Neural Networks 65 (2015) 44–52

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Attention modeled as information in learning multisensory integration

Johannes Bauer *, Sven Magg, Stefan Wermter

University of Hamburg, Department of Informatics, Knowledge Technology, WTM, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany

ARTICLE INFO

Article history:
Received 10 July 2014
Received in revised form 14 January 2015
Accepted 18 January 2015
Available online 2 February 2015

Keywords:
Attention
Multisensory integration
Superior colliculus
Self-organization

ABSTRACT

Top-down cognitive processes affect the way bottom-up cross-sensory stimuli are integrated. In this paper, we therefore extend a successful previous neural network model of learning multisensory integration in the superior colliculus (SC) by top-down, attentional input and train it on different classes of cross-modal stimuli. The network not only learns to integrate cross-modal stimuli, but the model also reproduces neurons specializing in different combinations of modalities as well as behavioral and neurophysiological phenomena associated with spatial and feature-based attention. Importantly, we do not provide the model with any information about which input neurons are sensory and which are attentional. If the basic mechanisms of our model – self-organized learning of input statistics and divisive normalization – play a major role in the ontogenesis of the SC, then this work shows that these mechanisms suffice to explain a wide range of aspects both of bottom-up multisensory integration and the top-down influence on multisensory integration.

1. Introduction

Natural multisensory integration (MSI) is an area of research which is as intriguing as it is broad. MSI is such an important part of sensory processing that it is present in virtually all organisms possessing multiple means of perception (Stein & Alex Meredith, 1993). Just how constitutive it is for our perception of the world is apparent from the curious effects which arise in the (rare) cases when it goes wrong, like the ventriloquism or the McGurk effects (Chen & Vroomen, 2013; McGurk & MacDonald, 1976). There are many aspects of MSI which can be studied, and many levels at which they can be studied: MSI can be studied in different species; involving different sensory modalities and different stimuli; in the time domain or in the spatial domain; on the physical, behavioral, or neural level; how it develops ontogenetically and phylogenetically; how it can be modeled and understood physiologically, mathematically, or algorithmically; how it behaves in isolation or in relation to higher cognitive functions.

Focussing on sensory input, we have recently presented a model of learning MSI in the superior colliculus (SC) which is based on the self-organizing map (SOM) algorithm (Bauer & Wermter, 2013; Bauer, Dávila-Chacón, & Wermter, 2014). In that model, neurons learn the firing statistics of each of their input neurons and use these statistics to approximately compute and encode the probability of a stimulus being in their receptive field. The output of the network is a population-coded approximation of a probability density function (PDF) for the position of a stimulus. We have shown (Bauer & Wermter, 2013; Bauer et al., 2014) that this model reproduces important aspects of natural MSI, namely the spatial principle, the principle of inverse effectiveness, and so-called optimal multisensory integration (Alais & Burr, 2004; King, 2013; Meredith & Stein, 1986; Stein & Stanford, 2008).

Like other models of the SC (or comparable MSI) (Beck et al., 2008; Deneve, Latham, & Pouget, 2001; Fetsch, DeAngelis, & Angelaki, 2013; Ohshiro, Angelaki, & DeAngelis, 2011; Ursino, Cuppini, Magosso, Serino, & Pellegrino, 2009), ours has so far been purely stimulus-driven. The models due to Anastasio and Patton (2003), Cuppini, Magosso, Rowland, Stein, and Ursino (2012), Martin, Meredith Alex, and Ahmad (2009), Pavlou and Casey (2010) and Rowland, Stanford, and Stein (2007) do include projections from cortical areas to the SC. However, both Anastasio and Patton (2003) and Martin et al. (2009) have modeled *only* the effect of cortical input on multisensory enhancement in the SC, leaving aside the topographic organization which is characteristic of SC neurons' receptive fields (RFs) (King, 2013; Sparks, 1988; Wallace & Stein, 1996). The models put forward by Cuppini et al. (2012) and Rowland et al. (2007), while modeling the effect of cortical input on multisensory integration in the SC, focus on replicating biology and

* Corresponding author. Tel.: +49 40 428 83 2522.
E-mail addresses: bauer@informatik.uni-hamburg.de (J. Bauer), magg@informatik.uni-hamburg.de (S. Magg), wermter@informatik.uni-hamburg.de (S. Wermter).

http://dx.doi.org/10.1016/j.neunet.2015.01.004

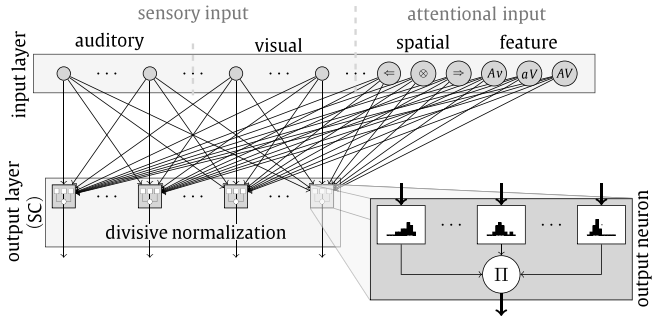

Fig. 1. Our network and structure of input.

refrain from interpreting the meaning of cortical input, network connectivity, and neural computations, functionally.

Our model was specifically developed with functionality and mathematical interpretation in mind: in our model of the SC, a self-organizing network learns a latent-variable model which it uses to infer the location of a stimulus from noisy, population-coded input. Its output approximates a population-coded PDF over that location. In this paper, we extend that model from a stimulus-driven model to one which also considers attentional input: specifically, we test the idea that effects of spatial and feature-based attention are based on very similar mechanisms (Maunsell & Treue, 2006). In fact, we model attentional input as just another source of input, indistinguishable to SC neurons from sensory input. We show that statistical self-organization, the basic mechanism of our artificial neural network (ANN) model, produces effects very similar to those of natural spatial and feature-based attention observed *in vivo*. It also naturally produces specialization to different stimulus combinations in SC neurons (Stein, 2012, chap.33; Wallace & Stein, 1996), a feature which has been interpreted in mathematical terms by Colonius and Diederich (2004) but whose development has not been modeled, to our knowledge.

2. The basis of our model: the network

The structure of our network[1] is shown in Fig. 1: all input neurons are modeled as part of one conceptual input layer regardless of their actual origin. The input layer is fully connected to the output layer. Neurons in the output layer self-organize to each have one preferred position of the input stimulus, and preferred stimulus positions are reflected in the network's topology. Each output neuron learns and maintains one histogram per input neuron which approximates the PDF of activities of that input neuron whenever the actual stimulus position is the output neuron's preferred stimulus position. This mechanism is to model the assumed capability of neurons to learn the statistical relationship between input activity and decision variables (Soltani & Wang, 2010; Yang & Shadlen, 2007). Each neuron computes the likelihood of its input activity under the hypothesis that the actual stimulus position is its preferred stimulus position.

Formally, let $\mathbf{o}$ be an output neuron and $\mathbf{i}$ an input neuron. Then $\mathbf{o}$ maintains a histogram which approximates the likelihood of different activities of $\mathbf{i}$ in case a stimulus is in $\mathbf{o}$'s preferred location $l_{\mathbf{o}}$. Let that histogram $h_{\mathbf{o},\mathbf{i}}$ be represented by the counts $h_{\mathbf{o},\mathbf{i},1}, h_{\mathbf{o},\mathbf{i},2}, \ldots, h_{\mathbf{o},\mathbf{i},n}$ for some large enough n. Then, given some activity $\mathbf{a}_{\mathbf{i}}$ of $\mathbf{i}$, the likelihood of that activity in case the true stimulus location L is $\mathbf{o}$'s preferred location $l_{\mathbf{o}}$ is approximated by:

$$p(\mathbf{a}_{\mathbf{i}}|L = l_{\mathbf{o}}) \simeq \frac{h_{\mathbf{o},\mathbf{i},\lfloor \mathbf{a}_{\mathbf{i}} \rfloor}}{\sum_{k=1}^{n} h_{\mathbf{o},\mathbf{i},k}}.$$

[1] The full code for network, experiments, and evaluation is available as Supplementary material (see Appendix A).

