

# Top-down control of learning in biological self-organizing maps

Thomas Trappenberg<sup>1,2</sup>, Pitoyo Hartono<sup>3</sup>, and Douglas Rasmusson<sup>1,4</sup>

<sup>1</sup> Dalhousie Neuroscience Institute, Halifax, NS, Canada

<sup>2</sup> Dalhousie University, Faculty of Computer Science, Halifax, Canada

<sup>3</sup> Future University, Hakodate Japan

<sup>4</sup> Dalhousie University, Department of Physiology and Biophysics, Halifax, Canada

**Abstract.** This paper discusses biological aspects of self-organising maps (SOMs) which includes a brief review of neurophysiological findings and classical models of neurophysiological SOMs. We then discuss some simulation studies on the role of topographic map representation for training mapping networks and on top-down control of map plasticity.

## 1 Introduction

Experience driven development and learning is of central importance in the brain, and self-organizing maps (SOMs) have long been a principle model for such systems (Kohonen, 1982; Willshaw & von der Malsburg, 1976). Work has continued in refining models for better matches with experiments and enhanced theoretical treatments (e.g. Tanaka, 1991). While map plasticity is generally celebrated as physical evidence of learning, learning-theoretic studies of the role of SOMs for learning on a cognitive level have been sparse.

This paper provides some discussions on modelling biological SOMs. We briefly point to some of the basic biological findings and literature on cortical maps, with some concentration on issues of top-down control of learning. We then review recent results by Zhou and Merzenich (2007) which demonstrate enormous changes in tonotopic maps of adult animals and the importance of behavioral relevant learning for map formation. The following brief review of biological models discusses the differences, and the relations, between some classical papers (Willshaw & von der Malsburg, 1976; Amari, 1977; Kohonen, 1982). Finally, we show that topographic map representations can assist learning of feed forward networks and address some questions of top-down control of map formation.

## 2 Topographic maps in the brain

Topographic maps are a common feature of cerebral cortical areas that process sensory information. The features that are mapped in the various sensory cortices differ somewhat, being mostly spatial in visual cortex (with visual space corresponding to position on the retina) and somatosensory cortex (position on the body surface) as compared to sound frequency in auditory cortex which results from a transformation of frequency to

position on the cochlea. An important characteristic of all maps and in all species studied is that there is significant distortion in the maps such that the parts of the sensory sheet that are biologically most important to the animal have a larger representational area. This is often called a magnification factor and results in the well-known distorted picture of the body representation in somatosensory cortex termed the homunculus in which the finger and lip representations are proportionally much larger than on the body itself. This enlargement of specific subregions in some animals has also provided important model systems that have facilitated the study of these maps, for example the high-frequency regions of auditory cortex of the bat that are involved in echolocation or the whisker region of rodent somatosensory cortex in which anatomically discrete barrels have a one-to-one correspondence to the large whiskers on the animals snout.

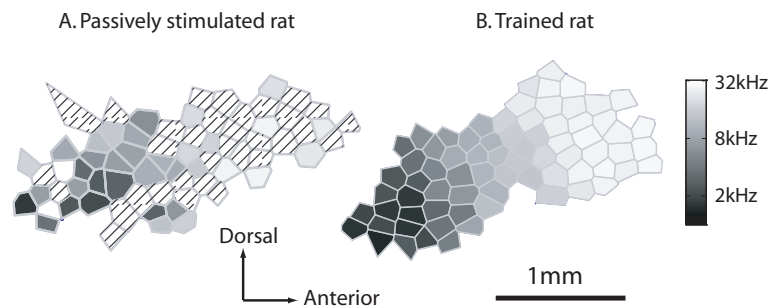
The original formation of topographic maps is largely a developmental problem involving chemical signaling between neurons in addition to network activity that results from sensory experience. While we can study how the framework of topographic maps can be laid down in the absence of neural activity (by applying drugs that block the electrical signals), it is impossible to do the reverse. Consequently, the role of network activity can best be studied in the more mature animal, after the developmental processes are complete, by perturbing some aspect of the map and observing how the mapping functions react. One approach is to remove one part of the input, say by denervating one part of the body or producing a small lesion in the cochlea or retina. A common finding of such experiments is that the deprived regions of cortex do not remain quiescent, but gradually begins to respond to stimulation of adjacent parts of the sensory sheet (Rasmusson, 1982; Robertson & Irvine, 1989; Kaas et al., 1990); it is often said that the map reorganizes itself. A related technique is to directly lesion part of the cortical map; again a consistent finding is that the surrounding regions gradually begin to respond to adjacent sensory inputs (Winship & Murphy, 2008). A second approach is to ask whether the map can be changed by increased neural activity, either by enhanced sensory input or by direct stimulation of a small part of the cortex. Here the usual finding is that the representational area enlarges with progressive stimulation, whether in the somatosensory cortex with finger stimulation (Jenkins, Merzenich, Ochs, Allard, & Guic-Robles, 1990), or the auditory system with repeated presentations of specific sound frequencies (Recanzone, Schreiner, & Merzenich, 1993).

