Intrinsic Plasticity in a Generative Model of V1

Cornelius Weber & Jochen Triesch

Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe University Frankfurt. Email: c.weber@fias.uni-frankfurt.de

Overview

Generative, sparse coding models explain the learning of localized edge detector neurons in primary visual cortex V1. Current models, which we use as a basis, work on just two time scales:

1 fast: the internal neurons' activations adapt to a visual stimulus;

2 slow: the weights adapt to the set of stimuli.

Here we distinguish a separate mechanism that may act on a medium time scale: the adaptation of a neuron's excitability, "intrinsic plasticity" (IP). Hereby a neuron maintains firing rate homeostasis in a dynamic environment.

3 We set a desired exponentially distributed firing rate, which represents sparse coding, and adapt neuronal transfer function parameters a and b so to maintain the desired rate distribution.

4 With natural image input, localized edge detectors emerge.

The novel time scale of the intrinsic plasticity parameters allows to explore visual aftereffects such as the tilt-aftereffect (TAE). In this effect, after a viewer adapts to a grid of a certain orientation, grids of a nearby orientation will be perceived as tilted away from the adapted orientation.

5 Our results show that adapting the neurons' gain- but not the threshold-parameter leads to quantitatively similar aftereffects as found psychophysically, without the use of horizontal connectivity. This happens, because neurons coding for the adapting stimulus attenuate their gain while others increase it.

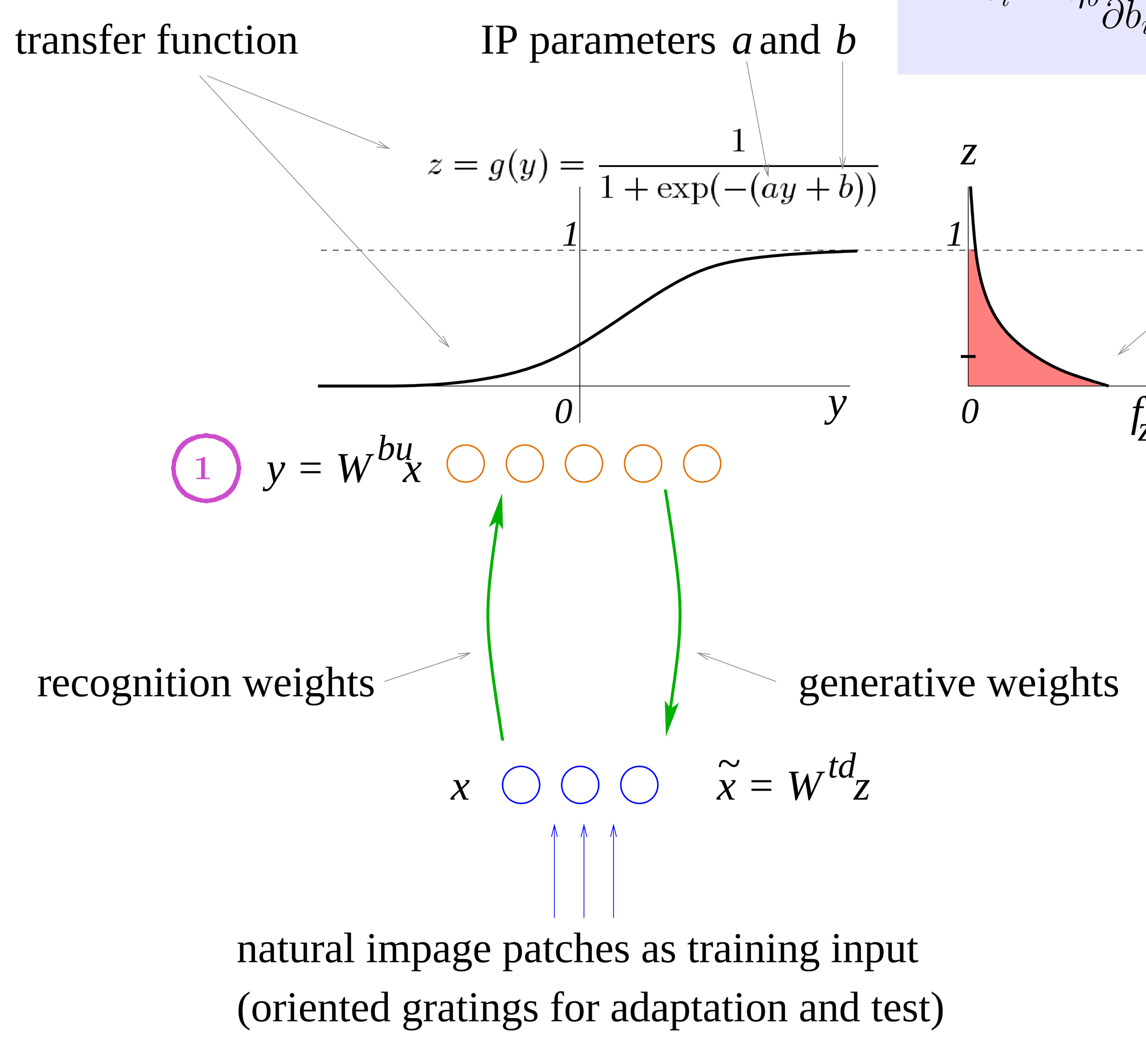
6 In contrast, long-term exposure to gratings of, e.g. 45° orientation yields enhanced representations of that orientation, while after training with natural images, the model also reproduces enhanced representations of vertical and horizontal edges, as experimentally observed.

