Supplementary Methods

General model structure

As described in a previous paper (Cuppini et al, 2010), the modelled cat's SC is represented by one area of multisensory neurons, which can receive excitatory stimuli from four different unisensory regions: visual inputs from AES (subregion AEV), auditory inputs from AES (subregion FAES), all non-AES visual inputs ("ascending" visual inputs), and all non-AES auditory inputs ("ascending" auditory inputs). All these areas are realized for simplicity by arrays of 100 units. Each unit should be interpreted as representing the collective activity of an ensemble of neurons, which have similar receptive fields and so exhibit similar behaviour in response to audio-visual stimuli. In evaluating the model performance, these responses can be compared to a single neuron's activity averaged over multiple trials. Each input unit is sensitive to restricted but overlapping regions of space. These regions are topographically organized, so that a stimulus activates a restricted population of adjacent units. Neurons belonging to the same input region are connected one another by means of lateral synapses, that can be excitatory, between nearby units, or inhibitory, among distant elements, according to a Mexican hat disposition. The unisensory regions and the SC area are interconnected through excitatory synapses.

As extensively discussed in the previous work (Cuppini et al. 2010), the model contains also four different populations of inhibitory interneurons, each receiving input from one specific unisensory region: cortical interneurons (Hv, Ha in Fig. 1) receive one to one synapses from neurons belonging to one of the AES subregions; interneurons in the ascending pathways (Iv, Ia in Fig. 1) receive projections from many elements of the unisensory non-AES areas. This widespread connectivity pattern reflects the synaptic architecture between the SC and the non-AES regions, and it prevents undesired integrative effects when AES is not active or its projections are still immature.

These populations of interneurons realize two different inhibitory mechanisms inside the network. In the first, inhibitory interneurons receiving inputs from non-AES sources (Iv, Ia) project both to the SC and inhibit one another. We assume that this ascending inhibitory mechanism is already present in the immature phase. This assumption has the effect that the stimulation of both non-AES sensory regions invokes a competition between the ascending inputs, so that the stronger overwhelms the weaker (i.e., a "winner takes all", WTA, competition). This mechanism has been introduced to simulate the lack of multisensory integration in the ascending route: in the absence of AES, SC neurons are driven by the non-AES input regions and respond only to the stronger of the two cross-modal stimuli. The second inhibitory mechanism is realized by interneurons stimulated

by AES (Hv, Ha in Fig. 1). We assume that these interneurons work only in the mature architecture (when the descending path is already developed), while in the immature phase these connections are still unformed and ineffective. Once the maturation took place, they send inhibitory synapses towards the SC and effectively eliminate the influence of non-AES excitatory inputs. The motivation of this inhibitory mechanism is that, in the adult, the AES dominates the SC response (Jiang et al. 2001; Stein et al. 2002; Stein 2005).

Lateral interactions between SC multisensory neurons: Extant data suggest that, in the neonatal, two cross-modal stimuli placed far apart (i.e., one inside, another outside the receptive fields) cause no significant suppression (see Wallace and Stein 1997). Conversely, a significant cross-modal suppression is documented in the adult (Kadunce et al. 1997). In our model, cross-modal suppression can be ascribed to lateral synapses in the SC (Magosso et al. 2008; Ursino et al. 2009; Cuppini et al. 2010). Accordingly, we assumed that there are no lateral connections between elements in the SC area at the beginning of training (i.e., in the immature phase). Moreover, for the sake of simplicity, in this work we did not perform the maturation of lateral synapses, supposing it occurs later than maturation of the descending path. Training lateral synapses may be the subject of a subsequent work.

Mathematical description

Notational conventions. In this section we refer to the different elements present in the model as follows:

Ca (cortical auditory): auditory AES (FAES) neurons:

Cv (cortical visual): visual AES (AEV) neurons;

Na: non-FAES auditory neurons;

Nv: non-AEV visual neurons;

Hv: inhibitory interneurons which receive input from AEV;

Ha: inhibitory interneurons which receive input from FAES;

Ia: inhibitory interneurons which receive input from the non-FAES auditory region;

Iv: inhibitory interneurons which receive input from the non-AEV;

Sm (superior colliculus multisensory): multisensory neurons in the SC

Single neurons are referenced with superscripts indicating their region and subscripts that indicate their position within that array (i.e., indicating their spatial position/sensitivity). $u(t)$ and $z(t)$ are used to represent the net input and output of a given neuron at time t, respectively. Thus, $z_i^h(t)$ represents the output of a unit receiving net input $u_i^h(t)$ at location i within array h at time t.

