

# Simulating the Effects of Cortical Feedback in the Superior Colliculus with Topographic Maps

Athanasios Pavlou and Matthew Casey

**Abstract**—The superior colliculus (SC) is a neural structure found in mammalian brains that acts as a sensory hub through which visual, auditory and somatosensory inputs are integrated. This integration is used to orient the eye's fovea towards a prominent stimulus, independently of which sensory modality it was detected in. A recently observed aspect of this integration is that it is moderated by cortical feedback. As a key sensorimotor function integrating low-level sensory information moderated by the cortex, studying the SC may therefore enable us to understand how natural systems prioritize sensory computation in real-time, possibly as a result of task dependent feedback. In this paper, we focus on such a biological model. From a computational perspective, understanding this combination of bottom-up processing with top-down moderation in a model is therefore appealing. We present for the first time a behavioral model of the SC which combines the development of unisensory and multisensory representations with simulated cortical feedback. Our model demonstrates how unisensory maps can be aligned and integrated automatically into a multisensory representation. Results demonstrate that our model can capture the basic properties of the SC, and in particular they show the influence of the simulated cortical feedback on multisensory responses, reproducing the observed multisensory enhancement and suppression phenomena compared to biological studies. This suggests that our unified competitive learning approach may successfully be used to represent spatial processing that is moderated by task, and hence could be more widely applied to other, task dependent processing.

## I. INTRODUCTION

The superior colliculus (SC) is a neural structure found in mammalian brains that acts as a sensory hub through which visual, auditory and somatosensory inputs are integrated creating multisensory representations of the sensory space [1]. The resulting multisensory topographic map is used to direct eye saccades and head movements [2] so that the fovea is focused on a stimulus rapidly. In particular, the SC prioritizes its output to respond to stimuli from the same spatial location, but registered from different senses (multisensory enhancement), whereas asynchronous or spatially unrelated multisensory inputs have lower priority (suppression). Interestingly, while this processing to shift gaze occurs in the midbrain, prior to cortical processing, the enhancement or suppression of multisensory stimuli is affected by descending afferents from cortical

areas, such as the visual, auditory and somatosensory divisions of the anterior ectosylvian sulcus in the cat [3,4]. The influence of these afferent connections is beginning to suggest that the integration of sensory stimuli in the SC depends upon behavioral context [4], demonstrating again the important nature of feedback within the brain (see [5,6] for two examples in vision).

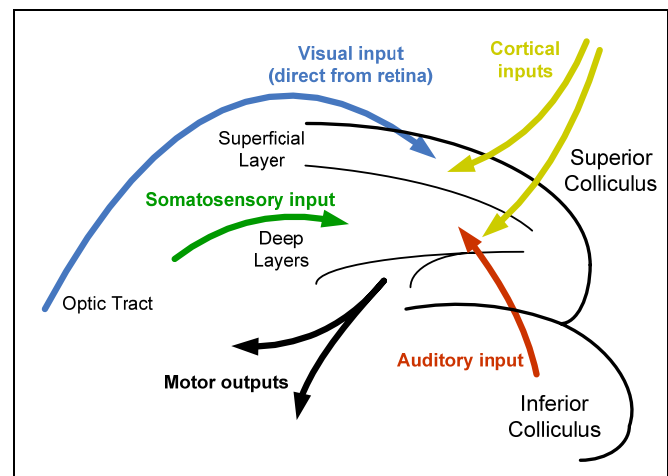


Fig. 1. Sensory and cortical information flows in the SC. The superficial layer of the SC is directly connected to the optic tract, while the deep layers receive auditory (from the inferior colliculus) and somatosensory input. Cortical feedback moderates the processing in the deep layer and influences the motor outputs.

As a mechanism for prioritizing the use of limited sensory resources, the SC is an evolutionary stable structure in vertebrates. It consists of a laminated structure of superficial and deep layers [2]. The superficial layers of the SC process visual stimuli directly from the optic tract to form a topographic map representing visual space. In the deep layers, auditory (from the inferior colliculus) and somatosensory representations of space are combined with the visual map to form a multisensory representation. This is used to form appropriate motor instructions to orient the eyes, head and body to focus on a selected stimulus, moderated by cortical feedback (Figure 1).

As a sensorimotor structure which is evolutionary stable, and which is moderated by cortical feedback, the SC is an interesting part of the brain to study computationally. As well as being sufficiently understood to permit the development of a computational model, the SC exhibits key principles that we wish to embody within artificial agents: seamless sensory integration coupled with motor control,

moderated by behavioral context. If we can understand and duplicate the key principles of how the SC can achieve these tasks, not only can we use the model to improve our neurophysiological knowledge, but we can perhaps also develop new computational paradigms for artificial agents which are capable of operating in real-time on complex, real-world stimuli.

A milestone towards this goal is to achieve the development of coordinate systems for each of the sensory modalities as well as their successful alignment and integration, moderated by cortical feedback. However, as yet, no sufficiently detailed model of the SC has been developed to achieve all of these. For example, until recently, most models have focused on the simulation of multisensory effects observed within the deep layers of the SC [7-9], rather than on the unisensory space. However, three recent models have looked more closely at unisensory representations [10,11] and development [12], respectively.

Magosso et al [10] considered the spatial properties of the sensory spaces, but not from a developmental perspective. They provided a modularized architecture of three neural networks. The first two represented unimodal visual and auditory neurons whilst the third represented multisensory neurons in the SC. Their model accounted for a number of behavioral multisensory phenomena within the SC, however it would be difficult to employ for practical use within artificial agents because it lacks the capability of developing coordinates which can then be aligned in later processing stages. A similar model was developed by Casey and Pavlou [11], who demonstrated how two unisensory maps could be coupled together through association to develop and align coordinate systems. Although their model was able to develop unisensory spaces and align their coordinates to the multisensory space, it employed multiple learning algorithms compromising its biological plausibility. Also, the multisensory map of their model was the product of a manual addition of the translated (aligned) unisensory maps and it did not explore how integration could be developed, or how the cortex can influence enhancement or suppression.

