A Possible Explanation of Oriented Bar Filling-in at the Blind-Spot in the light of Hierarchical Prediction Mechanism

 R Raman and S Sarkar[#]

Applied Nuclear Physics Division,
Saha Institute of Nuclear Physics, Kolkata, India
[#] *sandip.sarkar@saha.ac.in*

Abstract. Despite the absence of retinal input, the blind-spot gets filled up with surrounding visual attributes. This phenomenon is called the perceptual filling-in or completion. Various possible neural mechanisms have been proposed but most of these are not entirely consistent with the findings of brain activation and brain organization. In one study, recently, it has been shown that the filling-in of the shifting bar and the anisotropy in the filling-in of misaligned bars could be explained by incorporating a common general principle called hierarchical predictive coding (HPC). In this report, we have extended this proposition to explain the filling-in of oriented bars. We have considered a three level (LGN-V1-V2) HPC model network in which, the blind-spot was emulated by removing the corresponding feed-forward (LGN-V1) connection. We simulated the responses of predictive estimator (PE) neurons at blind-spot while stimulating the network with oriented bar stimuli. Results show that the filling-in is best for aligned bars but it faded away with increasing orientation and moreover, corners were not predicted in any orientation. Qualitatively, these results are consistent with the findings of psychophysical experiments. We discussed this phenomenon in the HPC framework and argue that in the absence feed-forward connections, the best prediction is dominated by learned statistical regularities that include an abundance of bar and similar structures. These results suggest that the filling-in process could be a manifestation of HPC.

1. Introduction

Because of the absence of photoreceptors at optics disc, the retina is unable to send the corresponding signal to the brain and thereby and therefore, some portion of the visual field is hidden. However, we never notice any odd patch due the hidden visual field, known as blind-spot, even in monocular vision, but rather we see the complete scene; filled up in accordance with the surrounding visual attributes [1]. This process is known as perceptual filling-in or simply filling-in. In addition to the blind spot, filling-in also occurs in other visual input deficit conditions, e.g. filling-in at the artificial and natural retinal scotoma [2, 3]. Filling-in also observed in visual illusions such as Neon colour spreading, Craik-O'Brien-Cornsweet illusion, Kanizsa shapes, and etc. and steady fixation condition like Troxler effect (for review see [4].

   Psychophysical and physiological studies suggest that the filling-in is an active process, where some neural processes are involved and mainly take place in the early visual cortex [5-7]. Studies show that neural activities, correlated to perceptual experiences, are evoked in the deep layer of primary visual cortex, when filling-in completion occurs [5, 6]. Matsumoto and Komatsu [7] demonstrated that some neuron in BS region in deep layer of primary visual cortex (BS neurons), which possess larger receptive fields that extend beyond the blind spot, exhibits non-linear elevated response when a long moving bar cross over the blind spot and perceptual completion occurs.

   Although some attempts have been made to understand the computational mechanism of general filling-in [8-13], little work has been devoted to the study of computational mechanism of filling-in

completion at the blind spot and the current knowledge is far from complete. Recently, it has been shown that the filling-in of the shifting bar and the anisotropy in the filling-in of misaligned bars could be predicted by incorporating HPC [14], which has, recently, been argued as a general coding principle of the visual cortex [15-22]. Here in this study, we have expanded the scope of HPC by demonstrating that the some results related to oriented bar filling-in could also be explained using this same general principle.

Functionality of a network in the HPC framework largely depends on a probabilistic hierarchical generative model and the efficient coding of natural images. In this framework, the visual system attempts to infer or estimate the properties of the world from signals coming from receptors [23-25]. This job is hypothesized to be completed by concurrent prediction-correction mechanism along the hierarchy of the visual system. Each higher visual area (say V2) attempt to predict the response at its lower area (say V1) on the basis of the learned statistical regularities, and send that prediction signal to the lower area by feedback connection. Lower area, in response, sends a residual error signal, signifying any mismatch, to the higher area by feed-forward connection. Neural networks, in this paradigm, are constantly attempting to match sensory input with the top-down predictions that are based on the learned statistical regularities of natural scenes. By using its best current predictions as the priors for the levels below and engaging in iterative estimation, the network allows priors and predictions to co-evolve to arrive at the best prediction.

