# A Parameter in the Learning Rule of SOM That Incorporates Activation Frequency

Antonio Neme[1,2] and Pedro Miramontes[2]

[1] Universidad Autónoma de la Ciudad de México, México, D.F., Department of Nonlinear Dynamics and Complex Systems, México
neme@nolineal.org.mx
[2] Universidad Nacional Autónoma de México, Facultad de Ciencias, México

**Abstract.** In the traditional self-organizing map (SOM) the best matching unit (BMU) affects other neurons, through the learning rule, as a function of distance. Here, we propose a new parameter in the learning rule so neurons are not only affected by BMU as a function of distance, but as a function of the frequency of activation from both, the BMU and input vectors, to the affected neurons. This frequency parameter allows non radial neighborhoods and the quality of the formed maps is improved with respect to those formed by traditional SOM, as we show by comparing several error measures and five data sets.

## 1 Introduction

Self-organizing map (SOM) is presented as a model of the self-organization of neural connections, what is translated in the ability of the algorithm to produce organization from disorder [1]. One of the main properties of SOM is its ability to preserve topographical relations present in input data in the output map [2], which is a desirable property for data visualization and clustering.

One main feature of the SOM is the ability to transform an incoming signal pattern of arbitrary dimension into a low-dimensional discrete map (usually of dimension one or two) and to adaptively transform data in a topologically ordered fashion [3, 4]. Each input data is mapped to a single neuron in the lattice, that with the closest weight vector to the input data, or best matching unit (BMU). The SOM preserves relationships during training through the neighbourhood function, which establishes the effect of the BMU to any other neuron. Weight neurons are updated accordingly to:

$$w_n(t+1) = w_n(t) + \alpha_n(t) h_n(g,t)(x_i - w_n(t)) \tag{1}$$

Where $\alpha(t)$ is the learning rate at time $t$ and $h_n(g,t)$ is the neighbourhood function from BMU neuron $g$ to neuron $n$ at time $t$. In general, neighbourhood function decreases monotonically as a function of the distance from neuron $g$ to neuron $n$. This decreasing property has been reported to be a necessary condition for convergence [5, 6]. The SOM tries to preserve relationships of input data by starting with a large neighbourhood and reducing the neighbourhood size during

S. Kollias et al. (Eds.): ICANN 2006, Part I, LNCS 4131, pp. 455–463, 2006.

the course of training [4]. It has been reported as well that the learning factor $\alpha$ should be a decreasing function [6].

As pointed out by Ritter [7], SOM and related algorithms share the idea of using a deformable lattice to transform data similarities into spatial relationships. The lattice is deformed by applying learning equation (1) to the neurons in the network. In this work, we propose an additional parameter that quantifies the influence of a BMU $n$ to the neurons in the network as a function of the number of times $n$ affects them as well as the influence of each data vector $m$ as a function of the number of times the BMU for $m$ affects the neurons. This frequency activation parameter allows non radial neighborhood which, as reported in the results, forms better maps, in terms of three error measures.

## 2 Related Work

Altough several modifications have been proposed to the SOM learning rule, they don't reflect, at least to our knowledge, the frequency of activation from other neurons. For example, Lee and Verleyen [8] propose the recursive Fisherman's rule and some hybrids from it that reflects an attenuation of the adaptation as the distance from the BMU to the affected neuron grows. The rules show a non radial neighborhood in the sense that the BMU pulls the direct neighbors and these neighbors pull farther neurons and so on, in a recursive manner.

Campoy and Vicente [9] proposed a residual activity memory for each neuron, so the SOM enlarges its temporal analysis capabilities, whereas one of the first works that incorporated a concept of memory for each neuron was in Chappell and Taylor [10], in which is defined an activation memory for each neuron, in order to define the new active neuron, and a modification of the selection mecanism is presented, so if the memory parameter is high, the previous winner neuron will win again unless another neuron matches very close the input data.

## 3 Frequency Function in the SOM's Learning Rule

In the traditional SOM, the BMU equally affects those neurons within its neighborhood. All neurons at the same distance (for the case of gaussian neighborhood) or inside the hypersphere (for the case of bubble neighborhood) are equally affected. We propose a modification to this scheme that includes a function of the relative frequency a given neuron $n$ is affected by each BMU $k$ or by each input vector $m$.