In contrast, Huo and Murray [12] used Spike Timing Dependent Plasticity (STDP) to model alignment (or rather realignment) of audio and visual maps in the SC with feedback. Their model consisted of representations for the retina, inferior colliculus (IC), which is responsible for generating an audio representation of space, and the SC. In the SC, a model interneuron was used to initiate learning in the IC. While this model perhaps lacks the scale of the previous two more practical models, it does provide a biologically plausible representation of plasticity that is robust to changing environmental conditions. Furthermore, it demonstrates the dependency and feedback between different parts of the midbrain: the IC and the SC. However, although this model does demonstrate the key characteristics of adaptive integration through feedback, it does not

sufficiently demonstrate how unisensory maps can develop and be aligned to produce multisensory integration moderated through cortical feedback.

In this paper, we report the development and evaluation of a biological model which brings together the key principles of the SC into a single architecture using a generalized associative learning model. The model includes: 1) development of unimodal topographic maps, 2) alignment of the unimodal maps and development of a multisensory representation under a common learning algorithm; 3) multisensory response suppression upon presentation of spatially non-coincident stimuli and 4) the effects of enabling and disabling cortical feedback on multisensory responses. While this model is evaluated on abstract stimuli in the same way as all previous models, it does demonstrate how a simple neuronal model can be used to process and integrate sensory stimuli seamlessly with the essential influence of the cortex, as compared to biological data.

## II. MODEL SPECIFICATION

Our modeling approach of the superior colliculus is comprised of three distinct layers. The architectural design simulates the actual SC structure which demonstrates increasing complexity from unisensory to multisensory neurons as information flows from the superficial to the deep layers. A similar approach with two unisensory maps being connected to a third multisensory map was also followed by Magosso et al [10]. Our architecture is a modification of the layered approach used by Armony et al [13], who used associative learning within a layered modularized architecture to model fear conditioning in the amygdala. Here, we extended this approach to be able to learn topography using Hebbian learning [11]. This results in modules at lower and higher levels forming their own coordinate systems, which is a key property of the SC.

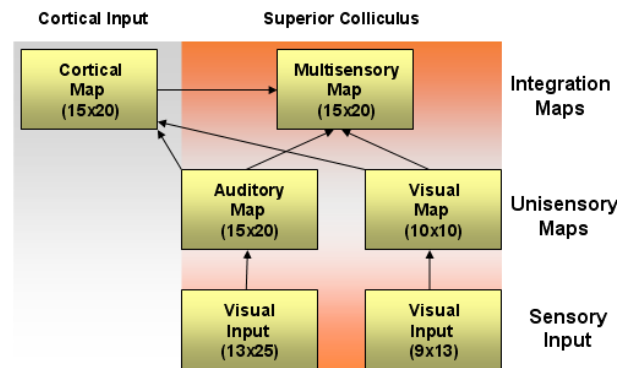


Fig. 2. Schematic of the superior colliculus model.

Figure 2 shows a schematic of the model, which consists of four layers. The first layer is the input representing the visual and auditory modality. The second is the unisensory layer where we have the auditory and the visual map. We note that for simplicity we do not implement a somatosensory input or map. Each of these maps is trained

using representations of their respective modality; the auditory input is only connected to the auditory map and the visual input to the visual map. The third layer represents a cortical map. While this is clearly a simplification of the extensive processing that may take place in several of the higher unisensory and multisensory cortical areas, here we are interested in what influence such a layer may have on forming a multisensory representation. Finally, the fourth layer is the integration map, which represents the multisensory space, moderated by the cortical feedback. The cortical map receives inputs from both the auditory and visual maps, whilst the multisensory map receives inputs from the cortical, auditory and visual maps. The layered representation allows us to study the behavior of the model in two phases: a) the unisensory phase where we can observe the development of topographic representations of the input space, b) the integration phase where we can study multisensory and cortical feedback effects.

Each of our modules is a grid of fully connected neurons. Training occurs per layer, so that the unisensory maps are developed first before continuing on to the integration maps. This is because we want first to achieve topographic organization of the sensory input. Then, the cortical map is trained using outputs from the auditory and visual maps. Finally, the multisensory map is trained using outputs from the cortical, auditory and visual maps. A neuron at location  $(i,j)$  on the grid outputs  $y$  given an  $m$ -dimensional input  $x$ :

$$u_{ij} = \sum_{k=1}^m x_k w_{kij}(t) \quad (1)$$

$$y_{ij} = \begin{cases} f(u_{ij}) & \text{if } \|c_{ij} - c_{win}\| < h(t) \\ f(u_{ij} - y_{win}) & \end{cases} \quad (2)$$

$$y_{ij} = \begin{cases} 1 & u \geq 1 \\ u & 0 < u < 1 \\ 0 & u \leq 0 \end{cases} \quad (3)$$