Assuming uncorrelated noise in input neurons, the likelihood of a given population activity $\mathbf{A} = \mathbf{a}_{\mathbf{i}_1}, \mathbf{a}_{\mathbf{i}_2}, \ldots, \mathbf{a}_{\mathbf{i}_m}$ of input neurons $\mathbf{i}_1, \mathbf{i}_2, \ldots, \mathbf{i}_m$ is

$$p(\mathbf{A}|L = l_{\mathbf{o}}) \simeq \prod_{t=1}^{m} p(\mathbf{a}_{\mathbf{i}_t}|L = l_{\mathbf{o}}).$$

If the locations of stimuli are uniformly distributed over the preferred locations $l_{\mathbf{o}_1}, l_{\mathbf{o}_2}, \ldots, l_{\mathbf{o}_q}$ of output neurons $\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_q$, then the probability of L being output neuron $\mathbf{o}$'s preferred location $l_{\mathbf{o}}$, given input population activity $\mathbf{A}$ is

$$p(L = l_{\mathbf{o}}|\mathbf{A}) = \frac{p(\mathbf{A}|L = l_{\mathbf{o}})}{\sum_{s=1}^{q} p(\mathbf{A}|L = l_{\mathbf{o}_s})}.$$

Thus, if we let the spontaneous output $\hat{\mathbf{a}}_{\mathbf{o}}$ of $\mathbf{o}$ in response to input population activity $\mathbf{A} = \mathbf{a}_{\mathbf{i}_1}, \mathbf{a}_{\mathbf{i}_2}, \ldots, \mathbf{a}_{\mathbf{i}_m}$ be

$$\hat{\mathbf{a}}_{\mathbf{o}} = \prod_{t=1}^{m} \frac{h_{\mathbf{o},\mathbf{i}_t,\lfloor \mathbf{a}_{\mathbf{i}_t} \rfloor}}{\sum_{k=1}^{n} h_{\mathbf{o},\mathbf{i}_t,k}}$$

and if we apply divisive normalization to get the stationary activity $\mathbf{a}_{\mathbf{o}}$ of $\mathbf{o}$:

$$\mathbf{a}_{\mathbf{o}} = \frac{\hat{\mathbf{a}}_{\mathbf{o}}}{\sum_{s=1}^{q} \hat{\mathbf{a}}_{\mathbf{o}_q}},$$

then the stationary population response approximates a PDF over the stimulus position L.

In our network, the histograms are filled using self-organized learning so that they reflect the statistics of the input neurons. The procedure is similar to that in the original SOM learning algorithm (Kohonen, 1995, p. 78–83): in every learning step, the network is presented with a newly generated input activity $\mathbf{A} = \mathbf{a}_{\mathbf{i}_1}, \mathbf{a}_{\mathbf{i}_2}, \ldots, \mathbf{a}_{\mathbf{i}_m}$. That output neuron with the strongest response to the input activity is chosen as the best-matching unit (BMU). All neurons update their histograms, with the update strength decreasing with distance from the BMU according to a function called the neighborhood interaction $f(\mathbf{o}, \mathbf{o}')$.

Specifically, let $\mathbf{a}_{\mathbf{i}}$ be the activity of input neuron $\mathbf{i}$, and let $\mathbf{o}_B$ be the BMU in learning step u. Then, for every output neuron $\mathbf{o}$ and input neuron $\mathbf{i}$, the histogram bin $h_{\mathbf{o},\mathbf{i},\lfloor \mathbf{a}_i \rfloor}$ is updated according to the learning rule:

$$h_{\mathbf{o},\mathbf{i},\lfloor \mathbf{a}_i \rfloor} \leftarrow h_{\mathbf{o},\mathbf{i},\lfloor \mathbf{a}_i \rfloor} + \alpha_u f(\mathbf{o}, \mathbf{o}_B),$$

where α_u is the update strength in learning step u. For the neighborhood interaction function $f(\mathbf{o}, \mathbf{o}')$, we chose a Gaussian function of the distance between the neurons $\mathbf{o}$ and $\mathbf{o}'$ in the network's grid:

$$f(\mathbf{o}, \mathbf{o}') = \exp\left(-\frac{d(\mathbf{o}, \mathbf{o}')^2}{\sigma^2}\right),$$

where $d(\mathbf{o}, \mathbf{o}')$ is the grid distance between neurons $\mathbf{o}$ and $\mathbf{o}'$, and σ is called the neighborhood interaction width. As training progresses, σ decreases such that fewer and fewer neurons are substantially affected by each update.

3. Training and testing the network with sensory and attentional input

In this paper, we extend our modeling to include top-down input in addition to bottom-up input. While extending the modeling, we preserve the network and algorithm. The goal is to test the hypothesis that the effects of attention can be explained (in part) by the same mechanisms used to model learning of multisensory integration. The only aspect we therefore change in our model is the nature of the input, which now is not only stimulus-driven, but also reflects higher cognitive processes.

3.1. Network input

Sensory input

The network was trained on simulated input consisting of 'sensory' and 'attentional' components (see Fig. 1). The sensory component was in itself separated into 'visual' and 'auditory' parts. Stimuli were defined by their location $L \in [0, 1]$ and their stimulus class $C \in \{Va, vA, AV\}$. The class determined the strength of the individual components. Stimuli of class Va ('visual') or AV ('audio-visual') had strong visual components. Stimuli of class vA ('auditory') or AV had strong auditory components. Concrete realizations l and c of the stochastic variables L and C were selected randomly and uniformly distributed in every step during training.

All sensory input neurons responded to a simulated stimulus at location l according to Poisson-noisy Gaussian tuning functions: each one of the $n_{\mathbf{i}} = 25$ auditory and visual input neurons $\mathbf{i}_{m,k}$, $m \in \{V, A\}$, $k \in [1 \ldots n_{\mathbf{i}}]$ had a preferred location

$$l_{\mathbf{i}_{m,k}} = \frac{k-1}{n_{\mathbf{i}}-1}.$$

Its Gaussian tuning function was centered around this preferred location. The activity $\mathbf{a}_{m,k}$ of $\mathbf{i}_{m,k}$ in response to a stimulus of class c at location l was then determined by the stochastic function

$$\mathbf{a}_{\mathbf{m},\mathbf{k}} \sim \text{Pois}\left(s(m,c) \times g_m \times \exp\left(-\frac{(l - l_{\mathbf{i}_{m,k}})^2}{\sigma_m^2}\right) + \nu_s\right), \qquad (1)$$

where

$$s(m,c) = \begin{cases} 1 & \text{if } m = V \wedge c \in \{Va, AV\} \\ 1 & \text{if } m = A \wedge c \in \{vA, AV\} \\ 0.5 & \text{otherwise.} \end{cases} \qquad (2)$$

Here, g_m and σ_m are the modality-specific gain and width of the tuning functions and $\nu_s = 3$ is the sensory background noise parameter.

The particular shape of the above tuning functions and the kind of noise is not important in the context of our model. However, Gaussian tuning functions are a simple choice and realistic in that they have a central peak, and fall off with distance from the center. Poisson-like noise has the property that the variance is proportional to the mean, which is true of the variability of actual neural responses (Tolhurst, Anthony Movshon, & Dean, 1983; Vogels, Spileers, & Orban, 1989).

More importantly, the gains ($g_v = 8, g_a = 7$) and widths ($\sigma_v = 0.05, \sigma_a = 0.06$) of the tuning functions were different in the two modalities, rendering auditory input less informative than visual input. This is to model the fact that auditory localization is generally less reliable than visual information for localization (Alais & Burr, 2004). See Section 3.3.5 for a discussion of the effects of different choices of parameters.