Before the precise mechanisms underlying these reorganizations can be uncovered, it is necessary to ask where exactly does plasticity occur, that is, where are the synaptic connections that are being changed during reorganization. The cortical responses are the result of a chain of three or more neurons proceeding from the sensory sheet to the cortex; at each level of processing there is a similar topographic map. It is possible, therefore, that plasticity occurs at one of these preceding levels and that the cortical response is simply a reflection of reorganization at these earlier stages. The cortex was the original focus of most early studies largely because it is more easily accessible. In fact, recent studies have demonstrated some degree of reorganization in the thalamus, which contains the final projection neurons that go to the cortex. The problem that arises here is that each of these sensory pathways is not only a bottom-up system with projections only going upstream. Feedback from the cortex to the thalamus has long been known and in fact outnumbers thalamocortical fibers by 10 to 1 (Steriade, Jones,

& McCormick, 1997). A new thalamic response might then be due to reorganization within the cortex relayed back to the thalamus or to plasticity within the thalamus itself.

### 3 Goal-directed learning controls topographic map plasticity

While most of the work on the formation of topographic maps is targeted towards early development, we concentrate on the ongoing refinements of cortical maps during adult plasticity. A nice demonstration of representational plasticity in adult mammals was recently given by Zhou and Merzenich (2007) and is shown in Fig. 1. They raised rat pups in a noisy environment that severely impaired the development of tonotopicity in the primary auditory cortex (A1), which lasted into adulthood. An example of such a map in an adult rat is shown in Fig. 1A. The hatched areas represent areas with neurons that showed abnormal, poor frequency tuning. These rats did not recover a normal tonotopic representation in A1 even though they were stimulated in adulthood with sounds of different frequencies. However, when the same sound patterns were used to train the rats in a discrimination task in order to get food reward, the developmentally degraded rats were able to recover a normal tonotopic map as shown in Fig. 1B. This example demonstrates that goal directed learning can influence map plasticity. A purely bottom-up stimulation of networks, which is the common focus in SOM modelling, is certainly not enough to explain the results by Zhou and Merzenich (2007). Before returning to questions on modelling these results, we first review classical SOM models in the next section.



**Fig. 1.** (A) Frequency map in A1 of a rat that was developmentally degraded by being raised in noisy environments. The hatched areas contain neurons with poor frequency tuning. (B) Tonotopic map of an adult rat that recovered with training on a frequency discrimination task. [Data courtesy of Xiaoming Zhou and Micheal Merzenich.]

