Adaptive Feedback Inhibition Improves Pattern Discrimination Learning

Frank Michler, Thomas Wachtler, and Reinhard Eckhorn

AppliedPhysics/NeuroPhysics Group, Department of Physics,
Philipps-University Marburg, Renthof 7, D-35032 Marburg, Germany
frank.michler@physik.uni-marburg.de
http://www.physik.uni-marburg.de

Abstract. Neural network models for unsupervised pattern recognition learning are challenged when the difference between the patterns of the training set is small. The standard neural network architecture for pattern recognition learning consists of adaptive forward connections and lateral inhibition, which provides competition between output neurons. We propose an additional adaptive inhibitory feedback mechanism, to emphasize the difference between training patterns and improve learning. We present an implementation of adaptive feedback inhibition for spiking neural network models, based on spike timing dependent plasticity (STDP). When the inhibitory feedback connections are adjusted using an anti-Hebbian learning rule, feedback inhibition suppresses the redundant activity of input units which code the overlap between similar stimuli. We show, that learning speed and pattern discriminatability can be increased by adding this mechanism to the standard architecture.

1 Introduction

1.1 Standard Architecture

Standard neural networks for unsupervised pattern recognition learning typically consist of adaptive forward connections and lateral inhibition (e.g. Fukushima 1975; Földiák 1990). Usually, the forward connections are modified using Hebbian learning rules: if pre- and postsynaptic activity is highly correlated, excitatory synapses are strengthened while inhibitory synapses are weakened. For excitatory synapses, Hebbian learning increases the correlation between pre- and postsynaptic activity and the connections grow infinitely. Connection strengths can be limited e.g. by using normalization mechanisms.

Lateral inhibitory connections introduce a *winner-take-all (WTA)* dynamics: if an output neuron is strongly activated, other output neurons receive strong inhibition and generate little or no output activity. WTA prevents the output neurons from being active all at the same time. When the lateral inhibitory connections are learned with an *anti-Hebbian* learning rule, as proposed by Földiák (1990), connections are strengthened if correlation between pre- and postsynaptic activity is high. Thus, strongly correlated output neurons will have strong inhibitory connections, which will reduce their correlation. This decorrelation can lead to a sparse representation of the input stimuli (Földiák,

F. Schwenker and S. Marinai (Eds.): ANNPR 2006, LNAI 4087, pp. 21–32, 2006.

1990). After self-organization, the neurons in the output layer of such networks should respond selectively to a single stimulus pattern or a subset of the training set, depending on the relation between the size of the stimulus set and the number of output neurons.

1.2 Improving Discrimination Performance with Feedback Inhibition

Consider a two layer network with an input and an output layer, and lateral inhibition between output neurons. What happens when the network is trained with a set of very similar stimuli? Typically the forward connections from the uninformative input neurons coding the overlap between stimuli will become much stronger compared to the connections coding features unique to certain stimuli (Fukushima, 1975; Földiák, 1990). Beyond a certain degree of stimulus similarity the output neurons only respond to the overlap, and thus fail to discriminate between the stimuli. Miyake and Fukushima (1984) proposed a mechanism to improve pattern selectivity fur such situations: they introduced a simple version of modifiable inhibitory feedback connections from the output units to the input units. These connections were paired with modifiable excitatory feedforward connections. When a feedforward connection was strengthened, the corresponding feedback connection was strengthened as well.

In this paper we show that this adaptive feedback inhibition can be generalized and adapted to a biologically more realistic network model with spiking neurons and *spike timing dependent plasticity (STDP)* based learning rule (Bi and Poo, 1998). We systematically varied the overlap between the patterns of the stimulus set and show how learning speed and selectivity increases after introducing modifiable inhibitory feedback connections.