Bottom-up visual and auditory projections to the biological SC have their origins mainly in the retina and the inferior colliculus, respectively (May, 2006; Stein, Stanford, & Rowland, 2014). The 'visual' and 'auditory' subpopulations in our simulation are intended to roughly correspond to these unisensory sources of afferents.

Attentional input

Apart from sensory input neurons, our model includes two types of input neurons from higher-level cognitive brain regions. The first type of what we will call 'attentional' input neurons encode information about the general region in which the stimulus is. These three neurons code for stimuli which are on the left ($\Leftarrow$), in the middle ($\otimes$), or on the right ($\Rightarrow$) of the simulated visual field. Another three neurons code for the type of stimulus. One neuron each codes for stimuli which are highly visible (Va), highly audible (vA), or both (VA).

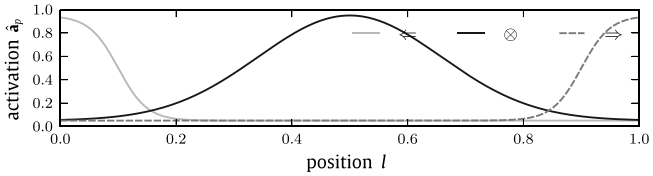

Fig. 2. Activation $\hat{\mathbf{a}}_p$ of attentional input neuron $\mathbf{i}_p$ for $p \in \{\Leftarrow, \otimes, \Rightarrow\}$.

We will call the former type 'spatial' input neurons and the latter type 'feature' input neurons. The intuition behind these additional input neurons is that we often have an expectation of which kind of stimuli we will be presented with. Often, we will expect a stimulus on the left or the right side of our visual field, and we will expect something that is very loud, or bright, or both. The encoding and activation of this knowledge (mostly in cortical areas) is represented in our model in the strongly simplified form of neurons whose activity is either 1 or 0 depending on whether the location or type of the expected stimulus is the one preferred by the respective conceptual input neuron.

Like sensory input, attentional input is modeled as stochastic, modeling non-determinism of ecological conditions, cognitive processes, and neural responses. More specifically, the activity of our attentional input neurons in every trial is modeled as a Bernoulli process whose parameter p depends on the location and class of the stimulus. The (deterministic) activation $\hat{\mathbf{a}}_{\Leftarrow}$, $\hat{\mathbf{a}}_{\otimes}$, and $\hat{\mathbf{a}}_{\Rightarrow}$ of the spatial input neurons $\mathbf{i}_{\Leftarrow}$, $\mathbf{i}_{\otimes}$, and $\mathbf{i}_{\Rightarrow}$, respectively, is modeled by the three functions:

$$\begin{aligned} \hat{\mathbf{a}}_{\Leftarrow} &= \frac{\upsilon}{1 + \exp((l - 0.1) * 40)} + \nu_c \\ \hat{\mathbf{a}}_{\Rightarrow} &= \frac{\upsilon}{1 + \exp(-(l - 0.9) * 40)} + \nu_c \\ \hat{\mathbf{a}}_{\otimes} &= \exp\left(\frac{-(l - .5)^2}{0.05}\right) * \upsilon + \nu_c, \end{aligned} \qquad (3)$$

where $\nu_c = 0.05$ and $\upsilon = 0.9$ are noise parameters. These seemingly complex functions are in fact just two sigmoidal functions which have large values to the left and to the right of the interval $[0, 1]$, respectively, and a Gaussian function centered around 0.5 (see Fig. 2). The activation of feature input neurons was simply $\hat{\mathbf{a}}_c = 1 - \nu_c$ whenever the actual stimulus class was c for $c \in \{aV, Av, AV\}$, and ν_c otherwise. Activity of each attentional input neuron was then stochastically computed from the activation:

$$\mathbf{a}_p \sim \text{Bern}(\hat{\mathbf{a}}_p), \quad \text{for } p \in \{\Leftarrow, \otimes, \Rightarrow, aV, Av, AV\}.$$

The SC receives descending projections from various areas in the cortex (Berson, 1988, chap. 2; Chabot, Mellott, Hall, Tichenoff, & Lomber, 2013; Ferraina, Paré, & Wurtz, 2002; May, 2006; Stein et al., 2014; Wallace & Stein, 1994). Some of those play a role in attention, like the frontal eye field (FEF), the dorsolateral prefrontal cortex (DLPFC), and the lateral intraparietal cortex (LIP) (Buschman & Miller, 2007; Kastner & Ungerleider, 2000). In cats, the anterior ectosylvian cortex (AES) plays an especially important role: its deactivation eliminates neurophysiological multisensory integration (Wallace & Stein, 1994) and drastically alters audio-visual orientation behavior (Wilkinson, Alex Meredith, & Stein, 1996). It has been implicated with selective attention (Dehner, Keniston, Clemo, & Meredith, 2004; Foxe, 2012, chap. 33), due to its effect on neural responses in the SC. Since orienting behavior is linked to attention (Ignashchenkova, Dicke, Haarmeier, & Thier, 2004; Kustov & Robinson, 1996, more recently), this implication is potentiated by the behavioral findings of Wallace and Stein (1994). In our model, 'attentional' input may relate, for example, to FEF, for more spatial input (Bruce, Goldberg, Bushnell, & Stanton, 1985), or AES, in cats, for more feature-related input (Dehner et al., 2004).

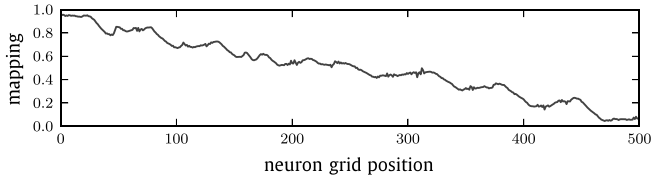

Fig. 3. Mapping of neurons to stimulus positions.

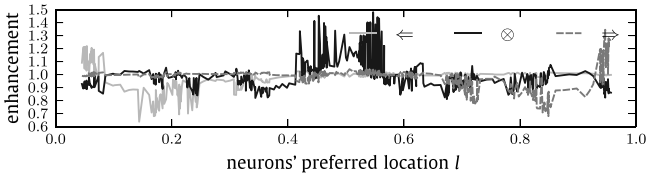

Fig. 4. Effect of spatial attention on neural responses. Average activation of neurons given attentional input coding for spatial classes ⇐( ), ⊗( ), ⇒( ) divided by average activation given zero attentional input.

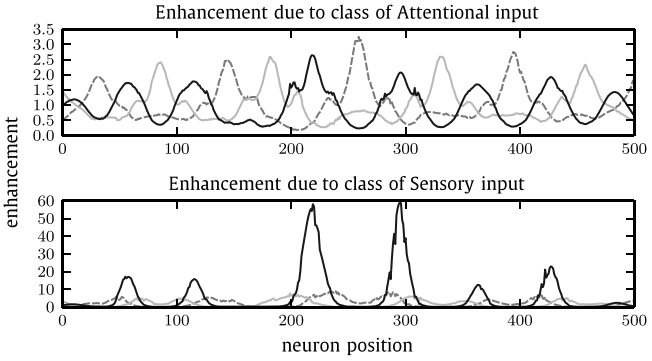

Fig. 5. Feature selectivity. Top: Average activation of neurons given *attentional* input coding for stimulus classes *Va* (, dashed), *vA* ( ), *VA* ( ) divided by average activation given zero attentional input. Bottom: Average activation of neurons given *Va* (, dashed), *vA* ( ), *VA* ( ) *sensory* input divided by average activation given *va* input.