If during the learning process $n$ is affected by two or more BMU (for one or more input vectors), it will not be affected the same by them (independently from the learning factor): the more a neuron $n$ is affected by $k$, the larger the strenght of its influence. A new parameter, the activation frequency, $\rho_n(k, m)$, that is a function of the number of times a neuron $n$ is affected by BMU $k$ or by input vector $m$ is incorporated to eq. (1). The weight modification rule is now:

$$w_n(t+1) = w_n(t) + \alpha_n(t) h_n(g,t) \rho_n(k,m)(x_i - w_n(t)) \tag{2}$$

In this model, every neuron $n$ maintains a record of the relative frequency by which it has been affected by each BMU $k$, $\Omega_n(k)$, defined as the number of times BMU $k$ has included $n$ in its neighborhood divided by the number of times $n$ has been affected by any BMU ($\sum_{j=1}^{|N|} \Omega_n(j)$, where $N$ is the number of neurons in the network). Also, $n$ has a record of the relative frequency it has been affected by each input vector $m$, $\beta_n(m)$, defined as the number of times vector $m$ has affected, through any BMU, $n$, divided by the number of times $n$ has been affected by any BMU (which is the same as the the number of times it has been affected by any input vector). For the gaussian neighborhood, we have defined $n$ is influenced by $k$ if $h_n(g,t) > 0.3$

Several frequency functions are proposed based on these two quantities and in the distance between BMU $k$ and neuron $n$, $d(n,k)$. The frequency parameter $\rho_n(k,m)$ varies as a function of $\Omega_n(k)$ and $\beta_n(m)$. We have found that $\rho_n(k,m)$ should be monotonic decreasing function with respect to $\Omega_n(k)$, as it is shown in eq. (3)-(6), so the formed maps present a lower error than the formed maps by eq (1).

$$(Rule\ 1) \qquad \rho_n(k,m) = \Omega_n(k) \times \frac{1}{d(n,k)} \tag{3}$$

$$(Rule\ 2) \qquad \rho_n(k,m) = \Omega_n(k) \times \beta_n(m) \times \frac{1}{d(n,k)} \tag{4}$$

$$(Rule\ 3) \qquad \rho_n(k,m) = \frac{1}{1 + e^{-\psi \Omega_n(k) \times \frac{1}{d(n,k)}}} \tag{5}$$

$$(Rule\ 4) \qquad \rho_n(k,m) = \Omega_n(k) \tag{6}$$

As an example of the behaviour of the rules, fig. (1) shows the case for rules (1) and (3). Two neurons, $i$ and $j$, such that $d(k,i) = d(k,j)$ will not be affected the same unless $\Omega_i(k) = \Omega_j(k)$. However, when $d(k,i)$ is large, the difference in the frequency activation woll be small. For low values of $d(k,i)$ the importance of $\Omega_i(k)$ becomes clear.

The proposed activation frequency functions modifies the natural neighborhood of BMU. For example, in fig. (2) it is shown the BMU and the affected neurons in four different time steps for a single data vector. It is observed that non radial and discontinuos neighborhoods are formed, which are not present in the traditiobal scheme. This discontinuity resembles the cortex activity patterns in mammals during several task processing [13].

## 4 Topological Preservation Quantization

To measure topological preservation, three metrics were applied. Those are the topographic error (TE) [11], error quantization (EQ) and preservation of original neighborhoods (VC) [12]. The first is defined as:

$$TE_t = \frac{1}{N} \sum_{k=1}^{N} \eta(x_k), \text{ where } \eta(x_k) = \begin{cases} 1, \text{ BMU and 2nd. BMU non adjacent} \\ 0, \text{ otherwise} \end{cases}$$

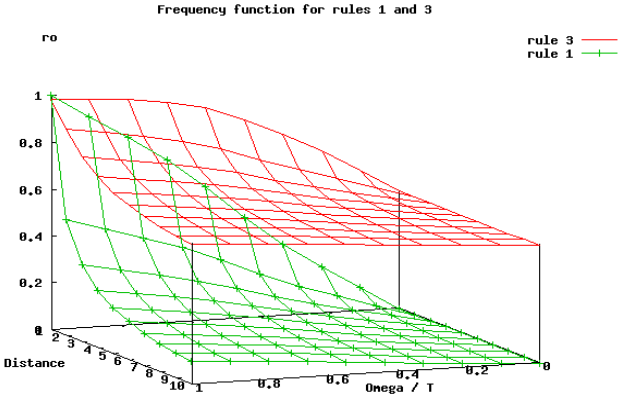


**Fig. 1.** Activation frequency function for rules 1, 3. For two neurons $i, j$ situated at the same distance, $\rho_i(\Omega_i(k)) > \rho_j(\Omega_j(k))$ if $\Omega_i(k) > \Omega_j(k)$.

