A Cognitive Model of Saliency, Attention, and Picture Scanning

Vassilis Cutsuridis

Abstract To view and understand the visual world, we shift our gaze from one location to another about three times per second. These rapid changes in gaze direction result from very fast eye movements called saccades. Visual information is acquired only during fixations, stationary periods between saccades. Active visual search of pictures is the process of active scanning of the visual environment for a particular target among distracters or for the extraction of its meaning. This article discusses a cognitive model of saliency, overt attention, and natural picture scanning that unravels the neurocomputational mechanisms of how human gaze control operates during active real-world scene viewing.

Introduction

When we visually explore our environment, our eyes do not move in smooth continuous movements. Instead, our eyes fixate to an object for a brief period of time (around 200–300 ms) before jumping to a new position in the visual field [20]. These rapid eye movements are called saccades. Saccades are stereotyped eye movements and are ballistic. Saccades can reach very high velocities, approaching 800°/s at their maximum [20]. The size of the saccade is 12–15° [20]. On the other hand, fixations are the

stationary periods between saccades. During a fixation, parts of the visual scene are brought to the eye's fovea, where the visual acuity is maximum. The duration of fixations when viewing pictures and scenes has been shown to have a skewed distribution with a mode at 230 ms, a mean at 330 ms, and a range from less than 50 to more than 1000 ms [24]. The mean fixation duration has been shown to increase as viewing continues [49].

Active visual search of a scene with multiple elements involves the coordination of target identification, target localization, and response generation. Many brain areas and multiple parallel routes are involved in active visual search of a scene. The brain areas involved in target identification and localization are the dorsal (location) and ventral (identity) pathways of the cortex. The cortical areas linked to response generation are the lateral intraparietal (LIP) area of the posterior parietal cortex (PPC), the frontal eye fields (FEF) of the frontal cortex, and the prefrontal cortex (PFC). Projections from FEF, LIP, and PFC to the superior colliculus (SC) either directly and/or through the basal ganglia structures [25, 26] are known to exist. Direct projection from the primary visual cortex (V1) to SC has been shown [1]. Lesion studies have shown that no single pathway is essential. The combined loss of both SC and FEF renders the animal unable to make saccades [41]. Inability in making saccades occurs also with lesions to both SC and V1 [34]. SC is the common recipient of excitation from the cortices, since stimulation of these regions no longer elicits saccades following SC ablation [41]. In turn, the intermediate and deep layers of the SC project to the brainstem saccade generators, although a direct FEF pathway to brainstem has been shown [23].

Many computational theories of visual search have been proposed over the years [12, 16, 19, 28, 30, 35, 48]. Most of these models, whether involving overt eye movements

V. Cutsuridis (✉)
Department of Computing Science and Mathematics,
University of Stirling, Stirling FK9 4LA, UK
e-mail: vcu@cs.stir.ac.uk

or covert shifts of attention, are based on the concept of a bottom-up saliency-map that biases attention [28, 30]. According to these models scanpaths (sequences of saccadic eye movements) are generated according to the value of saliency or conspicuity in the map. A winner-take-all competition between units in the map ensures that the most salient unit is gazed first, followed by the second most salient one and so forth. Inhibition, the exact nature of which is still unknown, ensures that the previously gazed region is not attended again for a period of 500–900 ms. Other computational models have emphasized the need of an interaction between a bottom-up stimulus-driven module and a top-down attentive module, which drives attention to specific regions-of-interest (ROIs) in the saliency map [12, 15, 19, 22, 39, 45, 48].

Recently, Cutsuridis [12] introduced a cognitive model of active visual search based on such interaction. The model offered a plausible hypothesis of how the participating brain areas work together to accomplish a scan of a scene within the allocated time (3–4 fixations per second). In this article, I will describe this model in more detail, describe its neurocomputational mechanisms and discuss its physiological implications by answering the following questions:

- How is a complex visual scene processed?
- How is the selection of one particular location in a visual scene accomplished?
- Does it involve bottom-up, sensory-driven cues or top-down, world knowledge expectations? Or both?
- How is the decision made when to terminate a fixation and move the gaze?
- How is the decision made where to direct the gaze in order to take the next sample?
- What are the neural mechanisms of inhibition of return?

