Biol Cybern (2008) 98:597–618
DOI 10.1007/s00422-008-0236-y

ORIGINAL PAPER

Biological Cybernetics

Inhibition, not excitation, is the key to multimodal sensory integration

Paul Friedel · J. Leo van Hemmen

Received: 18 March 2008 / Accepted: 28 April 2008

Abstract Multimodal neuronal maps, combining input from two or more sensory systems, play a key role in the processing of sensory and motor information. For such maps to be of any use, the input from all participating modalities must be calibrated so that a stimulus at a specific spatial location is represented at an unambiguous position in the multimodal map. Here we discuss two methods based on supervised spike-timing-dependent plasticity (STDP) to gauge input from different sensory modalities so as to ensure a proper map alignment. The first uses an excitatory teacher input. It is therefore called *excitation-mediated learning*. The second method is based on an inhibitory teacher signal, as found in the barn owl, and is called *inhibition-mediated learning*. Using detailed analytical calculations and numerical simulations, we demonstrate that inhibitory teacher input is essential if high-quality multimodal integration is to be learned rapidly. Furthermore, we show that the quality of the resulting map is not so much limited by the quality of the teacher signal but rather by the accuracy of the input from other sensory modalities.

Keywords Neuronal maps · Multimodal integration · Optic tectum · Superior colliculus · STDP · Barn owl · Receptive field · Spatial transformations · Neuronal modeling

P. Friedel (✉) · J. L. van Hemmen
Physik Department T35, Technische Universität München,
85748 Garching bei München, Germany
e-mail: pfriedel@ph.tum.de

J. L. van Hemmen
e-mail: lvh@tum.de

P. Friedel · J. L. van Hemmen
Bernstein Center for Computational Neuroscience - Munich,
Munich, Germany

1 Introduction

Neuronal maps play a crucial role in sensorimotor information processing. A neuronal map is a neuronal representation of the outside sensory world. Quite often this representation is such that neighboring neurons respond to sensory stimuli that are near to each other (van Hemmen 2002; Knudsen et al. 1987).

One advantage of information processing in neuronal maps seems to be wiring efficiency. Generally, one can assume that neurons coding for related events or properties will be interconnected, either directly or indirectly. If neurons coding for similar locations in sensory space are near to each other, wiring costs are minimal. A beautiful example of this is the connection between sensory and motor maps. In the Superior Colliculus (SC), the connection between sensory and motor maps has been studied in great detail (Knudsen 2002; Stein and Meredith 1993; Stein et al. 2004). Here the motor maps lie on top of the sensory input maps and are in spatial register. That is, sensory input is used to directly generate appropriate motor output; for example, in the form of a gaze shift or eye saccade (van Opstal and Munoz 2004; Stein and Meredith 1993; Stein et al. 2004). If, say, an interesting event is detected at 20° to the right of the animal, a gaze-shift of exactly 20° can be rapidly induced.

Another advantage of maps is found in *multi*modal integration, where several sensory input modalities merge so as to form a unified map of multisensory space (reviews in Calvert et al. 2004; Stein and Meredith 1993; Stein and Stanford 2008). In a multimodal map, most neurons respond to more than one sensory modality. This kind of map is omnipresent in the brain (Kaas and Collins 2004) and it has even been questioned whether genuine monosensory maps exist at all (Stein and Stanford 2008).

A multimodal neuron is defined by its responsiveness to sensory input from more than one modality. Several types of interaction are possible if multiple input modalities are stimulated simultaneously (Stanford et al. 2005; Stein and Stanford 2008), but usually the presence of multiple input modalities leads to an increase in firing rate, improving object identification and localization accuracy. This means that the response to two or more simultaneous stimuli is larger than that to either stimulus when presented alone. Alternatively, a response depression may be found. In this case, the firing rate decreases if several modalities are activated concurrently. A neutral response is of course also possible. Then a neuron does respond to several different modalities, but there is no clear response enhancement or reduction.

A general feature of multimodal maps is that all sensory modalities are spatially aligned with each other and with motor maps. This means that the spatial response of the neuron is not primarily defined by the type of input that is present, but rather by the *spatial location* of that input. To ensure proper multimodal interaction, input from all participating modalities must be calibrated so that a stimulus at a specific spatial location is represented at an unambiguous location on the multimodal map. Alignment of sensory and motor maps is not present at birth and must be *learned* through experience (Knudsen 2002; Stein and Stanford 2008).

In this paper, we will answer the question as to how sensory maps can be aligned to form a calibrated representation of multisensory space. There are actually two problems to be solved here. The first problem is that of *dynamic* adaptation. Since many animals can move their sensory organs (eyes, ears, whiskers, etc.) independently of their body, continuous dynamical map alignment is needed. The mechanism underlying this adaptation is not known, but retinotopic coordinate systems seem to play a key role (Gardner et al. 2008). Dynamical alignment will not be considered here. The second problem has to do with the *development* of multimodal alignment. Even if it is clear how dynamic alignment can be achieved, the question remains how static alignment is to be achieved in the first place.

Directly below, we present a short overview of the relevant literature. In Sect. 2, two models are discussed that can explain alignment of multisensory maps through spike-timing-dependent plasticity (STDP). The first model uses excitatory learning input (the EL model) and the second uses inhibitory learning input (the IL model). A complete mathematical treatment of the learning process is possible, based on the methods developed by Kempter et al. (1999). Such an analysis is presented in Appendix A and the ensuing considerations are illustrated and analyzed in Sect. 3. In Sect. 4 we present numerical simulations. Finally, we discuss our results in Sect. 5.

We will see that both models are very robust with respect to parameter variation and noise, but inhibition-mediated learning (IL) is much faster and more robust than excitation-mediated learning (EL). Furthermore, the quality of the resulting map is not limited by the quality of the teacher signal, but rather by the accuracy of the input from the other sensory modality.

1.1 Mechanisms in the development of multimodal integration

For many animals it has been shown that visual input is essential in guiding the development of other sensory systems and multimodal integration. Early work was done on hamster and ferret. Mooney et al. (1987) have found that somatosensory spatial receptive fields in the SC of hamsters are disorganized following destruction of visual input pathways. King et al. (1988) have shown that neonatal rotation or lateral deviation of the eye can lead to abnormal auditory spatial maps in the SC.

Similar findings were obtained in the cat (Wallace and Stein 2007; Wallace et al. 2004), the clawed frog (Claas 1994), and in snakes (Grace et al. 2001). In psychophysical experiments using congenitally blind and normally sighted subjects, it was shown that in humans multimodal integration also partially depends on visual input in early life for accurate development (Hötting et al. 2004; Putzar et al. 2007; Röder et al. 2004). Audio-visual integration in barn owls has been studied in great detail (for example reviewed in Knudsen 2002). This case will be discussed at length below.

It has been shown recently (Jiang et al. 2007) that visual cortex long-term potentiation and depression in mice develops sequentially. Plasticity is gradually lost in development and layers that come later in the processing pathway loose their plasticity later than earlier processing layers. Apparently, the visual cortex layers only settle into their final state after earlier layers have stabilized. The work by Jiang et al. (2007) also seems to support the idea that neuronal map formation is often guided by stabilized visual "teacher" input.

Much of the work on multimodal integration has been done in the SC or its non-mammalian homologue, viz., the optic tectum (OT). Both are known to be a major site of multimodal integration. Inputs from several sensory systems converge in the SC and are used to initiate motor responses (van Opstal and Munoz 2004; Stein and Meredith 1993; Stein et al. 2004).

1.2 Audio-visual integration in the barn owl

In the barn owl, the integration of visual and auditory information has been studied particularly well. The barn owl is an interesting object of study since it is a highly specialized auditory hunter and, in addition, its eyes have a fixed position in the head, preventing any problems related to dynamical coordinate transformations. A variety of experiments,

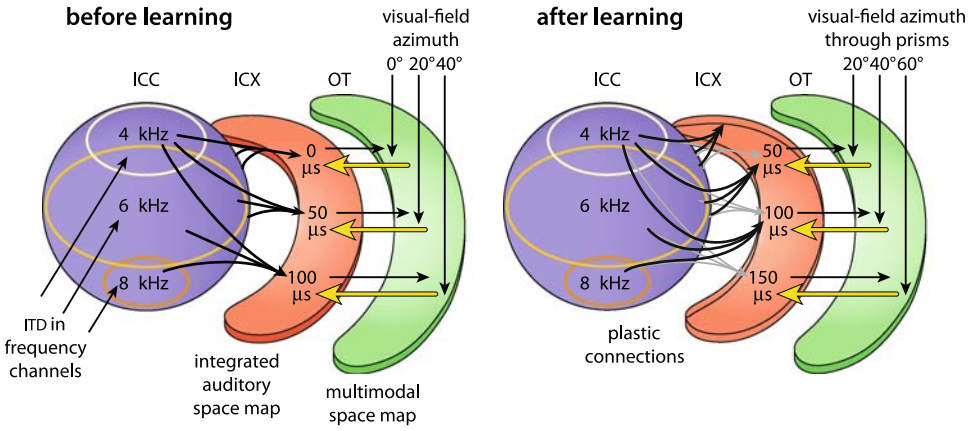

Fig. 1 Audio-visual integration in barn owls. In the "normal" situation (*left*), auditory positional information reaches the ICC in frequency-specific channels. Information on sound source localization is coded in the time-of-arrival difference (ITD) of the sound between the two ears. The frequency-specific maps from the ICC are integrated and relay to the ICX. From there, information is forwarded to the OT where it is integrated with visual input. There is a permanent feedback signal from the OT to the ICX (*open arrows*). If abnormal visual input is present (an azimuthal shift is introduced), the feedback signal is used to reorganize the synaptic connections between ICC and ICX so as to re-establish a match between the map in the ICX and the OT. Figure adapted from Knudsen (2002)

reviewed by Knudsen (2002), have shown that in barn owls integration of auditory and visual input is not functioning optimally in young hatchlings, but needs to be learned during several weeks of training in an environment with both visual and auditory stimulation.

Although it has been shown both theoretically and experimentally that an (imprecise) map of azimuthal sound source location can be learned in the absence of any visual input (Kempter et al. 2001b; Knudsen et al. 1991), a map of sound source elevation cannot be correctly learned without visual information during development. Proper integration of the auditory and visual modality is of course also impossible without visual input.

Input from the visual and auditory system in the barn owl converges in the OT. The last processing stages of the auditory pathway just before the OT are the central nucleus of the inferior colliculus (ICC) and the external nucleus of the inferior colliculus (ICX). Knudsen and his coworkers fitted barn owls with prisms displacing the view field of the animals. This was seen to lead to a change in the auditory input to the OT.

The synaptic connections between ICC and ICX changed in such a way that the auditory map in the ICX acquired the topography of the visual input (Fig. 1). The connections between ICX and OT did not change. That is, the multimodal map itself was not adapted, but rather the last input stage *before* multimodal integration was gauged to ensure proper alignment of the visual and auditory input to the multimodal map in the OT.

Furthermore, it has been shown that visual input provides a *topographic* teaching input to guide plasticity in the ICC–ICX pathway (Hyde and Knudsen 2001; Knudsen and Brainard 1991). This means that plasticity is not induced through a global "error signal" telling the neurons what to do, but rather on a per-neuron basis. It is even possible to induce differently-sized shifts in different parts of the map. Although the precise nature of this teaching signal has not been clarified, inhibitory gain control, or gating, seem to play an important role (Gutfreund et al. 2002; Winkowski and Knudsen 2006).

We will see below that the presence of inhibitory teacher input is essential for precise and quick tuning of input to multimodal maps.

2 Description of the models

2.1 Two possible models of map tuning

Recently, Davison and Frégnac (2006) presented a mechanism based on STDP that can account for map integration and coordinate transformation between different (sensory) maps. In their model, an array of output neurons receives excitatory input from an input modality and also from a set of teacher neurons. Because of the correlations between the firing rates of the input neurons and the teacher neurons, STDP causes strengthening of the "correct" input-to-output connections. In due time the output layer neurons learn to form a topographic spatial representation, even if the teacher input is removed later on.

Two different methods based on supervised STDP used to gauge input from different sensory modalities and ensure proper map alignment will be discussed here. The first model is very much like the one from Davison and Frégnac (2006)

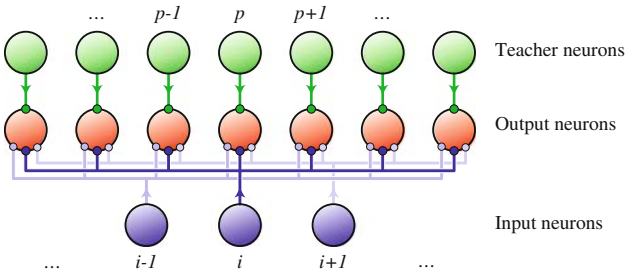

Fig. 2 Network topology of the models. There are three layers of neurons. The input layer (I, corresponding to ICC) and the teacher layer (T, corresponding to OT) fire at a rate that depends on the input position y. The output layer (O, corresponding to ICX) is driven by spikes from the input and teacher layer. The synaptic connections from input to output layer are subject to plastic weight changes and the connections from the teacher layer to the output layer remain fixed. Every output neuron receives spikes from one corresponding teacher neuron and from all input neurons

discussed above, using excitatory input and teacher neuron populations. The second model is inspired by the experimental results obtained in the barn owl and uses *inhibitory* teaching input instead of excitatory teaching input.