The strength of excitatory synapse between two neurons in different regions, at positions i (postsynaptic neuron) and j (presynaptic neuron), is denoted $W_{ij}^{h,k}$, where h and k represent the receiving and the projecting regions. The same convention is used for inhibitory connections but they are denoted by a capital K instead of W. The lateral (excitatory or inhibitory) connections in the unisensory regions are denoted $L_{i,j}^{h}$, where h is the region and the subscripts i and j represent the position of the target and projecting unit, respectively.

Unit model. The output of each unit in the network at each simulated moment in time is computed from its input through a static sigmoidal relationship and a first-order dynamics. Specifically, for a unit i in region s with time constant τ^s receiving net input $u_i^s(t)$ at a moment in time t, its output is determined by the following differential equation (Eq. (1)):

$$\tau^s \cdot \frac{d}{dt} z_i^s(t) = -z_i^s(t) + \varphi\left(u_i^s(t)\right) \quad (1)$$

where $\varphi\left(u^s(t)\right)$ is a sigmoidal function with parameters ϑ^s (the central point) and p^s, which sets the slope at the central point (Eq. 2):

$$\varphi\left(u^s(t)\right) = \frac{1}{1 + e^{-\left(u^s(t) - \vartheta^s\right) p^s}} \quad (2)$$

Thus, in this model, all units are initialized to an output of zero and their activities are normalized to a maximum of 1. The response of the unit for comparison to empirical data is taken as its output when the network reaches steady state in response to an external stimulus (see below).

Unisensory input regions. For simplicity, each unisensory input area is represented by an array of 100 units that receive input from external stimuli as well as from intrinsic lateral connections. External stimuli generate inputs that are functions of space (x) and time (t) to which a particular input area s is sensitive: $I^s(x,t)$. The receptive field of a generic unit i in an input area s is defined by a Gaussian function of space having a maximum amplitude R_0^s, center x_i, and standard deviation σ_i^s:

$$R_i^s(x) = R_0^s \cdot e^{-\frac{(x-x_i)^2}{2 \cdot \left(\sigma_R^s\right)^2}} \quad (3)$$

As a consequence of Eq. (3), a stimulus presented at a particular position x_i maximally excites unit i but can also excite adjacent units. The input that an external stimulus provides to a generic unit i in input area s, $r_i^s(t)$, is determined by summing the products of the receptive field and the input stimulus for each spatial location:

$$r_i^s(t) = \sum_x R_i^s(x) \cdot I^s(x,t)\Delta x \quad (4)$$

The input provided by intrinsic lateral connections, $l_i^s(t)$, is defined by the sum of the products of the weights of the lateral synapses and the output of the projecting units for each location:

$$l_i^s(t) = \sum_j L_{i,j}^s \cdot z_j^s(t) \quad (5)$$

Lateral connections are symmetric and their weights ($L_{i,j}^s$) are defined by a "Mexican hat" function derived by subtracting an inhibitory Gaussian function (max amplitude = L_{in}^s, std = σ_{in}^s) from an excitatory one (max amplitude = L_{ex}^s, std = σ_{ex}^s):

$$L_{i,j}^s = L_{ex}^s \cdot e^{-\frac{\left[d_x{}^2\right]}{2 \cdot (\sigma_{ex}^s)^2}} - L_{in}^s \cdot e^{-\frac{\left[d_x{}^2\right]}{2 \cdot (\sigma_{in}^s)^2}} \quad (6)$$

In this equation, d_x represents the distance between the projecting and target units. Units at the extreme ends of a linear array potentially might not receive the same number of connections as other units (e.g., there are no units to the "left" of i=1), and this can produce undesired border effects. To avoid this complication, the array is imagined as having a circular structure so that each unit within an area receives the same number of lateral connections:

$$d_x = \begin{cases} |i-j| & \text{if } |i-j| \le N/2 \\ N-|i-j| & \text{if } |i-j| > N/2 \end{cases} \quad (7)$$

The net input received by a unit at position i in a unisensory input area s, $u_i^s(t)$, is the sum of the inputs from the external stimulus (Eq. 4) and the intrinsic connections (Eq. 5):

$$u_i^s(t) = r_i^s(t) + l_i^s(t) \quad (8)$$

The output of these units in each unisensory input area is determined by Eq. 1, 2, and 8 where s is either *Ca*, *Cv*, *Na*, or *Nv*.