3.2. Training

We trained a network of $n_{\mathbf{o}} = 500$ output neurons extensively for 300 000 training steps (according to the procedure described in Section 2). Both values were chosen to be high enough to avoid artifacts like sampling error (too few neurons) or incomplete training (number of training steps). Smaller values easily yielded qualitatively similar results to the ones reported in the next section. The one distinct feature of our parameter setting was the minimum neighborhood width (of $0.01 < \frac{1}{n_{\mathbf{o}}}$) which we chose deliberately small. With a small neighborhood width, neurons which are close to each other are permitted to learn to respond to different stimuli. Given that training sets up a roughly topography-preserving mapping from data space into the grid while the neighborhood interaction is still large, we expected that neurons which were close to each other would learn to respond to different special cases of similar input. Specifically, we expected that they would self-organize to have similar preferred locations but different stimulus classes.

3.3. Results

3.3.1. Mapping

To determine the preferred location of each neuron, we simulated input at 50 000 positions, evenly spaced across the interval [0, 1]. At each location, we generated input for each stimulus class, and determined the BMU in response to that input. For each neuron $\mathbf{o}$ which had been BMU for input at locations $\{l_1, l_2, \ldots, l_k\} = L$, we chose the median of L as the empirical preferred value of $\mathbf{o}$. We chose the number of $50\,000 = 100 * n_{\mathbf{o}}$ to be sure that median was representative of the preferred location of each neuron. See Fig. 3 for the resultant mapping from neurons to locations. To read out decisions of the network given sensory and attentional input in our experiments, we determined the BMU and applied the mapping generated as described above.

3.3.2. Enhancement

Spatial attention can enhance the activity of SC neurons whose receptive fields overlap the attended region (Goldberg & Wurtz, 1972; Ignashchenkova et al., 2004). To demonstrate similar behavior in our network, we divided the mean activity of each neuron for trials in which 'attentional' input signaled a stimulus of spatial class ⇐ by the mean activity of that same neuron with zero attentional input (and the same for ⊗ and ⇒). Fig. 4 shows that activating the neurons coding for ⇐, ⊗, and ⇒ clearly enhanced mean activity in those neurons whose preferred values were in the respective region.

In contrast to spatial attention, feature-based attention enhances activity of neurons selective to the features attended to across the visual field (Born, Ansorge, & Kerzel, 2012; Maunsell & Treue, 2006). We tested whether this was also true for our network by simulating multisensory input at 100 regular positions between 0 and 1. For each of these positions, we generated 100 sensory input and corresponding spatial activations which we combined once with feature activations coding for each of the stimulus classes $c \in \{aV, Av, AV\}$ and for no stimulus class (av), respectively. From the network's output activation, we computed enhancement for each of the stimulus classes: for each output neuron $\mathbf{o}$, we selected those cases where the difference between the actual stimulus location l and $\mathbf{o}$'s empirical preferred value $l_{\mathbf{o}}$ was within ± 0.01. For each stimulus class, we divided the neuron's mean activity in cases where the attentional activity coded for that class by the mean activity in cases where the attentional activity did not code for any stimulus class. See the top graph in Fig. 5 for plots of enhancement for each of the stimulus classes. We can see that neurons specialized in *attentional* input coding for different stimulus classes.

3.3.3. Stimulus selectivity

To test whether neurons also specialized in different types of *sensory* input, and whether they generally specialized in the same kind of sensory and attentional input, we evaluated for each neuron the enhancement of activity due to *Va*, *vA*, and *VA sensory input* compared to av sensory input. Specifically, we divided, for each neuron, the mean activity given *Va*, *vA*, and *VA* sensory input by the mean activity given av input (when the stimulus was close to their preferred stimulus position, using the input and output activities generated to compute selectivity for *attentional* input, see Section 3.3.2, second part).

The bottom graph in Fig. 5 shows the result: we see, again, that neurons specialized in different kinds of input—this time, in different kinds of *sensory* input. A comparison of the two graphs in Fig. 5 also suggests that the same neurons were generally selective for sensory input from one combination of modalities and for attentional input coding for such a stimulus. Especially, neurons selective for *VA* stimuli were also selective for the corresponding attentional input. Note also that some neurons' responses were depressed by attentional activation coding for their non-preferred stimulus combination (values <1).

Since the same is a bit hard to see for *Va* and *vA* stimuli, in Fig. 5, the relationship between responsiveness to each combination of modalities and attentional enhancement is plotted in Fig. 6. What the figures show is that neurons which responded strongly to *Va* stimuli also tended to have their response enhanced by attentional input coding for *Va* input. More strikingly, their response was depressed by attentional *vA* input.

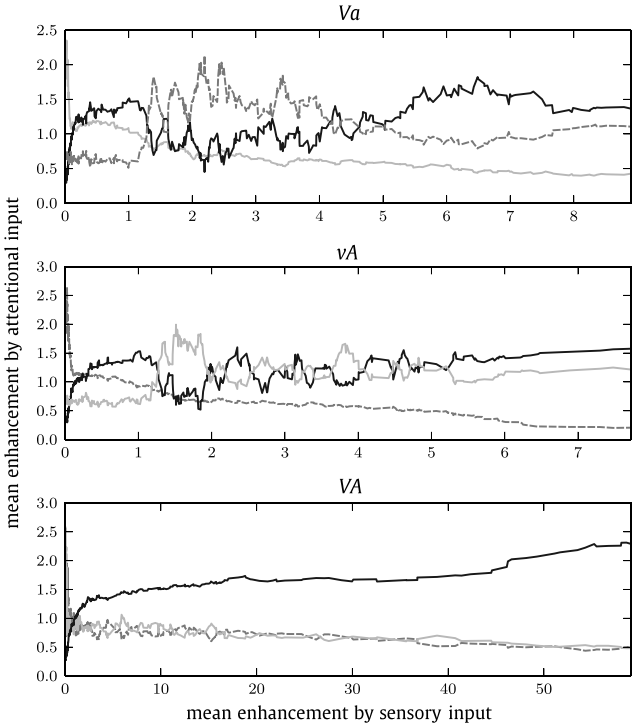

Fig. 6. Effect of Feature-based Attention related to Sensory Selectivity. *X*-Axis: mean response to *sensory* input of class *Va* (■, dashed), *vA* (■), *VA* (■), respectively, divided by mean response to *va* input, for each neuron. *Y*-Axis: mean response to *attentional* input of class *Va*, *vA*, *VA*, respectively, divided by mean response to *va* input, for each neuron. Smoothed for legibility (simple 10-step moving average).

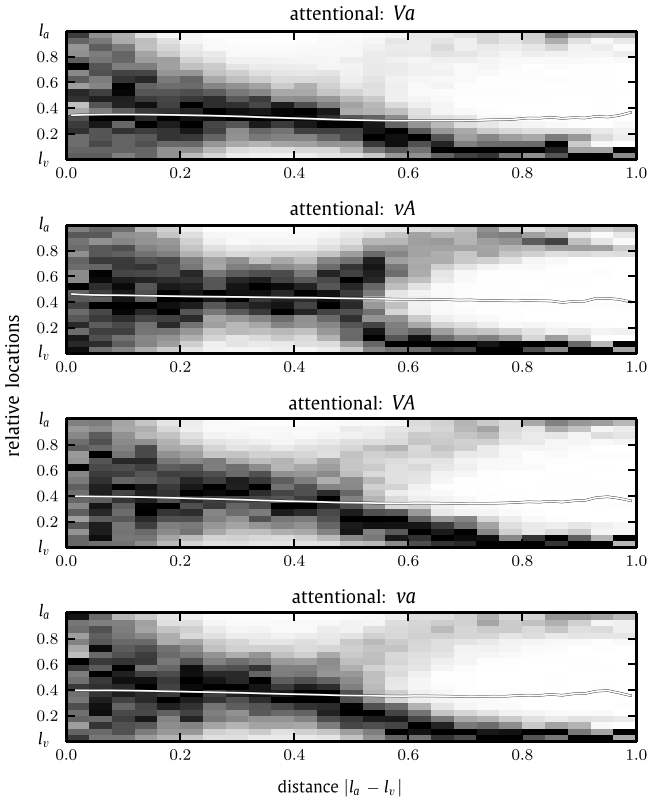

Fig. 7. Integration versus Decision by Relative Stimulus Distance. Gray scale: Frequency of relative localizations between visual (l_v) and auditory (l_a) sub-stimulus, depending on distance between l_v and l_a, given different attentional inputs. The values in each of the columns were normalized by dividing them by the maximum value in that column to improve legibility (darker: more frequent). White lines: Mean relative localization.