2.2 Network topology

For both models, we consider a network consisting of three neuron populations (Fig. 2). There is an *input* population (I) and a *teacher* population (T) that fire in response to an external sensory stimulus. These two populations are connected to an *output* population (O). The teacher neurons are topographically connected to the output neurons. Every teacher neuron connects to exactly one corresponding output neuron. At the beginning of the learning procedure the input population is connected to the output population through an all-to-all feedforward network of synapses. The connections between the teacher population and the output population can be either excitatory or inhibitory (see next section) and are considered as fixed. The synapses between the input population and the output neurons are subject to learning through STDP (see Sect. 2.4).

The teacher neurons exhibit a map-like response to an external sensory stimulus. That is, each teacher neuron (labeled by index p) has a preferred stimulus position x_p^T, that varies smoothly along the array of teacher neurons.[1] Without loss of generality, all stimulus positions are taken to lie between 0 and 1. The preferred stimulus position of a teacher neuron is then

$$x_p^T = \frac{p-1}{N^T-1}, \tag{1}$$

where $p \in [1, N^T]$ is the teacher neuron index, a positive integer, and N^T denotes the number of teacher neurons.

The input neurons (labeled by index i) are also tuned to specific directions, denoted by x_i^I. Although a topographical organization of the input neurons is mostly assumed here, this is certainly not a necessary condition (see Sect. 4.4).

Before learning starts, the synaptic connections are not yet structured and the output neurons do not have a directional preference. During learning, an output neuron acquires a preferred stimulus direction that is determined by the preferred direction of the teacher neuron connected to it. Since teacher neurons and output neurons are connected one-to-one, the output neurons are labeled by the same index (p) as the teacher neurons.

2.3 Sensory input and response of the output population

In the numerical simulations, all neurons are modeled as Poisson neurons; see Appendix B. Although this neuron model is very simple, it captures some very important aspect of neuron firing. First, neuronal firing is to some degree stochastic and, second, explicit time-dependence of firing probabilities is taken into account. Because of the simplicity of this neuron model, an exact mathematical understanding of the learning process is possible (Kempter et al. 1999).

It is important to note that the Poisson neuron model used here is *not* simply a rate coding. Since spike generation is explicitly taken into account, the theory and simulations are much richer than they would be with a pure rate-based modeling approach. In addition, the firing rate functions can have an arbitrary time dependence.

If sensory input is present at position $y \in [0, 1]$, the teacher and input neurons fire with a rate λ that depends on the neuron's preferred position

$$\lambda_i^I = A^I g(x_i^I, y) \quad \text{input population,} \tag{2a}$$
$$\lambda_p^T = A^T h(x_p^T, y) \quad \text{teacher population.} \tag{2b}$$

The constants A^I and A^T give the amplitude of the firing rate and g and h are normalized to have a maximal value of 1. The function g describes the receptive field properties of the input neurons. Typically, g depends on the difference between the input position y and the neuron's preferred position x_i^I, with its maximum for $x_i^I = y$.

As stated above, the teacher-output connections can be either excitatory or inhibitory.

[1] This paper restricts itself to one-dimensional maps. The approach can be generalized easily to include two-dimensional maps, but the computational costs of simulating learning in such a two-dimensional system become very large, due to the large number of synapses needed.

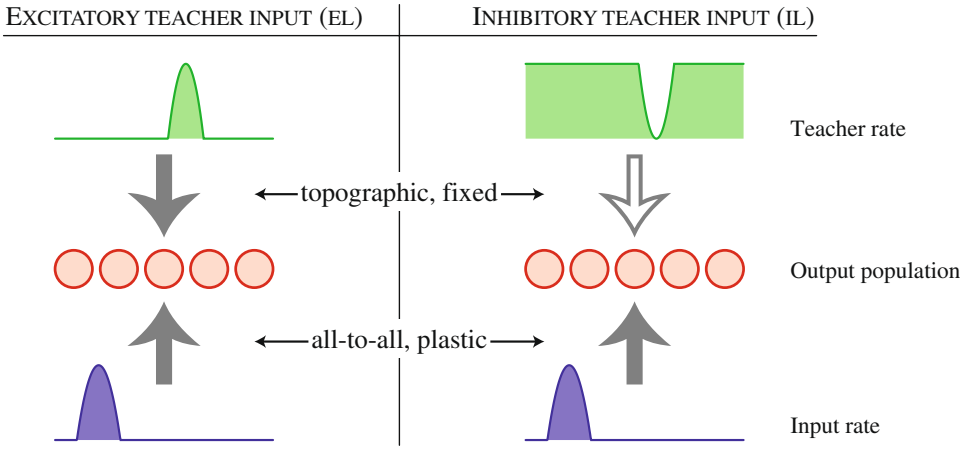

Fig. 3 Firing rates of input and teacher population. Both the input and the teacher population are organized in a map. For an external stimulus at a certain location, the firing rate strongly peaks at a well-defined location in the input and the teacher map. The input and teacher map can, however, be misaligned. The output population receives spikes from the input and from the teacher layer. *Left* In the EL model, the output neurons receive excitatory input from both the input and the teacher population. The output neurons corresponding to the teacher neurons with the highest rate will fire most and the connections from the active input neurons to this set of output neurons will therefore be strengthened by STDP. *Right* In the IL model, the output neurons are all silenced by the inhibitory teacher input, except for a small number of neurons receiving no inhibition. These neurons are driven by the input population and—as for the EL model—the connections between the active input neurons and the active output neurons are strengthened. Both models are able to form a high-quality map based on the input and the teacher signal. The resulting map is automatically aligned with that in the teacher population

2.3.1 Excitatory teacher input

In the excitatory case (EL), the function h behaves in the same way as described above for g, with a dependence upon the difference between input position and preferred position and maximal value for $x_p^T = y$ (Fig. 3, left). For a given stimulus position y, the output neurons corresponding to the teacher neurons with $x_p^T \approx y$ receive strong teacher input and the other output neurons remain silent. In addition, input from all input neurons will arrive at the output population. Since the input neurons with $x_i^I \approx y$ fire at a high rate, STDP selectively strengthens the connections between input neurons and output neurons that encode the same stimulus position y. After learning, a map will have been built in the output population. This map then responds correctly to the input population alone.

2.3.2 Inhibitory teacher input

In the case of inhibitory teacher input (IL), h behaves rather differently. The firing rate will be close to zero for teacher neurons with $x_p^T \approx y$ and maximal for neurons that code for any wrong position (Fig. 3, right). The output neurons are therefore strongly inhibited, except for a few neurons that correspond to the teacher neurons that are minimally active. This mechanism is called *selective disinhibition*. The output neurons that are not inhibited can fire freely and are now solely driven by the input population. Again, STDP potentiates connections between input neurons and output neurons that code the same position and weakens the wrong connections.

As will become clear below, both EL and IL allow for a topographical map to develop in the output population. This map is aligned with the teacher map and after learning the output neurons respond correctly to the input neurons even if no teacher signal is present. Still, the learning process in the two models differs in one important aspect. Because under IL guidance, the output population's firing is determined solely by the input population, the output neurons are trained to respond correctly to the input population *only*. In the case of EL, the output population learns to optimally react to a *combination* of input from the input population and the teacher population. This has two major consequences.

1. With IL, the firing rate of the output neurons is determined solely by the input population. This holds during training, but also if no teaching input is present anymore. That is, after an initial transient at the start of the learning process the output firing rate is always normalized; see for instance (Kempter et al. 1999, 2001a; Song et al. 2000). For EL, on the other hand, the output firing rate is determined by the input population *and* the teacher population. This means that the output rate may vary greatly, depending on whether teacher input is present or not.
2. If the temporal structure of the firing of the input neurons is nontrivial (viz., time coding rather than rate coding) it is advantageous to retain this temporal structure also in the output population. If the output neurons receive

excitatory input from both the input and the teacher population, the timing cues that are present in the input population spikes are "contaminated" by the simultaneous input from the teacher neurons. In the IL case the teacher input does not interfere with the input population signal. Any important temporal structure can therefore be retained in the output neuron firing statistics.

2.3.3 Output population firing rate

The rate function λ for the input population and the teacher population is given by Eqs. (2a) and (2b); see also Appendix B, in particular (47). The rate function of the output neurons depends on the firing of the input and teacher neurons. It is given by

$$\lambda_p^O(t) = \left[\sum_{i=1}^{N^I}\sum_f J_{ip}(t_i^f)\varepsilon^I(t-t_i^f) + J^T\sum_g \varepsilon^T(t-t_p^g)\right]^+. \tag{3}$$

The firing times of the input neurons are denoted by t_i^f and those of the teacher neurons by t_p^g. The strength of the synaptic connection from input neuron i to output neuron p is given by $J_{ip}(t)$, which is a function of time, and the strength of the connections between the teacher neuron p and the output neuron p is given by J^T. The postsynaptic response of the output neuron to an incoming spike from an input neuron is given by the kernel ε^I and ε^T is the response kernel for input from the teacher neurons. Since the teaching input may be negative ($J^T < 0$), the total rate function may become negative, which would not make sense for a probability density. To avoid this, half-wave rectification is applied, as indicated by $[\cdots]^+$. The operation of half-wave rectification is defined as

$$[f(x)]^+ := f(x) \to \begin{cases} 0 & f(x) < 0 \\ f(x) & f(x) \geq 0 \end{cases}. \tag{4}$$

2.4 Learning protocol

2.4.1 Learning trials

To numerically simulate the behavior of the models, the neurons are trained through many learning trials. In a single trial, an input position y is picked from a uniform distribution on $[0, 1]$ and the firing rates of the input and the teacher neurons in response to this input are calculated; see Sect. 2.6. The stochastic firing of the input and teacher population is then simulated during a trial of duration $\mathcal{T}$ and the spikes are sent to the output neuron population, which fires in response. The connections between input and output neurons are dynamically adjusted according to the learning rules specified below.

At regular times during learning, the quality of the map and the amount by which the weights have changed are measured as described in Sect. 2.5.

2.4.2 Plasticity rules

Learning in the brain occurs through spike-timing-dependent plasticity (STDP) (Bi and Poo 1998, 2001; Dan and Poo 2004; Gerstner et al. 1996; van Hemmen 2001; Kempter et al. 1999; Markram et al. 1997; Song et al. 2000; Zhang et al. 1998). Under normal circumstances and for excitatory synapses, presynaptic spikes that come before postsynaptic spikes lead to an increase in the synaptic weight (potentiation) and presynaptic spikes that arrive after a postsynaptic spike has been triggered lead to a weight decrease (depression). The learning window W (Gerstner et al. 1996) describes the dependence of the weight change on the difference between pre- and postsynaptic spike timing.

Many factors influence the precise result of any learning trial, but these influences do not radically alter the basic results of STDP-based learning without bells and whistles attached (Bi and Rubin 2005; Froemke et al. 2005; Lisman and Spruston 2005; Pike et al. 1999; Rubin et al. 2005; Tzounopoulos et al. 2004). This paper is therefore mainly restricted to learning with independent additive weight changes in dependence upon spike timing only. In Sect. 4.5 simulations are briefly discussed that take into account some of the parameters that influence learning in addition to spike timing.

The STDP rule used here consists of three parts,

1. If a presynaptic spike is emitted, the weight of the synaptic connection is changed by an amount ηw^{pre}.
2. If a postsynaptic spike is emitted, the weight of all synaptic input connections to this neuron changes by an amount ηw^{post}.
3. For every pair of pre- and postsynaptic spike, the synaptic efficacy changes by an amount $W(s)$ that depends on the difference $s = t_{\text{pre}} - t_{\text{post}}$ between the spiking times t_{pre} and t_{post}. The shape of the learning window is given by

$$W(s) = \begin{cases} \eta w^+ \dfrac{|s|}{(\tau^+)^2} e^{-|s|/\tau^+} & s < 0, \\ -\eta w^- \dfrac{s}{(\tau^-)^2} e^{-s/\tau^-} & s \geq 0. \end{cases} \tag{5}$$

The overall learning speed is set by the variable η, which should be small. The parameters w^{pre}, w^{post}, w^+ and w^- are all of order 1.

The synapses connecting input and output population have a starting weight of J_0^I and the weights are capped to always

stay between the values $J^I_{\min}$ and $J^I_{\min}$. The synapses between teacher and output population are *not* subject to learning.

Only nearest neighbor spike-pair interaction is considered. This seems biologically plausible, since each back-propagating action potential may be thought to reset the voltage in the dendritic spines. Moreover, only the first presynaptic spike after such a postsynaptic reset may be relevant due to saturation effects; see also Izhikevich and Desai (2003). The mathematical theory as developed in Kempter et al. (1999) is only valid for all-to-all spike pairings and for the sake of comparison, simulations are also carried out in which all spike pairs are taken into account; cf. Sect. 4.2.

2.5 Performance measures

To gain insight into the performance of the models, some measures are needed to assess the quality of the map in the output population and the speed of learning.

2.5.1 Map quality

To measure how well the output map performs for a given weight distribution the expected output firing rate without teaching input is calculated (see also Eq. (27)) for 100 input positions y_l ranging from 0 to 1 in equally-sized steps. The expectation of the output firing rate is

$$\nu^O_p(y) = \sum_i J_{ip}\nu^I_i(y). \tag{6}$$

In this equation, ν denotes a time-averaged firing rate over one learning trial of length $\mathcal{T}$. For each input position, the output neuron $p_{\max}$ with the highest firing rate is determined and the difference between its preferred position $x_{p_{\max}}$ and the input position y_l is obtained. The localization error E_{RMS} of the map is then the root-mean-square (RMS) value of this difference over all 100 input positions

$$E_{RMS} = \sqrt{\sum_l \frac{(x_{p_{\max}} - y_l)^2}{100}}. \tag{7}$$

A small value for E_{RMS} means that the map is of high quality.