Here,  $w_{kij}(t)$  represents the weight from input  $k$  for neuron  $(i,j)$  in the grid at time step  $t \geq 0$ . In equation 2 we consider a neuron within the winner area if its location  $(i,j)$  falls within the current radius value  $h(t)$  of the winning neuron. This implementation of lateral inhibition between neurons in the map is similar to Kohonen's SOM [14], and is sufficient to promote competition and allows the map to organize. We indicate  $c_{ij}$  and  $c_{win}$  the grid coordinates of the neuron  $(i,j)$  and the winning neuron respectively. The activation value of the winning neuron  $y_{win} = \max_{ij} f(u_{ij})$  inhibits neurons found outside of the winning area. During training, the winning neuron's area and the learning rate are reduced per epoch by a Gaussian neighborhood and logarithmic radius function:

$$h(t) = r_{min} + (r_{max} - r_{min}) e^{-\left(\frac{t/t_e}{2r_s^2}\right)} \quad (4)$$

$$\varepsilon(t) = l_{min} + (l_{max} - l_{min}) e^{-\left(\frac{(t/t_e)^2}{2l_s^2}\right)} \quad (5)$$

By doing so, each neuron's weights are gradually tuned to respond to inputs depending on their location in the input space. Here,  $r_{min}$ ,  $r_{max}$  and  $l_{min}$ ,  $l_{max}$  represent the minimum and maximum radii for the neighborhood and learning rate respectively with  $r_s$  and  $l_s$  being the bandwidth. The neighborhood and learning rate are updated per epoch thus the time step  $t_e$  for the Gaussian functions (4) and (5) is equal to the data set size.

We use the same set of parameters throughout all maps to ensure consistency during development. The auditory, cortical and multisensory maps have a size of 20 by 15 while the visual map a size of 10 by 10. These particular sizes were chosen for computational efficiency. A smaller visual map size is selected since the visual space representation is smaller than the auditory. In this way, the  $r_{max}$  is equal to the grid width which would be 20 for the auditory, cortical, and multisensory map whilst 10 for the visual map. The minimum radius  $r_{min}$  is 3 for all modules. We set the bandwidth to  $r_s = 300$ , and the maximum and minimum learning rates to  $l_{max} = 0.1$  and  $l_{min} = 0.001$ , respectively. Training lasts for 700 epochs with all stimuli being randomly presented per epoch. These parameters were selected by an in depth analysis [15]. The Hebbian rule variation we employ is used to update each weight. After the weight update we normalize all weights to avoid exponential growth.

$$w'_{kij}(t+1) = w_{kij}(t) + \varepsilon(t)x_k y_{ij} \quad (6)$$

$$w_{kij}(t+1) = \frac{w'_{kij}(t+1)}{\sum_{i=1}^m w'_{ij}(t+1)} \quad (7)$$

### III. EXPERIMENTS AND EVALUATION

We will first demonstrate the development of the unisensory maps and the creation of coordinate systems for each. Then we will evaluate whether the maps have developed a magnified central area, thus reflecting the dense (fovea) and non-dense (peripheral) areas of the input representations. Following this, we will examine coordinate alignment of the higher levels of the simulated cortical map as well as the multisensory map. Based on the developed topographic properties we will introduce example test data sets that comprise coincident, non-coincident and multi-stimuli data to evaluate the responses with respect to multisensory enhancement and suppression. Finally, by enabling and disabling cortical input to the multisensory map we will evaluate the effect that the former has on the manifestation of multisensory responses inspired by the observations of Stein and Meredith [1] from experiments on cats.

The input is a simplified representation of the visual and auditory space corresponding to spatially related stimuli. The inputs are in the form of Gaussian activity patterns (cf. [16]).

Such inputs provide us with a simple stimuli that is flexible enough to capture key properties, such as signal strength

(amplitude) and location. This helps us to obtain a clear understanding of the model's behavior. Each Gaussian activation pattern is centered at random intervals within a sensory space of elevation  $i$  and azimuth  $j$ . The values of an input  $x$  for a stimulus centered at  $(c,d)$  are given from:

$$x_{ij} = \lambda e^{-\left(\frac{(i-c)^2+(j-d)^2}{2\sigma^2}\right)} \quad (8)$$

with  $\lambda$  being the amplitude and  $\sigma$  the bandwidth. We introduce two pairs of bandwidth and amplitude per modality in order to represent dense and non-dense regions, as used in [11]. Greater bandwidth with smaller width activations are used for the dense, whilst lower bandwidth with larger width for non-dense firing regions. In this way we capture accuracy of stimulus representation depending on its location on the sensory space. This is based on sensory stimuli occurring approximately within the centre of the sensory space corresponding with the fovea [1]. An example of this input representation is seen in Figure 3.

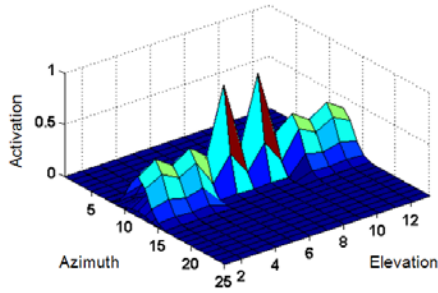


Fig. 3. A stripe of six auditory inputs. Each Gaussian activation pattern is a single input for the auditory (unisensory) map. The two patterns in the centre fall within the dense region hence they have higher amplitudes and smaller widths compared to the remaining four in the non-dense areas.

#### A. Development of Unisensory Maps

In this phase we train the modality specific maps. Each map is trained with its own data set. Both data sets used are Gaussian activation patterns. We use a 25 by 13 auditory and a 9 by 13 visual space that correspond to an azimuth and elevation range of  $[-180, 180]$ ,  $[-90, 90]$  and  $[-90, 90]$ ,  $[-65, 55]$ , respectively. These degrees correspond to approximate actual elevation and azimuth of sensory spaces [1]. The range of the input space results from using a discrete step size of 15. We choose this as a computational efficient size of input examples.