Interneuron populations. The four interneuron populations, each an array of 100 topographically-organized units, receive input from specific sensory input sources and send projections to the SC multisensory neurons and (in some cases) each other. Interneurons that receive input from AES areas have net inputs defined by the product of the activity of the topographically-aligned AES unit and the weight of the connection:

$$u_i^{Ha}(t) = W_i^{Ha,Ca} \cdot z_i^{Ca}(t) \quad (9)$$

$$u_i^{Hv}(t) = W_i^{Hv,Cv} \cdot z_i^{Cv}(t) \quad (10)$$

Every interneuron, belonging to the populations connected with non-AES areas, receives an input from every elements of the unisensory related area, defined by the sum of the products of the weights of the connections and the activity of the non-AES units for each location; it also receives an inhibitory input as the result of the WTA mechanism implemented by reciprocal synapses between interneurons at the same location in the two different populations. Their inputs are computed as follows (K denoting an inhibitory connection):

$$u_i^{Ia}(t) = \sum_j W_{ij}^{Ia,Na} \cdot z_j^{Na}(t) - K_i^{Ia,Iv} \cdot z_i^{Iv}(t) \qquad (11)$$

$$u_i^{Iv}(t) = \sum_j W_{ij}^{Iv,Nv} \cdot z_j^{Nv}(t) - K_i^{Iv,Ia} \cdot z_i^{Ia}(t) \qquad (12)$$

where the weights of connections from non-AES elements ($W_{ij}^{h,k}$) are defined by a Gaussian function (max amplitude = $L_{ex}^{h,k}$, std = $\sigma_{ex}^{h,k}$). The output of these units is determined by these input equations and Eq. 1 and 2 where s is Ha, Hv, Ia, and Iv, respectively.

SC multisensory units. SC multisensory units in the immature phase receive two types of inputs: weighted excitatory inputs from unisensory non-AES input areas (Na, Nv) and inhibitory inputs from the related interneuron populations (Ia, Iv). Initially lateral synapses are also present but not active. In the immature phase the patterns of synapses converging from elements of non-AES regions on a single SC neuron are assumed to be widespread; as described above, here, we made the hypothesis that the inhibitory interneurons present synaptic patterns matching those from the unisensory regions. This architecture reflects in-vivo findings of very large auditory and visual RFs in the neonate.

The direct synapses from AES are subject to the training phase. Their net inputs are computed as the sum of the products of synaptic weights and the output of the upstream unisensory neurons:

$$u_i^{Sm,Ca} = \sum_j W_{i,j}^{Sm,Ca} \cdot z_j^{Ca} \qquad (13)$$

$$u_i^{Sm,Cv} = \sum_j W_{i,j}^{Sm,Cv} \cdot z_j^{Cv} \qquad (14)$$

Given that they are initially ineffective, their initial net input to the SC is zero.

The non-AES inputs are subject to a multiplicative ("shunting") inhibition (as is the case in GABAa-mediated inhibition, see Koch (1998)): at the beginning of the training period only from the interneuron populations Ia and Iv, with matching location, and, when the circuitry acquires an adult-like configuration, also from interneurons stimulated by the AES cortex, producing more complicated input equations:

$$u_i^{Sm,Na}(t) = \left(\sum_j W_{ij}^{Sm,Na} \cdot z_j^{Na}(t) \right) \cdot \prod_j \left(1 - K_{ij}^{Sm,Ha} \cdot z_j^{Ha}(t)\right) \cdot \prod_j \left(1 - K_{ij}^{Sm,Hv} \cdot z_j^{Hv}(t)\right) \cdot \left(1 - K_i^{Sm,Iv} \cdot z_i^{Iv}(t)\right) \qquad (15)$$

$$u_i^{Sm,Nv}(t) = \left(\sum_j W_{ij}^{Sm,Nv} \cdot z_j^{Nv}(t) \right) \cdot \prod_j \left(1 - K_{ij}^{Sm,Hv} \cdot z_j^{Hv}(t)\right) \cdot \prod_j \left(1 - K_{ij}^{Sm,Ha} \cdot z_j^{Ha}(t)\right) \cdot \left(1 - K_i^{Sm,Ia} \cdot z_i^{Ia}(t)\right) \qquad (16)$$

where the weights of connections from non-AES elements ($W_{ij}^{h,k}$) are defined by a Gaussian

function (max amplitude = $L_{ex}^{h,k}$, std = $\sigma_{ex}^{h,k}$). The net input to a multisensory unit i is computed as

the sum of all of these inputs:

$$u_i^{Sm}(t) = u_i^{Sm,Ca}(t) + u_i^{Sm,Cv}(t) + u_i^{Sm,Na}(t) + u_i^{Sm,Nv}(t) \qquad (17)$$

Its output is computed from this input using Eq. 1 and 2 where $s=Sm$.