### 4 Basic cortical map models

A model for activity-dependent, self-organized development of topographic cortical maps was studied over 30 years ago by Willshaw and von der Malsburg (1976). They

considered a two dimensional cortical sheet as illustrated in Figure 2A. The states of the nodes in this cortical sheet represent activation rates in localized neuronal populations with a leaky integrator dynamics. Population models were derived by Wilson and Cowan (1972) with separate excitatory and inhibitory populations, and subsequently studied with centre-surround architectures (Wilson & Cowan, 1973). Amari (1977) abstracted these models further in two important ways, by combining the excitatory and inhibitory populations, and by using a continuous descriptions of the neural sheet. The dynamics of the neural sheet described by neural field equations,

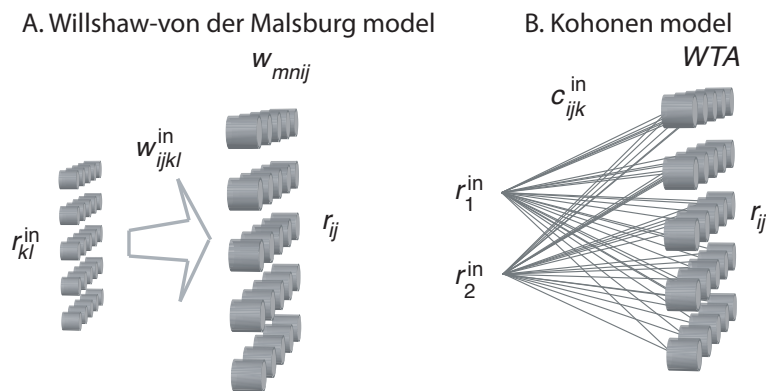
$$\tau \frac{\partial \mathbf{u}(\mathbf{x}, t)}{\partial t} = -\mathbf{u}(\mathbf{x}, t) + \int_{\mathbf{y}} \mathbf{w}(\mathbf{x}, \mathbf{y}) \mathbf{r}(\mathbf{y}, t) d\mathbf{y} + I^{\text{ext}}(\mathbf{x}, t) \quad (1)$$

$$\mathbf{r}(\mathbf{x}, t) = g(\mathbf{u}(\mathbf{x}, t)), \quad (2)$$

of which the equations by Willshaw and von der Malsburg are an discretized example. The general activation function  $g$ , which often has the form of a sigmoid, relates the internal state variable,  $u$ , to an externally observable rate variable,  $r$ . The integration kernel  $w$  describes the interaction within the cortical sheet (recurrent network), and has typically the characteristics of short-distance excitation and long-distance inhibition. For example, with a cortical sheet on a torus to minimize boundary effects, a common interaction kernel is a shifted Gaussian,

$$\mathbf{w}(|x - y|) = A_w \left( e^{-(x-y)^2/4\sigma^2} - C \right), \quad (3)$$

where  $A_w$ ,  $\sigma$ , and  $C$  are constants, and  $x$  and  $y$  are the locations of the interacting neural populations. This kernel can be learned through Hebbian learning (Stringer, Trappenberg, Rolls, & Araujo, 2002), but most studies of self-organizing maps consider this interaction kernel as fixed when learning the afferent (input) weights to this neural sheet. Neural field models are widely used to model reaction times in cognitive neuroscience (e.g. Trappenberg, 2009).



**Fig. 2.** Common architecture of self-organizing maps.

The activity-dependent self-organization of the inputs to the neural sheet is also based on Hebbian learning as outlined in (Willshaw & von der Malsburg, 1976). Briefly, the neural sheet implements a maximum-likelihood estimator (Wu, Nakahara, & Amari, 2001), where the estimate is given by a localized activity packet that develops through competitive neural field dynamics (Amari, 1977). This process is approximated by a winner-takes-all mechanism and a neighborhood function in Kohonen's formulation (Kohonen, 1982). The neural activity in the activity packet is then correlated with the input pattern through Hebbian learning, with most active nodes in the centre of the activity packet strengthened more than nodes in the periphery of the activity packet.

