Modeling Cortical Plasticity Based on Adapting Lateral Interaction

Joseph Sirosh and Risto Miikkulainen
Department of Computer Sciences
The University of Texas at Austin, Austin, TX–78712.
email: sirosh,risto@cs.utexas.edu

Abstract

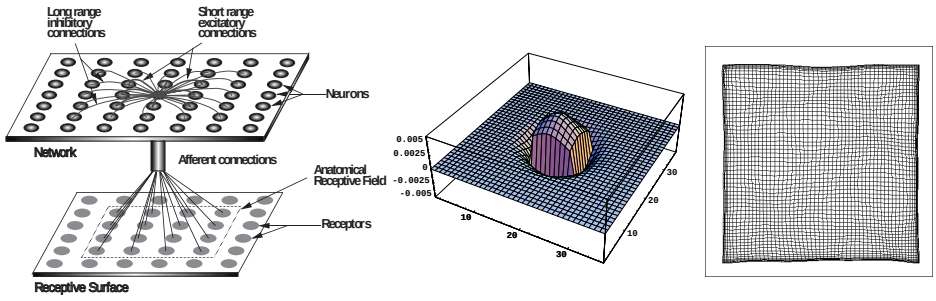
A neural network model called LISSOM for the cooperative self-organization of afferent and lateral connections in cortical maps is applied to modeling cortical plasticity. After self-organization, the LISSOM maps are in a dynamic equilibrium with the input, and reorganize like the cortex in response to simulated cortical lesions and intracortical microstimulation. The model predicts that adapting lateral interactions are fundamental to cortical reorganization, and suggests techniques to hasten recovery following sensory cortical surgery.

1 INTRODUCTION

The organization of the cerebral cortex was for a long time believed to be highly stable in adults. Much new evidence indicates that the adult cortex undergoes significant, often reversible, reorganization in response to various sensory and cortical manipulations such as lesions in the receptive surface and the cortex (for review see [1; 2]). The cortex appears to be a continuously adapting structure in a dynamic equilibrium with both the external and intrinsic input. This equilibrium is maintained by cooperative and competitive lateral interactions within the cortex, mediated by lateral connections. This article shows that adapting lateral interactions could be responsible for the experimentally observed plasticity of cortical maps as well.

Several aspects of cortical plasticity can be explained based on self-organizing neural network models of topographic maps [3; 5]. Such models assume predetermined lateral interactions within the cortex, and focus on the reorganization of the afferent connection weights. They demonstrate that the self-organization of afferent synapses alone (by Hebbian adaptation) can account for (1) the remapping of cortical topography following peripheral lesions, (2) the expansion of the cortical representation of a repetitively stimulated area of skin, and (3) the inverse relationship between cortical magnification (area of cortical map representing a unit area of the receptive surface) and receptive field size (area of receptive surface

To appear in *The Neurobiology of Computation: Proceedings of the Annual Computational Neuroscience Meeting* (Monterey, CA), 1994. Kluwer Academic Publishers.



(a) LISSOM architecture (b) Lateral interaction (c) Topographic map

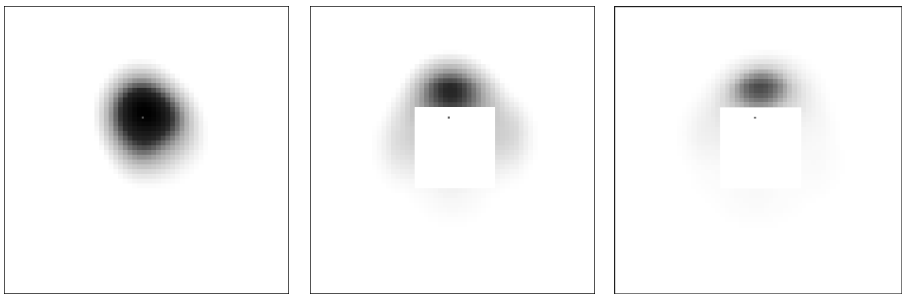
Figure 1: **Self-organization in a LISSOM network.** The afferent and lateral connections of one neuron are shown in (a). All connection weights start out random, but develop smooth profiles in the self-organizing process. The combined lateral excitation and inhibition of each neuron acquires a profile resembling a "Mexican hat" (b), and afferent weights organize across the network to form a topographic map (c). In (c), the center of gravity of the afferent weights of each neuron are plotted on the receptive surface, and the points of the neighboring neurons are connected by lines. The resulting smooth grid illustrates the topographic organization of the afferent weights.

driving a cortical neuron). However, several other aspects of cortical plasticity, such as the reorganization of the map in response to cortical lesions and intracortical microstimulation seem to involve adaptation of lateral connections as well, and cannot be explained by these models.