Training Phase

The synapses linking two neurons (say ij and hk) are modified as follows during the learning phase

$$W_{ij}(t+dt) = W_{ij}(t) + \gamma_{LTP}(t) \cdot \left(\frac{1 + sign(x_i - \vartheta_i)}{2} \right) \cdot (x_i - \vartheta_i) \cdot \left(\frac{1 + sign(x_j - \vartheta_j)}{2} \right) \cdot (x_j - \vartheta_j) +$$
$$+ \gamma_{LTD}(t) \cdot \left(\frac{1 + sign(x_i - \vartheta_i)}{2} \right) \cdot (x_i - \vartheta_i) \cdot \left(\frac{1 - sign(x_j - \vartheta_j)}{2} \right) \qquad (18)$$

where γ_{LTP} and γ_{LTD} represent learning factors for the Long Term Potentiation and Depression,
respectively.

In order to assign a value for the learning factors, γ_{LTP} and γ_{LTD}, in our model we assumed that the

overall sum of the synaptic weights from AES and targeting the SC are constrained to a maximum

value. This is implemented by steadily reducing the learning factor to zero when the actual sum of

these synaptic weights approaches its maximum saturation value. We have

$$\gamma_{LTP}(t) = \frac{1}{K_{LTP}^{Sm}} \cdot \left(W_{TOT\,max} - W_{TOT}(t) \right) \qquad (19)$$

$$\gamma_{LTD}(t) = \frac{1}{K_{LTD}^{Sm}} \cdot \left(W_{TOT}(t) - W_{TOT\,max} \right) \qquad (20)$$

where W_{TOTmax} is the maximum value allowed for the sum of synapses, and $\frac{1}{K_{LTP}^{Sm}}$ and $\frac{1}{K_{LTD}^{Sm}}$ are the

maximum learning factors (i.e., the learning factor when the overall sum is zero).

All of the other synapses targeting the SC, both from the interneurons stimulated by the AES cortex, and the ascending synapses from the non-AES unisensory regions, use a similar Hebbian learning rule:

$$W_{ij}^S(t+dt) = W_{ij}^S(t) + \gamma_{LTP}(t) \cdot \left(\frac{1+sign(x_i - \vartheta_i)}{2}\right) \cdot (x_i - \vartheta_i) \cdot \left(\frac{1+sign(x_j - \vartheta_j)}{2}\right) \cdot (x_j - \vartheta_j) + $$
$$+ \gamma_{LTD}(t) \cdot \left(\frac{1+sign(x_i - \vartheta_i)}{2}\right) \cdot (x_i - \vartheta_i) \cdot \left(\frac{1-sign(x_j - \vartheta_j)}{2}\right) \quad (21)$$

where γ_{LTP} and γ_{LTD} represent learning factors for the Long Term Potentiation and Depression, respectively.

Again the learning factors, γ_{LTP} and γ_{LTD}, are computed assuming that the synaptic weights cannot overcome a maximum saturation value by progressively reducing them to zero as synapses approach maximum saturation. We have

$$\gamma_{LTP} = \frac{1}{K_{LTP}^S} \cdot \left(W_{max}^S - W_{ij}^S\right) \quad (22)$$

$$\gamma_{LTD} = \frac{1}{K_{LTD}^S} \cdot \left(0 - W_{ij}^S\right) \quad (23)$$

where W_{max} is the maximum value allowed for each synapse, and $\frac{1}{K_{LTP}^S}$ and $\frac{1}{K_{LTD}^S}$ are the

maximum learning factors (i.e., the learning factor when the single synapse is zero).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1: Parameter values

Receptive Fields					
s = Cv, Ca, Nv, Na		**s = Cv, Nv**		**s = Ca, Na**	
R_0^s	1	σ_R^s	1 (1.8°)	σ_R^s	2.5 (4.5°)
Neurons **(s = Cv, Ca, Nv, Na)**					
τ^s	3 ms	ϑ^s	20	p^s	0.3
Interneurons **(s = Hv, Ha, Iv, Ia)**					
τ^s	3 ms	ϑ^s	3	p^s	1
Intra-area Synapses					
AEV		**Non-AEV**		**FAES**	**Non-FAES**