Using spiking neural network models aims towards an understanding of how pattern recognition problems could be solved in the brain. If a mechanism can not be implemented with biologically realistic spiking neurons, then it is unlikely that this mechanism is used in the brain. Furthermore spiking neurons provide for high temporal precision, which is relevant for real-world applications. This is the case e.g. for spatio-temporal pattern recognition or for audio patterns.

2 Model

2.1 Network Architecture

The network is organized in two layers of spiking neurons: the *input layer* U_0 and the *representation layer* U_1 (Fig. 1). There are excitatory forward connections from U_0 to U_1 and lateral inhibitory connections between all U_1 neurons. These connections are adapted due to the correlation between presynaptic and postsynaptic spikes with a Hebbian and anti-Hebbian learning rule, respectively (Section 2.3). So far this is the standard architecture for competitive learning. Additionally, we introduce modifiable inhibitory feedback connections from U_1 to U_0. These inhibitory connections are also adapted using an anti-Hebbian learning rule.

2.2 Model Neurons

As a spiking model neuron we use the two dimensional system of differential equations proposed by Izhikevich (2003):

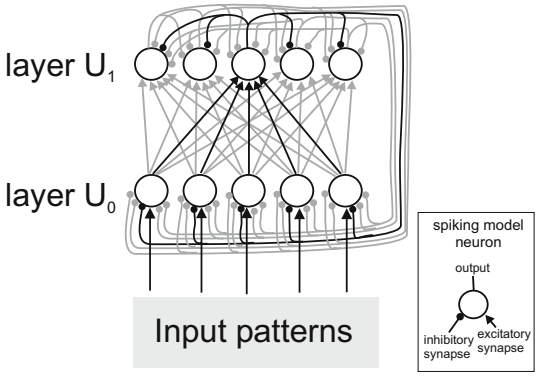

Fig. 1. Model architecture. The neurons of the input layer U_0 are activated when they are part of the current input pattern. U_0 neurons have modifiable excitatory connections to the representation layer U_1. U_1 neurons mutually inhibit each other. Additionally there are modifiable inhibitory feedback connections from U_1 to U_0. To better illustrate the network structure, connections from and to one of the neurons are plotted with black color while the other connections are plotted gray.

$$\frac{dV(t)}{dt} = 0.04V^2(t) + fV(t) + e - U(t) + I(t), \tag{1}$$

$$\frac{dU(t)}{dt} = a(bV(t) - U(t)) \tag{2}$$

with the auxiliary after-spike reseting:

$$\text{if } V(t) \geq 30mV, \text{ then } \begin{cases} V(t) \leftarrow c, \\ U(t) \leftarrow U(t) + d. \end{cases} \tag{3}$$

$V(t)$ and $U(t)$ are dimensionless variables. $V(t)$ represents the membrane potentials in mV. $I(t)$ is the synaptic input current. a, b, c, d, e and f are dimensionless parameters which determine the properties of the model neuron. In the simulations presented here we use a set of parameters which correspond to regular spiking cortical pyramidal neurons (example "L" in Izhikevich, 2004, a=0.02, b=-0.1, c=-55, d=6, e=108, f=4.1). The excitatory synaptic input I_e is modelled as a current injection with additional noise $\sigma(t)$. The inhibitory input I_i is modelled as a conductance based current. The excitatory synaptic input saturates at $I_{e,max}$. The inhibitory conductance saturates at $G_{i,max}$:

$$I = S_e(I_e) - S_i(G_i)(V - E_i), \tag{4}$$

$$S_e(I_e) = I_{e,max}\frac{I_e}{I_e + 1}, \tag{5}$$

$$S_i(G_i) = G_{i,max}\frac{G_i}{G_i + 1}, \tag{6}$$

$$\frac{d}{dt}I_e = -\frac{1}{\tau_e}I_e + \sum_{m=0}^{M-1} w_{e,m}\delta_m(t) + \sigma(t), \tag{7}$$

$$\frac{d}{dt}G_i = -\frac{1}{\tau_i}G_i + \sum_{m=0}^{M-1} w_{i,m}\delta_m(t). \tag{8}$$