Each input contains a dense and non-dense area representing fovea and peripheral areas of stimuli encoding. For the auditory input this area is within azimuth  $[-180, 180]$  and elevation  $[-30, 30]$  whilst for the visual with azimuth  $[-15, 15]$  and elevation  $[-15, 15]$  as observed by [1]. Thus, the dense areas fall within azimuth units 1-25 and elevation units 5-9 for the auditory modality, while approximately from azimuth 6-8 and elevation 4-6 for the visual modality. We note that for the visual modality the units correspond to an azimuth of  $[-15, 15]$  and elevation of  $[-10, 20]$  due to the step size of 15. In this way, dense area coverage is 38.5% of

the auditory and 7% of the visual unit space. The full details of the input specification can be seen in Table 1.

The number of training examples (6000 for auditory and 2400 for visual) was empirically chosen in order to achieve sufficient overlap, crucial for associative learning approaches, as the input centers are randomly presented within each of the respective input spaces. The number of auditory examples is larger than the visual examples reflecting the size difference of the auditory and visual input spaces.

In order to examine organization, we recorded the locations of the maximally activated neurons when inputting stripes of stimuli according to azimuth and elevation. These stimuli were different to the training set. As an example to illustrate the map organization, for the auditory maps we chose three input stripes between  $[-90, -45]$ ,  $[-30, 30]$  and  $[45, 90]$  of azimuth  $[-180, 180]$ . Also, for an elevation of  $[-90, 90]$  we used example azimuth stripes of  $[-180, -60]$ ,  $[-45, 60]$  and  $[75, 180]$ .

TABLE I  
TRAINING AND TESTING DATA FOR UNISENSORY MAPS

Modalities		Auditory	Visual
<b>Input Space</b>			
Non-dense	Elevation	$[-90, -30]; (30, 90]$	$[-65, -15]; (15, 55]$
	Azimuth	$[-180, 180]$	$[-90, -15]; (15, 90]$
	Gaussian	$\lambda = 0.5 \sigma = 30$	$\lambda = 0.5 \sigma = 30$
Dense	Elevation	$[-30, 30]$	$[-15, 15]$
	Azimuth	$[-180, 180]$	$[-15, 15]$
	Gaussian	$\lambda = 1 \sigma = 15$	$\lambda = 1 \sigma = 15$
Input dimension		325 (13 by 25)	117 (9 by 13)
<b>Unisensory Training and Testing</b>			
Whole area examples		4200	2100
Additional dense examples		1800	300
Total examples		6000	2400

As seen in Figure 4a, each auditory output stripe obtained a distinct location on the map. Our results show that 52.3% of the total number of maximally activated units belonged to the stripe of azimuth  $[-180, 180]$  and elevation  $[-30, 30]$ , which corresponds to the auditory dense area. The map therefore has a magnified representation of the input dense area on the auditory map since it only covers 38.5% of the input space. Similarly, we repeated the same process by presenting the visual maps with visual stimuli stripes (Figure 4b). The maximally activated number of neurons for the input dense area comprised 15%, whilst the input dense area was 7% of the total input space. The results from both maps shows that the unisensory maps can topographically organize the training examples (auditory and visual examples respectively) whilst also manifesting the magnification effect observed in the actual neural structure.

#### B. Multisensory Integration

In this section we evaluate the integration map responses. We are particularly interested in the multisensory map responses when presented with a) topographically coincident multisensory information (i.e. an auditory and visual stimulus originating from the same location) b) non-coincident stimuli, and c) unimodal stimuli. This is because

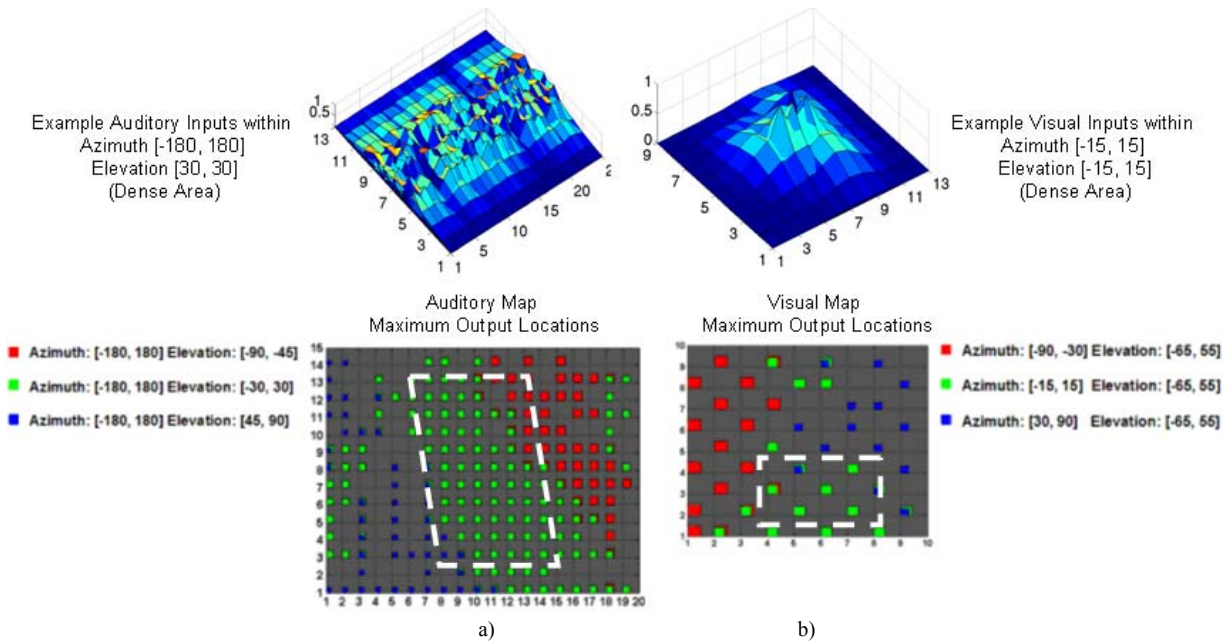


Fig. 4. Maximum neuron activation stripes of auditory and visual map: a) auditory map outputs according to elevation, and b) visual map outputs according to azimuth. The clusters on the maps formed by neighboring coordinate stripes are indicative of the developed topographical organizations. The dashed white rectangles indicate the approximate locations on the maps where the example inputs of the auditory and visual fovea are encoded.