3.3.4. Localization

Having tested the effect of attention on the network's activity, we next tested how this effect was reflected in decisions made using the network's responses. To do that, we simulated input in which the visual and auditory component had different locations l_v and l_a, respectively. Both components were strong uni-sensory components ($c = AV$), but each sensory component was combined once with attentional input coding for each of the stimulus classes and for no stimulus class. Using the empirical mapping of neurons to positions, we then derived a localization of the incongruent input.

Fig. 7 shows the distribution of relative localizations made by the network depending on the stimulus class represented by the feature-encoding input neurons. The individual graphs show histograms of the localization l_n of incongruent cross-modal stimuli, that is, cross-modal stimuli in which $l_v \neq l_a$, relative to the location of visual and auditory sub-stimuli l_v and l_a, depending on the absolute distance $|l_v - l_a|$. We see that attentional input coding for the stimulus class influences localization of incongruent audio-visual stimuli: at larger distances, visibly more stimuli were localized close to the auditory sub-stimulus if attentional content coded for a *vA* stimulus than in other conditions. Also, already at lower inter-stimulus distances, the mean of localizations in that condition is closer to the auditory stimulus. With attentional input coding for a *Va* stimulus, less stimuli were localized close to the auditory stimulus at large distances, and on average localizations were shifted towards the visual stimulus, compared to the other conditions.

Finally, to test whether spatial attention affected localization, we simulated incongruent audio-visual stimuli paired with spatial attention: in 10 000 steps, we simulated a visual stimulus in the left third of the interval [0, 1] and an auditory stimulus in the right third. We then combined the sensory input with attentional input coding for each combination of each of the spatial classes ⇐, and ⇒, and each of the stimulus classes *Va*, *vA*, *VA*, and *va*. After that, visual and auditory stimulus positions were switched, in every step, and combined with attentional input as above, giving us a total of 80 000 input activations. We found that the network localized the combined stimuli on average at a position of 0.397, relative to the interval $[l_v, l_a]$, as above, when spatial attention was on the side of the visual stimulus and 0.461 when it was on the side of the auditory stimulus. This means that spatial attention had a sizable effect on localization.

3.3.5. Parameters

All effects discussed in the next section were qualitatively robust under broad ranges of parameter settings. However, we did observe interesting quantitative effects due to tuning function parameters, which determined the information available for localization: information increased with lower background noise ν_s and greater gains g_a, g_v (Eq. (1)).

We ran experiments in which either the relative size of the sensory gains g_a, g_v was manipulated (Table 1(a)), they were jointly scaled, (Table 1(b)), or the baseline noise parameter ν_s was manipulated (Table 1(c)). For each experiment, we then computed the mean localizations given incongruent sensory and varying attentional input relative to the interval $[l_v, l_a]$, as in Section 3.3.4 (columns μ_{Va}, μ_{vA}, μ_{VA}, μ_{va} in Table 1). We also fitted two models to the distributions of relative localizations at different absolute distances $|l_a - l_v|$: one model was a simple Gaussian model, while the second was a mixture of two Gaussians whose respective modes were at the location of the visual stimulus, l_v, and the auditory stimulus, l_a. Thus, the first was an integration model, while the other was a stimulus selection model. We then used Akaike's information criterion (AIC) (Akaike, 1974; deLeeuw, 1992) to determine

Table 1
Comparison of Alternative Parameter Settings. Changing baseline noise levels and sensory gains affected the maximum distance at which stimuli were integrated and how strongly localization was influenced by attentional input. $a_c, c \in \{Va, vA, VA, va\}$: the least distance at which Akaike's information criterion was in favor of a stimulus selection model given attentional input of class c. $\mu_c, c \in \{Va, vA, VA, va\}$: mean of all relative localizations given c (analogous to y-axes in Fig. 7). **Bold rows:** same parameters as in the rest of the paper.

g_v, g_a	a_{Va}	a_{vA}	a_{VA}	a_{va}	μ_{Va}	μ_{vA}	μ_{VA}	μ_{va}
8.0, 5.0	0.629	0.591	0.619	0.611	0.241	0.280	0.252	0.250
8.0, 7.0	**0.663**	**0.653**	**0.649**	**0.643**	**0.364**	**0.458**	**0.405**	**0.407**
10.0, 7.0	0.685	0.665	0.651	0.663	0.281	0.331	0.301	0.303

(a) Alternative Relative Sensory Gains g_v, g_a

g_v, g_a	a_{Va}	a_{vA}	a_{VA}	a_{va}	μ_{Va}	μ_{vA}	μ_{VA}	μ_{va}
3.0, 2.6	0.591	0.649	0.617	0.599	0.341	0.532	0.412	0.423
4.0, 3.5	0.597	0.629	0.617	0.587	0.346	0.492	0.417	0.412
5.0, 4.4	0.589	0.655	0.681	0.635	0.349	0.482	0.412	0.412
6.0, 5.2	0.667	0.683	0.619	0.635	0.358	0.469	0.407	0.408
7.0, 6.1	0.697	0.667	0.621	0.647	0.366	0.448	0.410	0.406
8.0, 7.0	**0.663**	**0.653**	**0.649**	**0.643**	**0.364**	**0.458**	**0.405**	**0.407**
10.0, 8.8	0.705	0.709	0.655	0.689	0.383	0.447	0.413	0.414
12.0, 10.5	0.745	0.747	0.733	0.741	0.367	0.432	0.417	0.407
14.0, 12.2	0.727	0.737	0.733	0.733	0.380	0.451	0.424	0.420
16.0, 14.0	0.647	0.659	0.675	0.663	0.380	0.466	0.421	0.421
18.0, 15.8	0.747	0.745	0.703	0.735	0.381	0.452	0.406	0.410

(b) Scaled Sensory Gains g_v, g_a

ν_s	a_{Va}	a_{vA}	a_{VA}	a_{va}	μ_{Va}	μ_{vA}	μ_{VA}	μ_{va}
0.5	0.810	0.802	0.818	0.812	0.398	0.458	0.417	0.423
1.0	0.705	0.725	0.721	0.715	0.387	0.446	0.406	0.410
1.5	0.717	0.743	0.689	0.717	0.381	0.440	0.407	0.406
2.0	0.745	0.721	0.687	0.713	0.374	0.446	0.416	0.410
2.5	0.655	0.663	0.661	0.655	0.370	0.450	0.407	0.406
3.0	**0.663**	**0.653**	**0.649**	**0.643**	**0.364**	**0.458**	**0.405**	**0.407**
4.0	0.661	0.685	0.629	0.641	0.373	0.458	0.408	0.411
5.0	0.619	0.645	0.639	0.631	0.371	0.450	0.409	0.409
6.0	0.621	0.665	0.619	0.615	0.356	0.474	0.405	0.411
7.0	0.623	0.667	0.621	0.639	0.352	0.464	0.405	0.404
8.0	0.607	0.625	0.625	0.613	0.356	0.475	0.408	0.412

(c) Alternative Baseline Noise Levels ν_s

the least distance $|l_a - l_v|$ at which the stimulus selection model described the distribution of localizations better than the integration model (columns a_{Va}, a_{vA}, a_{VA}, a_{va} in subtables of Table 1).