2.5.2 Weight change and learning speed

The overall learning speed is set by the parameter η as described above. However, because the two models that are studied here have different connectivity properties, it is not possible to simply compare the performance of both networks for equal values of η. In the EL model, only a small number of teacher neurons are active simultaneously. In the IL model, a majority of the teacher neurons is always active; cf. Fig. 3. A measure of learning speed that is independent of this difference is therefore needed.

To obtain such a measure, the RMS distance d_{RMS} is defined as follows,

$$d_{RMS} = \sqrt{\sum_{i,p} \frac{[J_{ip}(t) - J_{ip}(t=0)]^2}{N^I N^T}}. \tag{8}$$

d_{RMS} measures the distance between the initial weight distribution and the weight distribution at some later time t. During learning d_{RMS} changes from 0 at $t = 0$ to some maximal value that is almost identical for both models. Let T be the time it takes to reach a value $d_{RMS} = 0.01$ (about 10% of its maximal value for the parameters used in this paper). Then we define the learning speed v_{learn} to be

$$v_{\text{learn}} := \frac{0.01}{T}. \tag{9}$$

Through the learning speed v_{learn} the EL and IL model performance can be compared in an *objective* way.

2.6 Specification of the model variables

The firing rate of the input and the teacher populations are given by a Gaussian dependence upon the difference between the input position y and the neuron's preferred position x^I_i or x^T_p.

For the EL model this means that the functions g and h in Eq. (2) are specified to be

$$\lambda^I_i(y) = A^I g(x^I_i, y) = A^I \exp\left[-\frac{(x^I_i - y)^2}{2(\sigma^I)^2}\right], \tag{10a}$$

$$\lambda^T_p(y) = A^T h(x^T_p, y) = A^T \exp\left[-\frac{(x^T_p - y)^2}{2(\sigma^T)^2}\right], \tag{10b}$$

where the Gaussian is just a specific choice, while for the IL model we take

$$\lambda^I_i(y) = A^I \exp\left[-\frac{(x^I_i - y)^2}{2(\sigma^I)^2}\right], \tag{11a}$$

$$\lambda^T_p(y) = A^T \left\{1 - \exp\left[-\frac{(x^T_p - y)^2}{2(\sigma^T)^2}\right]\right\}. \tag{11b}$$

The postsynaptic response of the output neurons to an incoming spike from the input population is taken to be an α function

$$\varepsilon^I(t) = \frac{t}{(\tau^I)^2}\exp(-t/\tau^I)\theta(t) \tag{12a}$$

while the response kernel for spikes from the teacher neurons is

$$\varepsilon^I(t) = \frac{t}{(\tau^T)^2}\exp(-t/\tau^T)\theta(t). \tag{12b}$$

The function θ is the Heaviside step function given by

$$\theta(t) := \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases}. \tag{13}$$

The preferred position of the teacher neurons is given by (1), viz.,

$$x_p^T = \frac{p-1}{N^T - 1}, \tag{14a}$$

and the preferred position of the input neurons is

$$x_i^I = \frac{i-1}{N^I - 1}. \tag{14b}$$

The learning speed parameter η varies for different simulations and its value is indicated where needed. In Table 1, the values of all further parameters are listed. Where different parameter values have been used, it is indicated in the text. It is important to realize that no parameter tuning is needed. Both the EL and IL models are robust against parameter variations and the values in Table 1 represent reasonable first guesses for the parameters.

Table 1 Model parameters as used in the analytical calculations and numerical simulations

Parameter	Value
Network layout	
Number of input neurons	$N^I = 100$
Number of teacher neurons	$N^T = 100$
Number of output neurons	$N^T = 100$
Initial strength of input→output synapses	$J_0^I = 0.1$
Minimal strength of input→output synapses	$J_{\min}^I = 0.0$
Maximal strength of input→output synapses	$J_{\max}^I = 0.25$
Strength of teacher→output synapses	J^T (EL) $= +1$
Strength of teacher→output synapses	J^T (IL) $= -1$
Postsynaptic response time to input spike	$\tau^I = 10$ ms
Postsynaptic response time to teacher spike	$\tau^T = 25$ ms
Learning parameters	
Learning trial length	$\mathcal{T} = 0.5$ s
Simulation time step	$\Delta t = 0.5$ ms
Weight change upon input spike	$w^{\text{pre}} = 1.5$
Weight change upon output spike	$w^{\text{post}} = -4.0$
Learning window potentiation amplitude	$w^+ = 4.0$
Learning window depression amplitude	$w^- = 1.0$
Learning window potentiation time	$\tau^+ = 20$ ms
Learning window depression time	$\tau^- = 40$ ms
Integral over learning window	$\widetilde{W} = 3\eta$ s
Integral over learning window and response kernel	$\overline{W} = 29.6\eta$
Input and teacher signal	
Input tuning curve amplitude	$A^I = 50\ \text{s}^{-1}$
Teacher tuning curve amplitude	$A^T = 100\ \text{s}^{-1}$
Input tuning curve width	$\sigma^I = 0.015$
Teacher tuning curve width	$\sigma^T = 0.025$

The value of the learning parameter η varies for different simulations. It is always explicitly specified where needed. The precise values of the parameters are not important. There is a wide range of parameters that enable stable learning; see also the discussion in Sect. 3

3 Mathematical analysis of learning

Taking advantage of the methods developed by Kempter et al. (1999) we can give an exact analytical description of the learning process. Such a description is admittedly somewhat technical. The problem is that the weight change of a single synapse is determined by the activity of both pre- and postsynaptic neuron. The postsynaptic activity is in turn determined by the activity of all presynaptic inputs and synaptic strengths inputting into this neuron. This means that the weight change of a synapse is interconnected with the weight changes of all other synapses contacting the same output neuron.

Here we only discuss the main results for the EL and IL model. A complete derivation of these results can be found in Appendix A.

3.1 Excitatory teacher input

To represent the evolution of all the synaptic weights J_{ip}, we write the weights as a vector $\mathbf{J}^{(p)}$. This vector contains all synaptic weights from the connections terminating at the output neuron with index p. As in (36), the rate of change of the components of this vector is given by a matrix-vector equation (cf. Eq. (36))

$$\mathrm{d}\mathbf{J}^{(p)}/\mathrm{d}t = \hat{\mathcal{A}}\mathbf{J}^{(p)} + \widehat{\mathbf{B}}^{(p)}. \tag{15}$$

The matrix $\hat{\mathcal{A}}$ has constant entries that do not depend on the output neuron p under consideration. The vector $\widehat{\mathbf{B}}^{(p)}$ is different for each output neuron p.

The entries in $\hat{\mathcal{A}}$ and $\widehat{\mathbf{B}}^{(p)}$ can be directly calculated from the parameters in Table 1. An important result is that the shape of the learning window is not important. Apart from the parameters $A^{I/T}$, $\sigma^{I/T}$, and $w^{\text{pre/post}}$ the entries are determined by

$$\widetilde{W} = \int_{-\infty}^{\infty} \mathrm{d}s\ W(s), \quad \text{and} \quad \overline{W} = \int_{-\infty}^{\infty} \mathrm{d}s\ W(s)\varepsilon^I(-s). \tag{16}$$

For the parameter set used here, the entries of $A^{I/T}$, $\sigma^{I/T}$, and $w^{\text{pre/post}}$ are explicitly given by

$$\hat{\mathcal{A}}_{[ij]} = \eta \left[-7.52 + 55.6\ \delta_{ij} + 199\ \exp\left(-\frac{(x_i^I - x_j^I)^2}{2\,(0.021)^2} \right) \right], \tag{17a}$$

$$\widehat{\mathbf{B}}^{(p)}_{[i]} = \eta \left[-22.2 + 484 \exp\left(-\frac{(x_i^I - x_p^T)^2}{2\,(0.029)^2} \right) \right]. \qquad (17b)$$

Note that all values are scaled by the learning parameter η. What, then, do Eqs. (15) and (17) mean?

For a wide range of parameter, the matrix $\hat{\mathcal{A}}$ has large positive values around the diagonal and small negative values away from the diagonal. The vector $\widehat{\mathbf{B}}^{(p)}$ has large positive values for $x_i^I \approx x_p^T$ and small negative values otherwise. If learning starts, the weight distribution is still unstructured. This implies that $\hat{\mathcal{A}}\mathbf{J}^{(p)} \approx \alpha \mathbf{J}^{(p)}$ for some number α that depends on the precise values of the parameters. $\hat{\mathcal{A}}$ causes all weights to increase or decrease by a fixed amount. For the parameters used here, α happens to be about zero and the initial weight evolution is dominated by $\widehat{\mathbf{B}}^{(p)}$. Even if weight evolution is *not* dominated by $\widehat{\mathbf{B}}^{(p)}$, $\hat{\mathcal{A}}$ does not cause structure formation during the initial stages of learning.

The vector $\widehat{\mathbf{B}}^{(p)}$ causes the weights for which $x_i^I \approx x_p^T$ to increase and the others to decrease. In this way, a seed of structure is generated and because of the diagonal form of $\hat{\mathcal{A}}$ this structure is further increased. At the end, all weights connecting input and output neurons with $x_i^I \approx x_p^T$ are maximal and all other weights vanish. Thus, the EL model indeed leads to correct structure formation and the development of a map in the output neuron population.

As stated above, we get the "right" learning for a wide range of parameters. Correct learning occurs because $\hat{\mathcal{A}}$ has a diagonal structure with negative entries off the diagonal. Similarly, $\widehat{\mathbf{B}}^{(p)} > 0$ for $x_i^I \approx x_p^T$ and $\widehat{\mathbf{B}}^{(p)} < 0$ otherwise. The conditions that must be satisfied by the parameters so as to ensure this behavior are as follows:

- $w^{\text{post}} < 0$. This condition must always be satisfied for weight stabilization to occur (Kempter et al. 1999).
- $w^{\text{pre}} > 0$. This condition makes biological sense, since synapses that receive much presynaptic input tend to be strengthened.
- $\overline{W} > 0$. This condition is always satisfied because of definition (16).
- $\widetilde{W}$ may be positive or negative. If it is negative, the absolute value may not be larger than $|\overline{W}|$. It needs, however, quite an extreme shape of the learning window for this to happen.
- $\sigma^{I/T} \ll 1$. This means that the input and teacher signal are localized sufficiently well. It is a mathematical statement of the fact that input and teacher signal are organized in a map-like fashion.

The above conditions are all rather weak. A very broad range of biologically realistic parameters can be chosen that satisfy them. The EL model is therefore very robust with respect to parameter variation but, as we will see shortly, it is not with respect to performance.

3.2 Inhibitory teacher input

For the IL model, an equation describing the weight change equivalent to Eq. (15) can also be given. However, the nonlinearity introduced by the rectification in Eq. (3) is problematic. For the EL model, this problem did not arise since the teacher weight J^T is positive and the output neurons' firing rate is already positive without rectification. It is possible to analytically treat the IL case but, to do so, we must take a slightly different form of teacher firing rate.

The input firing rate function is still given by (11a), but instead of being given by (11b), the teacher rate is taken to be

$$\lambda_p^T(y) = A^T \theta \left(\left| x_p^T - y \right| - \sigma^T \right). \qquad (18)$$

We now assume that A^T is very large. This means that the output neuron can fire freely if $\left| y - x_p^T \right| < \sigma^T$ (no teacher inhibition at all) and that the output neuron is completely silenced if $\left| y - x_p^T \right| > \sigma^T$.

Using this form of the teacher input and glancing (42), we can describe the weight evolution by

$$\mathrm{d}\mathbf{J}^{(p)}/\mathrm{d}t = \widehat{\mathcal{D}}^{(p)}\mathbf{J}^{(p)} + \widehat{\mathbf{E}} \qquad (19)$$

where the matrix $\widehat{\mathcal{D}}^{(p)}$ depends on the output neuron under consideration and the vector $\widehat{\mathbf{E}}$ is constant.

Just as above, the entries can be calculated as a function of the model parameters. The shape of the learning window is again found to be irrelevant and only $\widetilde{W}$ and $\overline{W}$ enter the equation. For the parameter values from Table 1 the matrix and vector components are

$$\begin{aligned} \widehat{\mathcal{D}}^{(p)}_{[ij]} = \eta \Big\{ & 27.8\,(\delta_{ij} - 0.135) \\ & \times \Big[\operatorname{erf}\left(47.1\,(x_p^T - x_j^I) + 1.18\right) \\ & - \operatorname{erf}\left(47.1\,(x_p^T - x_j^I) - 1.18\right) \Big] \\ & + 99.7 \exp\left[-\frac{(x_i^I - x_j^I)^2}{2\,(0.021)^2} \right] \\ & \times \Big[\operatorname{erf}\left(33.3\,(2x_p^T - x_j^I - x_i^I) + 1.67\right) \\ & - \operatorname{erf}\left(33.3\,(2x_p^T - x_j^I - x_i^I) - 1.67\right) \Big] \Big\}, \end{aligned} \qquad (20a)$$

$$\widehat{\mathbf{E}}_{[i]} = 2.82\,\eta. \qquad (20b)$$

Here "erf" denotes the error function. In contrast to the EL case the vector $\widehat{\mathbf{E}}$, being constant, plays no role in structure formation. The matrix $\widehat{\mathcal{D}}^{(p)}$ allows for more interesting behavior.

