

SACCADES GENERATION

From the Visual Input to the Superior Colliculus

Wahiba Taouali

Université Henri Poincaré - LORIA, Campus Scientifique, Vandoeuvre-lès-Nancy Cedex, France

Nicolas Rougier, Frédéric Alexandre

INRIA Nancy - Grand Est Research Center, Villers les Nancy Cedex, France

Keywords: Superior Colliculus, Dynamic Neural Field, Visual Attention, Ocular Saccades

Abstract: The superior colliculus is an important structure in the visuomotor pathway of mammals, that is known to be deeply involved in visual saccadic behavior. We present a model of this structure based on biological data, the specificity of which is related to the homogeneity of the underlying substratum of computation. This makes it more suitable to process massive visual flows on a distributed architecture, as it could be requested in a realistic task in autonomous robotics. The model presented here is embedded in the exogenous part of the visual pathway, from the retina to the superior colliculus.

1 INTRODUCTION

Displaying an intelligent behavior is often synonymous of intelligently exploiting the surrounding environment. In many animals and animats, this is massively performed through the visual channel. Particularly, two adaptive behaviors allow to decrease the huge amount of information brought by this channel: visual attention proposes a sequential processing of possible targets; saccadic movements orient the body and particularly the fovea on regions of interest in the visual scene. The premotor theory of attention (Rizzolatti et al., 1987) stipulates that there are common processes between these key behaviors and, more precisely, that they share common neuronal circuits: Attention would be pre-programming of a saccade. The importance of these visuomotor behaviors and the impact of the premotor theory certainly explains why there have been so many works to observe and model the superior colliculus (SC). Indeed, this small structure in the midbrain of mammals is known to be implicated in these behaviors. From an hodological viewpoint, it integrates visual information from many sources (cortical or not) in the brain and sends projections toward the brainstem premotor circuits that trigger saccades (Isa, 2002). From an anatomical viewpoint, it consists of a set of topological maps, mapping the surrounding space, from

visual to motor reference frames (Girard and Berthoz, 2005). And from a physiological viewpoint, its inactivation or electrical stimulation confirms its role in visual attention and saccades (Muller et al., 2005).

Many models have studied the SC and associated properties (cf. (Girard and Berthoz, 2005) for a review). We just mention here some models underlying the link to information flows and underlying behavior. The structure of the model described in (Findlay and Walker, 1999) underlines that the main task is to decide when and where the saccade must be performed. As a consequence, two hierarchical axes are defined. The When axis (corresponding to the FEF (Frontal Eye Field) area in the prefrontal cortex) decides when to leave the current fixation point, whereas the Where axis corresponds to the SC and implements a spatial competition between candidate targets. A double-axis model combining FEF and the SC is also proposed in (Kramer et al., 1999), to explain the integration of exogenous elements (external stimuli coming from the retina to the SC) and endogenous elements (internal expectancies or instructions elaborated in the prefrontal cortex). Later on, (Godijn and Theeuwes, 2002) proposed a competitive integration model based on strong experimental evidences at the behavioral level, indicating that all these elements (spatial vs temporal processing and integration of exogenous vs endogenous stimuli) can be integrated in a

unique map, seen as a model of the SC. This common saccade map also includes features generally reported as physiologically plausible in the SC: a local excitation in the map allowing to combine close stimuli and a wider inhibition mechanism to trigger a competition between far stimuli. This interaction scheme explains why it was possible to use such a formalism as Dynamic Neural Field (DNF) (Amari, 1977) to implement this kind of model, as it is also the case in (Trappenberg et al., 2001; Schneider and Erlhagen, 2002).

In summary, the SC is undoubtedly an important integrative structure, to be included in a cognitive neuroscience modeling approach of visuospatial behaviors. Most recent models give a stronger role to the SC at the price of a more complex internal functioning. Indeed, those models define several kinds of units, depending on their location on the map, which is not very consistent with the principle of homogeneity in DNF. More precisely, in (Trappenberg et al., 2001), the reported behavior is obtained with some units standing for the currently fixated stimulus (consequently in the fovea), other units representing potentially fixated stimuli in the periphery, both kinds sending inhibition to other units triggering saccades toward a target. These kinds of units, also exploited in the model by (Schneider and Erlhagen, 2002), are presented as representing respectively so-called fixation, build-up and burst neurons, which are sometimes reported as parts of the intermediate layer of the SC (Wurtz and Optican, 1994), though this is still to be clearly established. In (Godijn and Theeuwes, 2002) also, the substratum of computation is not homogeneous, since the sensitivity of units decreases with their eccentricity onto the map. This trick is used to reproduce the observation that the latency of a saccade toward a target, presented together with a distractor, is longer when the distractor is closer from the rostral zone of the SC (corresponding to the fovea).