which is simply the proportion of data vector for which the BMU and second best matching unit are not first-neighbors. The errror quantization is:

$$EQ = \frac{1}{N} \sum_{j=1}^{N} ||x_j - w_j||^2$$

The third metric is based on the neighbohood preservation quantization, which establishes that an input data vector $i$ has $k$ neighbors in its neighborhood $V_k$ in the original space and, if neighborhood preservation is complete, then, the BMU for $i$ has as its first $k$ active neurons those BMU for the data vectors in $V_k$.

$$VC = 1 - \frac{2}{Nk(2N-3k-1)} \sum_{i=1}^{N} \sum_{x_j \in V_k(x_i)} (r(x_i, x_j) - k)$$

where $N$ is the number of data vectors and $r(x_i, x_j)$ is the rank of $x_j$ when data vectors are ordered based on their distance from the data vector $x_i$ after projection. The closer VC is to 1, the better the map is. As VC is a function of $k$, we set a value for $k$ as $\frac{1}{10}$ of the size of each data set.

## 5 Results

The experiments were done in a 10x10 network. Two groups of experiments were done. In the first one, in order to test the sensivity of the proposed frequency functions to the initial values, several thousands of maps were formed($> 10000$), with different initial neighborhood width $h_0$, learning factor $\alpha_0$ and for a different number of epochs. Sensitivity results are presented in subsection 5.1. The second group of experiments establishes the ability of the proposed rules as a good alternative to form maps with a lower error that the traditional learning rule. In this group, several hundreds maps were formed, all with 1000 epochs, $\alpha_0 \in \{0.1, 0.2, 0.5, 0.7, 0.8, 0.9\}$ and $h_0 \in \{8, 9\}$, and with $\alpha_{1000} \in \{0.01, 0.001\}$, for the

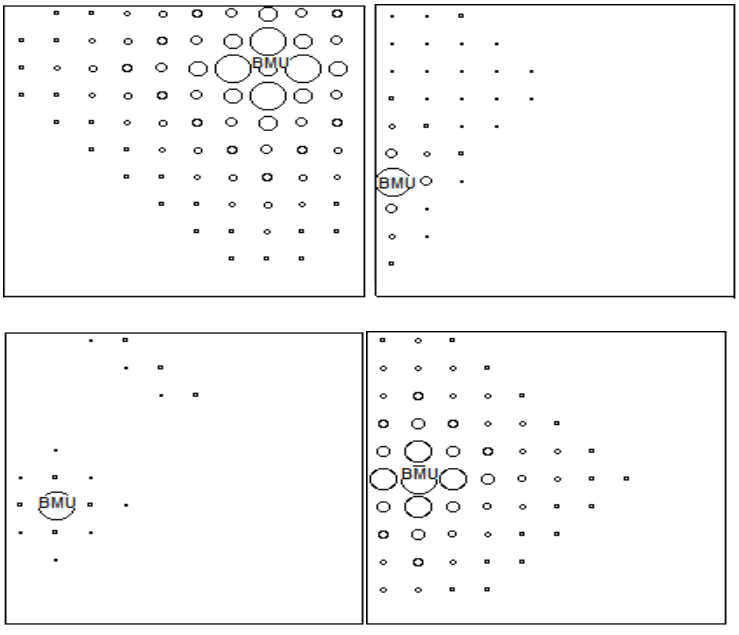


**Fig. 2.** BMU and affected neurons at $t = 1$, $t = 2$, $t = 3$ and $t = 9$ starting at top left for a given input vector from the spiral data set. Size of circumference is proportional to $\rho_i(k)$. In $t = 3$, there is a discontinuity in the area on influence for the BMU and for $t = 2$ there is a non radial neighborhood.

self-organizing stage, with an exponential decrease of the intermediate values for those parameters. In the second stage, convergence, the number of epochs was 10000, with initial values of $\alpha = 0.05$ and $h = 3$, exponentially decreasing until the final values $\alpha = 0.0001$ and $h = 0$. Subsection 5.2 shows convergence results.

### 5.1 Sensitivity

To test sensitivity to the initial conditions, SOM with several parameters were formed. Neighborhood width, $h_0$, was placed between 1 and 10, whereas the initial learning factor, $\alpha_0$, is in the close interval $[0.001, 1.0]$ and the number of epochs is situated between 1 and 130. The final values for neighborhood width is 1 and for the learning parameter is 0.001. Altough the number of maps may be insufficient to cover the whole range of combinations for the former parameters and the choises may be arbitrary, they cover a wide range of combinations.