Although $\widehat{\mathcal{D}}^{(p)}$ looks impressive, its behavior is relatively simple. For a large absolute value of its argument, the error function becomes either -1 or 1. This implies that the differences between the error functions in $\widehat{\mathcal{D}}^{(p)}$ vanish if the

separation between x_p^T on the one hand and $x_{i/j}^I$ on the other hand is large enough. If $x_j^I \approx x_p^T$ holds, the value $\widehat{\mathcal{D}}^{(p)}$ is positive if $i \approx j$ and negative otherwise. Calculating $\widehat{\mathcal{D}}^{(p)}\mathbf{J}^{(p)}$ shows that $\mathbf{J}_{[i]}^{(p)}$ is increased if $x_i^I \approx x_p^T$ and decreased otherwise.

It is still possible that $\widehat{\mathbf{E}}$ interferes with the weight evolution by increasing all weights. Since $\widehat{\mathbf{E}}$ is directly proportional to the product of A^I and σ^I this simply means that the input neurons should not fire too strongly. That is, A^I and σ^I should not be too large. If they would, all weights would evolve towards their maximal value because the input neurons are completely dominating the firing of the output neurons in spite of the inhibition from the teacher map.

The remaining conditions under which the analysis presented here is valid are identical to those discussed for the EL model. If these conditions are fulfilled, a stable weight distribution evolves with strong connection between input and output neurons if their preferred positions match and weak connections if the preferred positions do not match.

The above mathematical analysis shows that the synaptic weights converge to the correct set, leading to an accurate spatial representation in the output neurons. However, this analysis is limited to the case of all-to-all spike interactions, which need not be biologically realistic; see Sect. 2.4.2. Fortunately, numerical tests have shown that dropping all-to-all interactions does not alter the results obtained in this paper. Below, results of numerical simulations are discussed. These results are then used to further analyze the performance of the EL and IL model.

4 Numerical simulations

We now turn to discussing results of numerical simulations. The learning procedure of Sect. 2 has been implemented numerically using the C++ programming language. In the first two subsections, the results for EL and IL are compared with each other and with the analytical results from Sect. 3. The most important results will be that inhibition-mediated learning is faster and more accurate than excitation-mediated learning.

The third subsection discusses coordinate transformations between the input and the output map and, in addition, restructuring of maps is analyzed. Although both excitation- and inhibition-mediated learning are able to handle a coordinate transformation, only the IL model is able to drastically restructure *existing* maps. We therefore study the inhibition-based model in more detail in the last two subsections (Sects. 4.4 and 4.5).

4.1 Comparing excitatory and inhibitory teacher input

At the start our discussion, it is useful to get a basic idea of the typical course of the learning process. In Fig. 4 the weight distribution at various times during learning is displayed for EL with learning parameter $\eta = 3 \cdot 10^{-7}$.

Horizontally, the preferred position of the input neurons is shown and, vertically, the preferred position of the output neuron as dictated by the learning population. The synaptic strength of the input–output connections is gray-coded. At the lower right of the figure one sees the evolution of the localization error defined in Sect. 2.5. Clearly visible in the figure is the strengthening of the correct synaptic connections starting from a homogeneous synaptic strength distribution.

Typically, E_{RMS} first drops to a very low value as the weights on the central diagonal get strengthened. As the diagonal band of strong weights broadens, the quality deteriorates to a small extend. Finally, the accuracy stabilizes at about 2%.

Figure 5 shows the RMS deviation (Eq. (8)) of the weights from their starting value and the evolution of the localization error (Eq. (7)) for several values of η.

Although the learning parameter η varies by a factor of 100, all learning trials show the same behavior. After an initial drop of the localization error the quality slowly climbs

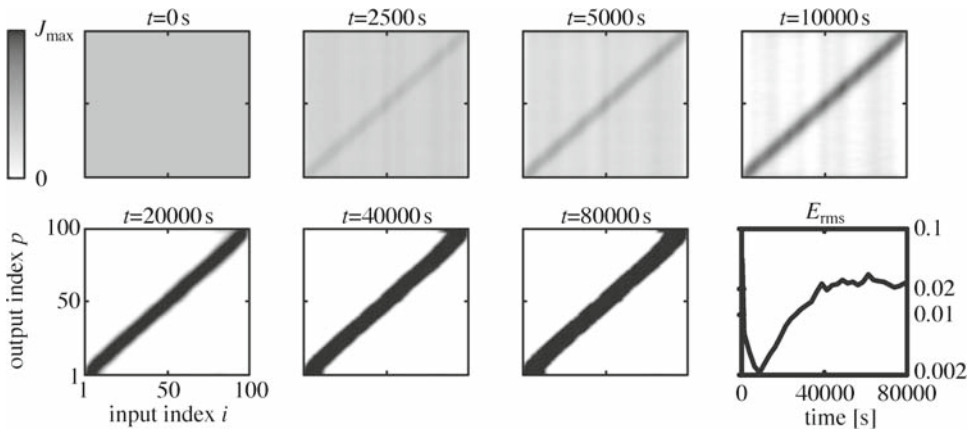

Fig. 4 Weight evolution for EL with $\eta = 3 \times 10^{-7}$. The development of a map-like organization is clearly visible as time progresses. The total learning time is in the order of days. The localization error quickly drops to a small value and then slowly approaches a steady state value, here about 2%

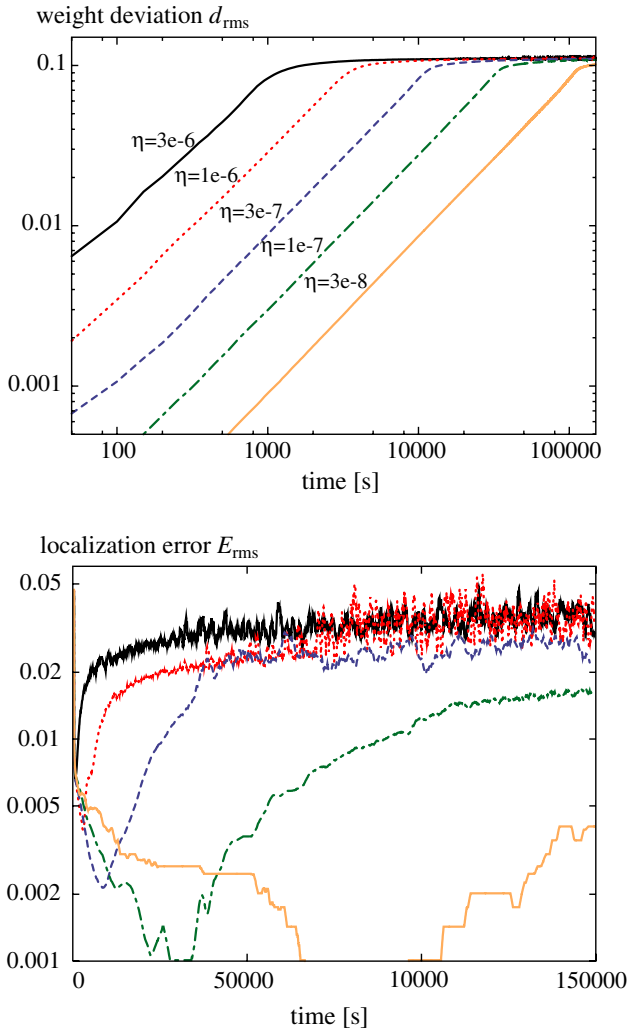

Fig. 5 The RMS deviation of the weights (*top*, according to (8)) and the RMS error (*bottom*, according to (7), with coding as in top) for several values of the learning parameter η in the EL case. The vertical scale (top & bottom) is logarithmic. For a large range of η values a steady increase in d_{RMS} is observed. The maximal value of d_{RMS} is always approximately 0.1. The final value of E_{RMS} decreases with the value of η, but even for relatively quick learning, the final error stays below 5%

towards its final value. If smaller values of η are used, the final map quality improves.

In Figs. 6 and 7 the same plots are shown for the case of inhibitory teacher input with $\eta = 3{\cdot}10^{-6}$. The results are very similar, but there is an important difference. For inhibitory teacher input, the localization error is much smaller than with excitatory teacher input, even if the learning speed is much larger. Just compare the time scales and the development of the error in Figs. 4 and 6.

As suggested by Figs. 4, 5, 6 and 7, the IL scheme is quicker and more accurate than the EL scheme. Figure 8 makes this difference explicit. If the final value of E_{RMS} is plotted against the learning speed as calculated by means of Eq. (9), there is a large difference between EL and IL. For IL, map quality is very good even for high learning speeds. In practice, the final map quality is reached within a couple of hours learning time. For EL, much slower learning is needed (in the order of days) for the model to remain stable and, even then the final map quality is poor as compared to the quality achieved by IL. Due to the finite number of neurons, the map quality cannot become arbitrarily small; there is a plateau at an error of $\approx 0.5\%$. Decreasing the learning speed any further to obtain a better map is impossible.

4.2 Comparison with analytical results

We now compare the analytical results from Sect. 3 with our numerical simulations. To arrive at the analytical result, the deterministic Eqs. (36) and (42) were integrated, taking into account that the weights must always stay within the range $[J^I_{\rm min}, J^I_{\rm max}]$. Figure 9 shows the comparison between analytical result and numerical simulation for both models.

The agreement between theory and numerical simulation is excellent, proving that the mathematical framework developed in Appendix A indeed provides a valid description.

Unfortunately, theory does not remain valid throughout the whole learning process. As soon as some of the weights reach the value $J^I_{\rm min}$ or $J^I_{\rm max}$, these weights are not allowed to decrease (increase) any further. This cutoff introduces a nonlinearity into the description of the weight change and the theory does not hold any more. The result is that the theoretical curve diverges from the numerical simulation results from $t \approx 2,500$ s for EL and $t \approx 400$ s for IL. Here the seconds are *formal* ones, depending on the timescale set by η.

4.3 Learning spatial transformations

Throughout this paper, it was tacitly assumed that the neuronal maps in the input and teacher population are similar. That is, both maps are organized in such a way that the preferred location increases smoothly from 0 to 1 from the leftmost to the rightmost neuron. In reality, this need not be the case.

A good example of such a flexible transformation is provided by the experiments of Knudsen and coworkers (Knudsen and Knudsen 1985; Knudsen and Knudsen 1990). If barn owls with normal visual and auditory capabilities are equipped with prism glasses that shift their visual field, the auditory map in the ICX changes to reflect this deviation. Below we have simulated a rather drastic version of this experiment. Whereas Knudsen et al. had only introduced a moderate shift in the visual input, we *completely* reversed the teacher input after learning has been completed and a stable map is formed.

The weight distribution after learning (see Fig. 6) is now the starting point for the simulations, but the teacher map will be inverted. That is, the teacher neurons now have a preferred position running from 1 to 0 instead of a preferred position

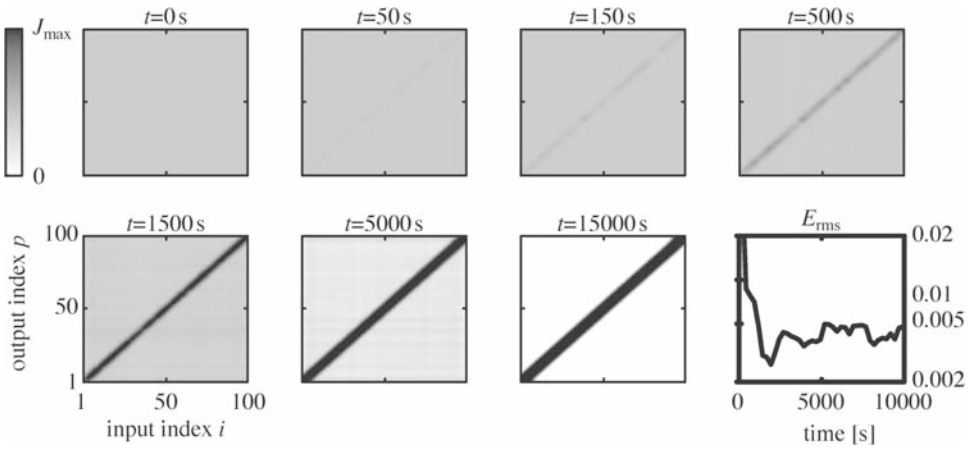

Fig. 6 Weight evolution for IL input with $\eta = 3 \cdot 10^{-6}$. As in the EL case, map formation is clearly visible. In contrast to Fig. 4, map evolution proceeds much faster than in the EL case. Only several hours of learning time are needed