This idea is based on the anatomical architecture of the visual system which is hierarchically organized and reciprocally connected [27]. Recently, several neuronal tuning properties in different visual areas such as the lateral geniculate nucleus (LGN), primary visual cortex (V1) and middle temporal level (MT) have been explained using this framework [19, 20]. For example, Rao [14] suggested that the extra-classical properties of neurons in V1 could be understood in terms of predictive-feedback signal from the secondary visual cortex (V2). We speculated that a similar computational mechanism could also explain the filling-in completion at the blind spot.

In this report, we have investigated the filling-in of oriented bars under a fresh perspective of HPC. We have considered a three level (LGN-V1-V2) HPC model network in which, the blind-spot was emulated by removing the corresponding feed-forward (LGN-V1) connection in early visual area. We simulated the responses of predictive estimator (PE) neurons at blind-spot while stimulating the network with oriented bar stimuli across the blind-spot. Results show that the filling-in is best for aligned bars but it faded away with increasing orientation. Additionally, we also observed that the corners were not predicted in any orientation. These results are consistent with the finding of psychophysical experiment [28].

We discussed this phenomenon in the HPC framework and argue that in the absence feed-forward connections, the best prediction is dominated by learned statistical regularities that include an abundance of bar and similar structures and anisotropy in orientation distribution. These results suggest that the filling-in process could be a manifestation of HPC.

2. Methods

2.1 Hierarchical Predictive coding of natural images

A general computational HPC architecture is illustrated in figure 1. In this framework, on the arrival of an input, predictor estimator modules (PE module) at each visual processing level generate the prediction (or estimate) on the basis of the learned statistical regularities. Each higher area (say V2) then sends these predictions to its immediate lower level (say V1) by feedback connections and in return receives the error signal, by feed-forward connections. An equilibrium state is achieved after several concurrent prediction-
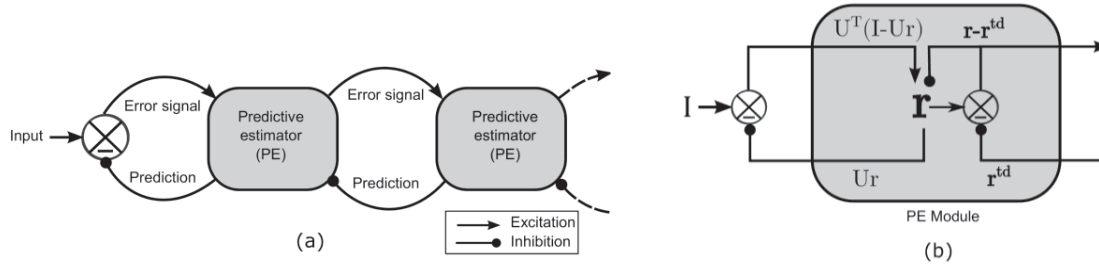
Figure 1. General HPC architecture (adopted from Rao.[15]). a) Illustration of connections between PE modules at different levels, b) Details of computation performed by each PE module.

correction cycles, where the estimate approximately matches the input signal caused by the stimuli. This optimum-estimate is regarded as a representation of the input at that level.

A single PE module (figure 2(b)) consists of, (i) Predictive estimator neurons (PE neurons) which represent the estimate of current input signal $\mathbf{I}$ with response vector $\mathbf{r}$ (state vector), (ii) neurons that carry prediction signal $U\mathbf{r}$ (for the input $\mathbf{I}$) to lower level by feed-back connections, whose synapse encode encoding efficacy matrix U, (iii) neurons that carry feed-forward error signal ($\mathbf{I}-U\mathbf{r}$) from lower level to higher level, whose synapses encoded rows of efficacy matrix U^T, and (iv) error detecting neurons which carry the residual error signal ($\mathbf{r}-\mathbf{r}^{td}$) to the higher level corresponding to the prediction $\mathbf{r}^{td}$ from the higher level.