The saturation constants were set to $I_{e,max} = 200$ and $G_{i,max} = 4.5$ to restrict excitatory and inhibitory input to a range where the numerical integration of the differential equations still works properly for $dt = 0.25ms$. The excitatory and inhibitory synaptic currents decrease exponentially with time constant τ_e and τ_i respectively, which were arbitrarily set to $5ms$. The biologically realistic range for the decay time constants of excitatory $AMPA$- and inhibitory $GABA_A$-currents is from 5 up to 50 ms. $w_{e,m}$ is the excitatory weight from the presynaptic neuron number m. $\delta_m(t)$ is 1 when a spike arrives at the presynaptic site, otherwise it is 0. E_i is the reverse potential for the inhibitory current which was chosen to be 10 mV lower then the resting potential.

2.3 Learning Rules

The synaptic weight $w_{m,n}$ of the connection from presynaptic U_0 neuron m to postsynaptic U_1 neuron n is adapted according to a Hebbian learning rule:

$$\frac{d}{dt}w_{m,n} = \delta_n(t)RL_{pre,m}L_{post,n}, \tag{9}$$

$$L_{pre,m} = \sum_{t_{sm}} e^{-\frac{t-t_{sm}}{\tau_{pre}}}, \tag{10}$$

$$L_{post,n} = \sum_{t_{sn}} e^{-\frac{t-t_{sn}}{\tau_{post}}}. \tag{11}$$

$\delta_n(t)$ is 1 when a spike occurs in the postsynaptic neuron n. t_{sm} and t_{sn} denote the times of the past pre- and postsynaptic spikes. When a spike occurs, the pre- or postsynaptic *learning potentials* $L_{pre,m}$ or $L_{post,n}$ are increased by 1. They exponentially decrease with time constant $\tau_{pre} = 20ms$ and $\tau_{post} = 10ms$. R is a constant corresponding to the learning rate and was tuned to allow for a weight change between 5 and 20 % after 10 stimulus presentations. For the excitatory connections from layer U_0 to U_1, we use a quadratic normalization rule:

$$w_{m,n}(t) = W\frac{w_{m,n}(t-dt)}{\sqrt{\sum_{m=0}^{M-1} w_{m,n}^2(t-dt)}}, \tag{12}$$

where W is a constant value to adjust the quadratic weight sum. This prevents infinite growing of weights and introduces competition between the input synapses of a postsynaptic neuron. Physiological evidence for the existence of such heterosynaptic interactions were found, e.g., by Royer and Paré (2003). W was set to a value which ensured a medium response activity at the beginning of the learning phase.

For the inhibitory connections we use the following anti-Hebbian learning rule:

$$\frac{d}{dt}w_{m,n} = R\big(\delta_n(t)L_{pre,m} - C\delta_m(t)w_{m,n}L_{post,n}\big), \tag{13}$$

$$L_{pre,m} = e^{-\frac{t-t_{sm}}{\tau_{pre}}}, \tag{14}$$

$$L_{post,n} = e^{-\frac{t-t_{sn}}{\tau_{post}}}. \tag{15}$$

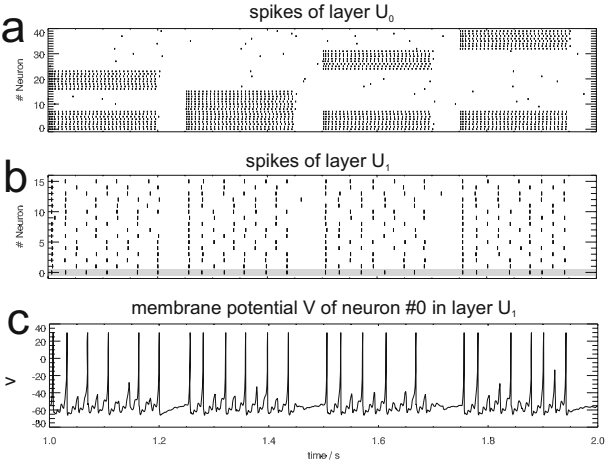

Fig. 2. Network without feedback inhibition, response before learning. a: Spikes of input layer U_0. b: Spikes of representation layer U_1. c: Membrane potential $V(t)$ of neuron #0 of U_1 (gray line in b).