For each one of the five data sets and for each set of values ($\alpha_0$ and $h_0$), a SOM were formed by the traditional SOM rule and by SOMs with the activation frequency functions described in (3) -(6), for both, bubble and gaussian neighborhood. Two data sets are bidimensional, one is five- dimensional (iris data set), one is 34-dimensional (ionosphere data set) and one is 64-dimensional (codon usage data set).

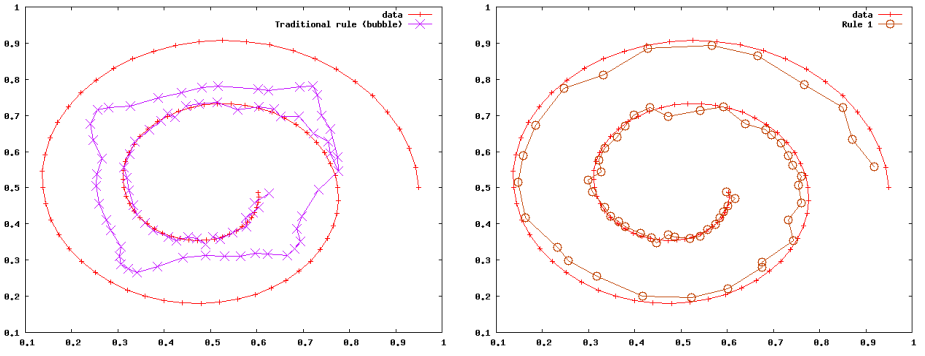


**Fig. 3.** Spiral data set. It is shown the SOM approximation for both, traditional rule (left) and for rule 1 (right, eq. (3)) after 10 epochs.

**Table 1.** Average TE, EQ and VC for the spiral and Henón data sets over all formed maps for the proposed frequency functions. Error is presented in pairs: (bubble neighborhood, gaussian neighborhood).

| Rule | Spiral | | | Henón | | |
|---|---|---|---|---|---|---|
| | TE | EQ | VC | TE | EQ | VC |
| Trad. | (0.17, 0.17) | (0.019, 0.018) | (0.71, 0.6) | (0.17, 0.29) | (0.044, 0.041) | (0.83, 0.73) |
| Rule 1 | (0.107, 0.23) | (0.010, 0.011) | (0.64, 0.66) | (0.21, 0.22) | (0.031, 0.03) | (0.9, 0.75) |
| Rule 2 | (0.106, .24) | (0.013, 0.013) | (0.64, 0.63) | (0.19, 0.31) | (0.028, 0.012) | (0.91, 0.89) |
| Rule 3 | (0.104, 0.261) | (0.012, 0.011) | (0.66, 0.62) | (0.173, 0.2) | (0.025, 0.016) | (0.89, 0.89) |
| Rule 4 | (0.102, 0.21) | (0.011, 0.015) | (0.67, 0.66) | (0.176, 0.21) | (0.026, 0.018) | (0.89, 0.89) |

**Table 2.** Average TE, EQ and VC for the iris and ionosphere data sets over all formed maps for the proposed frequency functions. Error is presented in pairs: (bubble neighborhood, gaussian neighborhood).

| Rule | Iris | | | Ionosphere | | |
|---|---|---|---|---|---|---|
| | TE | EQ | VC | TE | EQ | VC |
| Trad. | (0.241, 0.21) | (0.0673, 0.08) | (0.68, 0.715) | (0.17, 0.196) | (0.08, 0.061) | (0.73, 0.74) |
| Rule 1 | (0.206, 0.282) | (0.061, 0.079) | (0.83, 0.71) | (0.113, 0.195) | (0.081, 0.054) | (0.83, 0.75) |
| Rule 2 | (0.206, 0.272) | (0.062, 0.06) | (0.82, 0.84) | (0.11, 0.195) | (0.082, 0.055) | (0.83, 0.91) |
| Rule 3 | (0.235, 0.203) | (0.07, 0.053) | (0.8, 0.82) | (0.165, 0.145) | (0.082, 0.058) | (0.87, 0.89) |
| Rule 4 | (0.228, 0.21) | (0.06, 0.054) | (0.82, 0.88) | (0.165, 0.169) | (0.083, 0.0.69) | (0.83, 0.85) |

**Table 3.** Average TE, EQ and VC for the codon usage data set over all formed maps for the proposed frequency functions. Error is presented in pairs: (bubble neighborhood, gaussian neighborhood).