it has been established that coincident stimuli induce higher responses (enhancement) compared to non-coincident stimuli, which are often suppressed. In particular, we examine what the effect is of the cortical input to such responses as it is hypothesized that the feedback controls the responses of multisensory stimuli [4,17]. The integration maps are trained using outputs from the unisensory maps.

TABLE II  
TRAINING AND TESTING DATA FOR INTEGRATION MAPS

Integration Maps Training				
Type	Azimuth	Elevation	Auditory	Visual
Coincident	[-90, 90]	[-65, 65]	1500	1500
Non-coincident	[-90, 90]	[-65, 65]	500	500
Only Auditory	[-180, 180]	[-90, 90]	500	*
Only Visual	[-90, 90]	[-65, 65]	*	500
Total Concurrent Presentations			3000	

\* zero inputs

For this phase we create a new training data set. It contains a mixture of stimuli that fall: a) within identical locations for both input spaces (coincident input), b) within the input spaces of both modalities but in different locations (non-coincident), c) within the auditory space only with zero inputs from the visual modality, d) within the visual input space only, with zero inputs from the auditory modality. By zero input we define an example with zero activation to indicate the absence of a stimulus. The aim of this data set is to establish whether the integration maps are able to organize and produce multisensory responses. For this reason 50% of the examples used were coincident whilst the rest were of the remaining three types (17% each), obtaining a total of 3000 training examples (Table 2). We empirically chose these percentages in order to demonstrate that the simultaneous presentation of non-coincident input is not

adequate to disrupt the map organizations even if they comprised 50% of the total input examples. The inputs were presented to the unisensory maps. Their outputs were used to first train the cortical map. After training of the latter finished, we concatenated the outputs of the cortical and the unisensory maps and used them to train the multisensory map.

For testing we use examples with integer center coordinates. We do this in order to simplify the interpretation of results since examples at distinct locations have maximum amplitude of 1. Therefore, the term coincident examples from this point onwards refer to 117 example pairs (auditory and visual) that fall within the same locations of the auditory and visual maps. This number is derived from the size of the visual input space (13 by 9). Since the visual space is smaller than that of the auditory, we use the auditory spatial parameters for the multisensory space. The non-coincident inputs are examples that fall within the same space boundaries as the coincident but the pairs are not of the same location. Particularly, we are using patterns with distinct centers originating from opposite ends on the common space of the auditory and visual maps, converging towards the centre. Thus, we have again 117 examples, with example 59 being the only coincident input as it is the crossing-point of the modalities' patterns (Figure 5). By employing this structure for the non-coincident data we can observe in a controlled way how the coincident compare to the non-coincident activations, as they have in common one part per example pair (with the other part at a known location).

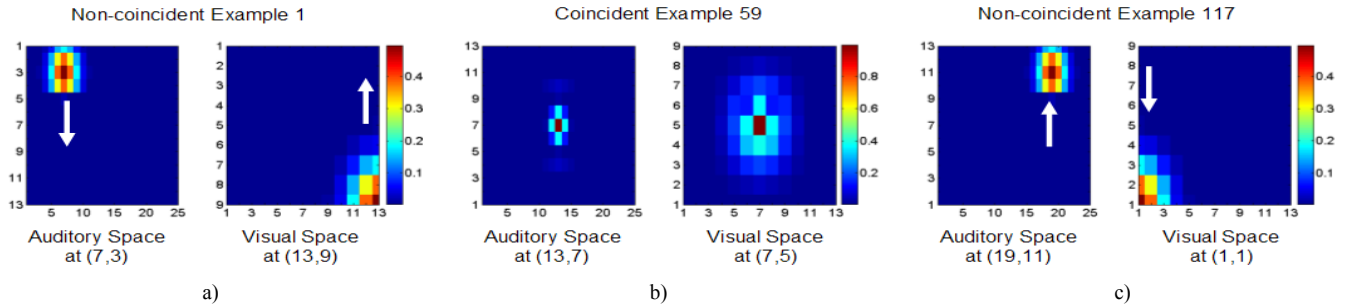


Fig.5. Maximum neuron activations corresponding to coincident and non-coincident stimuli. The white arrows indicate the direction of consecutive input pairs: (a) the first of the non-coincident input examples, (b) cross-over point of non-coincident input examples, and (c) the last non-coincident example.