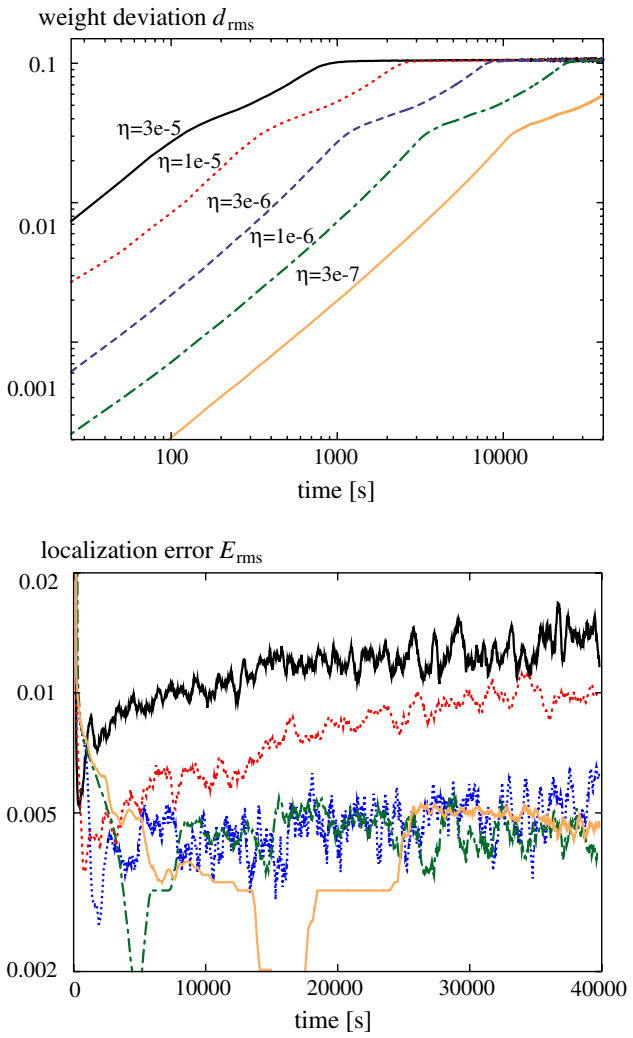

Fig. 7 The RMS deviation of the weights (*top*, according to (8)) and the RMS error (*bottom*, according to (7)) for several values of η in the IL case. Here too the vertical scales are logarithmic. The same picture as in Fig. 5 arises. The weight deviation converges to 0.1 and the localization error E_{RMS} decreases for smaller values of the learning speed. The final error ($< 2\%$ even for fast learning) is decidedly smaller than in the EL case

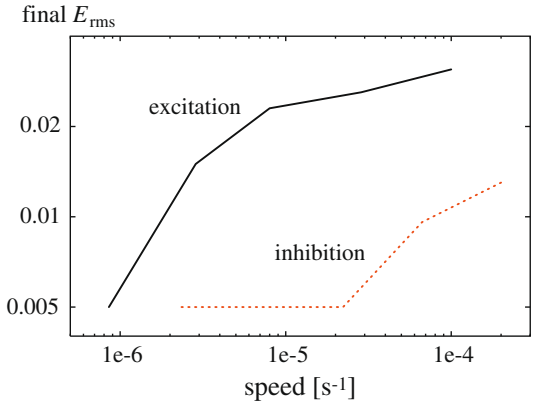

Fig. 8 Localization error versus learning speed for excitatory and inhibitory learning input (Eqs. (7) and (9)). The speed of the learning process basically measures the time it takes for the weight distribution to change significantly. Clearly, IL outperforms EL: at equal values of the learning speed, the quality of the map learned with the IL model is far better than the quality of the map formed with the EL model. This implies that the IL model is more robust than the EL model

running from 0 to 1. Based on the new teacher input, the output neurons must learn the correct transformation and the weights connecting input and output population must radically change in order to completely reverse the output map.

The result of this simulation using the IL model is displayed in Fig. 10. Clearly, dramatically altering an existing map is unproblematic with inhibition-mediated teacher input. The quality of the resulting inverted map is just as good as that of the original map.

For excitatory learning, the picture is quite different. Although the antidiagonal structure forms correctly, the original diagonal band never vanishes. This is shown in Fig. 11. Although there is no teacher input present to strengthen the old—now incorrect—connections, they stay in place solely because of the excitatory input they receive from the input population. If an inhibitory teacher signal is present, this ensures that the incorrect connections are actively weakened. Reorganizing the map is then unproblematic.

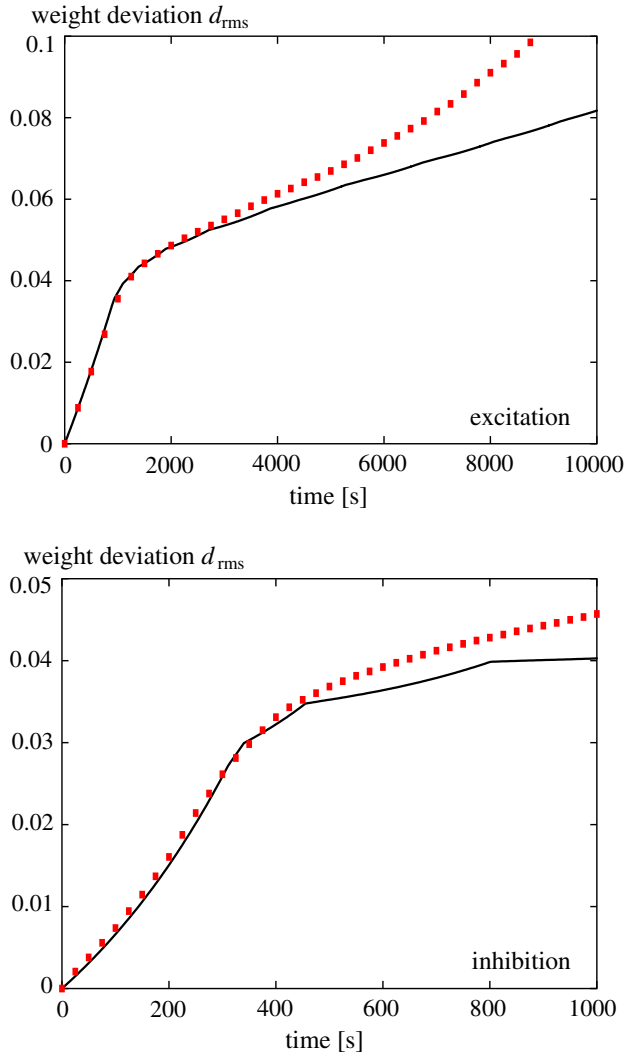

Fig. 9 Comparing analytical (*solid line*) and numerical (*squares*) results for EL (*top*) and IL (*bottom*). All spike pairs are taken into account instead of just neighboring spike pairs (see also the discussion in Sect. 2.4). For EL $\eta = 3 \cdot 10^{-7}$ and for IL $\eta = 3 \cdot 10^{-6}$. For both models, the analytical and numerical results fit very well. After the first synaptic strengths reach a value $J^I_{\max}$, the capping of the weights introduces a nonlinearity into the model and the theoretical description breaks down. This explains the deviation of the theoretical and numerical curves in these plots

Another example of a spatial transformation has been discussed by Davison and Frégnac (2006). They analyzed the transformation between two cortical maps, one of these containing the angle between upper and lower arm as measured by the proprioceptive system and another encoding the spatial position of the hand as perceived by the visual system. The coordinate transformation between these two maps is effectively a sine function. Davison and Frégnac showed that such a transformation can indeed be learned using their model, which is very much like the EL model discussed here. Below we demonstrate that such a transformation can also be achieved easily through the IL model. In this test the preferred position of the teacher population is not given by (14a), but instead by a sine function,

$$x^T_p = \frac{1}{2}\left[1 + \sin\left(2\pi \frac{p-1}{N^T - 1}\right)\right]. \tag{21}$$

The preferred position of the input population is still given by (14b). The result of the corresponding simulation is shown in Fig. 12. As was to be expected, learning a sinusoidal transformation is unproblematic in the IL model.

4.4 Influence of noise

The above comparison of EL and IL suggests that inhibitory learning is quite robust; even at high learning speed, stable learning occurs and radical alterations of the weight distribution are possible. Here the robustness of the IL model with respect to noise is tested in three separate test cases.

In the first test, noise is added to the firing rates of the input and teacher neurons as given by Eqs. (10) and (11). For each learning trial, when we present a stimulus at one particular position y, the input firing rate of each input neuron is changed according to

$$\lambda^I_i \to \lambda^I_i \cdot (1 + \chi_i) \tag{22}$$

with χ_i a Gaussian distributed stochastic variable with mean 0 and standard deviation 0.25. The teacher firing rate is changed in the same way.

In the second test, the initial weights are not all equal but instead the input weights J_{ip} are randomly chosen from a Gaussian distribution with mean J^I_0 and standard deviation $0.1 J^I_0$.

The third test assesses the influence of the absence of map-like organization in the input population. All input neurons do have a preferred position in this setup, but there is no topographical organization. Rather, the preferred direction of every input neuron is randomly taken from a uniform distribution on [0, 1].

The results of these tests are shown in Figs. 13 and 14. From Fig. 13 it is clear that a normal map is built if noise is added to the firing rates or if random input weights are used. This is confirmed by the localization error in Fig. 14. In both cases, the output neurons are able to form a high-quality map through STDP.

If the input neurons are not topographically organized the map quality may not become extremely high. Although most output neurons are strongly connected to the correct input neurons, some output neurons get input from the wrong neurons. This is due to the fact that some positions will be "overrepresented" and others will be "underrepresented". This may lead to a chain reaction in which wrong connections from an overrepresented input position start dominating. In the case shown here, the output positions $x^O_p \approx 0.03$ and $x^O_p \approx 0.32$ receive too much input from the wrong input

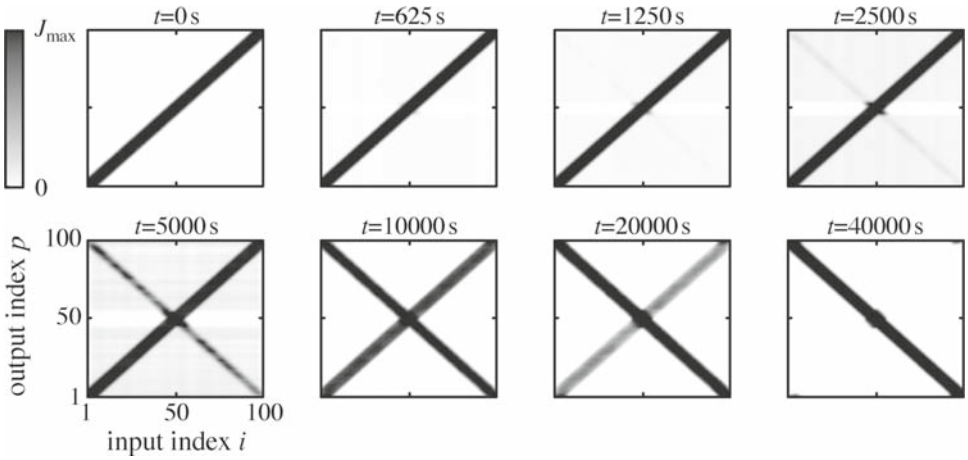

Fig. 10 Weight evolution after inverting the teacher map for the IL model. The weights change to form a radically altered map in the output population within just a couple of hours learning time. The quality of the newly developed map is nearly as good as the quality before restructuring

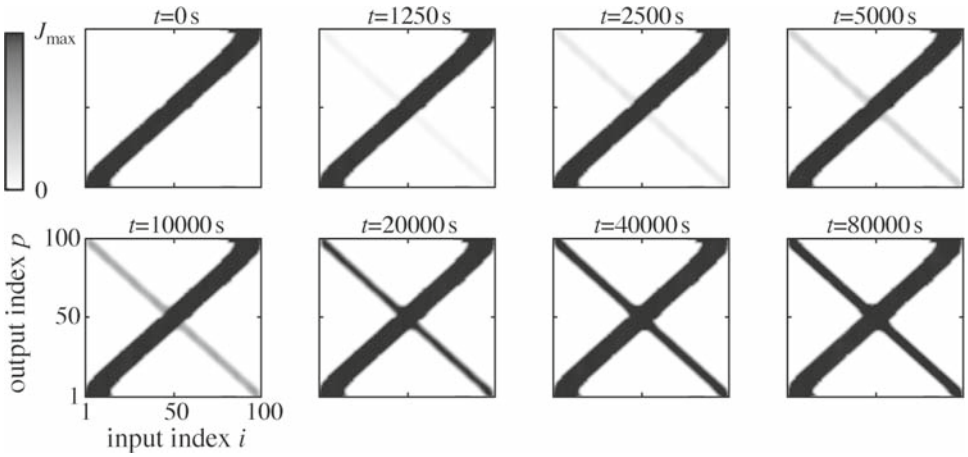

Fig. 11 Weight evolution after inverting the teacher map for the EL model. The weights change to form a new antidiagonal structure. The old main diagonal remains in place, however, and this leads to a useless map. Since there is no inhibition from the teacher map, the input population can sustain strong connections from the situation before reorganization

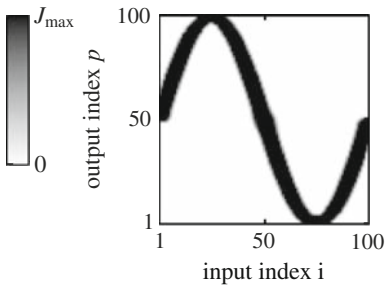

Fig. 12 Final weight distribution after a sine-transformation has been learned. Clearly, the IL model can be used to account for the learning of coordinate transformations between different modalities

neurons. This causes the poor quality of the output map. In Fig. 14 (right) the chain reaction can be recognized. At time $t \approx 18,000$ s the localization error suddenly increases. At this time, a stable seed with strong incorrect connections has formed.

If many more input neurons would be present, the change of some positions being under or overrepresented decreases and the map will have a much higher quality. If, in addition, learning is very slow, the chance that an initial seed of incorrect connections is formed is also smaller.

4.5 Influence of model parameters

As discussed in Sect. 2.4, linear and independent weight changes are generally used in this paper. This section further investigates the impact of the learning rules on the simulation results. In addition, the effect of varying the width of the tuning curves of the input and teacher neurons on the learning process is examined.