The equations are very similar to the Hebbian learning rule (equation 9) but with an additional depression term. The decay time constants of the learning potentials were set to $\tau_{pre} = 30ms$ and $\tau_{post} = 100ms$. C is a constant to adjust the ratio between potentiation and depression which determines the amount of inhibition. With lower C the inhibitory connections will be stronger. C was set to 0.005 for the feedback inhibition and 0.001 for the lateral inhibition. t_{sm} and t_{sn} denote the time of the last pre- and post- synaptic spike event respectively.

2.4 Stimuli

The input stimuli are binary spatial patterns that lead to additive modulation of excitatory synaptic current I_e (equation 4) of layer U_0 neurons:

$$I_e(t) = \sum_{i \in N} p_n^{k_i} I_0 \mathit{rect}\left(\frac{t - i\tau_1}{\tau_2}\right), \tag{16}$$

$$\mathit{rect}(t) = \begin{cases} 1 : |t| < 0.5 \\ 0 : otherwise\ . \end{cases} \tag{17}$$

$p_{k_i}^n$ is 1 if the neuron n is active for stimulus k_i, and 0 otherwise. I_0 is the input strength. τ_1 is the time difference between stimulus onsets, τ_2 is the duration of a single stimulus presentation (see Fig. 2 for an example). $k_1, k_2, ..., k_i$ is a random sequence of stimulus numbers.

For a systematic variation of the similarity between the input patterns, we constructed sets of stimuli as follows: each stimulus is a binary pattern P_k of N_{U_0} elements where N_{U_0} is the number of neurons in the input layer.

$$P_k = (p_1^k, p_2^k, p_3^k, ..., p_{N_{U_0}}^k), \tag{18}$$

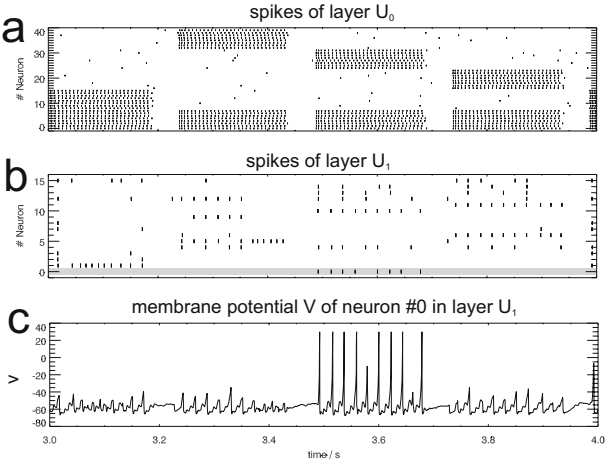

Fig. 3. Network without feedback inhibition, response after learning. a: Spikes of input layer U_0. b: Spikes of representation layer U_1. c: Membrane potential $V(t)$ of neuron #0 of U_1 (gray line in b).

$$p_m^k = \begin{cases} 1\ , m \leq n_o \\ 1\ , n_o + n_u(m-1) < m \leq n_o + n_u m \\ 0\ , otherwise\ . \end{cases} \tag{19}$$

$n_a = f_a N_{U_0}$ is the number and f_a the fraction of active neurons in each pattern. $n_o = f_o n_a$ is the number of neurons which are active in each pattern (overlap) and $n_u = n_a - n_o$ is the number of neurons which are unique for each pattern.