The Model

A graphical representation of the cognitive model is given in Fig. 1. The model proposes that an input image is processed in a bottom-up fashion, providing input to feature detectors, which in turn lead to the formation of salient maps (the *object* map, the *spatial* map, the *goals* map, and the *motor programs* map). Response generation is not achieved only by the degree of saliency as in Itti and Koch [28] and Koch and Ulman [30]. Adaptive resonance between salient maps is also needed. Resonance among the object, spatial, goals, and motor programs maps is achieved via a measure of degree of similarity, which depends on the amount of modulation the maps receive from the *overseer*. A winner-take-all competition between resonated salient representations ensures the salient representation that
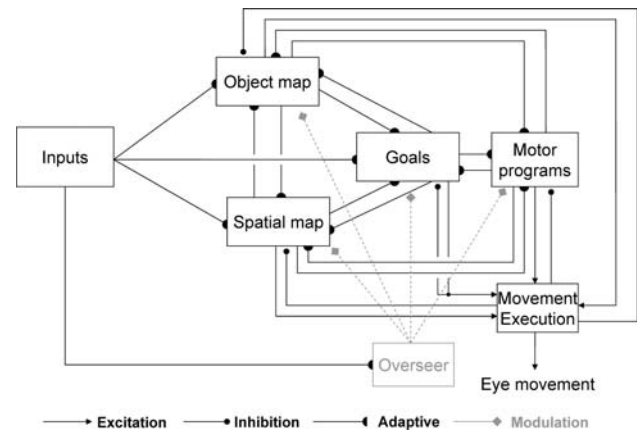

Fig. 1 Graphical representation of the cognitive model of saliency, overt attention, and natural picture scanning (see main text for details)

reached resonance first will be gazed first, followed by the second fastest, and so on. Once resonance is reached, a response (eye movement) is generated, which is sent to the motor execution module for execution. At the same time an inhibitory signal is sent back to the resonated salient maps that wipes out the representations that generated the previous response, thus allowing the second fastest representation in the queue to be expressed.

In order for the model to achieve such complicated processes, a number of modules with specific as well as distributed modules are required. The topology, interconnectivity, and proposed functionality of these modules are heavily supported by neuroscientific experimental evidence. I will describe these modules in great detail in the following section.

Input Module up to Object and Spatial Map Modules

The *input* module up to the formation of global saliency maps in both the dorsal (*space*) and the ventral (*object*) streams (see Fig. 2) is the same as in Itti and Koch [28] and Koch and Ulman [30]. Its functionality is to decompose an input image through several pre-attentive multi-scale feature detection mechanisms (sensitive to, for example, color, intensity, orientation, etc.) found in retina, lateral geniculate nucleus (LGN) of the thalamus, and primary visual cortical area (V1) and which operate in parallel across the entire visual scene, into two streams of processing, that is the dorsal for space and the ventral for object. Low-level vision features (e.g., orientation, brightness, color, hue, etc.) are extracted from the input image at several spatial scales using Gaussian pyramids, which consist of progressively low-pass filtering and subsampling of the input image. Pyramids have depth of n scales, where n is a free parameter taking integer values, providing horizontal and vertical image reduction factors ranging from 1:1 (level 0;
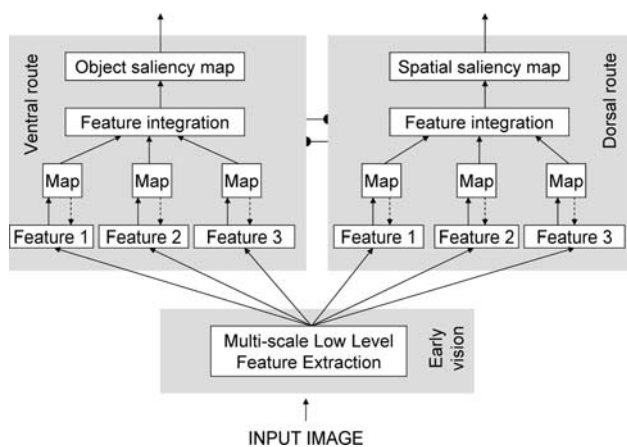
Fig. 2 Schematic of the information flow of the visual processing module (see main text for details)

the original input image) to 1:256 (level *n*). Each feature is computed in center-surround operation. Center-surround operations are implemented as differences between a fine and a coarse scale for a given feature.

Neurons in the feature maps in both dorsal and ventral streams then encode the spatial and object contrast in each of those feature channels. Neurons in each feature map spatially compete for salience, through long-range connections that extend far beyond the spatial range of the classical receptive field of each neuron. After competition, the feature maps in each stream are combined into a global saliency map, which topographically encodes for saliency irrespective of the feature channel in which stimuli appeared salient. In the model, the global spatial saliency map is assumed to reside in the PPC, whereas the global object saliency map resides in the ventral temporal cortex (TC). The speed of visual information processing from the early multi-scale feature extraction in the retina till the formation of global saliency maps in the dorsal PPC and ventral TC is 100–130 ms [47].

Goals Module

The *goals* module is represented by PFC cells. It receives a direct input visual signal from the early stages of visual processing (retina, LGN, V1) as well as from the FEF (motor plans), PPC (spatial representations), TC (object representations), and other brain areas [motivation (medial PFC), value representations (orbito-frontal cortex)]. Its role is to (1) send a *focus of attention* signal to every stage of the visual processing, which will amplify specific neuronal responses throughout the visual hierarchy, while at the same time will inhibit those of distracters, and (2) participate in the adaptive resonance process of the selectively tuned via modulation from the *overseer* module target

(spatial and object) and motor plan salient representations in the PPC, TC, and FEF.