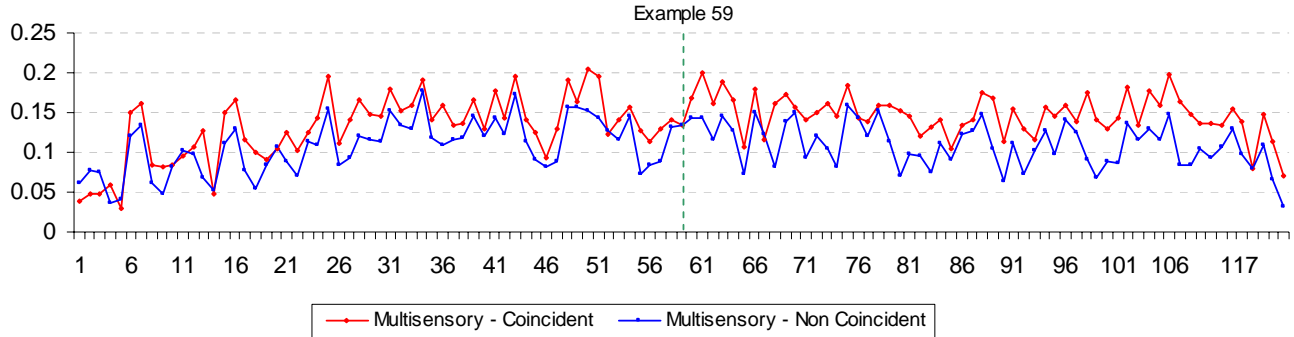


Fig. 6. Maximum neuron activations for coincident (red line) and non-coincident data (blue line) for the multisensory map with cortex enabled. Test data were 117 examples having distinct centers on the auditory and visual space.

For the auditory only inputs we use distinct centers throughout the entire auditory input space (25 by 13) whilst setting zero amplitude inputs at the corresponding visual space. This yields a total of 325 auditory examples. Similarly for the visual only inputs, we use distinct centers throughout the entire visual space (9 by 13), with the auditory inputs set to zero amplitudes.

We proceed to compare the multisensory map activations when presented with coincident and non coincident inputs. As seen in Figure 6 the coincident inputs induce higher outputs than the non coincident ones. This is attributed to the multisensory map developing weight formations that respond maximally to coincident stimuli. In other words, non-coincident inputs did not fall within the developed neuron weights and this resulted in an overall reduced neuron response. Throughout the examples, coincident values are higher (mean 0.138) compared to the non-coincident (mean 0.108) for the multisensory map (with statistical significance  $p < 3.757e-011$ ). This verifies the fact that the map learns to respond maximally to coincident auditory-visual stimuli whilst showing suppressed activations to non coincident ones. In this way we were able to reproduce a well known phenomenon of the SC where locally or temporally disparate stimuli demonstrate suppressed activations [2]. We note that in Figure 6, example 59 is the crossing point of the non-coincident data we used, and hence the activation values are the same.

Our final task was to investigate whether the multisensory map's outputs were affected by enabling and disabling the cortex. This is shown in Figure 7 where the outputs of the

117 locally coincident multisensory inputs (audio and visual) are presented. We observe that activations drop when the cortex input is disabled from a mean of 0.138 to a mean of 0.112 ( $p < 1.7868e-009$ ). This shows that the magnitude of the multisensory map's activations is dependent on the input received by the cortex module. Neurobiological findings also support that by disabling connectivity of cortical areas to the SC, such as the anterior ectosylvian sulcus (AES) and the rostral lateral suprasylvian sulcus (rLS), effects multisensory enhancement whilst leaving relatively unaltered modality specific responses [17,18]. Alvarado et al. [17] is one of the latest studies to have shown this by measuring responses of a multisensory neuron when presented with cross-modal (auditory and visual) coincident stimuli. They demonstrated that the *superadditivity* phenomenon of a multisensory neuron (using three different intensity levels of the visual stimulus effectiveness) is lost when the AES and rLS are deactivated, and reappear when they are reactivated. Although in our experiment we did not use different effectiveness levels and were not concerned with the level of enhancement (additivity or superadditivity) we managed to reproduce maximal responses only when the cortex map was enabled. This is summarized in Table III compared to the responses from [17], where enhancement changes between +10% to -59.06% are recorded after deactivating and activating AES and rLS. This was also apparent in our model with an activation change of -18.84% after disabling the cortex module and comparing the multisensory map responses to the ones recorded when it was enabled.

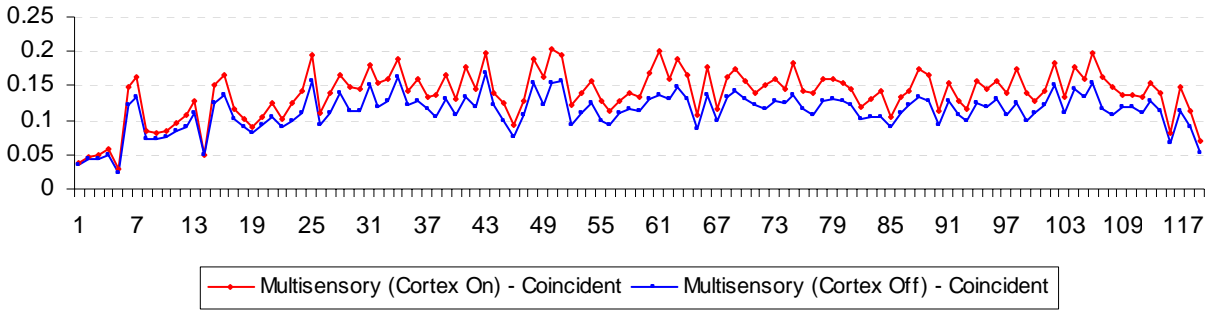


Fig. 7. Maximum neuron activations of multisensory map for coincident audio-visual data with cortex on (red line) and off (blue line). Activating the cortex (red) clearly produces higher responses compared to it being deactivated (blue). Test data were 117 examples having distinct centres on the auditory and visual space.