2.5 Performance Measure

In order to quantify the ability of the network to discriminate between the stimuli, we simulated a test phase after every learning phase. In the test phases the network was stimulated with the same input patterns as in the learning phases. We calculated the preferred stimulus κ_n and a selectivity index η_n for every U_1 neuron:

$$\kappa_n = \big\{k : R_{n,k} = max(\{R_{n,1}, ..., R_{1,K}\})\big\}, \tag{20}$$

$$\eta_n = \frac{R_{n,\kappa_n}}{\sum_{k=0}^{K} R_{n,k}} - \frac{1}{K}\ . \tag{21}$$

K is the number of stimuli. κ_n is the number of the stimulus which evokes the maximal response in U_1 neuron n. The selectivity index η_n is 0 if all stimuli evoke the same response $R_{n,k}$, which means that this neuron bears no information about the identity of the stimulus. The maximum selectivity is $\frac{K-1}{K}$ when only one stimulus evokes a response but the others do not. From the following test phase we calculated how the activity of the U_1 neurons predict the identity of the input patterns: for each stimulus onset we derived the response $r_{n,i}$ for every U_1 neuron (number of spikes in a specified interval after stimulus onset), where j is the number of the current stimulus. Combining

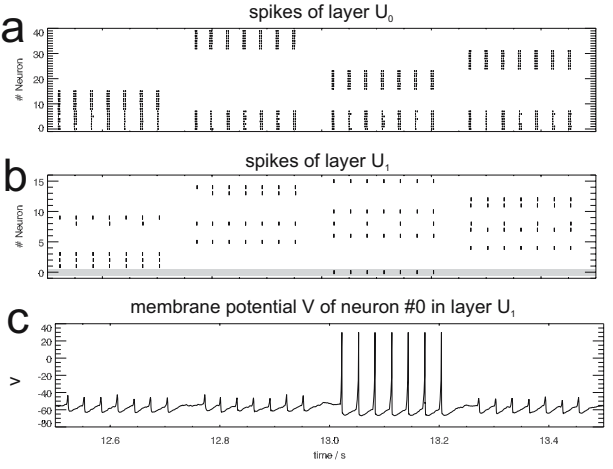

Fig. 4. Network with feedback inhibition, response after learning. a: Spikes of input layer U_0. b: Spikes of representation layer U_1. c: Membrane potential $V(t)$ of neuron #0 of U_1 (gray line in b). The feedback inhibition circuit causes rhythmic spike patterns in both layers.

these responses with the preference and the selectivity of the neurons, we calculated the stimulus ν_j predicted by this network activity:

$$\nu_j = \left\{k : \xi_k = max(\{\xi_1, ..., \xi_K\})\right\}, \tag{22}$$

$$\xi_k = \sum_{n\epsilon\{i:\kappa_i=k\}} \eta_n r_{n,k} \ . \tag{23}$$

If $\nu_j = j$ then the prediction is correct, otherwise it is false. The performance ρ is then $\rho = \frac{n_{hit}}{n_{hit}+n_{fail}}$ where n_{hit} is the number of correct predictions and n_{fail} the number of mistakes. The chance level is $\frac{1}{K}$.

3 Results

First we demonstrate the properties of the network without feedback inhibition for a stimulus set with little overlap (50%). The number of stimuli was $K = 4$. The numbers of neurons were: $N_{U_0} = 40$ and $N_{U_1} = 16$. Before learning, the network responds unselectively to the input stimuli (Fig. 2). The network quickly converges to a selective state: for each stimulus there is at least one U_1 neuron that selectively responds to it (Fig. 3).

When we systematically increased the overlap between the elements of the stimulus set the network needed longer to reach a selective state. When the overlap was very high it completely failed to discriminate between the stimuli (Fig. 5).

When we added the modifiable inhibitory feedback connections, the network took less time steps to reach a selective state. Even for high overlap, where it had failed without feedback inhibition, the network learned a selective representations (Fig. 6). Furthermore, the feedback inhibition causes rhythmic spike patterns in both layers and synchronizes the activated neurons (Fig. 4).