	AEV		**Non-AEV**		**FAES**		**Non-FAES**
L_{ex}^s	5.4	L_{ex}^s	1.2	L_{ex}^s	5.4	L_{ex}^s	3
σ_{ex}^s	1.8 (3.24°)	σ_{ex}^s	2.5 (4.5°)	σ_{ex}^s	2.5 (4.5°)	σ_{ex}^s	1.5 (2.7°)
L_{in}^s	18	L_{in}^s	1	L_{in}^s	15	L_{in}^s	3
σ_{in}^s	8 (14.4°)	σ_{in}^s	6 (10.8°)	σ_{in}^s	12 (21.6°)	σ_{in}^s	37.4 (67.32°)

Inter-area Excitatory Synapses							
$W_i^{Hv,Cv}$	15	$W_i^{Ha,Ca}$	14	$W_{TOT\,max}$	40	ϑ	0.12
$K_{LTP}^{Sm,Cv}$	700	$K_{LTD}^{Sm,Cv}$	700	$K_{LTP}^{Sm,Ca}$	1500	$K_{LTD}^{Sm,Ca}$	1500
$L_{ex}^{Iv,Nv}$	8	$\sigma_{ex}^{Iv,Nv}$	1.5 (2.7°)	$L_{ex}^{Ia,Na}$	4	$\sigma_{ex}^{Ia,Na}$	2 (3.6°)
$L_{ex}^{Sm,Nv}$	5.8	$\sigma_{ex}^{Sm,Nv}$	2 (3.6°)	$L_{ex}^{Sm,Na}$	2	$\sigma_{ex}^{Sm,Na}$	20 (36°)

$W_{\max}^{Nv}$	6	$W_{\max}^{Na}$	3.2	K_{LTP}^{S}	1500	K_{LTD}^{S}	1500
Inter-area Inhibitory Synapses							
Visual Interneurons (H_v, I_v)				**Auditory Interneurons (H_a, I_a)**			
ϑ	0.2	$W_{\max}^{S}$	1	K_{LTP}^{S}	200	K_{LTD}^{S}	1500
$K_i^{Ia,Iv}$	33			$K_i^{Iv,Ia}$	33		
$K_i^{Sm,Iv}$	1			$K_i^{Sm,Ia}$	1		

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References in Supplementary material

Cuppini C, Ursino M, Magosso E, Rowland BA, Stein BE (2010) An emergent model of multisensory integration in superior colliculus neurons. Front Integr Neurosci 22: 4-6. doi: 10.3389/fnint.2010.00006

Jiang W, Wallace MT, Jiang H, Vaughan JW, Stein BE (2001) Two cortical areas mediate multisensory integration in superior colliculus neurons. J Neurophysiol 85: 506 –522.

Kadunce DC, Vaughan JW, Wallace MT, Benedek G, Stein BE (1997) Mechanisms of within- and cross-modality suppression in the superior colliculus. J Neurophysiol 78: 2834–2847.

Koch C (1998) Biophysics of computation: Information Processing in Single Neurons. New York, Oxford University Press.

Magosso E, Cuppini C, Serino A, Di Pellegrino G, Ursino M (2008) A theoretical study of multisensory integration in the superior colliculus by a neural network model. Neural Networks 21: 817–829.

Stein BE (2005) The development of a dialogue between cortex and midbrain to integrate multisensory information. Exp. Brain Res. 166: 305–315.

Stein BE, Wallace MW, Stanford TR, Jiang W (2002) Cortex governs multisensory integration in the midbrain. Neuroscientist 8: 306–314.

Ursino M, Cuppini C, Magosso E, Serino A, Di Pellegrino G (2009) Multisensory integration in the superior colliculus: a neural network model. J Comput Neurosci 26: 55–73.

Wallace MT, Stein BE (1997) Development of multisensory neurons and multisensory integration in cat superior colliculus. J Neurosci 17: 2429–2444.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supplementary figures