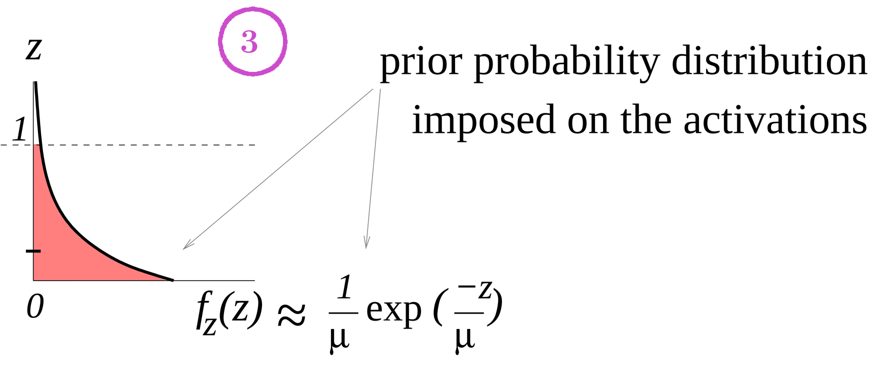
Model

Adaptation of a and b:

$$\Delta a_i = \eta_a \frac{\partial}{\partial a_i} d(f_z(z_i) \| f_{\exp}(z_i)) \;=\; \eta_a \Big(\frac{1}{a_i} + y_i - 2 y_i z_i - \frac{1}{\mu} y_i z_i + \frac{1}{\mu} y_i z_i^2 \Big)$$

$$\Delta b_i = \eta_b \frac{\partial}{\partial b_i} d(f_z(z_i) \| f_{\exp}(z_i)) \;=\; \eta_b \Big(1 - 2 z_i - \frac{1}{\mu} z_i + \frac{1}{\mu} z_i^2 \Big)$$





Sleep phase, using random binary "phantasy" data $\tilde{z}$ to adapt W^{bu}:

$$\hat{z} = \tilde{z} - g(W^{bu}\tilde{x})$$
$$\Delta w_{ij}^{bu} \approx \hat{z}_i \tilde{x}_j$$

with $\sum_j (w_{ij}^{bu})^2 = 1$ for each neuron i

2

Wake phase, using data $\vec{x}$ to adapt W^{td}:

$$\hat{x} = \vec{x} - W^{td}\vec{z}$$
$$\Delta w_{ji}^{td} \approx \hat{x}_j z_i$$