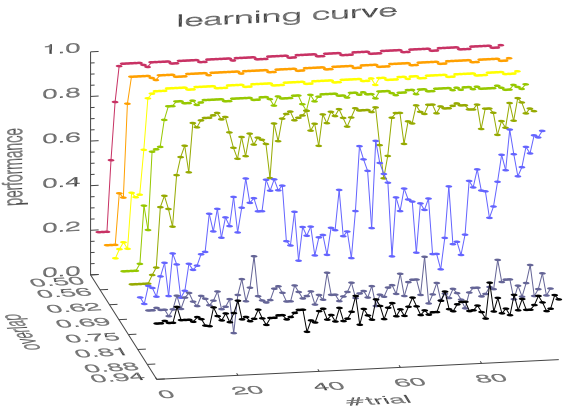

Fig. 5. Learning curves without feedback inhibition. A *trial* consisted of 40 stimulus presentations. For overlap up to 75% the network quickly learned a selective representation. For higher overlap it took longer training time to reach a selective state. For overlap higher than 88% the network stayed in an unselective state. Input strength: $I_0 = 0.008$.

Because the feedback inhibition reduces the spiking activity in U_0, we compensated this effect by increasing excitatory input strength I_0 (see equation 16) when turning on the feedback inhibition. To make sure that the differences in learning speed and learning performance were not caused by these parameter changes, we systematically tested the effect of different input strengths. We calculated a performance index for each I_0 value by averaging the performance values for the second half of learning trials over all overlap levels. Without feedback inhibition the maximum performance of the network (at $I_0 \approx 0.008$) was still lower than the maximum performance of the network with feedback inhibition (Fig. 7).

4 Discussion

Our simulations show that in a network of spiking neurons adaptive feedback inhibition can speed up learning of selective responses and enable discrimination of very similar input stimuli. The mechanism works as follows: While the network is in an unselective state, the correlation between the output units and these input units which code the overlap ($p_1^k ... p_{n_o}^k$ in Eq. 18) is higher than the correlation between the output units and the input units which are unique for different patterns. Therefore, the inhibitory connections to the input neurons representing the overlap will grow stronger and the redundant activity will be reduced. In contrast, the input neurons coding the difference between the stimuli receive less inhibition. Thus, the network can use the discriminative information carried by these neurons to learn a selective representation.

The network parameters were chosen in a biologically realistic range. The input strength I_0 and the feed forward weight sum W were set to obtain reasonable firing rates. The learning parameters that control the inhibitory connections (C, τ_{pre}, τ_{post}) must be guanrantee a substantial amount of inhibition. Overall the mechanism doesn't

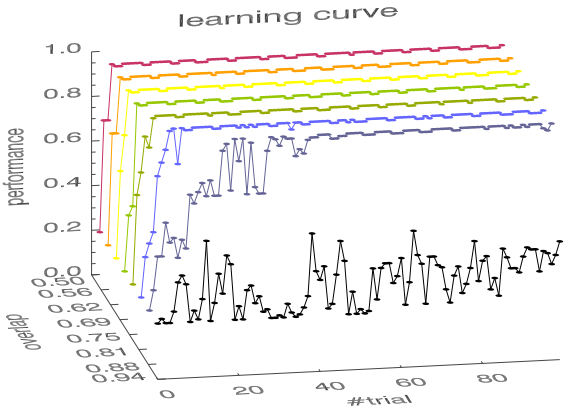

Fig. 6. Learning curves with feedback inhibition ; a *trial* consisted of 40 stimulus presentations. For the low overlap stimulus sets (50% - 81%) the network converged to a selective state faster than without feedback inhibition. Even for very high overlap (94%) the network still learned some selectivity. Input strength: $I_0 = 0.016$.

depend on the precise values of the parameters. Small or medium parameter changes do not qualitatively alter the properties of the network.