TABLE III  
RESPONSE COMPARISON TO ALVARADO ET AL. [4,17] RESULTS

Modality Specific	Multisensory Neurons		Unisensory Neurons	
Alvarado et al. 2009 [4]				
Stimuli	Visual 1	Visual 2	Visual 1	Visual 2
Deactivate FAES and AEV				
Increase	7.32%	0%	6.45%	3.23%
<b>No Change</b>	<b>24.39%</b>	<b>28.57%</b>	<b>83.87%</b>	<b>90.32%</b>
Decrease	<b>68.29%</b>	<b>71.43%</b>	9.68%	6.45%
Our Results				
Multisensory map activations percentage change after turning off cortex				
Auditory activations magnitude within unisensory space				-13.11%
Auditory activations magnitude within multisensory space				-18.78%
<b>Cross Modal</b>		<b>Multisensory Neuron Response Magnitude</b>		
Alvarado et al. 2007 [17]				
Visual + Auditory Stimuli (Using alternate visual stimulus effectiveness)				
Deactivated AES and rLS		Enhancement between 48% to 70%		
Activated AES and rLS		Enhancement between 43% to 171%		
Enhancement Change		Between +10% to -59.06%		
Our Results				
Multisensory map activations percentage change after turning off cortex				
Visual-auditory activations magnitude with cortex module off				-18.84%

To further illustrate the correspondence of our model with neurobiological recordings (summarized in Table III) we illustrate the multisensory map's responses when presenting it with auditory input only. Inputs are located at distinct locations throughout the whole auditory input space (325 examples of distinct centers resulting from a 25 by 13 auditory space). We observed that auditory input responses of the multisensory map that fell within the unisensory space were effected less (-13.11%) by turning the cortex off (Table III). On the other hand, greater activation differences (-18.78%) were observed for auditory inputs falling within the multisensory space (Figure 8). Thus the overall mean value of the activations with the cortex on and off (0.048 and 0.040 respectively with  $p < 0.021$ ) was mostly affected by the multisensory rather than the unisensory space activations. A similar effect was observed when presenting only visual inputs (mean values of 0.050 and 0.042 having the cortex on and off respectively, with  $p < 0.001$ ). Therefore our model's outputs are comparable with neurobiological findings after cortical area deactivation (FAES and AEV) where multisensory neurons responded more (decrease) and unisensory neurons less (no change) even to presentations of a single modality stimulus [4,17] (Table III).

#### IV. CONCLUSION

In this paper we presented a model of the superficial and deep layers of the SC using a modularized associative learning architecture. Our model provides a complete functional representation of the SC, which includes unisensory topographic maps and coordinate alignment of these maps to form a multisensory space. We demonstrated how topographic unisensory maps can be developed for the auditory and visual modalities using abstract stimuli. Although the organization was not perfect, clusters of spatially similar stimuli are formed clearly within each map. This includes greater representation of the dense (fovea) and non-dense (peripheral) areas of each sensory input, commensurate with biological data. Having developed these unisensory representations, we then input these to a simple representation of cortical processing. This, together with the unisensory map outputs, was then used to form a multisensory map that exhibited behavior found in the SC. In particular, when a non-spatially coincident stimuli is input to the auditory and visual maps, the multisensory signal is suppressed, while spatially coincident stimuli are enhanced. While the enhancement demonstrated in the model is not superadditive as found in the real SC, neither suppression or enhancement is achieved without cortical feedback. Here, we found that the cortex only affected multisensory map responses for inputs that fell within the multisensory space. These findings reflect some of the core properties of the SC.

This model for the first time shows how each aspect of the SC may be implemented using a uniform architecture. Our model demonstrates how unisensory representations of the auditory and visual spaces can be developed through a process of competition. These topographic maps, which have their own coordinate systems, are aligned through the same process of learning to form a multisensory space. This space in turn is moderated by cortical feedback which is essential to cause increased activations for coincident multimodal, versus non-coincident or unimodal stimuli, albeit without the property of inverse effectiveness (cf. [19]).

With more and more complex models of the brain and sensory processing now being developed (cf. [20]), we must also ask what this model contributes computationally. For the most part, modeling the computational properties of

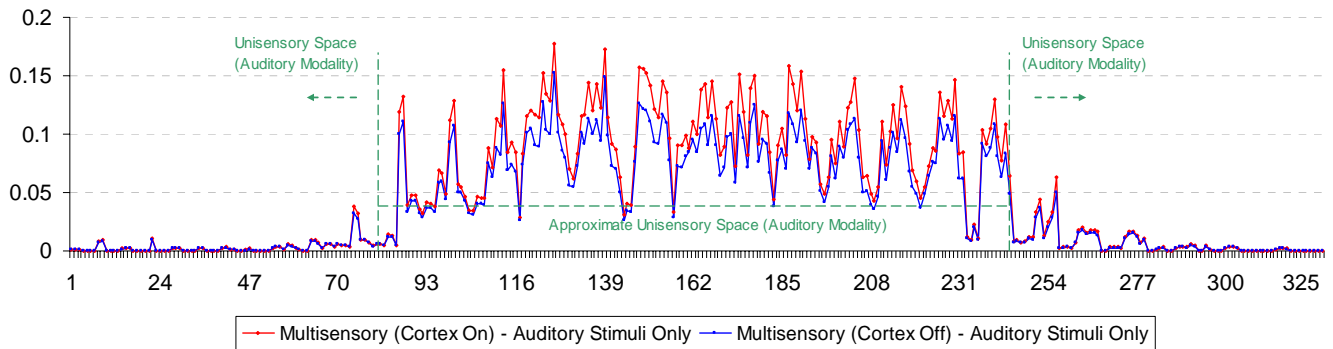


Fig. 8. Maximum neuron activations of the multisensory map for auditory data only with cortex on (red line) and off (blue line). The cortex effects responses of auditory inputs to multisensory neurons only when falling within the multisensory space. For inputs that fall within the unisensory space, no changes were observed by having the cortex on or off. Test data were 325 examples having distinct centers on the auditory space.