Overseer Module

At the same time and in a parallel manner, the retinal multi-scale low-level features propagate to the upper layers of the SC, which in turn provide the sensory input to the substantia nigra pars compacta (SNc) and ventral tegmental area (VTA) (see Fig. 3). Recent neuroanatomical evidence has reported a direct tectonigral projection connecting the deep layers of the SC to the SNc across several species [5, 17, 33, 37]. This evidence is confirmed by neurophysiological recordings in freely behaving animals [4, 37].

The SNc and VTA comprise the *overseer* module of the model. Both SNc and VTA contain the brain's dopaminergic (DA) neurons, which have been implicated in signaling reward prediction errors used to select actions that will maximize the future acquisition of reward [42] as well as the progressive movement deterioration of patients suffering from Parkinson's disease [7–9, 13, 14]. The conduction latency of the signal from the retina to SC and from there to SNc is 70–100 ms, whereas the duration of the DA phasic response is ∼100 ms (see Fig. 4 and Redgrave et al. [38]).

The SC-activated SNc DA neurons broadcast neuromodulatory signals to neurons in PFC, FEF, PPC, and TCs [7]. In brief, the source of the DA fibers in cerebral cortex was found to be the neurons of the SNc and the VTA. DA afferents are densest in the anterior cingulate (area 24) and
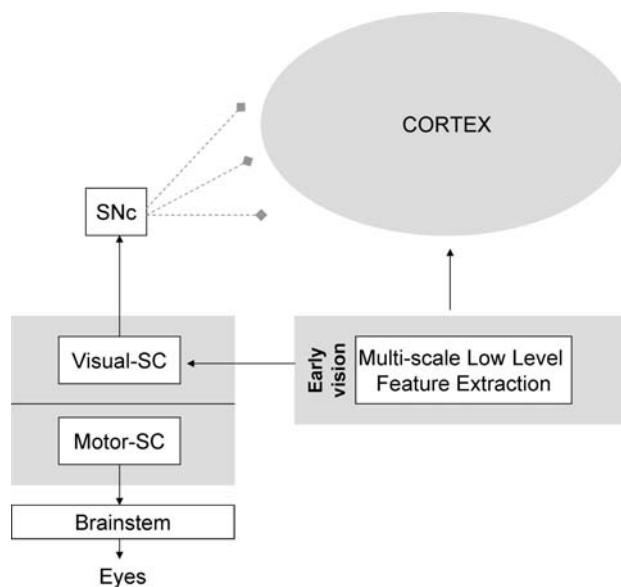

Fig. 3 Schematic of the information flow from the early vision to (1) cortex and (2) visual superior colliculus (SCv) and SNc. SNc, in turn, broadcasts modulatory signals to the cortex, which facilitate the decision-making process
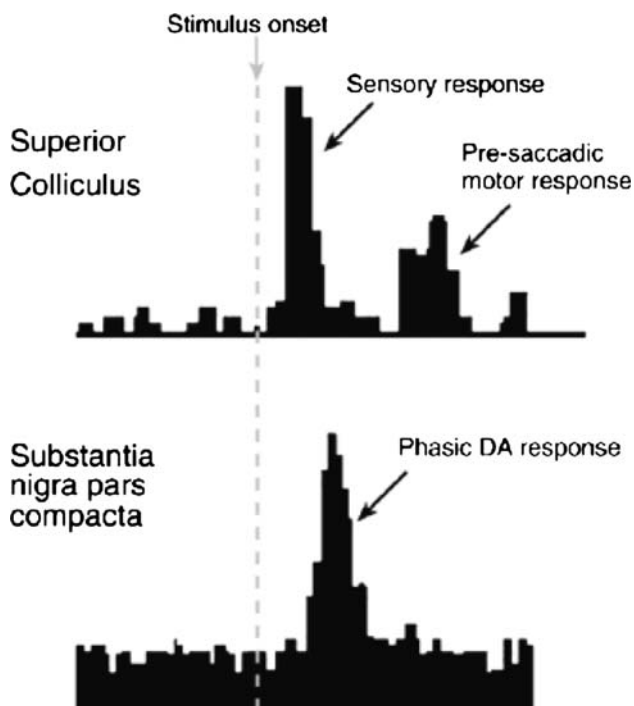
Fig. 4 Relative timing of peri-stimulus histogram responses in the SC and SNc evoked by an unexpected visual stimulus. Responses are aligned to stimulus onset. **a** Activity in the SC is characterized by an early sensory response (latency ~ 40 ms) followed by a later motor response (latency ~ 200 ms). The latter is responsible for driving the orienting gaze-shift to bring the stimulus onto the fovea (reprinted with permission from Redgrave et al. [38], Fig. 3, p. 326, Copyright© 2008 Elsevier). **b** The phasic dopaminergic response (latency ~ 70 ms) occurs after the collicular sensory response but prior to its pre-saccadic motor response (reprinted with permission from Redgrave et al. [38], Fig. 3, p. 326, Copyright© 2008 Elsevier)