Unsurprisingly, more information in the visual or less in the auditory modality (larger gain g_v, lower gain g_a) caused localizations to generally move towards the visual stimulus in incongruent conditions (see Table 1(a)). More interestingly, the amount of sensory information was reflected in the maximum distance at which stimuli were integrated: what we can see in Tables 1(b) and (c) is a tendency for the mean of localizations to move towards the visual stimulus in *Va* conditions and towards the auditory stimulus in *vA* conditions with *less* sensory information (columns μ_{Va}, μ_{vA} in Tables 1(b) and (c)). Also, the network tends to stop integrating and start selecting one of the sub-stimuli earlier with less sensory information (strong background noise ν_s, low sensory gains g_v, g_a) than with more sensory information (smaller values in columns a_{Va}, a_{vA}, a_{VA}, a_{va}).

Unfortunately, as we can see, it is hard to make out a consistent pattern in the relationship between the amount of sensory information, attentional input, and integration versus stimulus selection. While there are appreciable differences between the columns a_{Va}, a_{vA}, a_{VA}, and a_{va} of Tables 1(b) and (c), these differences do not coherently point into one direction. To be able to make a statement about the effect of sensory information on that of attentional input on integration and stimulus selection, many more simulations would be necessary. Additionally, a statistic different from the one used here, the minimum distance between l_v and l_a at which AIC favors the stimulus selection model, may be more appropriate for our purposes. Since the focus of this study is more on qualitative effects of attention than on quantitative differences with varying parameter settings, we leave these aspects for future work.

3.4. Discussion

Figs. 5 and 6 show clearly that some neurons reacted much more strongly to attentional and sensory input related to one stimulus class than others. As can be seen in Fig. 5, neurons whose activity was strongly enhanced by *Va*-class stimuli were different from those whose enhancement for *vA*-class stimuli was strong, and vice-versa. This enhancement was reflected in the decision made by the network: attentional input coding for a *vA* stimulus led to substantially more localizations close to the auditory substimulus than attentional input coding for any other stimulus class. This can be seen in Fig. 7, where the upper 'arm' of the distribution at greater inter-stimulus distances has visibly more weight for attentional *vA* input, and in the mean of localizations, which is closer to the visual stimulus at all distances (see also columns μ_{Va}, μ_{vA}, μ_{VA}, μ_{va} in Table 1).

We relate these effects to those of feature-based attention: attention focused on the visual features of an object enhances the activity of neurons across the visual field in whose receptive fields is a stimulus with the attended features if they are sensitive to those features. On the behavioral side, attending to certain stimulus features will increase detection of objects with these features (Andersen, Müller, & Hillyard, 2009; Born et al., 2012; Maunsell & Treue, 2006). Similarly, activating the attentional content coding for a stimulus with high auditory and low visual salience enhanced the activity of specific neurons in our network and it increased the likelihood for the network to choose the location of the auditory substimulus over that of the visual sub-stimulus.

Moreover, like in the experiments by Jack and Thurlow (1973) and Warren, Welch, and McCarthy (1981), semantic content changed the extent to which stimuli in different modalities were integrated: depending on the type of stimulus cognitive content coded for, the localization of cross-sensory stimuli was shifted towards the visual component, towards the auditory component, or in between. This is reflected in the mean relative localizations under the different conditions, visualized in Fig. 7, and in the respective columns in Table 1.

Mechanistically, the effects described above emerge because competitive learning leads to specialization among neurons such that different neurons react to different stimuli. Each neuron specializes in stimuli from a specific position in simulated space, and, to varying extent, to a specific stimulus combination. SOM-style self-organization tries to embed the topology of data space into the network's grid. Since data space is two-dimensional (stimulus position versus stimulus type) but the grid only has one dimension, this cannot succeed completely. One of the dimensions – generally the one describing less variance in the data – would have been ignored by the network if we had kept the neighborhood size during learning above a certain threshold. Intentionally decreasing the neighborhood size to a very small number allowed the network to have some non-monotonicity in the mapping (see Fig. 3), as it were, an effect similar to what Kohonen (1995, p. 87 *f*) calls 'zebra stripes.'

Miikkulainen, Bednar, Choe, and Sirosh (2005, p. 62 *f*) call this effect 'folding' and they showed how it can produce structures resembling ocular dominance stripes or stripes of neurons selective for different stimulus orientations in the visual cortex. Ocular dominance stripes are also present naturally in the SCs of monkeys (Pollack & Hickey, 1979) and they have been shown to arise in the tecta of tadpoles when they are implanted with a third eye (Law & Constantine-Paton, 1981). In our context, multiple neurons came to code for the same location, but combined with a different stimulus class.

Specialization of neurons not only in stimuli from some direction but also of a certain stimulus class implements an important feature of natural multisensory integration. Wallace and Stein (1996) have found that not all SC neurons react to stimuli in all or

even more than one sensory modality. This has been modeled computationally by Colonius and Diederich (2004) who make a normative argument for why there are uni-sensory neurons in the SC. That argument goes along the lines that a neuron which uses only evidence from one sensory modality to decide whether a stimulus is in its receptive field is not affected by noise in any other modality. Our model produces such a specialization, as can be seen in Figs. 5 and 6, and it makes this argument more specific: according to our account, a mixture of uni-sensory and multisensory neurons effectively evaluates hypotheses about stimulus combinations and stimulus locations. It then chooses that stimulus combination and location which are most consistent with the evidence. In this context, cognitive content (attention) can either be seen as additional evidence or, equivalently, as a prior over stimulus locations and combinations.

Together, these findings show that attentional input to the SC needs no different wiring from that of sensory input to have the neurophysiological and behavioral effects seen in experiments, which is the main result of this paper. Of course, this does not preclude the possibility that goal-directed learning may play a role. Weber and Triesch (2009) have shown how essentially unsupervised learning can be extended and combined with mechanisms from reinforcement learning to emphasize learning of goal-relevant over goal-irrelevant features. Similarly, if there is a goal-directed feedback signal to the SC, that feedback signal could modulate the unsupervised training process. What we show here is that in our model neither feedback is needed to produce the neurophysiological and behavioral effects shown here, nor do projections from different sources of input need to be treated differently in the overall architecture or in how they are treated by integrative SC neurons.

Our model fits in well with the view recently expressed by Krauzlis, Bollimunta, Arcizet, and Wang (2014) that attention may be not so much a mechanism *causing* certain behavioral and neural phenomena, but an effect *emerging* from the need for effective information processing. Krauzlis et al. (2014) argue that for example the SC is involved in regulating spatial attention behaviorally, but neural activity related to selective attention in visual cortex remains after collicular deactivation. Furthermore, even animals without a well-developed neocortex or even SC show signs of selective attention. Since no single brain region or circuit seems necessary for an organism to exhibit behavioral effects of attention, Krauzlis et al. (2014) argue that attention and its known neural correlates emerge simply because effective biological (and artificial) information processing requires state estimation. The estimated state at any point then modulates action and perception. We would add that zooming into loci which seemingly evolved to implement state estimation, like the SC, may show that, there again, attention is not an inbuilt mechanism but an emergent effect resulting from neurons using all available information to accomplish their function in the best possible way.

A prediction of our model is the existence of neurons whose activity is *depressed* by a strong stimulus in their non-preferred modality, even if that stimulus is in their receptive field. This effect has not been observed experimentally, to our knowledge. We see a number of possible reasons for this: first, depression has not been studied as extensively as enhancement. Second, it is hard to precisely determine the best stimulus location of a neuron and thus to tell with any confidence whether the (perceived) location of an auditory stimulus is exactly at the same location as the visual stimulus. This is especially true for a neuron which does not respond strongly to an auditory stimulus to begin with. Third, it might be that ecologically sensory noise is so great relative to sensory information that depression vanishes or at least becomes hard to detect (see Section 3.3.5). In that case, depression due to congruent stimuli in a non-preferred modality would be more likely to develop under unusually noiseless conditions. Finally, it may just be that the neural implementation does not permit this kind of depression. If sensory noise typically is high relative to sensory information, then that depression would be weak and therefore its behavioral benefits could become negligible. In that case, it could be economical to completely prune connections to input neurons from the non-preferred modality, thereby eliminating the small amount of depression that would be there otherwise.