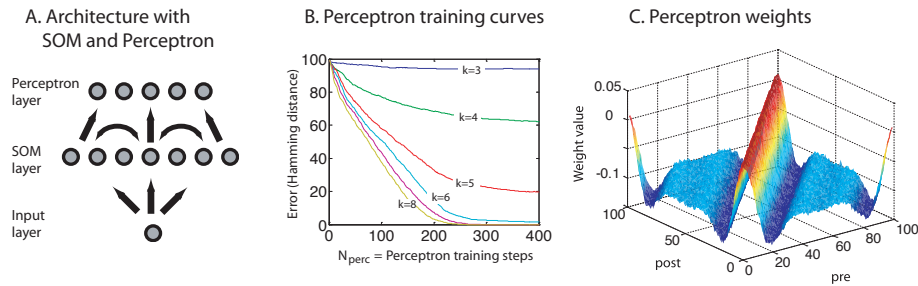
A final approximation in Kohonen's formulation of SOMs is the replacement of distributed activity patterns for feature values in the input sheet with a direct representation of feature values in an input vector (Figure 2B). Note that we discuss here the case where the dimensionality of the feature space is the same as the dimensionality of the feature map, although SOMs can be used to map high dimensional feature spaces into lower dimensional maps (Obermeyer, Ritter, & Schulten, 1990). The activity packet that develops through the neural dynamics determines the tuning curves in neural sheets (see Trappenberg, 2009 for an illustration). Therefore, the radial basis function nodes in Kohonen's network can be seen as a model of unimodal tuning curves, and the input weights in Kohonen's formulation correspond to the centre of the tuning curve. This centre of the tuning curve is often called the preferred response (orientation in V1) of a neuron.

Kohonen's formulation, while more abstract, has several advantages to the model by Willshaw and von der Malsburg. Integrating the neural field dynamics is usually very time consuming, and the replacement of neural field representations with tuning curves reduces the computational burden dramatically. Furthermore, topographic organizations are easily observable when plotting centers of tuning curves, which are directly modelled in Kohonen's model. This information has first to be decoded in the neural field formulation. However, the neural field dynamics need to be considered when studying the dynamics of responses in neural tissue. Also, it is not clear how multiple concurrent objects in a sensory field can be simulated in Kohonen's model.

## **5 The importance of topographic representations in learning mapping functions**

To demonstrate how topographic representations can help train mapping networks, we provide here a basic example of a system that must learn to solve a decoding task. The decoding task is to convert an analog signal to a digital representation (binary output vector). The system includes SOM in Kohonen's formulation, which receives the analog input, and a simple perceptron to map SOM states to the desired local, binary representation as shown in Figure 3A. The SOM was then trained on random examples of analog input signals for a specific number of training steps before using this representation to train the perceptron.

Such a decoding task can, in principle, be solved with a perceptron with one hidden layer since such machine learning systems are universal approximators (Hornik, Stinchcombe, & White, 1989). Thus, why should we complicate the system with a SOM layer?



**Fig. 3.** Example of a decoder resulting from training a perceptron on self-organized representations. (A) Illustration of the architecture. (B) Learning curves for the perceptron with different number of learning steps in the SOM. (C) Weight values of the perceptron with well organized SOM and trained perceptron.

To demonstrate this we use a simple perceptron (no hidden layers) as the output module of the network. Avoiding hidden layers makes it harder to solve the mapping problem. Also, a simple perceptron can be viewed as a reinforcement learner, which is the desired architecture to model the results of Zhou and Merzenich (2007).

Training curves for the perceptron are shown Figure 3B. The different curves correspond to different *number of steps* ( $= 2^k$ ) of training the SOM before training the perceptron. We found that some organization was necessary for the perceptron to learn, and that further self-organization helped the perceptron to reduce its training time to reach a desired accuracy. We found similar results in more complicated mapping tasks in which multilayer perceptrons were used as the output layer (not shown here).

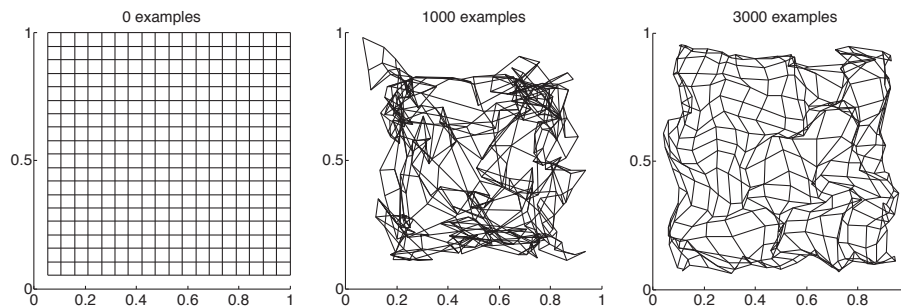
Given enough training steps for SOM learning and perceptron learning, the decoder was able to accurately learn the decoding task. The weights of the perceptron after training are shown in Figure 3C. The pattern of weights has a centre-surround organization. This is a consequence of extended activity packets in the SOM layer and the forced, localized output representation. As some of the neighboring nodes are activated in the SOM layer, the output layer learns to suppress these activation through centre-surround inhibition.