the motor areas (areas 4, 6, and SMA), where they display a tri-laminar pattern of distribution, predominating in layers I, IIIa, and V–VI. In the granular prefrontal (areas 46, 9, 10, 11, 12), parietal (areas 1, 2, 3, 5, 7), temporal (areas 21, 22), and posterior cingulate (area 23) cortices, DA afferents are less dense and show a bi-laminar pattern of distribution in the depth of layers I and V–VI. The lowest density is in area 17, where the DA afferents are mostly restricted to layer I.

The role of the DA broadcasting signals is to selectively tune by increasing the signal-to-noise ratio of the goals, spatial, object, and motor program salient representations and to ensure their between resonance (see *decision-making* module for details).

Decision-Making Module

The decision to where to gaze next is determined by the coordinated actions of the *focus of attention*, *overseer*, *object and spatial maps*, *motor programs*, and *movement execution* modules in the model (see Fig. 5). More
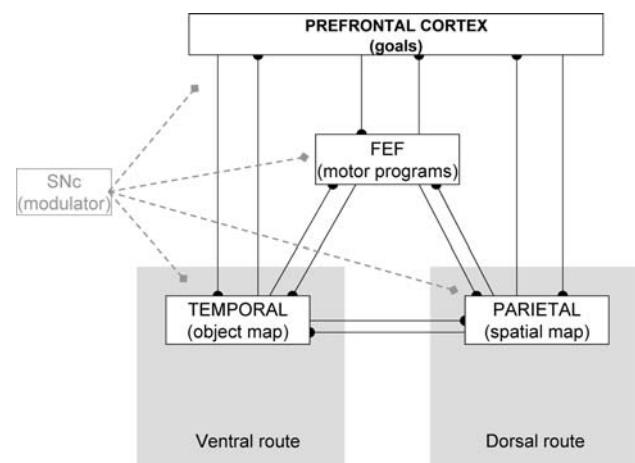

Fig. 5 Schematic of the information flow of the model's decision-making module. Bidirectional adaptive connections among the PFC, the FEF, temporal and parietal cortical areas work like an ART network. Neuromodulatory signals from SNc to PFC, FEF, TC, and PPC act like ART's vigilance parameter, selectively tuning the goals, motor programs, object, and spatial map salient representations

specifically, bottom-up, top-down, and reset mechanisms represented by the complex and intricate feedforward, feedback, and horizontal circuits of PFC, PPC, TC, FEF, motor SC, and the brainstem are making decisions. Adaptive reciprocal connections between (1) PFC and PPC, (2) PFC and TC, (3) PFC and FEF, (4) FEF and PPC, (5) FEF and TC, and (6) PPC and TC operate exactly as the comparison and recognition fields of an ART (Adaptive Resonance Theory) system [2].

In its most basic form, an ART system consists of two interconnected fields of neurons: the comparison field and the recognition field. The comparison field responds to input features, whereas the recognition field responds to categories of the comparison field activity patterns. Bidirectional connections between the two fields are adaptive (modifiable). Neurons in the recognition field compete with each other in a recurrent on-center off-surround fashion. Inhibition from the recognition field to the comparison field shuts off most of the comparison field activity, if the input mismatches the active category's response. If the match is close, enough of the comparison field nodes excited by both the input and the active category node overcome the non-specific inhibition of the recognition field. If on the other hand mismatch occurs, the recognition field inhibition shuts off the active category node as long as the current input is present. Matching occurs when sufficient correspondence between comparison and recognition field patterns is greater than a parameter value called vigilance.

In the model, the ART's vigilance parameter is represented by the broadcasted DA reinforcement teaching signals. High and intermediate levels of DA ensure the formation of fine and coarse categories, respectively,

whereas low values of DA (low signal-to-ratio signals) ensure that non-relevant representations and plans perish.

The reciprocal connections between (1) PFC, PPC, and TC and (2) PFC and FEF allow for the amplification of the spatial, object, and motor representations pertinent to the given context and the suppression of the irrelevant ones, whereas the reciprocal connections among the FEF, PPC, and TC ensure for their groupings.

Decisions in the model are made from the interplay of a winner-take-all mechanism in the spatial, object, and motor salient maps between the selectively tuned by DA and resonated spatial, object, and motor representations [7–9, 13, 14] and a reset mechanism due to a feedback signal from the SC to FEF [43], PFC, PPC, TC, and SNc [38] analogous to the IOR in Itti and Koch [27], which suppresses the last attended location and executed motor plan from their saliency maps and allows for the next salient motor plan to be executed.