A new model of cortical self-organization called LISSOM (Laterally Interconnected Synergetically Self-Organizing Map: [6; 7]) was developed to explain how afferent and lateral connections could self-organize cooperatively and simultaneously to form topographic maps. The maps formed by LISSOM are continuously adapting structures in a dynamic equilibrium, and susceptible to changes in the distribution of external and intrinsic inputs. As a result, the model can account not only for the plasticity due to reorganizing afferent synapses, but also plasticity due to adapting lateral connections. This article demonstrates how the self-organizing process can model the reorganization of the cortex after cortical lesions and intracortical microstimulation.

2 THE LISSOM MODEL

The LISSOM network is a sheet of interconnected neurons (figure 1a). Through the afferent connections, neurons receive input from a receptive surface. In addition, each neuron has reciprocal excitatory and inhibitory lateral connections with other neurons. Lateral excitatory connections are short-range, connecting only close neighbors, but lateral inhibitory connections run for long distances. Initially, all connection weights are random. Each neuron develops an initial response as a weighted sum of the activation in its afferent input connections. The lateral interactions between neurons then focus the initial activation pattern into a localized response on the map (see appendix for the exact equations governing the process). After the pattern has stabilized, all connection weights are modified according to the same Hebbian rule, and normalized so that the sum of the weights of each connection type (afferent, lateral excitatory and lateral inhibitory) are constant. The process is repeated for each input. As self-organization progresses, the afferent connections develop smooth localized weight profiles and



(*a*) Activity before lesion (*b*) Immediately after (*c*) After 500 iterations

Figure 2: **How response patterns change after a cortical lesion.** The activity of neurons across the network are shown for the same input before the lesion (*a*), immediately after (*b*) and a few hundred adaptation steps later (*c*). The lesioned area is seen as a white square with no activity in figure (*b*). Immediately after the lesion, the activity spreads out to neurons that were previously inactive and therefore, the functional loss appears less severe than expected. As lateral connections reorganize (figure 3), the unmasked activity decreases because of increased lateral inhibition.

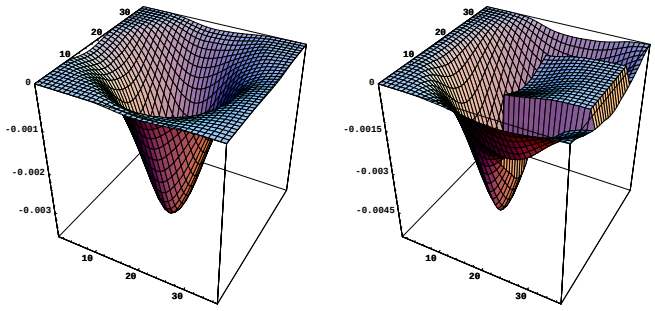
form a topographic map of the receptive surface (figure 1c). At the same time, the lateral connections develop smooth "Mexican hat" shaped profiles (figure 1b).

3 MODELING CORTICAL LESION PLASTICITY

To study the effects of cortical lesions, a small set of neurons in the organized network are made unresponsive to input. Three phases of reorganization are observed, like in the somatosensory cortex [2]. Immediately after the lesion, the receptive fields (RFs) of neurons in the perilesion zone enlarge. The lesion reduces the inhibition of the perilesion neurons, and unmasks previously suppressed input activation. In effect, the perilesion neurons immediately take over representing part of the input to the lesioned region, and the apparent loss of receptive surface representation is smaller than expected based on the prelesion map (figure 2b).

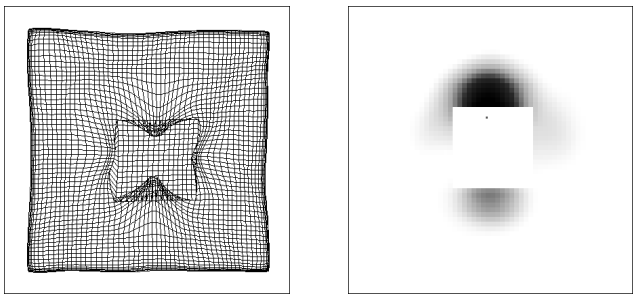
The lesion disrupts the dynamic equilibrium of the network, and both lateral and afferent connections of the active neurons adapt to compensate for the lesion. Neurons close to the lesion boundary encounter a large imbalance of lateral interaction in their neighborhood, with no lateral activation from inside the lesion and normal activation from outside. As a result, the lateral connection weights to the lesioned area decrease to zero, and by Hebbian adaptation and normalization, all the lateral weights rapidly redistribute to the the lesion's periphery. Neurons at the lesion boundary have the largest number of inhibitory connections from the lesioned zone; therefore, the reorganization of inhibition is especially pronounced at the boundary (figure 3). As a result, in the second phase the lateral inhibition very rapidly becomes strong outside the lesion, and the previously unmasked activity is partly suppressed (figure 2c). This produces an apparent outward shift of perilesion receptive fields.