| Rule | TE | EQ | VC |
|---|---|---|---|
| Trad. | (0.175, 0.12) | (0.24, 0.14) | (0.72, 0.72) |
| Rule 1 | (0.103, 0.1) | (0.22, 0.17) | (0.79, 0.77) |
| Rule 2 | (0.1, 0.098) | (0.21, 0.17) | (0.82, 0.76) |
| Rule 3 | (0.167, 0.256) | (0.29, 0.289) | (0.83, 0.82) |
| Rule 4 | (0.167, 0.2) | (0.19, 0.17) | (0.82, 0.71) |

As the analysis of data is extensive, only some results are presented (the whole set is available from the authors). In tables (1) - (3), the average error for all maps formed by the proposed frequency functions, as well as for those generated by traditional SOM with both, bubble and gaussian neighborhood, including all initial conditions for $\alpha_0$, $h_0$ and the number of iterations, are presented, for each one of the five data sets.

Fig. (4) shows the TE as a function of number of epochs for the traditional SOM and for SOM's with the frequency functions proposed in eqs. (3) -(6) for four of the data sets. It can be seen that the maps formed by the proposed rules are less sensitive to initial conditions that those formed by eq. (1). From fig. (3), it is clear that for the spiral data set the proposed functions form maps that folds more accurately to data than those formed by the traditional SOM.

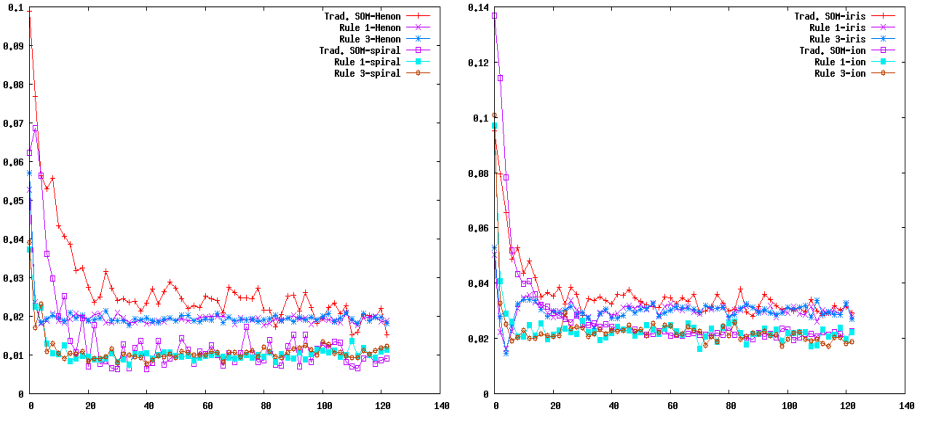


**Fig. 4.** TE for rules 1 and 3 and for the traditional learning rules as a function of the number of iterations for the spiral, Henón, iris and ionosphere data sets

### 5.2 Convergence

Once the sensitivity was analyzed, the properties of the proposed rules for maps suitable for data visualization (low error) were studied. In tables (4) and (5) it is shown the average error for several maps ($> 500$) for a larger number of epochs and two-stages differentiation, in contrast to what was done in the previous subsection. It is observed that the error measures are, in general, lower for those maps formed with the proposed rules. Also, a map formed by traditional rule and another formed by rule 1 are shown in fig.3.

**Table 4.** Average TE, EQ and VC obtained after two stages of trainning of the SOM, for the spiral and Henón data sets over all formed maps for the proposed frequency functions. Error is presented in pairs: (bubble neighborhood, gaussian neighborhood).

| Rule | Spiral | | | Henón | | |
|---|---|---|---|---|---|---|
| | TE | EQ | VC | TE | EQ | VC |
| Trad. | (0.01, 0.009) | (0.0001, 0.0001) | (0.92, 0.93) | (0.018, 0.016) | (0.0015, 0.0012) | (0.91, 0.93) |
| Rule 1 | (0.01, 0.008) | (0.00003, 0.00002) | (0.94, 0.95) | (0.013, 0.005) | (0.001, 0.0008) | (0.9, 0.91) |
| Rule 2 | (0.02, 0.018) | (0.00002, 0.00001) | (0.95, 0.96) | (0.012, 0.01) | (0.001, 0.0004) | (0.9, 0.92) |
| Rule 3 | (0.015, 0.019) | (0.00001, 0.00001) | (0.95, 0.95) | (0.013, 0.014) | (0.0006, 0.0004) | (0.89, 0.92) |
| Rule 4 | (0.018, 0.012) | (0.00002, 0.00001) | (0.96, 0.96) | (0.014, 0.012) | (0.001, 0.0003) | (0.92, 0.94) |