4.1 Comparison to Other Models

Miyake and Fukushima (1984) had already proposed a inhibitory feedback mechanism and showed how it could be included in their Cognitron model. They demonstrated the increased selectivity using stimulus pairs with up to 50% spatial overlap. As our simulations show, such an amount of overlap can still be separated using a network without feedback inhibition (Fig. 5).

Spratling (1999) had proposed a *pre-integration lateral inhibition* model. In this model for example an output neuron O_i which has strong excitatory connection from input neuron I_j will have strong inhibitory influence on the excitatory connections from I_j to the other output neurons $O_{k\neq i}$. Spratling and Johnson (2002) showed that *pre-integration lateral inhibition* can enhance unsupervised learning. Spratling (1999) argues against the feedback inhibition model, that an output neuron cannot entirely inhibit the input to all other neurons without entirely inhibiting its own input. van Ooyen and Nienhuis (1993) point out a similar argument: With feedback inhibition the Cognitron model fails to elicit sustained responses for familiar patterns, because the corresponding input activity is deleted. But these drawbacks do not hold in our dynamic model: After strong activation of an output neuron O_i, the feedback inhibition will suppress the input and thus prevent all output neurons from firing including O_i. Inhibition is reduced, and excitatory input can grow again. Thus, for sustained input, the inhibitory feedback generates rhythmic chopping of both input and output layer neurons (Fig. 4). The strongest activated output neurons are able to fire output spikes before inhibition grows, while weakly activated output neurons are kept subthreshold. Furthermore, the

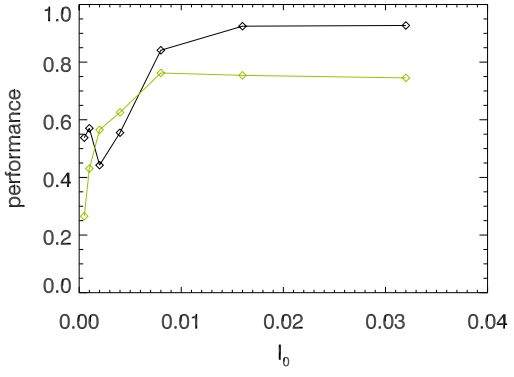

Fig. 7. Performance depends on input strength I_0. The data points show mean performance values, averaged over all overlap values and the second half of the learning trials. Black: Performance with feedback inhibition. Green (gray): Performance without feedback inhibition. Note that with feedback inhibition the network reaches higher performance values (90% compared to 75%).

common feedback inhibition tends to synchronize the activity of these input neurons which are part of the recognized pattern. Such a synchronization has been proposed to support object recognition through dynamic grouping of visual features (see e.g. Eckhorn, 1999; Eckhorn et al., 2004). In the model presented here, synchronization occurs as a consequence of successful pattern recognition.

The adaptive feedback inhibition model is in line with predictive coding models (Rao and Ballard, 1997). These models are based on the working principle of extended Kalman filters, where a prediction signal is subtracted from the input. Thus, in these models the predicted (expected) information is suppressed. This approach is the opposite to the *Adaptive-Resonance-Theorie (ART)*, which is based on enhancement of predicted information (Grossberg, 2001).

4.2 Physiological Equivalent

What could be a physiological basis for the proposed feedback inhibition mechanism? The main input to a cortical area arrives in layer 4 (Callaway, 1998). For example, layer 4 of the primary visual cortex receives input from the thalamic relay neurons of the lateral geniculate nucleus (LGN). Neurons in layer 2/3 have more complex receptive fields. They represent the main output of a cortical module to other cortical areas (Callaway, 1998). Thus, layer U_0 of our model corresponds to cortical layer 4 and layer U_1 to cortical layer 2/3.