4.5.1 Multiplicative weight change

Instead of linearly adding weight changes, one might use a learning window in which the weight changes depend on the actual strength of the weight itself see e.g., van Rossum et al. (2000). A regularly used learning rule consists of taking a "multiplicative" learning window. This learning window has

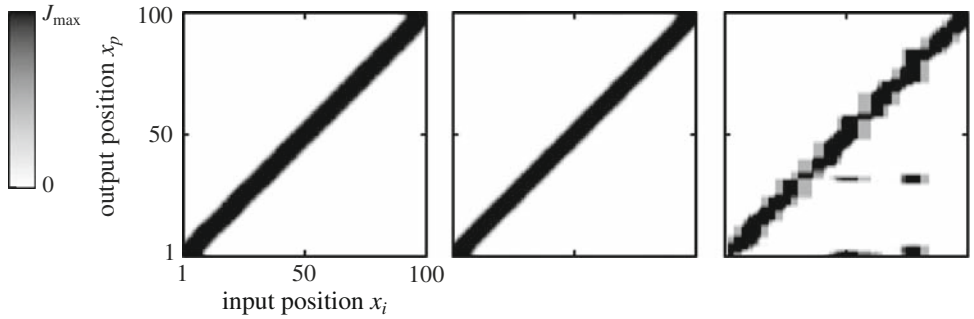

Fig. 13 Final weight distribution for several learning runs including noise. *Left* noise in input and teacher rate; *Middle*: random initial synaptic weights; *right* random preferred positions. In all cases the distribution of the synaptic strengths guarantees the existence of a neuronal map in the output layer, although the map is less precise if the input population is not organized in a map

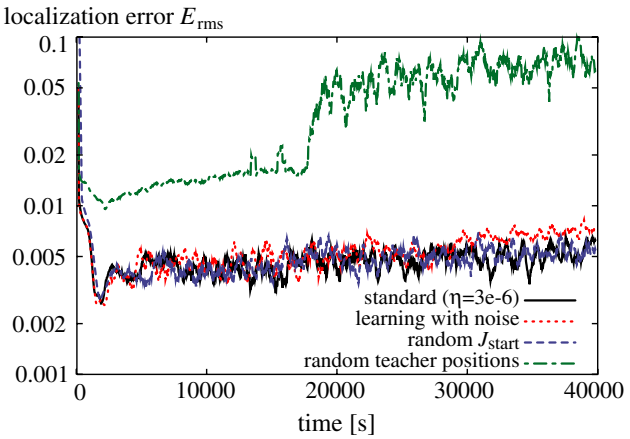

Fig. 14 Localization error E_{RMS} if sources of noise are taken into account. Noise in the input and teacher signals and random initial weights do not influence the final quality of the map. If the input neurons are not organized in a map, but rather have random preferred positions, quality is not so good. Still, a localization error below 10% is attained

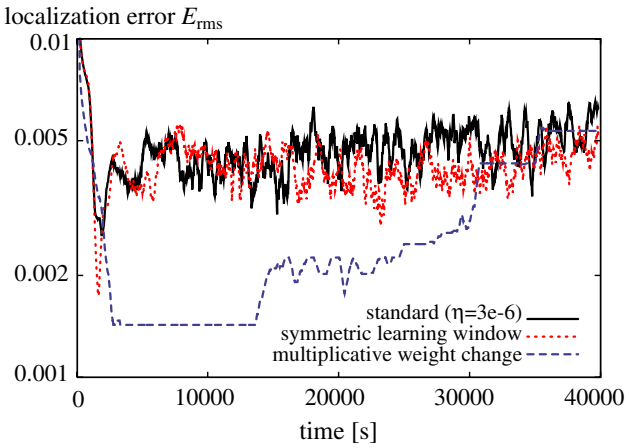

Fig. 15 Localization error E_{RMS} for different learning rules. The final localization error does not depend on the precise nature of the learning rule. This finding confirms the mathematical analysis, which predicts that only *integrals* over the learning window are needed to describe the learning process

the following form:

$$W(s) = \begin{cases} w^+ \dfrac{|s|}{(\tau^+)^2} e^{-|s|/\tau^+} \cdot (J^I_{\max} - J) & s < 0 \\ -w^- \dfrac{s}{(\tau^-)^2} e^{-s/\tau^-} \cdot J & s \geq 0 \end{cases} . \quad (23)$$

All parameters used here are identical to those given in Table 1. The learning window above ensures that the size of weight increases tends to zero as the weight approaches its maximum and that the amplitude of weight decreases becomes zero as the weight goes to zero. In this way a *soft bound* on the synaptic weight is achieved.

4.5.2 Symmetric weight change

Another possibility is to take a learning window that does not behave in a strictly Hebbian way (see overview in Abbott and Nelson 2000; Roberts and Bell 2002). Instead of increasing the weight if presynaptic spikes precede postsynaptic spikes and decreasing the weight if a postsynaptic spike precedes a presynaptic spike, a symmetric learning window is taken

$$W(s) = w^+ \frac{1}{\sqrt{2\pi}\tau^+} e^{-s^2/2(\tau^+)^2} - w^- \frac{1}{\sqrt{2\pi}\tau^-} e^{-s^2/2(\tau^-)^2} . \quad (24)$$

Again all parameters are as given in Table 1. Through this window, the weight is increased if pre- and postsynaptic spikes occur in close temporal distance and if the temporal distance is large the weight is decreased. Although this learning rule cannot be used to learn precise temporal structure, correlations between input and output signal can still be learned.

The results of simulations using these two alternative learning windows are shown in Fig. 15. There is no influence of altering the learning rules on the quality of the resulting map. Although the learning speed is different for the different rules, the final localization error approaches the same

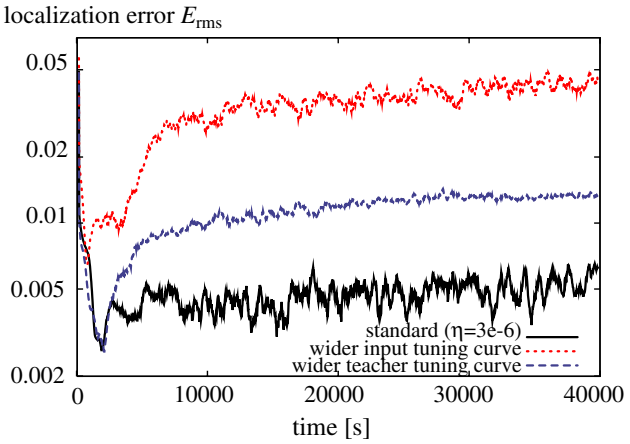

Fig. 16 Localization error E_{RMS} for different widths of the tuning curves. A twofold increase of the width of the teacher tuning curve leads to about the same increase in the localization error. If the width of the input tuning curve is doubled, the error increases approximately fivefold. The quality of the map in the output population thus depends more strongly on the quality of the input map than on the quality of the teacher signal

value. In Appendix A it is shown that indeed learning only depends on the integral over the learning window and on the convolution of the learning window with the output neuron postsynaptic response to input spikes. This simulation confirms that the precise shape of the learning window is irrelevant.

4.5.3 Tuning curve width

As a final numerical experiment, we have investigated the effect of the accuracy of the input and the teacher neurons. It would be expected that a wider tuning curve increases the learning speed (see (35) and (44)), and at the same time decreases the map quality. The behavior of the IL model has been tested for two cases.

In the first case the width of the input tuning curve (σ^I) was doubled, leaving all other parameters constant. In the second the teacher tuning curve width (σ^T) was doubled. The results are shown in Fig. 16.

The figure indicates a clear increase in the localization error if the tuning curve of the input or teacher neurons is wider. If the tuning curve of the teacher neurons is made twice as large, E_{RMS} increases by about the same factor. Changing the input tuning curve has a much larger effect. If σ^I is doubled, the localization error of the output map increases approximately five-fold. Perhaps surprisingly, the quality of the output map is therefore mainly determined by the accuracy of the input neurons and *not* by the precision of the teacher signal.

5 Discussion

5.1 Inhibitory or excitatory teacher input?

In this paper we have discussed two different possibilities to gauge neuronal maps through STDP, viz., through an excitatory and an inhibitory teacher system. Both models can account for map formation based on a topographic teacher template. The first model that suggests itself, using excitatory learning input (EL), is clearly not the most efficient learning paradigm. If inhibitory teacher input (IL) is used the quality of the map and the learning speed can be expected to be much better. This is because not only strengthening of the correct synaptic connections between input and output layer occurs, but the incorrect connections are also actively suppressed.

Regarding the EL model, the results in this paper confirm the work done by Davison and Frégnac (2006). Robust map formation is possible and the learning of spatial transformations is no problem. Restructuring of existing maps was not considered by Davison and Frégnac, but we have shown here that radically altering a completely developed map is not possible by means of an excitatory teacher signal.

The IL model is robust against noise (Sect. 4.4) and learning is flexible with regard to parameter variation (Sect. 4.5). Spatial transformations can easily be learned and completely developed maps can be radically restructured, as has also been demonstrated biologically (Sect. 4.3, compare Knudsen and Knudsen 1985; Knudsen and Knudsen 1990). This property of the model matches the finding that neuronal maps retain some degree of plasticity even after the developmental stage (Kaas 1991; Knudsen and Knudsen 1990; Linkenhoker and Knudsen 2002).

Although such an inhibition-mediated learning paradigm need of course not be realized in the brain at every location where neuronal maps are learned, evidence from the barn owl brain (Gutfreund et al. 2002; Hyde and Knudsen 2001; Knudsen and Brainard 1991) strongly suggests that at least in this animal a topographic and inhibitory teacher signal is used.

Especially if map changes must be radical and rapid, the learning algorithm must be very stable. The IL mechanism fulfills this requirement. A complete reorganization of the map such as shown in Fig. 10 poses no problem for the IL model but is *not* possible through the EL scheme. With excitatory teacher input it is of course possible to strengthen new synaptic connections, but since the wrong connections are not actively weakened they will continue to exist alongside the newly strengthened ones.

From the work presented here it can therefore be predicted that especially if the spatial accuracy and the learning speed of neuronal maps need to be high, inhibition-mediated learning can be expected to play a key role, whereas

excitatory-based learning may still do a good job if demands are not so high.

5.2 Influence of learning rules

Although Hebbian-style plasticity can be implemented in several ways, STDP is the only *local* mechanism that prevents runaway firing rates and at the same time is able to learn precise temporal input correlations (Abbott and Nelson 2000; Gerstner et al. 1996; Kempter et al. 1999, 2001a; Song et al. 2000). As argued above in Sect. 2.4.2, incorporating more complicated behavior into the learning rules is not expected to change the results; see Sects. 2.4 and 4.5 and also (Pfister and Gerstner 2006; Standage et al. 2007).

By using a learning protocol that is radically different from the one proposed in this paper, it might be possible to obtain good results for EL-like models. A possibility is to include global inhibition or a constant decay term in the weight evolution. Another option is to use a learning rule that depends on post-synaptic bursting (Pike et al. 1999). A third possibility is to use a negative value for w^{pre}. We do not feel that these scenarios are satisfying. The first "solution" violates a very important learning principle. STDP is a *local* mechanism and constant inhibition or weight decay terms destroy this desirable feature of our models. The second and third possibility to "fix" the EL model require *precise* parameter tuning, which is biologically unrealistic.

Another question of interest is whether to use all-to-all spike pairings or just neighboring spike pairs. It was argued above that using just neighboring spike pairs is biologically more plausible; see also discussion in Izhikevich and Desai (2003). Taking, however, all-to-all instead of finite range spike pairings does not change the results of the simulations. Learning proceeds a bit faster if all spike pairs are taken into account, but the map quality again stabilizes at the same level as in the standard case (data not shown).

The shape of the learning window does not strongly influence the learning process. The mathematical analysis has shown that only the integral over the learning window and the convolution of the learning window with the postsynaptic response of the output neurons to input spikes play a role in the dynamics, as is highlighted by (16) and confirmed by numerical simulations (Fig. 15).

The fact that the precise shape of the learning window is unimportant has to do with the lack of temporal structure in the input and teacher signal. If for a given input site y the input and teacher rate are constant—as chosen here—the output rate (27) will be constant as well. If the firing rates have a richer temporal structure, the mathematical description from Appendix A breaks down. Preliminary simulations suggest, however, that the output population can learn a map based on temporal features of the input signal. Obviously, this only works for inhibitory teacher input, since excitatory teacher input would interfere with the signal from the input neurons and the temporal structure would be destroyed. Accordingly, the precise form of the learning window *does* play an important role in learning.

The mathematical discussion explicitly shows the dependence of the learning process upon the model parameters (Sect. 3 and Appendix A). Especially the width and the amplitude of the tuning curves of the input and teacher neurons are important (as is clear from Eqs. (35) and (44)). A wide and high-amplitude tuning curve allows for faster learning, but of course the resulting map quality is worse, as is brought out by Fig. 16.

The effect of the tuning curve width on the accuracy of the resulting map is different for the teacher and the input populations. The dominant influence in determining map accuracy is not the tuning width of the teacher neurons but rather the precision of the input neurons. It should be possible to experimentally verify this prediction by designing experiments in which learning takes place while the response of one of the input modalities is artificially blurred. For vision and audition this could for example be achieved by equipping animals with distorting glasses or ear plugs.

5.3 Origin of the teacher map

One of the questions that remains to be answered is why the visual system always seems to function as a teacher for the other modality. An important property of the visual system is that a spatial map is automatically induced through the input mapping upon the retina *by a lens*. The retina is therefore *intrinsically* ordered in a map-like fashion. Activity waves spreading over the retina have been found in the embryos of many species (Wong 1999). Such retinal waves, occurring regularly on a time scale of seconds to minutes, provide coherent input to the visual system, aiding the development of high-quality spatial maps. Even if the animal does not use its eyes, or opens them relatively late in development (as many mammals do), a spatial template from the visual system would be present to guide multimodal integration.