## 6 Modelling top-down control of SOM plasticity

The above example demonstrate the usefulness of topographic organizations in intermediate layers of mapping networks. The next question is how top-down control of map plasticity can be implemented. One possible direction, which we are currently exploring, is using the error signal of the output layer as modulation signal for learning.

Two key components of the behavioral task by Zhou and Merzenich are attention and motivation. In the mammalian brain, both of these functions involve analysis in systems outside the traditional sensory systems and would contribute to sensory plasticity in a top-down manner. One example is the basal forebrain neurons that use acetylcholine as a neurotransmitter to project to many regions in cortex. Acetylcholine has

been implicated in many models of plasticity (Rasmusson, 2000) and these neurons are preferentially activated during tasks with high attentional demands (Sarter, Bruno, & Givens, 2003). The prefrontal cortex is one higher cognitive region that regulates activity of the basal forebrain neurons and could thus provide modulatory acetylcholine to the sensory cortices in such a top-down manner (Rasmusson, Smith, & Semba, 2007).



**Fig. 4.** Examples of organization in a Kohonen map. The map is started from a perfect organization and then trained for 1000 training steps in the presence of dominating noise which deteriorates the map considerably. After this, the map is further trained with an enhanced input signal relative to noise, which helps to recover the map.

The following simulations of experiments similar to the ones by Zhou and Merzenich were done with a standard Kohonen model. The activation of a node  $i$ ,  $\mathbf{r}_i$ , was determined by the tuning curve,  $t(\mathbf{x}, c_i)$ , of the neuron with preferred feature  $c_i$

$$\mathbf{r}_i = \eta + bt(\mathbf{x}, c_i), \quad (4)$$

in addition to some uniformly distributed noise,  $\eta$ . The strength value  $b$  was altered in the experiments to simulate attentional processes. To represent the initial noise-reared case, we simulated the map development with an environment in which the response of each node in the SOM layer was noise-dominated by setting  $b = 0.1$ . With this parameter, no map development was achieved. Similar, even if we start the system with a perfect topographic organization (left graph in Figure 4), the map deteriorates with training (middle graph).

After these initial 1000 training steps, we simulate the effect of top-down control resulting from goal-directed training in the experiments in two ways. While standard procedures in SOM development call for a reduction of the learning rate to stabilize learning (stability over plasticity), and to narrow the neighborhood function to allow better local organizations after an initial global phase, we keep the learning rate and width of the neighborhood function constant throughout the simulations. This can be seen as an increase in the case when the parameters are lowered with time. There are several possible explanations for ongoing adult plasticity. For example, it is possible that the apparent slowdown in adult plasticity might only be a reflection of the dynamic reinforcement of existing representations after the brain has developed a model that can

sufficiently model world states. But it is also possible that neural plasticity is actively modulated. The current model does not distinguish between these possible sources of adult plasticity. The second important ingredient is the boosting of the signal relative to noise by using  $b = 10$ . This reflects the outcome of goal-directed learning via motivation and attention that facilitates the enhancement of the relevant signals. With these changes, the maps starts to organize with continuous SOM learning. After 2000 further training steps, the map recovered dramatically, as shown in the right graph of Figure 4.

## 7 Conclusion

We discussed biological SOMs with a focus on ongoing adult plasticity and top-down control. New experimental findings have clearly demonstrated that there can be considerable changes in cortical maps resulting from goal-directed learning, and that attentional processes are important for map formation.