Motor Programs Module

In this module, the global spatial and object saliency maps formed in the PPC and TC, respectively, are transformed in their corresponding global saliency motor programs maps. The motor saliency program module is assumed to reside in the FEF of the frontal lobes [46]. Reciprocal connections among PPC, TC, and FEF ensure the sensorimotor groupings of the spatial and object representations with their corresponding motor programs.

Movement Execution Module

The motor program that has won the winner-take-all competition in the FEF field propagates to the intermediate and deep layers of SC and the brainstem (*movement execution module*), where the final motor command is formed. This final motor command instructs the eyes about the direction, amplitude, and velocity of movement. Once, the motor program arrives in the SC, inhibitory feedback signals propagate from the SC to PFC, FEF, PPC, and TC in order to reset these fields and set the stage for the salient point to gaze to. The speed of processing from the input image presentation till the generation of an eye movement is ∼220–250 ms [11].

Bringing Everything Together

Once an input image is presented *three parallel and equally fast processing modes of actions* are initiated (see Fig. 6). In the first mode of action (*visual processing*; see Fig. 6a), pre-attentive multi-scale feature detection and extraction mechanisms sensitive to different features (e.g.,

color, intensity, orientation, etc.) operating in parallel at the level of the retina, LGN, and V1 start to work. From the level of V1 and on the features are separated into two streams: the dorsal for space processing and the ventral for object processing. At the end levels of the visual hierarchy, the PPC and TC lie, where global saliency maps for space and object are formed. In the second mode of action (*neuromodulation*; see Fig. 6b), the retinal signal activates the phasic reinforcement teaching (dopamine) signals via the visual layers of the SC [17]. In turn, the phasic DA teaching signals will be broadcasted to the whole cortex (PFC, FEF, PPC, and TC) and will selectively tune the responses of different neuronal populations in these areas according to previous similar acquired experiences. In the third mode of action (*focus of attention*; see Fig. 6c), the retinal signal will travel a long distance to PFC, where will activate the recognition neuronal populations. The recognition neuronal populations will send/receive top-down/ bottom-up feedback/feedforward signals to/from the spatial, object, and motor saliency maps of the PPC, TC, and FEF. All three modes of action take the same amount of time (∼130 ms) [21, 38, 47].

In the next step, the spatial and object salient maps will go through a sensory-motor transformation to generate their corresponding motor salient maps at the FEF level. Reciprocal connections among PPC, TC, and FEF will bind the perceptual and motor salient maps together. While this transformation and grouping is taking place, attentional and reinforcing teaching signals from the PFC and SNc, respectively, will amplify/selectively tune the neuronal responses at the PFC, PPC, TC, and FEF levels. A winner-take-all mechanism in these fields will select the most salient *and* resonated spatial, object, and motor program representations. The selected motor program will then be forwarded to the motor execution areas (SC and brainstem) where the final motor command will be formed and the eye movement will be generated. The speed of processing from the start of the attentive resonance, selective tuning and motor program formation, selection, and execution takes another ∼100–120 ms (a total of ∼220–250 ms from input image presentation to eye movement execution) [11].