As early as 1976, Willshaw and Von der Malsburg (1976) showed that layers of neurons can self-organize into topographical maps, provided that a small set of correctly-organized neurons exists. In this way, the intrinsic topography from the retina might dictate the organization and alignment of all multimodal maps in the brain.

Acknowledgments The authors thank Matthieu Gilson, Andreas Sichert, and Christine Voßen for useful comments on the manuscript. P.F. received funding from the Deutsche Forschungsgemeinschaft (DFG grant no. He 3252/4) and the Bernstein Center for Computational Neuroscience – Munich.

A Analytical description of multimodal sensory integration

We now analyze the learning process mathematically. In so doing, we follow the methods developed by Kempter et al. (1999). The first step is to focus on a single stimulus that is applied at position y. The evolution of the synaptic weights is given by the *learning equation*. For the network architecture used in this paper it reads

$$\mathrm{d}J_{ip}(t)/\mathrm{d}t = w^{\mathrm{pre}}\nu_i^I(y) + w^{\mathrm{post}}\nu_p^O(y) + \int_{-\infty}^{\infty} \mathrm{d}s\; W(s)C_{ip}(s,t,y). \tag{25}$$

The learning parameters w^{pre}, w^{post} and the learning window W have been defined in Sect. 2.4. The quantities ν_i^I and ν_p^O are the time-averaged expectation values of the input rate and the output rate, respectively. $C_{ip}(s,t)$, finally, is the time-averaged expectation value of the correlation between the spike trains of input neuron i and output neuron p,

$$C_{ip}(s,t) = \overline{\left\langle S_i^I(t+s)S_p^O(t)\right\rangle} \tag{26}$$

where the brackets denote the expectation value over stochastic realizations of the Poisson process and the overbar denotes the time average over one learning trial of length $\mathcal{T}$. Furthermore, S_i^I denotes the spike train of the input neuron i and S_p^T is the spike train of the teacher neuron p.

The time-averaged mean input firing rate is just $\nu_i^I(y) = A^I g(x_i^I, y)$ from (2a). The time-averaged mean output rate is

$$\begin{aligned}\nu_p^O(y) &= \overline{\left[\sum_i J_{ip}(t)\int_0^\infty \mathrm{d}t'\,\varepsilon^I(t')\lambda_i^I(t-t') + J^T\int_0^\infty \mathrm{d}t'\,\varepsilon^T(t')\lambda_p^T(t-t')\right]^+}\\ &= \left[\sum_i J_{ip}(t)\nu_i^I + J^T\nu_p^T\right]^+.\end{aligned} \tag{27}$$

The second equality holds because the firing rate functions $\lambda^{I/T}$ are constant in this case and $\int_0^\infty \mathrm{d}t\;\varepsilon^{I/T} = 1$.

The expectation of the correlation between input and output spike train is more difficult to obtain. The definition results in

$$\left\langle S_i^I(t+s)S_p^O(t)\right\rangle = \left\langle S_i^I(t+s)\left[\sum_j J_{jp}S_j^I(t) + J^T S_p^T(t)\right]^+\right\rangle. \tag{28}$$

Depending on whether the teacher input is excitatory ($J^T > 0$) or inhibitory ($J^T < 0$), a different result is obtained for the right-hand side of (28).

A.1 The learning equation for excitatory teacher input

A.1.1 Form of the learning equation

When the teaching input is excitatory ($J^T > 0$), the rectification in (28) does nothing since the argument is already positive. The correlation can then be calculated straightforwardly (for detailed methods see Kempter et al. 1999), giving

$$C_{ip}(s,t) = \nu_i^I\left(J^T\nu_p^T + J_{ip}(t)\varepsilon^I(-s) + \sum_j J_{jp}(t)\nu_j^I\right). \tag{29}$$

The learning Eq. (25) then becomes

$$\begin{aligned}\mathrm{d}J_{ip}/\mathrm{d}t = &\sum_j \underbrace{\nu_j^I\left(w^{\mathrm{post}} + \widetilde{W}\nu_i^I + \delta_{ij}\overline{W}\right)}_{A_{ij}} J_{jp}\\ &+ \underbrace{w^{\mathrm{pre}}\nu_i^I + J^T(w^{\mathrm{post}} + \widetilde{W}\nu_i^I)\nu_p^T}_{B_{ip}}\end{aligned} \tag{30}$$

with the definitions

$$\widetilde{W} = \int_{-\infty}^{\infty}\mathrm{d}s\; W(s), \quad \overline{W} = \int_{-\infty}^{\infty}\mathrm{d}s\; W(s)\varepsilon^I(-s). \tag{31}$$

For every output neuron p, the result is a matrix-vector equation describing the weight evolution of its input connections:

$$\mathrm{d}\mathbf{J}^{(p)}/\mathrm{d}t = \mathcal{A}\mathbf{J}^{(p)} + \mathbf{B}^{(p)}. \tag{32}$$

The vector $\mathbf{J}^{(p)}$ contains all the input connections to the output neuron p and the components of the matrix $\mathcal{A}$ and the vector $\mathbf{B}^{(p)}$ are taken from (30). The notation stresses that $\mathbf{B}^{(p)}$ depends on which output neuron p is considered. $\mathcal{A}$ is the same for all output neurons.

The matrix $\mathcal{A}$ and vector $\mathbf{B}^{(p)}$ give the response of the system to sustained input at some particular input position y. To find the average learning result over all positions, the expectation value of (32) with respect to y is to be calculated. Such an expectation value, denoted by a hat, is given by

$$\hat{f} = \int_0^1 \mathrm{d}y\;\rho(y)f(y), \tag{33}$$

where $\rho(y)$ specifies the probability distribution of the variable y. Taking this average is allowed if learning proceeds slowly enough.

The assumption that all positions are equally likely (justified by the learning protocol, see Sect. 2.4) gives $\rho(y) = 1$ and for the components of $\mathcal{A}$ and $\mathbf{B}^{(p)}$ this leads to

$$\hat{\mathcal{A}}_{[ij]} = \left(w^{\text{post}} + \delta_{ij}\overline{W}\right)\xi_j^1 + \widetilde{W}\xi_{ij}^2, \tag{34a}$$

$$\widehat{\mathbf{B}}_{[i]}^{(p)} = w^{\text{pre}}\xi_i^1 + w^{\text{post}}J^T\xi_p^3 + \widetilde{W}J^T\xi_{ip}^4 \tag{34b}$$

where δ_{ij} is the Kronecker delta ($\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$) while

$$\xi_i^1 = \int_0^1 \mathrm{d}y\ \nu_i^I(y), \tag{35a}$$

$$\xi_{ij}^2 = \int_0^1 \mathrm{d}y\ \nu_i^I(y)\nu_j^I(y), \tag{35b}$$

$$\xi_p^3 = \int_0^1 \mathrm{d}y\ \nu_p^T(y), \tag{35c}$$

$$\xi_{ip}^4 = \int_0^1 \mathrm{d}y\ \nu_i^I(y)\nu_p^T(y). \tag{35d}$$

The complete evolution of the weights for excitatory learning input is thus described by

$$\mathrm{d}\mathbf{J}^{(p)}/\mathrm{d}t = \hat{\mathcal{A}}\mathbf{J}^{(p)} + \widehat{\mathbf{B}}^{(p)} \tag{36}$$

together with (34) and (35).

A.1.2 Interpretation of the learning equation

The complete Eq. (36) can be solved exactly through Duhamel's formula,

$$\mathbf{J}^{(p)} = \mathcal{S}\exp(\mathbf{\Lambda}t)\mathcal{S}^{-1}\left(\mathbf{J}_{t=0}^{(p)} + \hat{\mathcal{A}}^{-1}\widehat{\mathbf{B}}\right) - \hat{\mathcal{A}}^{-1}\widehat{\mathbf{B}}. \tag{37}$$

In this solution, $\mathcal{S}$ is the matrix containing as columns the eigenvectors of $\hat{\mathcal{A}}$ and $\mathbf{\Lambda}$ the matrix containing the eigenvalues of $\hat{\mathcal{A}}$ on its diagonal. The matrix $\mathcal{S}$ diagonalizes $\hat{\mathcal{A}}$ and therefore $\mathcal{S}^{-1}\hat{\mathcal{A}}\mathcal{S} = \mathbf{\Lambda}$.

Unfortunately, this solution is of limited use, since the maximal and minimal values of $\mathbf{J}^{(p)}$ are not taken into account. It therefore makes more sense to focus on the values of the ξ's in (35) to understand how the learning Eq. (36) behaves. Because the input neurons and the teacher neurons are tuned to specific positions, the averaged firing rates $\nu_i^I(y)$ and $\nu_p^T(y)$ are large if the neuron's preferred position matches y ($x_i^I \approx y$ or $x_p^T \approx y$) and zero if the positions do not match. If the width of the tuning curves is not too large, the integrals in (35) can be extended to range from $-\infty$ to ∞, since only a small domain lying completely within [0, 1] will give a significant contribution anyway.

The value of ξ_i^1 is then just a constant independent of i that is given by the total area under the tuning curve of the input neurons. Similar reasoning applies to ξ_p^3. The value of ξ_{ij}^2 sharply peaks for $x_i^I \approx x_j^I$ (meaning that $i \approx j$) and if the input neurons are organized topographically, this implies that ξ_{ij}^2 is a square matrix with large values around the main diagonal and values very close to zero elsewhere. The same reasoning applies to ξ_{ip}^4. The matrix $\{\xi_{ip}^4\}$ has large entries when $x_i^I \approx x_p^T$ and very small values if this is not the case.

For the input and teacher firing rates defined in (10) the following expressions are found, which are discussed in Sect. 3,

$$\xi_i^1 = A^I\sigma^I\sqrt{2\pi}, \tag{38a}$$

$$\xi_{ij}^2 = (A^I)^2\sigma^I\sqrt{\pi}\exp\left(-\frac{(x_i^I - x_j^I)^2}{4\,(\sigma^I)^2}\right), \tag{38b}$$

$$\xi_p^3 = A^T\sigma^T\sqrt{2\pi}, \tag{38c}$$

$$\xi_{ip}^4 = A^I A^T\frac{\sigma^I\sigma^T}{\sqrt{(\sigma^I)^2+(\sigma^T)^2}}\sqrt{2\pi}\exp\left(-\frac{(x_i^I - x_p^T)^2}{2\,[(\sigma^I)^2+(\sigma^T)^2]}\right). \tag{38d}$$

A.2 The learning equation for inhibitory teacher input

A.2.1 Form of the learning equation

If $J^T < 0$ the rectification in (28) leads to problems. Since rectification is a nonlinear operation, it cannot simply be pulled out of the expectation value.

This problem can be solved by taking a slightly different form for the teacher signal. Instead of taking (11b) we assume that inhibition is very large for inputs $\left|y - x_p^T\right| > \sigma^T$ so that the output neuron is perfectly silent. For $\left|y - x_p^T\right| < \sigma^T$ the teacher signal vanishes completely and the output signal is only determined by the positive contributions from the input neurons. This is called *shunting inhibition*. Then the correlation term becomes

$$\left\langle S_i^I(t+s)S_p^O(t)\right\rangle = \left\langle S_i^I(t+s)\sum_j J_{jp}S_j^I(t)\right\rangle\theta\left(\sigma^T - \left|x_p^T - y\right|\right), \tag{39}$$

with θ as the Heaviside step function. The temporally averaged mean is given by (using again the methods outlined in Kempter et al. 1999)

$$C_{ip}(s,t) = \nu_i^I\left(J_{ip}(t)\varepsilon^I(-s) + \sum_j J_{jp}(t)\nu_j^I\right)\theta\left(\sigma^T - \left|x_p^T - y\right|\right). \tag{40}$$

The learning Eq. (25) now becomes

$$\mathrm{d}J_i^{(p)}/\mathrm{d}t = \sum_j \underbrace{\nu_j^I\left(w^{\text{post}} + \widetilde{W}\nu_i^I + \delta_{ij}\overline{W}\right)\theta\left(\sigma^T - \left|x_p^T - y\right|\right)}_{D_{ijp}} J_{jp} + \underbrace{w^{\text{pre}}\nu_i^I}_{E_i}. \tag{41}$$

The position-averaged learning equation can be written

$$\mathrm{d}\mathbf{J}^{(p)}/\mathrm{d}t = \widehat{\mathcal{D}}^{(p)}\mathbf{J}^{(p)} + \widehat{\mathbf{E}}, \tag{42}$$

and the expectation values over the input y are now

$$\widehat{\mathcal{D}}_{[ij]}^{(p)} = \left(w^{\text{post}} + \delta_{ij}\overline{W}\right)\zeta_{jp}^1 + \widetilde{W}\zeta_{ijp}^2, \tag{43a}$$

$$\widehat{\mathbf{E}}_{[i]} = w^{\text{pre}}\zeta_i^3 \tag{43b}$$

where

$$\zeta_{jp}^1 = \int_0^1 \mathrm{d}y\ \nu_j^I(y)\,\theta\left(\sigma^T - \left|x_p^T - y\right|\right), \tag{44a}$$

$$\zeta_{ijp}^2 = \int_0^1 \mathrm{d}y\ \nu_i^I(y)\nu_j^I(y)\,\theta\left(\sigma^T - \left|x_p^T - y\right|\right), \tag{44b}$$

$$\zeta_i^3 = \int_0^1 \mathrm{d}y\ \nu_i^I(y). \tag{44c}$$

A.2.2 Interpretation of the learning equation

To understand the complete learning Eq. (42) in the case of inhibitory teacher input, a closer look at the ζ's as defined in (44) is needed. As in the case of excitatory teacher input, the integrals in (44) are extended to run from $-\infty$ to ∞. This leads to

$$\zeta_{jp}^1 = \int_{x_p^T-\sigma^T}^{x_p^T+\sigma^T} \mathrm{d}y\ \nu_j^I(y), \tag{45a}$$

$$\zeta_{ijp}^2 = \int_{x_p^T-\sigma^T}^{x_p^T+\sigma^T} \mathrm{d}y\ \nu_i^I(y)\nu_j^I(y), \tag{45b}$$

$$\zeta_i^3 = \int_{-\infty}^{\infty} \mathrm{d}y\ \nu_i^I(y). \tag{45c}$$

It can be seen that ζ_{jp}^1 peaks for $x_j^I \approx x_p^T$ and is very small elsewhere. Furthermore, ζ_{ijp}^2 has a maximal value if $x_i^I \approx x_j^I \approx x_p^T$ and is nearly zero otherwise. Finally, ζ_i^3 is constant.