We have tested our model under a range of parameters, manipulating the amount of information in the simulated input on the location of the stimulus. We found that the network stopped integrating incongruent cross-sensory stimuli at greater inter-stimulus distances when trained and tested with *more* sensory information than with *less* sensory information. Our explanation for this behavior is that an output neuron **o** which had learned that strong activity of some input neuron **i** almost always indicated a stimulus close to **i**'s preferred location did not as easily discount **i**'s strong activity in the incongruent condition as noise as did neurons which had learned that the difference between driven and spontaneous input activity was low. For a better intuition, imagine looking for a police car in a very crowded street. If you know there are many cars that look similar to a police car and many sounds that are similar to a police car siren, then you will be more inclined to ignore parts of auditory or visual information and focus on stimuli which are overall more salient, or more in line with your expectation, than if there is only one police-car-like object and only one sound that may be a siren.

We present our model as a model of the SC. As we have demonstrated in the previous sections, it reproduces the convergence of primary and secondary sensory information in that brain region, the SC's topographic organization, and its unsupervised adaptation to stimulus statistics. Also, we have previously (Bauer et al., 2014) shown that it can reproduce the spatial principle and the principle of inverse effectiveness, as well as maximum likelihood estimator (MLE)-like behavioral multisensory integration which is presumably caused by the neural processes in the SC. The SC is *one* multisensory region in the brain, whose input–output behavior is particularly well understood, and knowledge we glean about it can inform research on others (Stein, 2012, chap.33). Thus, one might wonder whether our model of the SC also fits some of the other cases of multisensory integration.

On the one hand, there are other brain regions which perform MSI more or less similar to the SC, like parts of AES (Stein & Stanford, 2008), which are visual, auditory, and somatosensory; medial superior temporal area (MSTd), in which vestibular and visual cues converge (Duffy, 1998); and, sub-cortically, regions in putamen, in which there is somatosensory and visual convergence (Graziano & Gross, 1993). On the other hand, MSI in these brain regions differs from that in SC in some respects. For example, the organizing principle of AES is much more specificity to modalities and much less retinotopy than in SC (Clarey & Irvine, 1990; Dehner et al., 2004; Alex Meredith, 2004, chap. 21; Olson & Graybiel, 1987). Also, superadditivity does at least not seem to be as common in MSTd as in SC (although that might be due to the stimuli used in related studies of the two regions) (Morgan, DeAngelis, & Angelaki, 2008). Therefore, it could be a fruitful effort to study in detail which aspects of our model apply to other multisensory brain regions than the SC. In particular, it would be very interesting to see which *changes* to our model would be necessary, since these might point to important differences between the neural input to the SC and that to other brain regions, or to mechanisms which are at work in SC but not elsewhere, and vice-versa.

4. Conclusion

In this paper, we have presented a model of learning of MSI in the SC. We have shown that enhancement of neural responses due

to spatial and feature-based attention can arise naturally from statistical self-organization. It is especially interesting that the model presented here is an extension of a previous model which considered only bottom-up multisensory processing: first, a simple and natural extension of the domain of a model, if successful, corroborates that model, and thus our success at including attentional input along with sensory input in our model of learning MSI in the SC strengthens it. Second, the extension does not require any new mechanisms and treats attentional and sensory input identically, thus showing that the effects it reproduces can be explained by the same basic mechanisms as stimulus-driven MSI. This view fits in well with previous work by Krauzlis et al. (2014) who argued that attention may, on the network level, be seen not as a cause but an effect of the ecological requirements on information processing. Finally, the probabilistic origins of the network at the basis of our model suggest an elegant functional interpretation of both the specialization of neurons in different modality combinations and spatial and feature-based attention: according to this interpretation, self-organization produces a latent-variable model in which each neuron represents in its activity the probability of a specific hypothesis about the position and the quality of a stimulus in terms of modality combination. Attentional input can then be seen either as additional evidence or as a prior on either one of the dimensions of the latent variable.

Acknowledgments

This work is funded in part by the DFG German Research Foundation (grant #1247) – International Research Training Group CINACS (Cross-modal Interactions in Natural and Artificial Cognitive Systems).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.neunet.2015.01.004.

References

Akaike, Hirotugu (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Alais, David, & Burr, David (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262.

Anastasio, Thomas J., & Patton, Paul E. (2003). A two-stage unsupervised learning algorithm reproduces multisensory enhancement in a neural network model of the corticotectal system. *The Journal of Neuroscience*, *23*(17), 6713–6727.

Andersen, Søren K., Müller, Matthias M., & Hillyard, Steven A. (2009). Color-selective attention need not be mediated by spatial attention. *Journal of Vision*, *9*(6).

Bauer, Johannes, & Wermter, Stefan (2013). Learning multi-sensory integration with self-organization and statistics. In Artur Garcez, Luis Lamb, and Pascal Hitzler (Eds.), *Ninth international workshop on neural-symbolic learning and reasoning* NeSy'13, (pp. 7–12), August.

Bauer, Johannes, & Wermter, Stefan (2013). Self-organized neural learning of statistical inference from high-dimensional data. In F. Rossi, (Ed.), *International joint conference on artificial intelligence* IJCAI, (pp. 1226–1232), August.

Bauer, Johannes, Dávila-Chacón, Jorge, & Wermter, Stefan (2014). Modeling development of natural multi-sensory integration using neural self-organisation and probabilistic population codes. *Connection Science*, 1–19.

Beck, Jeffrey M., Ma, Wei J., Kiani, Roozbeh, Hanks, Tim, Churchland, Anne K., Roitman, Jamie, Shadlen, Michael N., Latham, Peter E., & Pouget, Alexandre (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, *60*(6), 1142–1152.

Berson, David M. (1988). Retinal and cortical inputs to cat superior colliculus: composition, convergence and laminar specificity. In *Progress in brain research*: *vol. 75* (pp. 17–26). Elsevier.

Born, Sabine, Ansorge, Ulrich, & Kerzel, Dirk (2012). Feature-based effects in the coupling between attention and saccades. *Journal of Vision*, *12*(11).

Bruce, Charles J., Goldberg, Michael E., Bushnell, M. Catherine, & Stanton, Gregory B. (1985). Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. *Journal of Neurophysiology*, *54*(3), 714–734.

Buschman, Timothy J., & Miller, Earl K. (2007). Top-Down versus Bottom-Up control of attention in the prefrontal and posterior parietal cortices. *Science*, *315*(5820), 1860–1862.

Chabot, Nicole, Mellott, Jeffrey G., Hall, Amee J., Tichenoff, Emily L., & Lomber, Stephen G. (2013). Cerebral origins of the auditory projection to the superior colliculus of the cat. *Hearing Research*, *300*, 33–45.

Chen, Lihan, & Vroomen, Jean (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception, & Psychophysics*, *75*(5), 790–811.

Clarey, Janine C., & Irvine, Dexter R. F. (1990). The anterior ectosylvian sulcal auditory field in the cat: II. A horseradish peroxidase study of its thalamic and cortical connections. *The Journal of Comparative Neurology*, *301*(2), 304–324.

Colonius, Hans, & Diederich, Adele (2004). Why aren't all deep superior colliculus neurons multisensory? A Bayes' ratio analysis. *Cognitive, Affective & Behavioral Neuroscience*, *4*(3), 344–353.

Cuppini, Cristiano, Magosso, Elisa, Rowland, Benjamin A., Stein, Barry E., & Ursino, Mauro (2012). Hebbian mechanisms help explain development of multisensory integration in the superior colliculus: a neural network model. *Biological Cybernetics*, *106*(11–12), 691–713.