Fig. S1 - SC targeting synapses in the newborn(A) and after a prolonged development(B). (in figures only non-FAES, left panels, and FAES synapses, rights panels, are shown. Non-AEV synapses are similar but narrower than non-FAES; AEV synapses are inactive as FAES). X-axes and y-axes represent the position of the pre-synaptic and post-synaptic neuron respectively, while the color denotes the strength of each synapse. Thus, each row represents the synapses that target one specific SC neuron. In the newborns (panels A) the SC receives effective (but weak) synapses only from non-AES areas (specifically, non-FAES for auditory and non-AEV for visual inputs). These inputs provide the sole sensory drive to SC neurons. Connections from AES (i.e., AEV, visual area and FAES auditory area) are not effective. After the development (panels B), a high number of SC multisensory neurons have become capable of integrating cross-modal stimuli. These neurons present mature AES-SC synapses, and pruned non-AES connections; these synaptic patterns led to a contraction in their RFs. In this phase of development there are few non-integrative multisensory neurons in the SC, characterized by widespread projections from non-AES areas (light-grey stripes in the top left figure), and non-effective synapses from AES.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
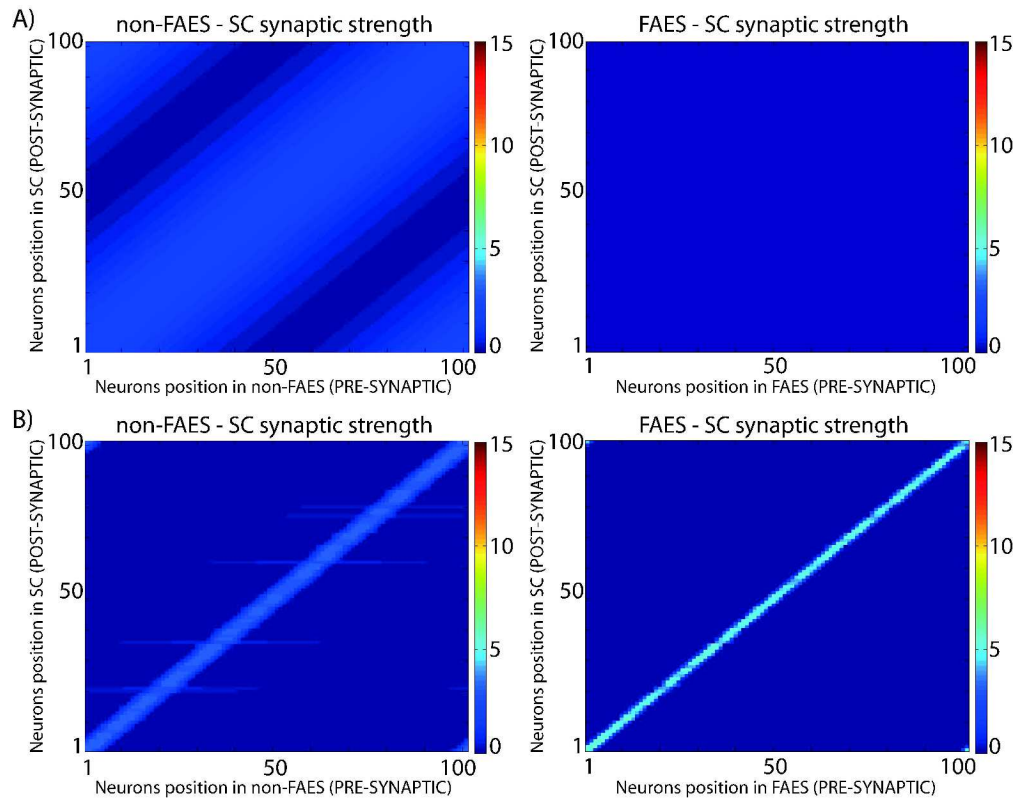48
49
50
51
52
53
54
55
56
57
58
59
60

Fig. S1 - SC targeting synapses in the newborn(A) and after a prolonged development(B). (in figures only non-FAES, left panels, and FAES synapses, rights panels, are shown. Non-AEV synapses are similar but narrower than non-FAES; AEV synapses are inactive as FAES). X-axes and y-axes represent the position of the pre-synaptic and post-synaptic neuron respectively, while the color denotes the strength of each synapse. Thus, each row represents the synapses that target one specific SC neuron. In the newborns (panels A) the SC receives effective (but weak) synapses only from non-AES areas (specifically, non-FAES for auditory and non-AEV for visual inputs). These inputs provide the sole sensory drive to SC neurons. Connections from AES (i.e., AEV, visual area and FAES auditory area) are not effective. After the development (panels B), a high number of SC multisensory neurons have become capable of integrating cross-modal stimuli. These neurons present mature AES-SC synapses, and pruned non-AES connections; these synaptic patterns led to a contraction in their RFs. In this phase of development there are few non-integrative multisensory neurons in the SC, characterized by widespread projections from non-AES areas (light-grey stripes in the top left figure), and non-effective synapses from AES.
265x210mm (600 x 600 DPI)