In this paper, we present a model of the SC, based on DNF formalism, with an identical functioning rule for all the units in the map. Obtaining such an homogeneous substratum can yield fully distributed computation, which is important to design models that can be used online in robotic visuomotor tasks. Finally, the proposed model is not an isolated structure but is a function of exogenous information flows and associated geometrical properties.

2 MODEL

We designed a three layers neural network model of the visual pathway for generating a saccade from the

retina (R) to the superior colliculus (SC) through the primary visual cortex (V1):

- Retina (R, 256×512 units) receives visual input from a CCD camera.
- Visual cortex (V1, 256×256 units) implements the actual cortical magnification.
- Superior colliculus (SC, 63×63 units) is the place where salient locations enters competition.

The retina model is restricted to the right visual field as it is known to be the case in mammals visual pathway (left visual field projects to right colliculus and right visual field projects to left colliculus). We used an image size of 512×512 pixels and fed the retina with a normalized gray-level image of size 512×256 pixels.

2.1 Cortical Magnification

The retina, the sensory input space, has a complex structure composed of layers of neurones. The phenomena of vision begins in the layer of photo-receptors. Then the flow is transmitted to the ganglion cells which are large nerve cells whose cylinders form the optic nerve. Due to the non homogeneous repartition of photo-receptors on the (human) retina surface, visual acuity decreases from the center of the retina (fovea) to the periphery. This property is attributed to a variation in the density of photo-receptors that decreases from the center to the periphery (Marilly et al., 1999). Consequently, the foveal region benefits from a much higher resolution than peripheral regions and this property is preserved along the visual pathway up to early visual areas (Purves, 2004). This is referred as cortical magnification. To analyze this magnification in a quantitative way, a coordinate system is often defined in the visual field. The coordinates that are best suited to the visual system are polar coordinates (ρ, ϕ) that characterizes a position in the visual field by its eccentricity ρ from the center of gaze and its polar angle ϕ is measured for example in relation to the lower vertical meridian. We can therefore define a retinotopic map which corresponds to the spatial transformation of the image by the spatial arrangement of the grid of neurons. It is often approximated by a log-polar transformation of the spherical image centered on the eye (Robinson, 1972). We used a simplified model of the retina considering only the photo-receptors layer. And for computational reasons (speed), we did not enforce the non-uniform repartition of photo-receptors on the retina surface but we modeled instead a uniform distribution of neurons onto the retina associated with a deformed polar coordinate system as proposed by

(Ottes et al., 1986). Each cortical visual cell is supposed to be connected to a single or several photoreceptor cells, with respect to a logpolar deformation, that form its receptive field. So the non uniformity is caused by the changing size of the receptive fields. We used equations mapping retinotopic polar coordinates (ρ, ϕ) onto V1 Cartesian coordinates (x, y) . These equations were first introduced by (Ottes et al., 1986):

$$x = B_x \ln\left(\frac{\sqrt{\rho^2 + 2A\rho|\cos(\phi)| + A^2}}{A}\right) \quad (1)$$

$$y = B_y \arctan\left(\frac{\rho \sin(\phi)}{\rho|\cos(\phi)| + A}\right) \quad (2)$$

with $A = 3^\circ$, $B_x = 1.4mm$, $B_y = 1.8mm$. These parameters have been chosen to fit the stimulation map of the SC given by (Robinson, 1972). A neuron in the visual cortex fires an action potential when a visual stimulus appears within its receptive field. But for any given neuron, it may respond best to a subset of stimuli within its receptive field corresponding to its preferred direction. Neurons with similar tuning properties (what the neurons respond to) tend to cluster together but the exact structure is still unclear. Then, it is acceptable to assume that V1 has a retinotopic map similar to the collicular motor map in (Bear et al., 1996). It means that a cell at a given position (x, y) in the V1 map is activated by retinal cells in positions (ρ, ϕ) according to given equations. One result of this deformation is that the same stimulus causes a large activation in the V1 map if it is located near the fovea and smaller activation in peripheral positions (fig1). Visual receptors of V1 have been modeled in two dimensions corresponding to an eye visual hemifield with no connection between the different receptors.