The above discussions provide some background and simulations with a basic model. In particular, we simulated two factors which seem necessary to explain the results of experiments by Zhou and Merzenich. These two factors are the continuation of sensible plasticity thresholds, and the enhancement of task-relevant activity. The demonstration in this paper leave, of course, many open questions. For example, the differences of the contributing factor should be studied in more detail to guide further experimental work. Also, the experiments by Zhou and Merzenich not only show a deficit in tonotopic map organization, but also in the development of tuning curves with clear preferred frequency tuning. Such details can only be studied with more detailed models such as the Willshaw-von der Malsburg model. A better knowledge of such learning mechanisms could ultimately help to enhance teaching methods or to develop effective pharmacological interventions to promote recovery after brain injuries.

## Acknowledgment

This research was supported by NSERC (Canada). TT would like to thank Takayuki Nakata for inspiring discussions while at Future University, Hakodate.

## References

- Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27, 77–87.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359 - 366.
- Jenkins, W., Merzenich, M., Ochs, M., Allard, T., & Guic-Robles, E. (1990). Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. *J. Neurophysiol.*, 63, 82-104.
- Kaas, J., Krubitzer, L., Chino, Y., Langston, A., Polley, E., & Blair, N. (1990). Reorganization of retinotopic cortical maps in adult mammals after lesions of the retina. *Science*, 248, 229-231.



- Kohonen. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59-69.
- Obermeyer, K., Ritter, H., & Schulten, K. (1990). A principle for the formation of the spatial structure of cortical feature maps. *Proc. Natl. Acad. Sci., USA*, 87, 8345-8349.
- Rasmusson, D. (1982). Reorganization of raccoon somatosensory cortex following removal of the fifth digit. *J. Comp. Neurol.*, 205, 313-326.
- Rasmusson, D. (2000). The role of acetylcholine in cortical synaptic plasticity. *Behav. Brain Res.*, 115, 205-218.
- Rasmusson, D., Smith, S., & Semba, K. (2007). Inactivation of prefrontal cortex abolishes cortical acetylcholine release evoked by sensory or sensory pathway stimulation in the rat. *Neuroscience*, 149, 232-241.
- Recanzone, G., Schreiner, C., & Merzenich, M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.*, 13, 87-103.
- Robertson, D., & Irvine, D. (1989). Plasticity of frequency organization in auditory cortex of guinea pigs with partial unilateral deafness. *J. Comp. Neurol.*, 282, 456-471.
- Sarter, M., Bruno, J., & Givens, B. (2003). Attentional functions of cortical acetylcholine inputs: what does it mean for learning and memory? *Neurobiol. Learning Mem.*, 80, 245-256.
- Steriade, M., Jones, E., & McCormick, D. (1997). *Thalamus*. Elsevier, Amsterdam.
- Stringer, S., Trappenberg, T., Rolls, E., & Araujo, I. (2002). Self-organising continuous attractor networks and path integration: One-dimensional models of head direction cells. *Network: Computation in Neural Systems*, 13, 217-242.
- Tanaka, S. (1991). Theory of ocular dominance column formation. *Biological Cybernetics*, 64, 263-272.
- Trappenberg, T. (2009). Computational modelling in behavioural neuroscience: Closing the gap between neurophysiology and behaviour. In D. H. E. Mavritsaki (Ed.), (chap. Decision making and population decoding with strongly inhibitory neural field). Psychology Press, London.
- Willshaw, D. J., & von der Malsburg, C. (1976). How patterned neural connexions can be set up by self-organisation. *Proc Roy Soc B*, 194, 431-445.
- Wilson, H., & Cowan, J. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, 12, 1-24.
- Wilson, H., & Cowan, J. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13, 55.
- Winship, I., & Murphy, T. (2008). In vivo calcium imaging reveals functional rewiring of single somatosensory neurons after stroke. *J. Neurosci.*, 28, 6592-6606.
- Wu, S., Nakahara, H., & Amari, S.-I. (2001). Population coding with correlation and an unfaithful model. *Neural Computation*, 13, 775 - 797.
- Zhou, X., & Merzenich, M. M. (2007). Intensive training in adults refines all representations degraded in an early postnatal critical period. *Proc. Natl. Acad. Sci.*, 104, 4423-4428.