Using the rates from (11) to explicitly calculate the ζ's gives

$$\zeta_j^1 = A^I\sigma^I\sqrt{\pi/2} \times\left[\mathrm{erf}\left(\frac{x_p^T - x_j^I + \sigma^T}{\sqrt{2}\sigma^I}\right) - \mathrm{erf}\left(\frac{x_p^T - x_j^I - \sigma^T}{\sqrt{2}\sigma^I}\right)\right], \tag{46a}$$

$$\zeta_{ijp}^2 = (A^I)^2\sigma^I\sqrt{\pi/4}\exp\left(-\frac{(x_i^I - x_j^I)^2}{4\,(\sigma^I)^2}\right) \times\left[\mathrm{erf}\left(\frac{2x_p^T - (x_i^I + x_j^I) + 2\sigma^T}{2\,\sigma^I}\right) - \mathrm{erf}\left(\frac{2x_p^T - (x_i^I + x_j^I) - 2\sigma^T}{2\,\sigma^I}\right)\right], \tag{46b}$$

$$\zeta_i^3 = A^I\sigma^I\sqrt{2\pi}. \tag{46c}$$

B Poisson neuron

The Poisson neuron is a simple model neuron with stochastic firing dynamics. The neuron lacks a threshold and may therefore be seen as biologically unrealistic. On the other hand, precisely because of its lacking threshold, the model is linear and can be described very easily mathematically (van Hemmen 2001; Kempter et al. 1999), see also Tuckwell (1988). On the other hand, due to (48) it does incorporate refractory behavior.

The Poisson neuron model is defined by the following properties:

– The probability P that a spike occurs in the small time interval $[t, t+\delta t)$ is given by

$$P_{\text{spike in }[t,\,t+\delta t)} = \lambda(t)\delta t. \tag{47}$$

– The probability of getting more than one spike during an interval of length δt is of order $o(\delta t)$, that is,

$$\frac{P_{\text{more than one spike}}}{\delta t} \to 0 \quad \text{for } \delta t \to 0. \tag{48}$$

– The probability of firing a spike does not depend on the previous spiking history of the neuron.

Although this neuron model is very simple, it captures some very important aspects of neuron firing. First, neuronal firing is to some degree stochastic and second, explicit time-dependence of firing probabilities is taken into account. Because of the linearity of (47), an exact mathematical understanding of the learning process is possible (Kempter et al. 1999).

It is to be stressed that the Poisson model is *not* simply equivalent to a rate coding. Since spike generation is explicitly taken into account, the theory and simulations are much richer than they would be with a pure rate-based modeling approach. In addition, the firing rate functions are allowed to have an arbitrary time dependence. The firing rate can therefore vary on a very small time scale, whereas a rate-based description always uses time-averaged firing rates.

The expectation value of the output from a Poisson neuron can be calculated analytically. If a spike occurring at time $t = t_0$ of the Poisson neuron gives rise to a response $h(t - t_0)$, a post-synaptic current, the total response is given by van Hemmen (2001)

$$\int_{-\infty}^{t} \mathrm{d}s\ h(t-s)\lambda(s). \tag{49}$$

If h is chosen so that causality is ensured, i.e., $h(t) = 0$ for $t < 0$, the expectation of the output is given by

$$\int_{-\infty}^{\infty} \mathrm{d}s\, h(t-s)\lambda(s). \tag{50}$$

Since a Poisson neuron fires stochastically two realizations of the firing dynamics governed by some particular $\lambda(t)$ will always differ. The variance of the output response can also be calculated explicitly (van Hemmen 2001) and is given by

$$\int_{-\infty}^{\infty} \mathrm{d}s\, h^2(t-s)\lambda(s). \tag{51}$$

References

Abbott LF, Nelson SB (2000) Synaptic plasticity: taming the beast. Nat Neurosci 3:1178–1183

Bi GQ, Poo MM (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci 18:10464–10472

Bi GQ, Poo MM (2001) Synaptic modification by correlated activity: Hebb's postulate revisited. Ann Rev Neurosci 24:139–166

Bi GQ, Rubin J (2005) Timing in synaptic plasticity: from detection to integration. Trends Neurosci 28:222–228

Calvert G, Spence C, Stein BE (eds) (2004) The handbook of multisensory processes. MIT Press, Cambridge

Claas B (1994) Removal of eyes in early larval stages alters the response of the clawed toad, *Xenopus laevis*, to surface waves. Physiol Behav 56:423–428

Dan Y, Poo MM (2004) Spike timing-dependent plasticity of neural circuits. Neuron 44:23–30

Davison AP, Frégnac Y (2006) Learning cross-modal spatial transformations through spike timing-dependent plasticity. J Neurosci 26:5604–5615

Froemke RC, Poo MM, Dan Y (2005) Spike-timing-dependent synaptic plasticity depends on dendritic location. Nature 434:221–225

Gardner JL, Merriam EP, Movshon JA, Heeger DJ (2008) Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. J Neurosci 28:3988–3999

Gerstner W, Kempter R, van Hemmen JL, Wagner H (1996) A neuronal learning rule for sub-millisecond temporal coding. Nature 383:76–78

Grace MS, Woodward OM, Church DR, Calisch G (2001) Prey targeting by the infrare-imaging snake *Python*: effects of experimental and congenital visual deprivation. Behav Brain Res 119:23–31

Gutfreund Y, Zheng W, Knudsen EI (2002) Gated visual input to the central auditory system. Science 297:1556–1559

van Hemmen JL (2000) Theory of synaptic plasticity. In: Moss F, Gielen S (eds) Neuro-informatics, neural modelling, handbook of biological physics, vol 4. Elsevier, Amsterdam, pp 771–823

van Hemmen JL (2002) The map in your head: How does the brain represent the outside world?. Chem Phys Chem 3:291–298

Hötting K, Rösler F, Röder B (2004) Altered auditory-tactile interactions in congenitally blind humans: An event-related potential study. Exp Brain Res 159:370–381

Hyde PS, Knudsen EI (2001) A topographic instructive signal guides the adjustment of the auditory space map in the optic tectum. J Neurosci 21:8586–8593

Izhikevich EM, Desai NJ (2003) Relating STDP to BCM. Neural Comp 15:1511–1523

Jiang B, Treviño M, Kirkwood A (2007) Sequential development of long-term potentiation and depression in different layers of the mouse visual cortex. J Neurosci 27:9648–9652

Kaas JH (1991) Plasticity of sensory and motor maps in adult mammals. Ann Rev Neurosci 14:137–167

Kaas JH, Collins CE (2004) The resurresction of multisensory cortex in primates: Connection patterns that integrate modalities. In: Calvert G, Spence C, Stein BE (eds) The handbook of multisensory processes, Chap. 17. MIT Press, Cambridge, pp 285–293

Kempter R, Gerstner W, van Hemmen JL (1999) Hebbian learning and spiking neurons. Phys Rev E 59:4498–4514

Kempter R, Gerstner W, van Hemmen JL (2001) Intrinsic stabilization of output rates by spike-based Hebbian learning. Neur Comp 13:2709–2741

Kempter R, Leibold C, Wagner H, van Hemmen JL (2001) Formation of temporal-feature maps by axonal propagation of synaptic learning. Proc Natl Acad Sci USA 98:4166–4171

King AJ, Hutchings ME, Moore DR, Blakemore C (1988) Developmental plasticity in the visual and auditory representations in the mammalian superior colliculus. Nature 332:73–76

Knudsen EI (2002) Instructed learning in the auditory localization pathway of the barn owl. Nature 417:322–328

Knudsen EI, Brainard MS (1991) Visual instruction of the neural map of auditory space in the developing optic tectum. Science 253:85–87

Knudsen EI, Knudsen PF (1985) Vision guides the adjustment of auditory localization in young barn owls. Science 230:545–548

Knudsen EI, Knudsen PF (1990) Sensitive and critical periods for visual calibration of sound localization by barn owls. J Neurosci 10:222–232

Knudsen EI, du Lac S, Esterly SD (1987) Computational maps in the brain. Ann Rev Neurosci 10:41–65

Knudsen EI, Esterly SD, du Lac S (1991) Stretched and upside-down maps of auditory space in the optic tectum of blind-reared owls; Acoustic basis and behavioral correlates. J Neurosci 11:1727–1747

Linkenhoker BA, Knudsen EI (2002) Incremental learning increases the plasticity of the auditory space map in adult barn owls. Nature 419:293–296

Lisman J, Spruston N (2005) Postsynaptic depolarization requirements for LTP and LTD: A critique of spike timing-dependent plasticity. Nat Neurosci 8:839–841

Markram H, Lübke J, Frotscher M, Sakmann B (1997) Regulation of synaptic efficacy by coincidence of postsynaptci APs and EPSPs. Science 275:213–215

Mooney RD, Klein BG, Rhoades RW (1987) Effects of altered visual input upon the development of the visual system and somatosensory representations in the hamster's superior colliculus. Neurosci 20:537–555

van Opstal AJ, Munoz DP (2004) Auditory-visual interactions subserving primate gaze orienting. In: Calvert G, Spence C, Stein BE (eds) The handbook of multisensory processes, Chap. 23. MIT Press, Cambridge, pp 373–393

Pfister JP, Gerstner W (2006) Triplets of spikes in a model of spike timing-dependent plasticity. J Neurosci 26:9673–9682

Pike FG, Meredith RM, Olding AWA, Paulsen O (1999) Postsynaptic bursting is essential for 'Hebbian' induction of associative long-term potentiation at excitatory synapses in rat hippocampus. J Neurophysiol 518:571–576

Putzar L, Goerendt I, Lange K, Rösler F, Röder B (2007) Early visual deprivation impairs multisensory interactions in humans. Nat Neurosci 10:1243–1245

Roberts PD, Bell CC (2002) Spike timing dependent synaptic plasticity in biological systems. Biol Cybern 87:392–403

Röder B, Rösler F, Spence C (2004) Early vision impairs tactile perception in the blind. Curr Biol 14:121–124

van Rossum MCW, Bi GQ, Turrigiano GG (2000) Stable Hebbian learning from spike timing-dependent plasticity. J Neurosci 20:8812–8821

Rubin JE, Gerkin RC, Bi GQ, Chow CC (2005) Calcium time course as a signal for spike-timing-dependent plasticity. J Neurophysiol 93:2600–2613

Song S, Miller KD, Abbott LF (2000) Competitive Hebbian learning through spike-timing-dependent synaptic pasticity. Nat Neurosci 3:919–926

Standage D, Jalil S, Trappenberg T (2007) Computational consequences of experimentally derived spike-time and weight dependent plasticity rules. Biol Cybern 96: 615–623

Stanford TR, Quessy S, Stein BE (2005) Evaluating the operations underlying multisensory integration in the cat superior colliculus. J Neurosci 25:6499–6508

Stein BE, Meredith MA (1993) The merging of the senses. MIT Press, Cambridge

Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. Nat Rev Neurosci 9:255–266

Stein BE, Jiang W, Stanford TR (2004) Multisensory integration in single neurons of the midbrain. In: Calvert G, Spence C, Stein BE (eds) The handbook of multisensory processes, Chap. 15. MIT Press, Cambridge, pp 243–264

Tuckwell HC (1988) Introduction to Theoretical Neurobiology. Cambridge studies in theoretical neurobiology. Cambridge University Press, Cambridge

Tzounopoulos T, Kim Y, Oertel D, Trussel LO (2004) Cell-specific, spike timing-dependent plasticities in the dorsal cochlear nucleus. Nat Neurosci 7:719–725

Wallace MT, Stein BE (2007) Early experience determines how the senses will interact. J Neurophysiol 97:921–926

Wallace MT, Perrault TJJr, Hairston WD, Stein BE (2004) Visual experience is necessary for the development of multisensory integration. J Neurosci 24:9580–9584

Willshaw DJ, von der Malsburg C (1976) How patterned neural connections can be set up by self-organization. Proc Roy Soc Lond B 194:431–445

Winkowski DE, Knudsen EI (2006) Top-down gain control of the auditory space map by gaze control circuitry in the barn owl. Nature 439:336–339

Wong ROL (1999) Retinal waves and visual system development. Ann Rev Neurosci 22:29–47

Zhang LL, Huizong WT, Holt CE, Harris WA, Poo MM (1998) A critical window for cooperation and competition among developing retinotectal synapses. Nature 395:37–44