cybern. 35:63
[10] Amari S. (1980), Bull. Math. Biol. 42:339
[11] Suga N., Jen P.H. (1976), Science 194:542
[12] Blasdel G.G., Salama G. (1986), Nature 321:579
[13] Linsker R. (1986), PNAS 83:8779
[14] Pearson J.C. et al. (1987), J. Neurosci. 12:4209
[15] Kohonen T. (1982a), Biol. Cybern. 43:59
[16] Kohonen T. (1982b), Biol. Cybern. 44:135
[17] Martinetz T. et al. (1988), In: SGAICO-Proceedings "Connectionism in Perspective", Zürich, p. 403
[18] Ritter H., Schulten K. (1986), Biol. Cybern. 54:99
[19] Ritter H., Kohonen T. (1989), Biol. Cybern. 61:241
[20] Merzenich M.M. et al.(1984), J. Comp. Neu. 224:591
[21] Merzenich M.M. et al.(1983), Neurosci. 10, 3:639
[22] Kohonen T., Mäkisara K. (1986), Neural Networks for Computing, AIP Conf. Proc., Snowbird, p. 271