Dehner, Lisa R., Keniston, Leslie P., Clemo, Ruth R., & Meredith, Alex A. (2004). Cross-modal circuitry between auditory and somatosensory areas of the cat anterior ectosylvian sulcal cortex: A 'new' inhibitory form of multisensory convergence. *Cerebral Cortex*, *14*(4), 387–403.

deLeeuw, Jan (1992). Introduction to Akaike (1973) information theory and an extension of the maximum likelihood principle. In Samuel Kotz, & Norman L. Johnson (Eds.), *Breakthroughs in statistics, springer series in statistics* (pp. 599–609). New York: Springer-Verlag.

Deneve, Sophie, Latham, Peter E., & Pouget, Alexandre (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, *4*(8), 826–831.

Duffy, Charles J. (1998). MST neurons respond to optic flow and translational movement. *Journal of Neurophysiology*, *80*(4), 1816–1827.

Ferraina, Stefano, Paré, Martin, & Wurtz, Robert H. (2002). Comparison of Cortico-Cortical and Cortico-Collicular signals for the generation of saccadic eye movements. *Journal of Neurophysiology*, *87*(2), 845–858.

Fetsch, Christopher R., DeAngelis, Gregory C., & Angelaki, Dora E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience*, *14*(6), 429–442.

Foxe, John J. (2012). The interface of multisensory processing and selective attention. In Stein (2012), (pp. 337–343).

Goldberg, Michael E., & Wurtz, Robert H. (1972). Activity of superior colliculus in behaving monkey. ii. effect of attention on neuronal responses. *Journal of Neurophysiology*, *35*(4), 560–574.

Graziano, Michael S., & Gross, Charles G. (1993). A bimodal map of space: somatosensory receptive fields in the macaque putamen with corresponding visual receptive fields. *Experimental Brain Research*, *97*(1), 96–109.

Ignashchenkova, Alla, Dicke, Peter W., Haarmeier, Thomas, & Thier, Peter (2004). Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. *Nature neuroscience*, *7*(1), 56–64.

Jack, Charles E., & Thurlow, Willard R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual and Motor Skills*, *37*(3), 967–979.

Kastner, Sabine, & Ungerleider, Leslie G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*(1), 315–341.

King, Andrew J. (2013). Multisensory circuits. In *Neural circuit development and function in the brain* (pp. 61–73). Oxford: Academic Press.

Kohonen, Teuvo (1995). Self-Organizing Maps. In *Springer series in information sciences*. Berlin: Springer-Verlag.

Krauzlis, Richard J., Bollimunta, Anil, Arcizet, Fabrice, & Wang, Lupeng Attention as an effect not a cause. Trends in Cognitive Sciences, June 2014.

Kustov, Alexander A., & Robinson, David L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, *384*(6604), 74–77.

Law, Margaret I., & Constantine-Paton, Martha (1981). Anatomy and physiology of experimentally produced striped tecta. *Journal of Neuroscience*, *1*(7), 741–759.

Martin, Jacob G., Meredith Alex, A., & Ahmad, Khurshid (2009). Modeling multisensory enhancement with self-organizing maps. *Frontiers in Computational Neuroscience*, *3*.

Maunsell, John H., & Treue, Stefan (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322.

May, Paul J. (2006). The mammalian superior colliculus: laminar structure and connections. In *Progress in brain research*: *vol. 151* (pp. 321–378). Elsevier.

McGurk, Harry, & MacDonald, John (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

Meredith, Alex M., & Stein, Barry E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, *365*(2), 350–354.

Alex Meredith, M. (2004). Corticocortical connectivity of Cross-Modal circuits. In Gemma Calvert, Charles Spence, & Barry E. Stein (Eds.), *The handbook of multisensory processes* (pp. 343–355). The MIT Press.

Miikkulainen, Risto, Bednar, James A., Choe, Yoonsuck, & Sirosh, Joseph (2005). *Computational maps in the visual cortex* (2005 edition). New York, USA: Springer.

Morgan, Michael L., DeAngelis, Gregory C., & Angelaki, Dora E. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, *59*(4), 662–673.

Ohshiro, Tomokazu, Angelaki, Dora E., & DeAngelis, Gregory C. (2011). A normalization model of multisensory integration. *Nature Neuroscience*, *14*(6), 775–782.

Olson, Carl R., & Graybiel, Ann M. (1987). Ectosylvian visual area of the cat: Location, retinotopic organization, and connections. *The Journal of Comparative Neurology*, *261*(2), 277–294.

Pavlou, Athanasios, & Casey, Matthew (2010). Simulating the effects of cortical feedback in the superior colliculus with topographic maps. In *The 2010 International joint conference on neural networks, IJCNN* (pp. 1–8). IEEE.

Pollack, Jay G., & Hickey, Terry L. (1979). The distribution of retino-collicular axon terminals in rhesus monkey. *The Journal of Comparative Neurology*, *185*(4), 587–602.

Rowland, Benjamin A., Stanford, Terrence R., & Stein, Barry E. (2007). A model of the neural mechanisms underlying multisensory integration in the superior colliculus. *Perception*, *36*(10), 1431–1443.

Soltani, Alireza, & Wang, Xiao-Jing (2010). Synaptic computation underlying probabilistic inference. *Nature Neuroscience*, *13*(1), 112–119.

Sparks, David L. (1988). Neural cartography: Sensory and motor maps in the superior colliculus. *Brain, Behavior and Evolution*, *31*(1), 49–56.

Stein, Barry E., & Alex Meredith, M. (1993). *The merging of the senses. cognitive neuroscience series* (1 edition). MIT Press.

Stein, Barry E., & Stanford, Terrence R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(5), 406.

Stein, Barry E., Stanford, Terrence R., & Rowland, Benjamin A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, *15*(8), 520–535.

Stein, Barry E. (2012). Early experience affects the development of multisensory integration in single neurons of the superior colliculus. In *The new handbook of multisensory processing* Stein (2012), (pp. 589–606).

Stein, Barry E. (Ed.) (2012). *The new handbook of multisensory processing*. Cambridge, MA, USA: The MIT Press.

Tolhurst, David J., Anthony Movshon, J., & Dean, Andrew F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, *23*(8), 775–785.

Ursino, Mauro, Cuppini, Cristiano, Magosso, Elisa, Serino, Andrea, & Pellegrino, Giuseppe (2009). Multisensory integration in the superior colliculus: a neural network model. *Journal of Computational Neuroscience*, *26*(1), 55–73.

Vogels, Rufin, Spileers, Werner, & Orban, Guy A. (1989). The response variability of striate cortical neurons in the behaving monkey. In *Experimental brain research*: *vol. 77* (pp. 432–436). Springer-Verlag.

Wallace, Mark T., & Stein, Barry E. (1994). Cross-Modal synthesis in the midbrain depends on input from cortex. *Journal of Neurophysiology*, *71*(1), 429–432.

Wallace, Mark T., & Stein, Barry E. (1996). Sensory organization of the superior colliculus in cat and monkey. In *Progress in brain research*: *vol. 112* (pp. 301–311). Elsevier.

Warren, David H., Welch, Robert B., & McCarthy, Timothy J. (1981). The role of visual-auditory compellingness in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics*, *30*(6), 557–564.

Weber, Cornelius, & Triesch, Jochen (2009). Goal-directed feature learning. In *2009 International joint conference on neural networks* (pp. 3319–3326). IEEE.

Wilkinson, Lee K., Alex Meredith, M., & Stein, Barry E. (1996). The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. *Experimental Brain Research*, *112*(1), 1–10.

Yang, Tianming, & Shadlen, Michael N. (2007). Probabilistic reasoning by neurons. *Nature*, *447*(7148), 1075–1080.