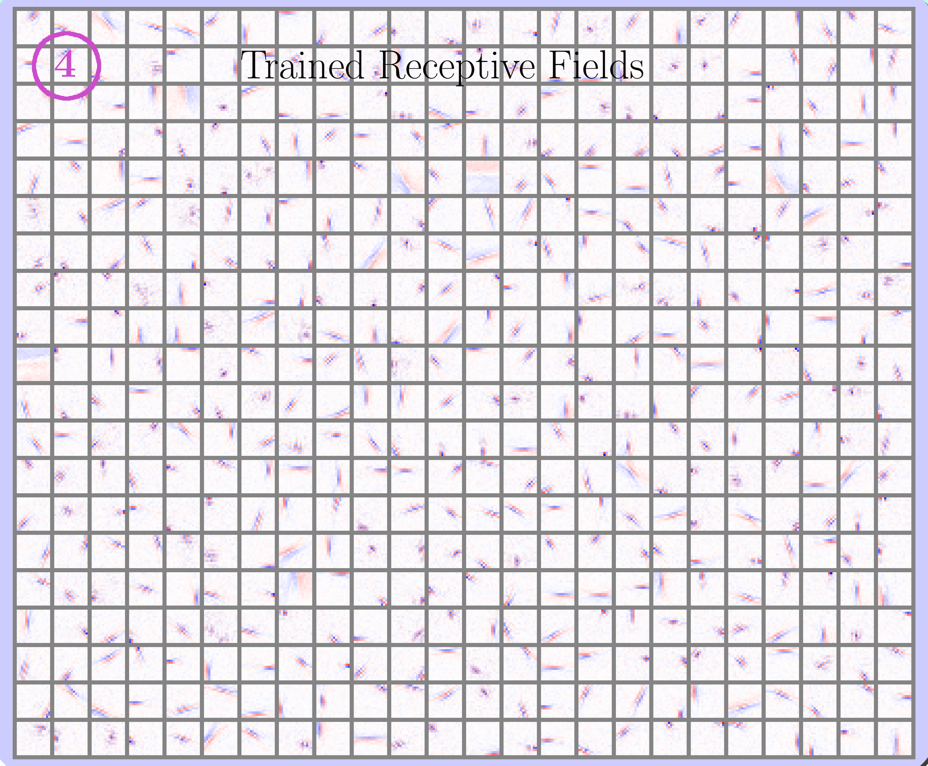
with $\sum_i (w_{ji}^{td})^2 = 1$ for each input neuron j

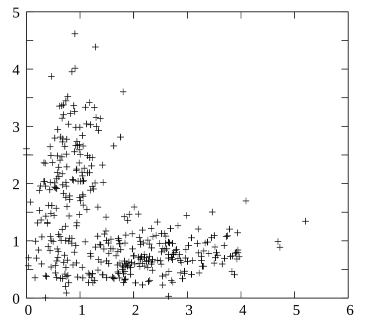
Discussion

- Our model has realistic assumptions, learns realistic V1 neuron receptive fields, and accounts for adaptation on a time scale of tens of seconds, including the TAE.
- It has relatively few parameters and is easy to use.
- It is information theoretically founded, yet has local learning rules.
- The IP adaptation works also with Olshausen's generative network algorithm (result not shown). The weight normalization used here does away with Olshausen's weight decay parameter.
- IP parameter adaptation could alternatively be regarded to happen on a time scale of ∼48 hours, to account for Turrigiano's experiments (but then no explaination of the TAE).
- Gain adaptation accounts for the TAE. Unlike the gain parameters, the neurons's threshold parameters are all very similar. They may be unmodifiable.
- Bednar & Miikkulainen's model explains the TAE with rapidly strengthened inhibitory horizontal V1 connections. Activity-induced inhibition is hard to distinguish from rate adaptation. With horizontal weights, our model may also replicate the indirect TAE.

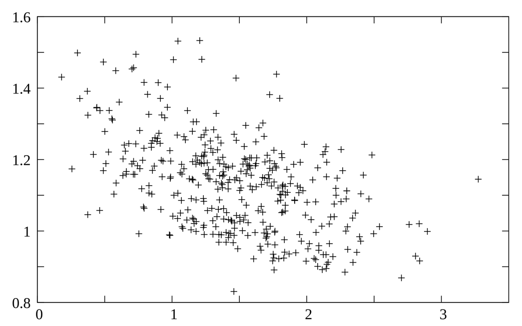
Acknowledgments

We acknowledge financial support by the European Union through projects FP6-2005-015803 ("Daisy"), MEXT-CT-2006-042484 ("PLICON") and by the Hertie Foundation.

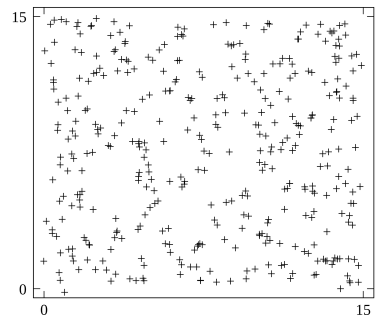




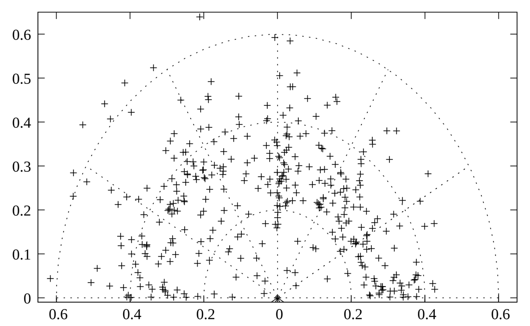
a-half-axis vs. b-half-axis of the elliptical gabor envelope fits to the receptive fields: the RF's tend to be elliptical.