2.2 Network dynamics and learning rule

In the Bayesian framework (maximum a posteriori (MAP) approach), the dynamics and the learning can be derived by maximizing the posterior probability $P(\mathbf{r},\mathbf{r}^{td},U\,|\,\mathbf{I})$, which is equal to the product of $P(\mathbf{I}\,|\,\mathbf{r},U)$, $P(\mathbf{r}^{td}\,|\,\mathbf{r})$ and prior probabilities $P(\mathbf{r})$ and $P(U)$, with respect to $\mathbf{r}$ and U respectively. By assuming $P(\mathbf{I}\,|\,\mathbf{r},U)$ and $P(\mathbf{r}^{td}\,|\,\mathbf{r})$ as Gaussians of zero mean and variances σ^2 and σ_{td}^2 respectively, the total coding length E can be written as (for details, see [15], [21]),

$$E = \frac{1}{\sigma^2}(\mathbf{I}-U\mathbf{r})^T(\mathbf{I}-U\mathbf{r}) + \frac{1}{\sigma_{td}^2}(\mathbf{r}-\mathbf{r}^{td})^T(\mathbf{r}-\mathbf{r}^{td}) + g(\mathbf{r}) + h(U) \qquad (1)$$

here, $g(\mathbf{r})$ and $h(U)$ are the negative log of prior probabilities $P(\mathbf{r})$ and $P(U)$ respectively. Minimizing the coding length E, with respect to $\mathbf{r}$ provides the dynamics of PE module and it is equivalent to recurrently achieving an internal state matching the "sensory driven" input. The visual representation of the predication $U\mathbf{r}$, is regarded as the representation of our perception. Furthermore, the minimization of coding length E, with respect to U provides the learning rule for basis matrix U. A kurtotic prior distribution ($P(r_i)=exp(-\alpha log(1+r_i^2))$) is chosen for the prior probability for embedding sparse representation [25].

3.   Simulation

In this study, we have utilised a three level linear HPC network (similar to [14]), where level 1 (equivalent to V1) consists of 9 PE modules. These modules receive input from level 0 (LGN) and send the output to

the sole module at level 2 (equivalent to V2). Therefore, the PE module at level 2 receives input from all the nine level 1 PE modules and sends back the feedback signal to all of them. This architecture is based on the fact that the visual area higher in hierarchy operates on a higher spatial scale. Each PE module at level 1 consists of 64 PE neurons, 144 Prediction carrying neurons, 64 afferent error carrying neurons and 64 error detecting neurons for conveying the residual error to level 2. The layer 2 module consists of 169 PE neurons, 576 prediction carrying neurons and 169 error carrying neurons. Six natural images of size 512×512 pixels were used for training after pre-processing with a low-pass filter of cut-off frequency 200 cycles/image. Variance normalized 1000 batches of 100 image patches of size 30×30 pixel, which were extracted from randomly selected locations from the randomly selected pre-processed images, were given as input to the network. For each batch of image patches, the network was allowed to achieve steady states and the average of these states was used to update the efficacy of neurons initially assigned random value (for details see [25, 26]).

To mimic the blind spot the feed-forward connection in a certain area was removed from the model network, which was pre-trained with usual feed-forward connections. The removal was implemented by setting the efficacy of early feed-forward (level 0 - level 1) neurons to zero. This "pre-training" preserved the experimentally observed binocular nature of these active neurons in the deep layer (5/6) (for the detail see [14]).