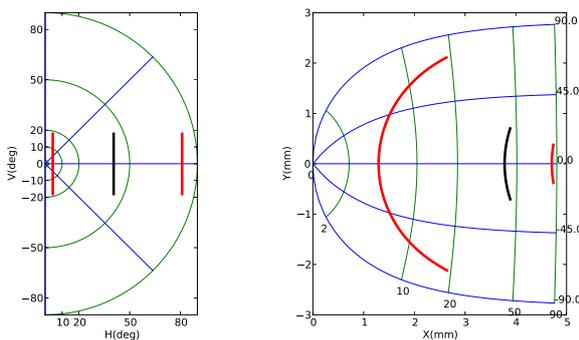


Figure 1: Cortical magnification from the retina to the visual cortex distorts geometrical properties of the image while keeping neighborhood relationship.

2.2 Dynamic Neural Field Theory

Collicular population (the motor layer of one superior colliculus) has been modeled with respect to the dynamical neural field theory (Wilson and Cowan, 1973; Amari, 1977; Taylor, 1999) that describes the evolution of a neural population using equation (see (Rougier and Vitay, 2006) for details):

$$\tau \frac{\partial u(\mathbf{x}, t)}{\partial t} = -u(\mathbf{x}, t) + \int w(\mathbf{x} - \mathbf{y}) f(u(\mathbf{y})) d\mathbf{y} + h + I(\mathbf{x}, t) \quad (3)$$

where \mathbf{x} denotes a location onto the SC; t is time; $u(\mathbf{x}, t)$ denotes the membrane potential of a neural population at point \mathbf{x} and time t ; τ is the temporal decay of synapses, f is a sigmoid function computing the mean firing rate, w is a neighborhood function, $s(\mathbf{x})$ is the input received at position \mathbf{x} and h is the mean neuron threshold. w has been set as a difference of Gaussian (DoG) with short-range excitations and long range inhibitions following anatomical and physiological data as reported in (Munoz and Istvan, 1998) (see also figure 2):

$$w(\mathbf{x} - \mathbf{y}) = A e^{-\frac{|\mathbf{x} - \mathbf{y}|^2}{a^2}} - B e^{-\frac{|\mathbf{x} - \mathbf{y}|^2}{b^2}} \quad (4)$$

and f has been set as a simple rectification of \mathbf{x} . The input $I(\mathbf{x}, t)$ is a direct one-to-one relationship according to V1 and SC respective sizes.

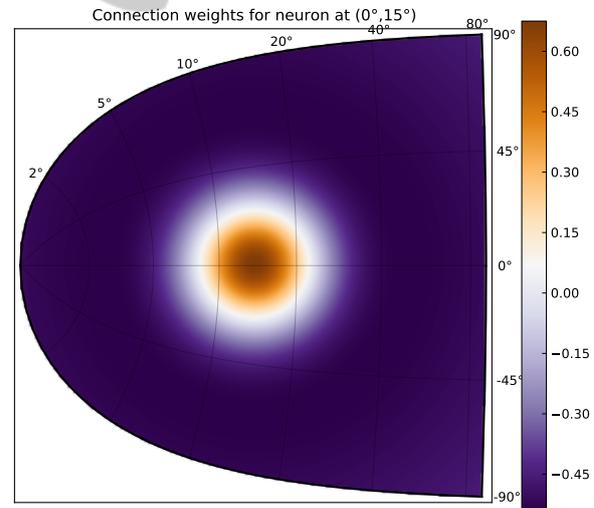


Figure 2: The lateral weights connecting neurons from the SC is a difference of Gaussian with short-range excitations and long range inhibitions following anatomical and physiological data. The figure displays incoming weights from all SC neurons to the neuron at position $(0^\circ, 15^\circ)$.