Even after the lateral connections reorganize, the remaining unmasked input activation causes an imbalance in the network. Such activation forces the afferent weights to reorganize and respond better to inputs that were previously stimulating the lesioned zone. Gradually, the representation of the receptive surface within the lesion zone is taken over by the neurons around it (figure 4), and the cortical lesion is partly compensated for.



(a) Inhibition before lesion (b) After 500 steps

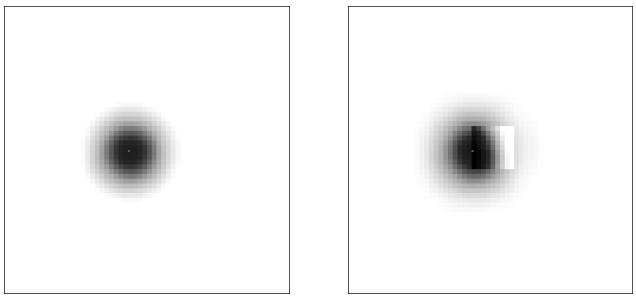
Figure 3: **Reorganization of lateral inhibition at the lesion boundary.** The inhibitory connections of a neuron at the boundary of the lesion are shown. The neuron has 40×40 connections, and the prelesion inhibition is circularly symmetric around the neuron (a). Shortly after the lesion, the inhibitory weights from the lesioned neurons decrease to zero. Because the total inhibitory weight is kept constant by weight normalization, the inhibition concentrates in the connections outside the lesioned zone, and the trough becomes deeper (b).



(a) Reorganized map (b) Final activity

Figure 4: **Topography and activity in the reorganized network.** Several thousand adaptation steps after the lesion, afferent weights of the perilesion neurons have spread out into the area previously represented by the lesioned neurons. Though lateral inhibition is still stronger in the perilesion area, the input activation after reorganization overcomes the inhibition, and neurons at the boundary of the lesion become more responsive to inputs previously stimulating lesioned neurons.

The LISSOM model suggests two techniques to accelerate recovery following surgery in the sensory cortices. Normally, the recovery time after cortical surgery would include an initial period of regression due to the reorganization of inhibition, and gradual and slow compensation afterward. The first phase of regression could be ameliorated if a transient blocker of inhibitory neurotransmitters were applied locally around the surgical area. Neurons around the surgical area would then fire intensively because of reduced inhibition, and afferent connections would adapt rapidly to compensate for the lesion. Though the inhibition would strengthen when the blockade goes away, the pace of recovery would have been hastened. Secondly, the topographic map could be shifted as in figure 4 even before surgery. This preshifting could be achieved by intensive and repetitive stimulation of the area expected to lose sensation and by sensory deprivation of its surroundings. The receptive fields would then have to move less to reach the final state, and the recovery would be faster.



(a) Activity before ICMS (b) After ICMS removal

Figure 5: **How activity patterns change after ICMS.** Activity patterns in the network are shown in the vicinity of the ICMS region before (a) and immediately after the stimulation (b), for the same input. After training with ICMS, activity is enhanced and sharpened within the ICMS region (the square area outlined by the activity) because of the reorganized lateral connections. With time, the effects of ICMS get reversed as connections adapted back to the original state.

The effects of simulated cortical lesions are reversible, and the original topographic representation would be regained if the lesioned neurons were restored to normal. Thus the topographic map is dynamic and maintained in a dynamic equilibrium with external and internal inputs. To illustrate how the equilibrium is altered when the internal inputs change, we simulated the effects of intracortical microstimulation (ICMS).

4 MODELING ICMS PLASTICITY

Intracortical microstimulation involves applying a high-frequency electrical current to a small region of the cortex using a microelectrode. In the somatosensory cortex, after a few hours of ICMS stimulation, the receptive fields of the stimulated neurons became clustered and overlapping [4]. Because there is no peripheral stimulation, such clustering must be solely due to the adaptation of intrinsic connections.