Recently, Redgrave and Gurney [37] reported that the duration of the phasic DA signal (reinforcement teaching signal in this model) is ∼100 ms and it precedes the first eye movement response (see Fig. 4). This finding validates the model's assumption about a co-active reinforcing teaching signal with the resonated attention and motor plan selection. All these mechanisms are reset by a feedback excitatory signal from the SC (*movement execution module*) to the inhibitory neurons of FEF, PFC, PPC, TC, and SNc (*all other model modules*), which in turn inhibit and hence prevent the previously selected targets, objects, and plans from being selected again (see Fig. 6d).
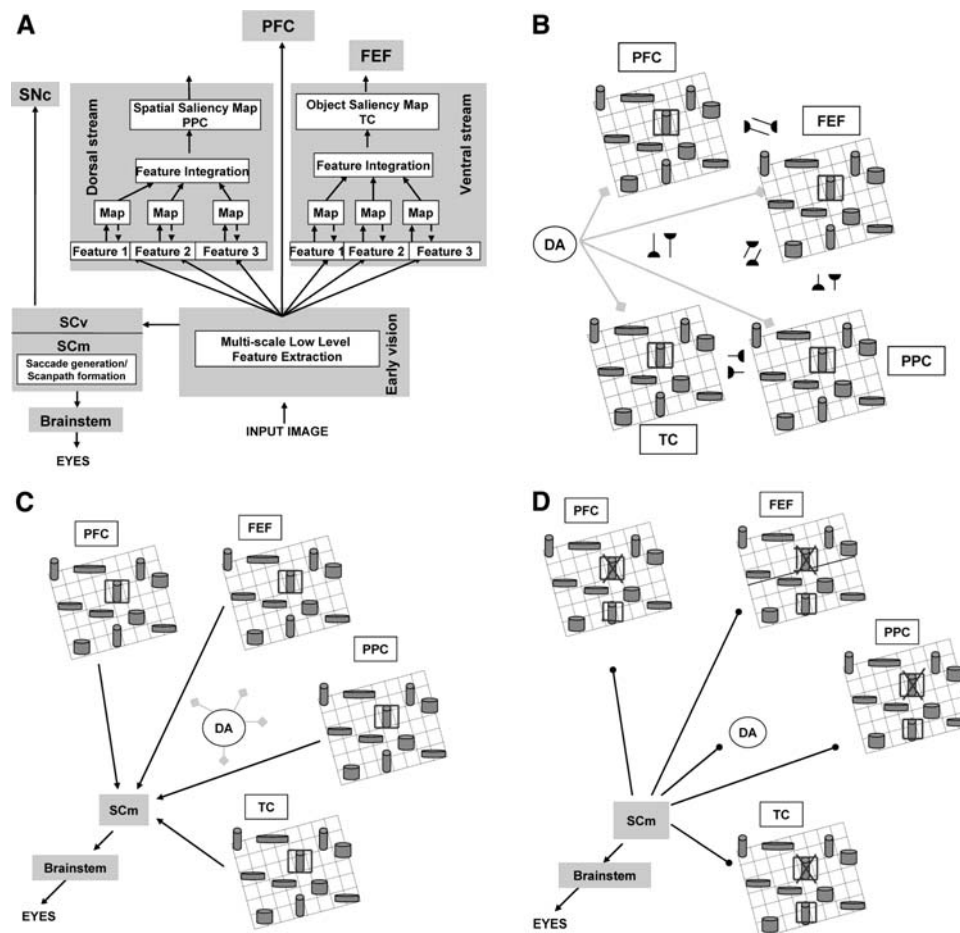
Fig. 6 Information processing stages of the natural scene viewing cognitive model. **a** Visual processing stage. Once an input image is presented three parallel and equally fast processing pathways get activated: (1) Visual hierarchy pathways till the level of PPC (space) and TC (object), (2) sensory activated by the visual SC (SCv) SNc (dopamine) system, and (3) direct visual input to PFC. **b** DA broadcasting teaching signals to PFC, PPC, TC, and FEF. Different neuronal populations receive different levels of DA. High and intermediate DA values result in "sharp tuned" neuronal responses, whereas low DA values result in "broadly tuned" neuronal responses. Neuronal responses are depicted by gray-colored towers in each brain area. The height of each tower represents the neuronal amplitude activation, whereas the width of each tower represents the degree of tuning. **c** Feedforward activation of the motor SC (SCm) by FEF,

PFC, PPC, and TC. *Dark gray square* surrounding the response of a neuronal population represents the winner salient and resonated according to some value of vigilance (DA signal) representation in each brain area. **d** Reset mechanism by feedback inhibitory projections from the SCm to SNc, FEF, PFC, PPC, and TC. Reset mechanism prevents previously selected representation (*dark gray crossed square*) and allows all other resonated neuronal population responses to compete each other for selection. Bottom tower surrounded by *dark gray square* represents the winner salient and resonated representation. *PFC* prefrontal cortex, *PPC* posterior parietal cortex, *TC* temporal cortex, *FEF* frontal eye fields, *DA* dopamine, *SC* superior colliculus, *SCv* visual superior colliculus, *SCm* motor superior colliculus, *SNc* substantia nigra pars compacta

Discussion

What Have We Learned from the Model?

The model presented herein is a cognitive model of picture scanning based on the interaction of bottom-up stimulus-driven saliency maps of object identity and location and a top-down focus-of-attention signal, which drives attention to specific ROIs in the picture. Picture scanning in the model was a set of mechanisms that helped optimize the search processes inherent in perception, cognition, and

action. Four main classes of mechanisms have been detailed: *saliency*, *focus of attention*, *resonance*, and *reset*. Each mechanism included a number of more specific mechanisms.

The *saliency* mechanism operated the same way as in the model of Itti and Koch [27]. Neural substrates of saliency maps have been found throughout the dorsal and ventral visual streams, the PPC, FEF, and PFC [3, 31, 46].

The *focus-of-attention* mechanism included the more specific mechanisms of *amplification* of relevant information and the *suppression* of irrelevant ones throughout the

visual and the motor programs fields. Experimental [3, 36, 40] and computational [16, 22, 45, 48] studies have confirmed the presence of such a signal in the brain.

The *resonance* mechanism worked as the matching process among the salient representations of the object, spatial, and motor programs maps based on the *focus-of-attention* mechanism generated by the goals module and the *DA modulation* mechanism of the overseer module, which worked like a vigilant parameter of an ART network [2]. The representations that reached resonance first were the ones that were gazed first, followed by the second fasters, and so on.