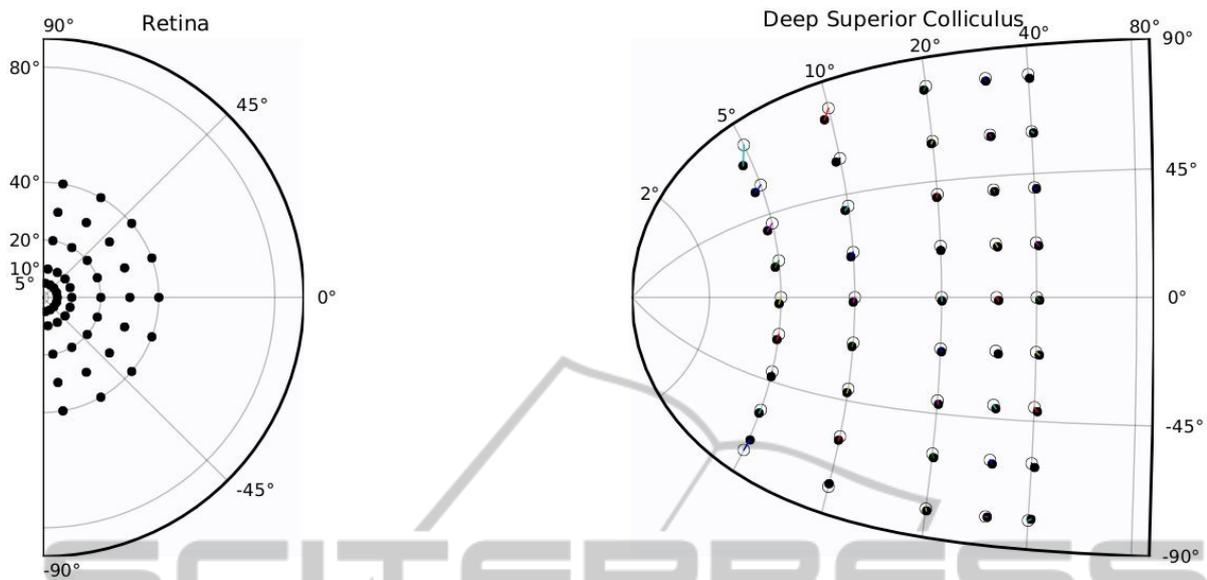


Figure 3: Accuracy of the model of the superior colliculus has been measured using a set of retina targets (black dots on left figure) that have been sequentially presented to the SC model. For each target and after convergence (difference of activity between time t and time $t + dt$ is negligible), the center of mass of the collicular activity has been decoded and represented as a circle on right figure (black dots represent the actual projection of the target in collicular coordinates).

3 RESULTS

3.1 Output Decoding

One of the questions related to the superior colliculus concerns the proper way to decode the output. Since the amplitude and direction of a saccade depend on the activity of the neural population in the deep SC (Sparks et al., 1990), different ways of SC output evaluation have been proposed in the past:

- winner-take-all where the most active site indicates the direction
- summation (McIlwain, 1976; Sparks et al., 1976) where all activities of active neurons are summed with weights determined by their individual labels
- weighted average (Lee et al., 1988) using a normalization according to the number of active neurons

These three evaluation schemes are equivalent in the case of a normally activated population but differ when there is a deactivation or an over-activation. We've retained the last decoding scheme because we modeled the superior colliculus using a dynamic neural field and it is thus ensured that a stereotyped activity profile emerges anytime corresponding to the most salient location of the V1 area. Furthermore, this stereotyped activity possesses a Gaussian shaped two-dimensional profile and it is possible to find its center

of mass. We have been testing the accuracy of this coding scheme by feeding the model with standard Gaussian shaped stimuli (aperture 0.016°) at different locations (see figure 3). Despite the magnification effect, one can see that the model has a high precision in the standard saccadic range (-30° to $+30^\circ$, 0 to 50). We also tested the inactivation of a subpart of the colliculus to check that we obtain both hypo-metric and hyper-metric saccades as reported in (Robinson, 1972) but those results are not presented in this paper.

3.2 Natural Images Processing

We have also tested the model using natural images taken from a color CCD camera. No image processing has been performed on the image but a conversion to a gray-level representation. Figure 4 exhibits an example where a subpart of a computer keyboard has been shot. This allows to illustrate the main feature of the proposed model. If one look closely at the half retina representing the keyboard (upper left part of the figure), one can see that several letters (O, P, L, M) are eligible for attention focus and an ocular saccade. However, the retinotopic projection onto the model of the V1 area reduces quite naturally this set to letters O and L. The model of the SC is thus confronted with a choice between these two locations and the dynamic field theory, as it has been introduced in the previous section, ensures that only one location remains after competition. However it is hard to spec-