Among direct input from thalamic relay neurons, layer 6 neurons receive feedback connections from layer 2/3. In visual area V1 they project back to the LGN but also have collaterals which project to layer 4, where they mainly target inhibitory interneurons (Beierlein et al., 2003). Thus, the anatomy of the neocortex provides the necessary connections for adaptive feedback inhibition: *layer 4 → layer 2/3 → layer 6 → inhibitory*

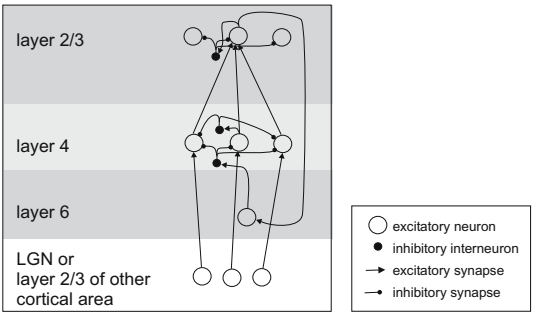

Fig. 8. Possible microcircuit underlying selective feedback inhibition: information enters the cortical module via layer 4, layer 2/3 learns selective representation of input patterns and projects back to layer 6, layer 6 neurons have projections to inhibitory interneurons in layer 4

interneurons of layer 4. This microcircuit could provide the basis for the suppression of uninformative input activity (Fig. 8).

We have shown, that adaptive feedback inhibition can increase learning speed and improve discrimination of highly similar patterns. For simplicity, we used a small set of simple stimulus patterns. The proposed mechanism can also be used for recognition of more complex patterns (e.g. 3d visual objects), if it is incorporated in a hierarchical multi-layer network architecture with feedback inhibition from higher to lower layers.

Acknowledgements

This work was supported by DFG grant EC 53/11.

References

Beierlein, M., Gibson, J. R., Connors, B. W. (2003). Two dynamically distinct inhibitory networks in layer 4 of the neocortex. Journal of Neurophysiology 90, 2987–3000.

Bi, G., Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. The Journal of Neuroscience 18 (24), 10464–10472.

Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. Annual Review of Neuroscience 21, 47–74.

Eckhorn, R. (1999). Neural mechanisms of scene segmentation: Recordins from the visual cortex suggest basic circuits for linking field models. IEEE Transactions on Neural Networks 10 (3), 464–479.

Eckhorn, R., Bruns, A., Gabriel, A., Al-Shaikhli, B., Saam, M. (2004). Different types of signal coupling in the visual cortex related to neural mechanisms of associative processing and perception. IEEE Transactions on Neural Networks 15 (5), 1039–1052.

Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. Biological Cybernetics 20, 121–136.

Földiák, P. (1990). Forming sparse representations by local anti-hebbian learning. Biological Cybernetics 64, 165–170.

Grossberg, S. (2001). Linking the laminar circuits of visual cortex to visual perception: Development, grouping and attention. Neuroscience and Biobeavioral Revies 25, 513–526.

Izhikevich, E. M. (2003). Simple model of spiking neurons. IEEE Transactions on Neural Networks 14 (6), 1569–1572.

Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? IEEE Transactions on Neural Networks 15 (5), 1063–1070.

Miyake, S., Fukushima, K. (1984). A neural network model for the mechanism of feature-extraction. A self-organizing network with feedback inhibition. Biological Cybernetics 50, 377–384.

Rao, R. P. N., Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Computation 9, 721–763.

Royer, S., Paré, D. (2003). Conservation of total synaptic weight through balanced synaptic depression and potentiation. Nature 422, 518–522.

Spratling, M. W. (1999). Pre-synaptic lateral inhibition provides a better arcitecture for self-organizing neural networks. Network: Computation in Neural Systems 10, 285–301.

Spratling, M. W., Johnson, M. H. (2002). Pre-integration lateral inhibition enhances unsupervised learning. Neural Computation 14 (9), 2157–2179.

van Ooyen, A., Nienhuis, B. (1993). Pattern recognition in the neocognitron is improved by neuronal adaptation. Biological Cybernetics 70, 47–53.