**Table 5.** Average TE, EQ and VC obtained after two stages of trainning of the SOM, for the iris and codon usage data sets over all formed maps for the proposed frequency functions. Error is presented in pairs: (bubble neighborhood, gaussian neighborhood).

| Rule | Iris | | | Codon usage | | |
|---|---|---|---|---|---|---|
| | TE | EQ | VC | TE | EQ | VC |
| Trad. | (0.15, 0.105) | (0.05, 0.048) | (0.85, 0.85) | (0.137, 0.231) | (0.26, 0.25) | (0.79, 0.77) |
| Rule 1 | (0.152, 0.1) | (0.035, 0.035) | (0.84,0.85) | (0.072, 0.045) | (0.25, 0.22) | (0.72, 0.79) |
| Rule 2 | (0.14, 0.1) | (0.04, 0.038) | (0.91, 0.91) | (0.088, 0.069) | (0.27, 0.08) | (0.88, 0.87) |
| Rule 3 | (0.129, 0.12) | (0.04, 0.038) | (0.89, 0.91) | (0.085, 0.12) | (0.29, 0.29) | (0.8, 0.89) |
| Rule 4 | (0.095, 0.114) | (0.002, 0.001) | (0.93, 0.92) | (0.01, 0.012) | (0.23, 0.22) | (0.94, 0.96) |

## 6 Discussion and Conclusions

An activation frequency parameter for the weight update equation is proposed. This parameter is a function of the activation frequency from a given BMU as

well as from the relative frequency of influence of an input vector, through any BMU, to neurons within its neighborhood. Distance between BMU and neurons may also important for this parameter. This parameters add some differential influence between BMU and equally distant neurons, driven by how much those neurons are being affected by other BMUs.

The fact that the proposed rules form non radial neighborhoods gives biological plausibility, due to the fact that a neuron affects differentially other neurons not only based on their distance but in the frequency with which one of them affects the other.

Several experiments show that the error measures in the maps formed with some of the proposed activation frequency functions are lower than those formed by the traditional SOM, and, for the two-dimensional data sets, it is observed that the formed maps fold more accurately to data than those formed by the traditional SOM rule. However, we believe this results could be improved by identifying other activation frequency schemes, such as simmulated annealing. It could be of interest to study the mathematical properties of those functions as they seem to be important for the map formation.

## References


1. Cottrell, M. Fort, J.C., Pagés, G. Theoretical aspects of the SOM algorithm. Neurocomputing **21** (1998) 119-138.
2. Kirk, J. Zurada, J. A two-stage algorithm for improved topography preservation in self-organizing maps. Int. Con. on Sys., Man and Cyb. **4** (2000) 2527-2532.
3. Haykin, S. Neural Networks, a comprehensive foundation. 2nd. ed. Prentice Hall. 1999.
4. Kohonen, T. Self-Organizing maps. 3rd. ed. Springer-Verlag. 2000.
5. Flanagan, J. Sufficiente conditions for self-organization in the SOM with a decreasing neighborhood function of any width. Conf. of Art. Neural Networks. Conf. pub. No. 470 (1999)
6. Erwin, Obermayer, K. Schulten, K. Self-organizing maps: Ordering, convergence properties and energy functions. Biol. Cyb. **67** (1992) 47-55
7. Ritter, H. Self-Organizing Maps on non-euclidean Spaces Kohonen Maps, 97-108, Eds.: E. Oja and S. Kaski, 1999
8. Lee, J., Verleysen, M. Self-organizing maps with recursive neighborhood adaption. Neural Networks **15** (2002) 993-1003
9. Campoy, P., Vicente, C. Residual Activity in the Neurons Allows SOMs to Learn Temporal Order LNCS 3696 (2005) 379-384
10. Chappell, G., Taylor, J. The temporal Kohonen map. Neural Networks **6** (1993) 441-445
11. Kiviluoto, K. Topology preservation in Self-Organizing maps. Proc. ICNN96, IEEE Int. Conf. on Neural Networks.
12. Venna, J., Kaski, S. Neighborhood preservation in nonlinear projection methods: An experimental study.
13. Lamm, C. Leodolter, U., Moser, E., Bauer, H. Evidence for premotor cortex activity during dynamic visuospatial imaginery. Neuroimage **14** (2001) 268-283