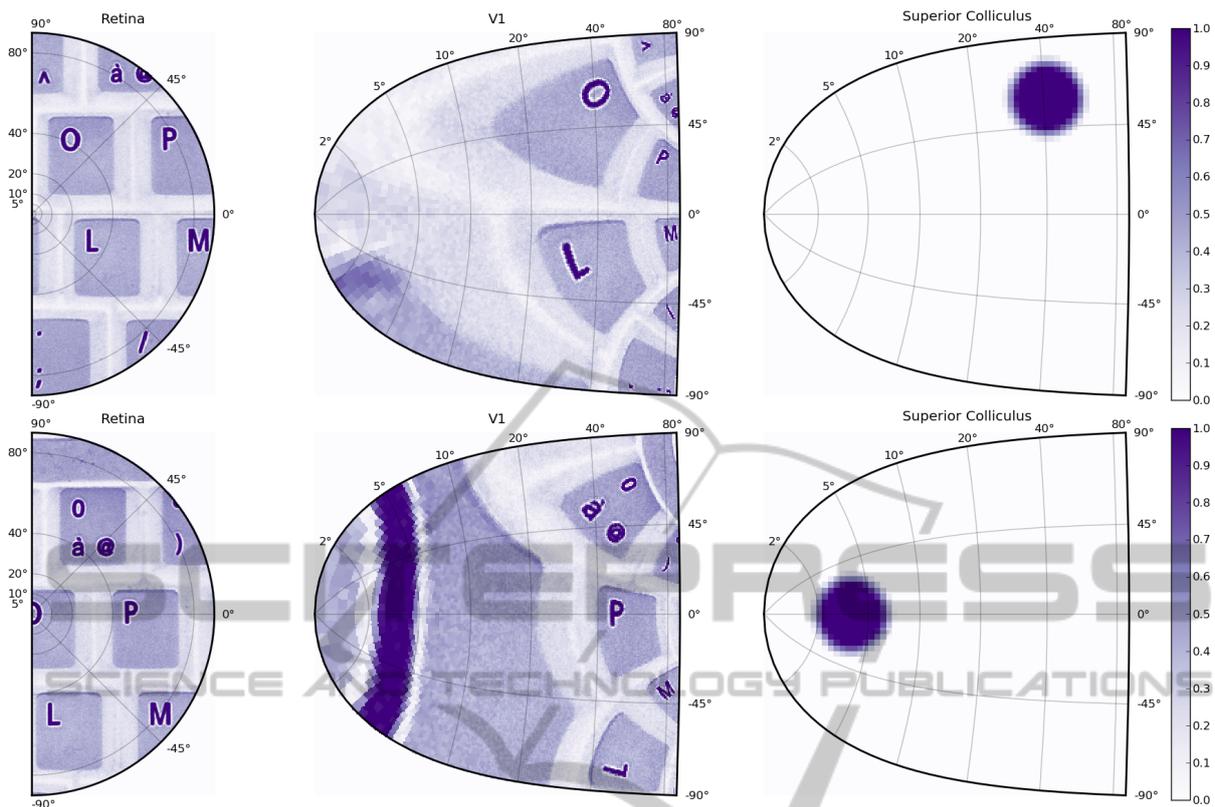


Figure 4: An image of a computer keyboard has been captured using a color camera (resolution 1024×1280) and transformed into a normalized gray level image. **Upper figure.** The right half of the image is presented to the half retina area which in turn feeds the V1 area where retinotopy is applied following equation 2. The colliculus area enters a competition stage where most salient locations are eligible for final activation and after some iterations, the competition ends up onto the O letter that is thus considered the most salient location of the visual scene according to its location and activation. **Lower figure.** A saccade has been simulated to center the O letter into the center of the fovea and the colliculus now focuses onto a subpart of the O letter that appears to be the newly most salient location of the new visual scene.

ify the exact conditions that make the model to focus on the O instead of the L letter in the given example and the spatially compact shape of the O is certainly to be taken into account. This example also underlies the inherent difficulty in temporally organizing ocular saccades without any top-down control. If we were to let the model only reacts to its environment, it would certainly focus on the most salient location without ever exploring other point of interest (from a behavioural point of view). If the actual saccade brings into view another salient location, the model would jump again to the new location (provided we inhibited the foveal region to prevent the model to be stuck forever on this single location) but in such a case, nothing would prevent the model from going to location A then location B and the again location A, being trapped in a cycle. Exploring the whole visual scene thus requires some kind of top down control to be able to dynamically inhibit visited location once they have been focused in order to favor other loca-

tions. This is out of scope of the present article but this has been already made in a less realistic model (Fix et al., 2006).

4 DISCUSSION

In many visuomotor behaviors, the superior colliculus is reported as an important multimodal map, integrating various kinds of exogenous and endogenous information. Most of the models of this structure insist on its internal structure and function, but rarely on its implication in the main information flows, which is important from a behavioral point of view.