sensory processing has focused on cortical columns, which integrate various levels of processing with feedback. Typically, such models treat low level structures simply, either as inputs (the retina) or to route signals (the thalamus). From a computational perspective, models of the SC demonstrate that there is more to computational modelling than the cortex. Although the cortex is clearly an important area to study and model, we must not forget the role that the midbrain and other lower structures play, particularly because they seem to be an integral part of processing, as demonstrated through the likes of cortical feedback to these structures. In our model, we have demonstrated how such low level structures can perform important and non-trivial functions which are important enough to be actively influenced by cortical processing. Through a unified, if simplistic learning paradigm, we have shown how different processing levels in the brain can be simulated. The next stage of this work is to expose this model to real-world stimuli to determine if it can process more complex inputs of images from video and audio from microphones in real-time.

#### REFERENCES

- [1] B.E.Stein and M.A.Meredith, *The Merging of the Senses*. Cambridge, MA.: A Bradford Book, MIT Press, 1993.
- [2] A.J.King, "The Superior Colliculus," *Current Biology*, vol. 14, pp. R335-R338, 2004.
- [3] B.E.Stein, W.Jiang and T.R.Stanford, "Multisensory Integration in Single Neurons of the Midbrain," G.A.Calvert, C.Spence, and B.E.Stein, Ed. Cambridge, MA.: A Bradford Book, MIT Press, 2004, pp. 243-264.
- [4] J.C.Alvarado, T.R.Stanford, B.A.Rowland, J.W.Vaughan and B.E.Stein, "Multisensory Integration in the Superior Colliculus Requires Synergy among Corticocollicular Inputs," *The Journal of Neuroscience*, vol. 29, pp. 6580-6592, 2009.
- [5] M.Corbetta and G.L.Shulman, "Control of Goal-directed and Stimulus-driven Attention in the Brain," *Nature Reviews Neuroscience*, vol. 3, pp. 215-229, 2002.
- [6] L.A.Notman, P.T.Sowden and E.Özgen, "The Nature of Learned Categorical Perception Effects: A Psychophysical Approach," *Cognition*, vol. 95, pp. B1-B14, 2005.
- [7] S.Grossberg, K.Roberts, M.Aguilar and D.Bullock, "A Neuronal Model of Multimodal Adaptive Saccadic Eye Movement Control by Superior Colliculus," *Journal of Neuroscience*, vol. 17, pp. 9706-9725, 1997.
- [8] T.J.Anastasio, P.E.Patton and K.Belkacem-Boussaid, "Using Bayes' Rule to Model Multisensory Enhancement in the Superior Colliculus," *Neural Computation*, vol. 12, pp. 1165-1187, 2000.
- [9] P.E.Patton and T.J.Anastasio, "Modeling Cross-Modal Enhancement and Modality-Specific Suppression in Multisensory Neurons," *Neural Computation*, vol. 15, pp. 783-810, 2003.
- [10] E.Magosso, C.Cuppini, A.Serino, G.Di Pellegrino and M.Ursino, "A Theoretical Study of Multisensory Integration in the Superior Colliculus by a Neural Network Model," *Neural Networks*, vol. 21, pp. 817-829, 2008.
- [11] M.C.Casey and A.Pavlou, "A Behavioral Model of Sensory Alignment in the Superficial and Deep Layers of the Superior Colliculus," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN) 2008*, 2008, pp. 2751-2756.
- [12] J.Huo and A.Murray, "The Adaptation of Visual and Auditory Integration in the Barn Owl Superior Colliculus with Spike Timing Dependent Plasticity," *Neural Networks*, vol. 22, pp. 913-921, 2009.
- [13] J.L.Armony, D.Servan-Schreiber, J.D.Cohen and J.E.LeDoux, "Computational Modeling of Emotion: Explorations Through the Anatomy and Physiology of Fear Conditioning," *Trends in Cognitive Sciences*, vol. 1, pp. 28-34, 1997.
- [14] T.Kohonen, "Self-Organized Formation of Topologically Correct Feature Maps," *Biological Cybernetics*, vol. 43, pp. 59-69, 1982.
- [15] A.Pavlou, "Computational Modelling of Visual Fear Conditioning," Unpublished doctoral thesis, University of Surrey, Guildford, UK, 2009.
- [16] J.L.McClelland, A.Thomas, B.D.McCandliss and J.A.Fiez, "Understanding Failures of Learning: Hebbian Learning, Competition for Representational Space and Some Preliminary Experimental Data," J.Reggia, E.Ruppin, and D.Glanzman, Ed. The Netherlands: Elsevier Science BV, 1999, pp. 75-80.
- [17] J.C.Alvarado, T.R.Stanford, J.W.Vaughan and B.E.Stein, "Cortex Mediates Multisensory but not Unisensory Integration in Superior Colliculus," *Journal of Neuroscience*, vol. 27, pp. 12775-12786, 2007.
- [18] M.T.Wallace and B.E.Stein, "Cross-modal Synthesis in the Midbrain Depends on Input from Cortex," *Journal of Neurophysiology*, vol. 71, pp. 429-432, 1994.
- [19] B.A.Rowland, T.R.Stanford and B.E.Stein, "A Model of the Neural Mechanisms Underlying Multisensory Integration in the Superior Colliculus," *Perception*, vol. 36, pp. 1431-1443, 2007.
- [20] E.Izhikevich and G.M.Edelman, "Large-Scale Model of Mammalian Thalamocortical Systems," *Proceedings of the National Academy of Sciences of the USA*, vol. 105, pp. 3593-3598, 2008.