Finally, the *reset* mechanism was initiated immediately after the final motor command was sent to the eyes and it worked as a global inhibitory signal that wiped out all relevant to motor response cortical representations and ensured that these representations were not selected for another 500–900 ms. That is, the reset mechanism worked as the neural substrate of the *inhibition-of-return* mechanism observed experimentally in Klein [29].

Future Work

Work is currently underway to test the active visual search performance of the current model with simple and complex natural images and movies. A particularly interesting extension of the model is how it may resolve conflicts and generate a gaze when two different sets of representations reach resonance at the same time. Cutsuridis et al. [6, 10, 11] have shown that such conflict resolution can occur at the motor execution level (motor SC) through a simple competition between decision signals. Recent experimental evidence has shown that conflict resolution may also be resolved more centrally in anterior cingulate and prefrontal cortices [18]. Finally, another interesting extension of the model is how previous experiences and strategies may bias the selection process of the next gaze [32, 44].

References

1. Berman RA, Wurtz RH. Exploring the pulvinar path to visual cortex. Prog Brain Res. 2008;171:467–73.
2. Carpenter GA, Grossberg S. Adaptive resonance theory. In: Arbib MA, editor. The handbook of brain theory and neural networks. 2nd ed. Cambridge: MIT Press; 2003. p. 87–90.
3. Chelazzi L, Duncan J, Miller EK, Desimone R. Responses of neurons in the inferior temporal cortex during memory guided visual search. J Neurophysiol. 1998;80(6):2918–40.
4. Coizet V, Comoli E, Westby GW, Redgrave P. Phasic activation of substantia nigra and the ventral tegmental area by chemical stimulation of the superior colliculus: an electrophysiological investigation in the rat. Eur J Neurosci. 2003;17(1):28–40.
5. Comoli E, Coizet V, Boyes J, Bolam JP, Canteras NS, Quirk RH, et al. A direct projection from the superior colliculus to substantia nigra for detecting salient visual events. Nat Neurosci. 2003;6(9):974–80.
6. Cutsuridis V, Kahramanoglou I, Perantonis S, Evdokimidis I, Smyrnis N. A biophysical model of decision making in an antisaccade task through variable climbing activity. In: Duch W, et al., editors. ICANN 2005. LNCS, vol. 3695. Berlin: Springer; 2005. p. 205–10.
7. Cutsuridis V, Perantonis S. A neural network model of Parkinson's disease bradykinesia. Neural Netw. 2006;19(4):354–74.
8. Cutsuridis V. Neural model of dopaminergic control of arm movements in Parkinson's disease bradykinesia. In: Kollias SD, Stafylopatis A, Duch W, Oja E, editors. ICANN 2006. LNCS, vol. 4131. Heidelberg: Springer; 2006. p. 583–91.
9. Cutsuridis V. Does reduced spinal reciprocal inhibition lead to co-contraction of antagonist motor units? A modeling study. Int J Neural Syst. 2007;17(4):319–27.
10. Cutsuridis V, Kahramanoglou I, Smyrnis N, Evdokimidis I, Perantonis S. A neural variable integrator model of decision making in an antisaccade task. Neurocomputing. 2007;70(7–9):1390–402.
11. Cutsuridis V, Smyrnis N, Evdokimidis I, Perantonis S. A neural network model of decision making in an antisaccade task by the superior colliculus. Neural Netw. 2007;20(6):690–704.
12. Cutsuridis V. A bio-inspired system architecture of an active visual search model. In: Kurkova V, Neruda R, Koutnik J, editors. ICANN 2008, LNCS vol. 5164. Berlin: Springer; 2008. p. 248–57.
13. Cutsuridis V. Neural network modeling of voluntary single joint movement organization. I. Normal conditions. In: Chaovalitwongse WA, Pardalos P, Xanthopoulos P, editors. Computational neuroscience. Berlin: Springer-Verlag; 2010.
14. Cutsuridis V. Neural network modeling of voluntary single joint movement organization. II. Parkinson's disease. In: Chaovalitwongse WA, Pardalos P, Xanthopoulos P, editors. Computational neuroscience. Berlin: Springer-Verlag; 2010.
15. Deco G, Schürmann B. A neuro-cognitive visual system for object recognition based on testing of interactive attentional top-down hypotheses. Perception. 2000;29(10):1249–64.
16. Desimone R, Duncan J. Neural mechanisms of selective visual attention. Ann Rev Neurosci. 1995;18:193–222.
17. Dommett E, Coizet V, Blaha CD, Martindale J, Lefebre V, Walton N, et al. How visual stimuli activate dopaminergic neurons at short latency. Science. 2005;307(5714):1476–9.
18. Egner T, Hirsch J. Cognitive control mechanisms resolve conflict through cortical amplification of task relevant information. Nat Neurosci. 2005;8(12):1784–90.
19. Fazl A, Grossberg S, Mingolla E. View-invariant object category learning, recognition, and search: how spatial and object attention are coordinated using surface-based attentional shrouds. Cogn Psychol. 2009;58(1):1–48.
20. Findlay JM, Gilchrist ID. Active vision: the psychology of looking and seeing. Oxford: Oxford University Press; 2003.
21. Foxe JJ, Simpson GV. Flow of activation from V1 to frontal cortex in humans. Exp Brain Res. 2002;142:139–50.
22. Hamker FH. The re-entry hypothesis: the putative interaction of the frontal eye field, ventrolateral prefrontal cortex and areas V4, IT of attention and eye movement. Cereb Cortex. 2005;15:431–47.
23. Hanes DP, Wurtz RH. Interaction of frontal eye field and superior colliculus for saccade generation. J Neurophys. 2001;85:804–15.
24. Henderson JM, Hollingworth A. The role of fixation position in detecting scene changes across saccades. Psychol Sci. 1999;50:243–71.