The model of the superior colliculus presented in this paper has two specificities: (i) It is based on a bio-inspired homogeneous local functioning rule. (ii) It is linked with the external world through an information flow coming from the retina. In ongoing and future works, these specificities will be developed fur-

ther. Concerning the first aspect, the model has been designed together with colleagues from neuroscience and its justification with regard to biological data is being also prepared for publication. Concerning the second aspect, we have only considered for the moment exogenous inputs coming from the retina. Further works will consider endogenous inputs conveying such information as instructions, goals or motivations from other neural structures.

REFERENCES

- Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2):77–87.
- Bear, M., Connors, B., and Paradiso, M. (1996). *Neuroscience: Exploring the Brain*. Lippincott Williams & Wilkins.
- Findlay, J. and Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, 22(4):661–674.
- Fix, J., Vitay, J., and Rougier, N. (2006). A computational model of spatial memory anticipation during visual search. In *Anticipatory Behavior in Adaptive Learning Systems*.
- Girard, B. and Berthoz, A. (2005). From brainstem to cortex: computational models of saccade generation circuitry. *Progress in Neurobiology*, 77(4):215–251.
- Godijn, R. and Theeuwes, J. (2002). Programming of endogenous and exogenous saccades: evidence for a competitive integration model. *Journal of experimental psychology: human perception and performance*, 28(5):1039–1054.
- Isa, T. (2002). Intrinsic processing in the mammalian superior colliculus. *Current opinion in Neurobiology*, 12(6):668–677.
- Kramer, A., Irwin, D., Theeuwes, J., and Hahn, S. (1999). Oculomotor capture by abrupt onsets reveals concurrent programming of voluntary and involuntary saccades. *Behavioral and Brain Sciences*, 22(4):689–690.
- Lee, C., Rohrer, W., and Sparks, D. (1988). Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature*, 332(6162):357–360.
- Marilly, E., Mercier, A., Coroyer, C., Faure, A., and Cachard, O. (1999). Propriétés d'un pr-processeur de vision fovale. *Dix-septième colloque GRETSI*.
- McIlwain, J. (1976). Large receptive fields and spatial transformations in the visual system. *Int. Rev. Physiol.*, 10:223–248.
- Muller, J., Philiastides, M., and Newsome, W. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proceedings of the National Academy of Sciences*, 102(3):524–529.
- Munoz, D. and Istvan, P. (1998). Lateral inhibitory interactions in the intermediate layers of the monkey superior colliculus. *J. Neurophysiol.*, 79:1193–1209.
- Ottes, F., Gisbergen, J. V., and Eggemont, J. (1986). Visuomotor fields of the superior colliculus: a quantitative model. *Vision Res*, 26(6):857–873.
- Purves, D. (2004). *Neurosciences*. De Boeck, second edition.
- Rizzolatti, G., Riggio, L., Dascola, I., and Umil, C. (1987). Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25(1):31–40.
- Robinson, D. (1972). Eye movements evoked by collicular stimulation in the alert monkey. *Vision Research*, 12(11):1795–1808.
- Rougier, N. and Vitay, J. (2006). Emergence of attention within a neural population. *Neural Networks*, 19(5):573–581.
- Schneider, S. and Erlhagen, W. (2002). A neural field model for saccade planning in the superior colliculus: speed-accuracy tradeoff in the double-target paradigm. *Neurocomputing*, 44-46:623–628.
- Sparks, D., Holland, R., and Guthrie, B. (1976). Size and distribution of movement fields in the monkey superior colliculus. *Brain Res.*, 113(1):21–34.
- Sparks, D., Lee, C., and Rohrer, W. (1990). Population coding of the direction, amplitude, and velocity of saccadic eye movements by neurons in the superior colliculus. *Cold Spring Harbor symposia on quantitative biology*, 55:805–811.
- Taylor, J. (1999). Neural bubble dynamics in two dimensions: foundations. *Biological Cybernetics*, 80:393–409.
- Trappenberg, T., Dorrison, M., Munoz, D., and Klein, R. (2001). A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience*, 13(2):256–271.
- Wilson, H. and Cowan, J. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biological Cybernetics*, 13(2):55–80.
- Wurtz, R. and Optican, L. (1994). Superior colliculus cell types and models of saccade generation. *Current Opinion in Neurobiology*, 4(6):857–861.