To simulate the effect of ICMS, a restricted region of the organized network was set to maximum activity for several steps. Because there was no afferent input, afferent connections remained unchanged (equation 2), and only the lateral connections in the ICMS region adapted. Neurons at the boundaries of the ICMS region developed asymmetric lateral interaction profiles, with strong excitation and inhibition towards the center of the region. Therefore, activity patterns were enhanced and sharpened within the ICMS region even when the original center of activity was outside the zone (figure 5). As a result, the receptive fields of neurons in the close vicinity of the ICMS region appeared to shift into the region, as was observed in the somatosensory cortex [4].

5 CONCLUSION

The LISSOM model demonstrates that not only the self-organization of topographic maps, but also many aspects of cortical plasticity can be explained based on the simultaneous adaptation of afferent and lateral connections. The simulated reorganizations are reversible, and demonstrate how a topographic map can be maintained in a dynamic equilibrium with extrinsic and intrinsic inputs.

The model suggests that functional recovery after cortical surgery may be hastened by blocking lateral inhibition locally in the cortex and by forced presurgical reorganization of cortical topographic maps.

APPENDIX: ACTIVATION AND ADAPTATION MECHANISMS

The neural activation is calculated from the sum of external, lateral excitatory and lateral inhibitory input:

$$\eta_{ij}(t) = \sigma\left(\sum_{r_1,r_2} \xi_{r_1,r_2}\mu_{ij,r_1r_2} + \gamma_e \sum_{k,l} E_{ij,kl}\eta_{kl}(t-\delta t) - \gamma_i \sum_{k,l} I_{ij,kl}\eta_{kl}(t-\delta t)\right), \quad (1)$$

where $\eta_{ij}(t)$ is the activity of neuron (i,j) at time step t, σ is a piecewise linear approximation to the sigmoidal activation function, ξ_{r_1,r_2} is the activation of a retinal receptor (r_1,r_2), μ_{ij,r_1r_2} is the afferent weight of neuron (i,j) from (r_1,r_2), $E_{ij,kl}$ is the excitatory lateral connection weight from neuron (k,l) to neuron (i,j), $I_{ij,kl}$ is the inhibitory lateral connection weight and γ_e, γ_i are scaling factors on the excitatory and inhibitory weights.

Connection weights are adapted according to the normalized Hebbian rule:

$$w_{ij,mn}(t+1) = \frac{w_{ij,mn}(t) + \alpha\eta_{ij}X_{mn}}{\sum_{mn}[w_{ij,mn}(t) + \alpha\eta_{ij}X_{mn}]}, \quad (2)$$

where $w_{ij,mn}$ is the afferent or lateral connection weight (μ_{ij,r_1r_2}, $E_{ij,kl}$ or $I_{ij,kl}$), α is the learning rate for each type of connection (α_a for afferent, α_E for excitatory, and α_I for inhibitory) and X_{mn} is the presynaptic activity (ξ_{r_1,r_2} for afferent, η_{kl} for lateral).

Acknowledgements

This research was supported in part by National Science Foundation under grant #IRI-9309273. Computer time for the simulations was provided by the Pittsburgh Supercomputing Center under grants IRI930005P and TRA940029P.

References

[1] Jon H. Kaas. Plasticity of sensory and motor maps in adult animals. *Annual Review of Neuroscience*, 14:137–167, 1991.

[2] M. M. Merzenich, G. H. Recanzone, W. M. Jenkins, and K. A. Grajski. Adaptive mechanisms in cortical networks underlying cortical contributions to learning and nondeclarative memory. In *Cold Spring Harbor Symposia on Quantitative Biology, Vol. LV*, pages 873–887, Cold Spring Harbor, NY, 1990.

[3] K. Obermayer, H. J. Ritter, and K. J. Schulten. Large-scale simulation of a self-organizing neural network. In *Proceedings of the International Conference on Parallel Processing in Neural Systems and Computers*, New York, 1990. Elsevier.

[4] G. H. Recanzone, M. M. Merzenich, and H. R. Dinse. Expansion of cortical representation of a specific skin field in primary somatosensory cortex by intracortical microstimulation. *Cerebral Cortex*, 2:181–196, June 1992.

[5] Helge J. Ritter, Thomas Martinetz, and Klaus J. Schulten. *Neural Computation and Self-Organizing Maps: An Introduction.* Addison-Wesley, Reading, MA, 1992.

[6] Joseph Sirosh and Risto Miikkulainen. Cooperative self-organization of afferent and lateral connections in cortical maps. *Biological Cybernetics*, 71(1):66–78, 1994.

[7] Joseph Sirosh and Risto Miikkulainen. A unified neural network model for the self-organization of topographic receptive fields and lateral interaction. Technical Report AI94-213, Department of Computer Sciences, The University of Texas at Austin, 1994.