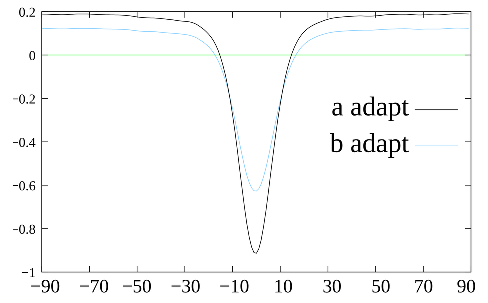
The distribution of the IP parameter a (y-axis) over the receptive field size (x-axis).



The centers of the receptive fields in the model retina: homogeneous distribution.

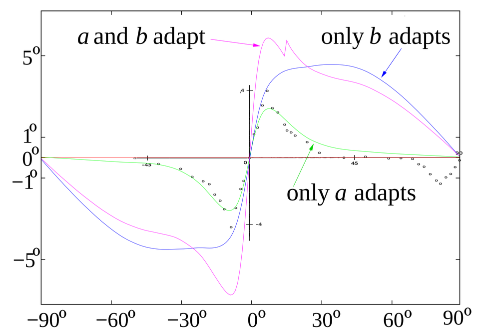


The orientation- (angle) / frequency- (radius) distribution of the receptive fields: all orientations and a broad frequency range are covered.

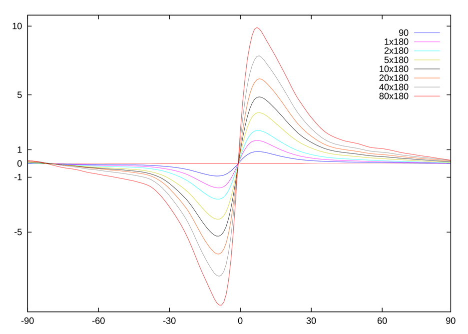


Adaptation (change) of IP parameters a and b when a grating of one orientation (center on x-axis) is shown for a while.

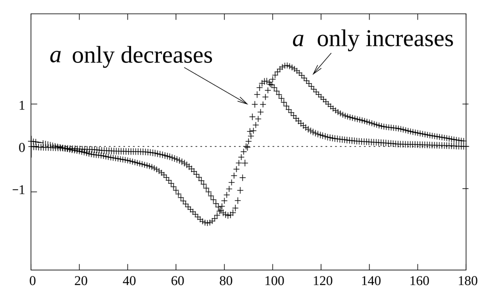
5



Tilt aftereffect: x-axis is the test orientation; middle the adapting orientation. y-axis is the change of the perceived orientation. Only the green curve is realistic (only a allowed to adapt).

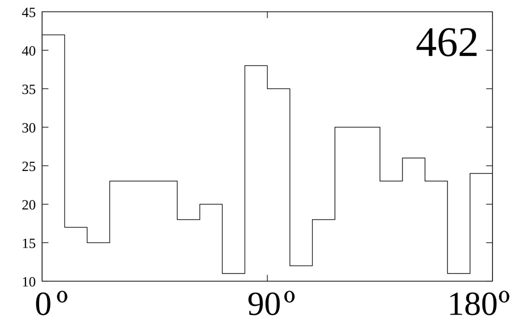


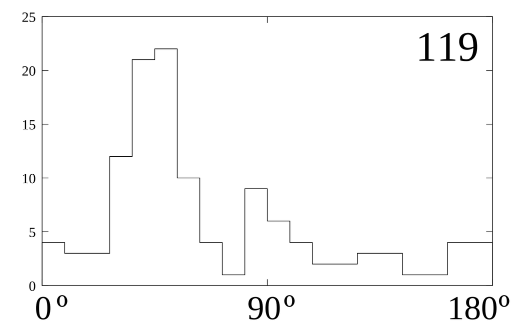
Adaptation times double from curve to curve: the model TAE does not saturate.



Is the TAE due to gain decrease or increase? If a is allowed only to increase (which happens away from the adapting orientation) then also the maximum of the effect shifts furhter away.

6





Histograms of the neurons' preferred orientations obtained from Gabor fits. Top: the map trained from natural images has slight preference for cardinal orientations. Bottom: after additional learning with gratings of 45° orientation, neurons preferentially code these orientations, but fewer can be fitted by Gabors.