25. Hikosaka O, Wurtz RH. Visual and oculomotor functions of monkey substantia nigra pars reticulate. J Neurophys. 1983;49:1230–301.
26. Hikosaka O, Takikawa Y, Kawagoe R. Role of the basal ganglia in the control of purposive saccadic eye movements. Physiol Rev. 2000;80:954–78.
27. Itti L, Koch C. Computational modelling of visual attention. Nat Neurosci. 2001;2:194–203.
28. Itti L, Koch C. A saliency based search mechanism for overt and covert shifts of visual attention. Vision Res. 2000;40:1489–506.
29. Klein RM. Inhibition of return. Trends Cogn Sci. 2000;4(4): 138–47.
30. Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol. 1995;4:219–27.
31. Kusunoki M, Gottlieb J, Goldberg ME. The lateral intraparietal area as a salience map: the representation of abrupt onset, stimulus motion and task relevance. Vision Res. 2000;40:1459–68.
32. Lleras A, Von Mühlenen A. Spatial context and top-down strategies in visual search. Spat Vis. 2004;17(4–5):465–82.
33. McHaffie JG, Jiang H, May PJ, Coizet V, Overton PG, Stein BE, et al. A direct projection from superior colliculus to substantia nigra pars compacta in the cat. Neuroscience. 2006;138(1):221–34.
34. Mohler CW, Wurtz RH. Role of striate cortex and superior colliculus in visual guidance of saccadic eye movements in monkeys. J Neurophyiol. 1977;40:74–94.
35. Olshausen BA, Anderson CH, van Essen DC. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. J Neurosci. 1993;13(11): 4700–19.
36. Reynolds JH, Desimone R. The role of neural mechanisms of attention in solving the binding problem. Neuron. 1999;24(1): 19–29.
37. Redgrave P, Gurney K. The short latency dopamine signal: a role in discovering novel actions. Nat Neurosci. 2006;7:967–75.
38. Redgrave P, Gurney K, Reinolds J. What is reinforced by the phasic dopamine signals? Brain Res Rev. 2008;58(2):322–39.
39. Rybak IA, Gusakova VI, Golovan AV, Podladchikova LN, Shevtsova NA. A model of attention-guided visual perception and recognition. Vision Res. 1998;38(15–16):2387–400.
40. Schall JD, Hanes DP, Thompson KG, King DJ. Saccade target selection in frontal eye field of macaque. I. Visual and premovement activation. J Neurosci. 1995;15:6905–18.
41. Schiller PH, True SD, Conway JL. Deficits in eye movements following frontal eye field and superior colliculus ablations. J Neurophys. 1980;44:1175–89.
42. Schultz W. Predictive reward signal of dopamine neurons. J Neurophys. 1998;80:1–27.
43. Sommer MA, Wurtz RH. Frontal eye field neurons orthodromically activated from the superior colliculus. J Neurophys. 1998;80:3331–3.
44. Tavassoli A, Linde I, Bovik AC, Cormack LK. Eye movements selective for spatial frequency and orientation during active visual search. Vision Res. 2009;49(2):173–81.
45. Taylor JG, Hartley M, Taylor N, Panchev C, Kasderidis S. A hierarchical attention-based neural network architecture, based on human brain guidance, for perception, conceptualisation, action and reasoning. Image Vis Comput. 2009;27:1641–57.
46. Thompson KG, Bichot NP. A visual saliency map in the primate frontal eye field. Prog Brain Res. 2005;147:251–62.
47. Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature. 1996;381(6582):520–2.
48. Tsotsos JK, Culhane S, Wai W, Lai Y, Davis N, Nuflo F. Modeling visual attention via selective tuning. Artif Intell. 1995;78(1–2):507–47.
49. Viviani P. Eye movements in visual search. Cognitive, perceptual and motor control aspects. In: Kowler E, editor. Eye movements and their role in visual and cognitive processes. Amsterdam: Elsevier; 1990. p. 353–93.