4. Results

Initially, the model network was trained with binocular connections. This resulted in the Gabor like receptive field (RF) at level 1 whereas, level 2 learned more abstract RF (e.g. corner, long bar etc.) similar to those reported in earlier studies [14,25,26]. After the completion of the learning process, the network
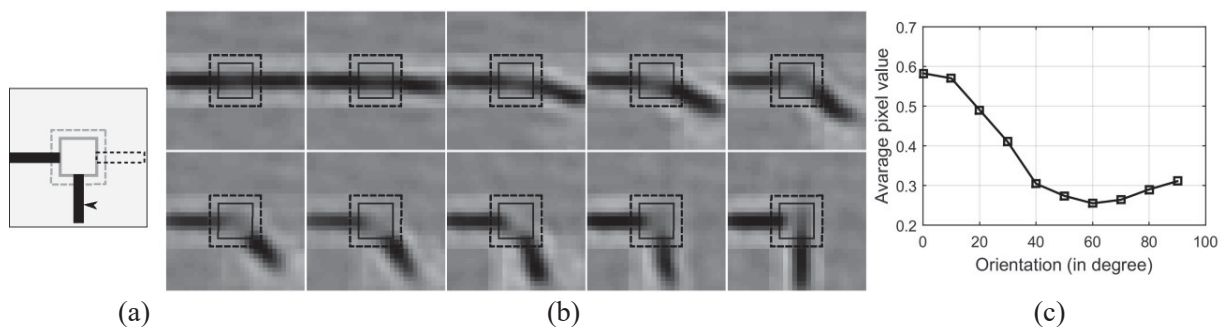

(a)   (b)   (c)

Figure 2. Results of oriented bar filling-in. a) Stimulus used for the study is shown. One part of the stimulus, the horizontal black bar, was fixed. The other part was rotated clockwise from its initial horizontal position (dotted bar) to the vertical position in steps of 10 degree. The central rectangle is BS and the outer rectangle (dotted) is the receptive field. Rectangular regions are shown for illustration only, b) Perceptual images, generated from the response of PE neurons at level 1, corresponding to all ten orientations of stimulus are shown, c) Average response at the centre (2x2) of the BS region in (b) is plotted as a function of orientation angle.

was modified to accommodate the blind spot before stimulated with the bar stimuli shown in figure 2(a). One part of the bar was fixed at a position outside of blind spot. Orientation of the other part was varied from horizontal to vertical position producing ten different input stimuli. We recorded the response PE neurons at level 1 to these stimuli, and generated the corresponding 'perceptual images', which we defined as reconstruction $U\mathbf{r}$. These 'perceptual images' (figure 2(b)) demonstrate that the filling-in is best when

the bars were aligned (horizontal) and with the increasing mismatch in orientation, the filling-in deteriorates. This is illustrated in more detail in figure 2(c), where the average pixel value at the centre of the BS region is plotted as a function of the orientation difference. This plot demonstrates that filling-in deteriorates at a faster rate till 40 degree and for further increase in angle, it still deteriorates but with a slower rate till 60 degree difference in orientation, Thereafter, a little increase is also visible in the plot. The observed decrease in filling-in corroborates the experimental findings reported in [28]. Additionally, in none of these results (figure 2(b)), corners are visible, which leads to the conclusion that corners were never anticipated in the top-down expectation. Regarding the increase in filling-in beyond 60 degree, it is not possible to make any concluding remark and this discussed in detail in the discussion section.

5. Discussion

In this study, we have investigated the computational aspects of oriented bar filling-in at the blind spot. We postulated that this could be understood in the HPC paradigm inside the visual cortex and therefore, conducted simulation studies on the three level HPC model network modified in accordance with the blind-spot architecture. Our studies show that the filling-in completion, which occurs in the case of perfectly aligned bar segments, also happens with a small degree of misalignment, but the extent of filling-in completion deteriorates with increasing misalignment in orientation (figure 2).
These results show good agreement with the physiological and psychophysical results reported earlier [28] and can be discussed in the context of HPC. When an input stimulus was presented in and around the blind spot, higher areas (V2) generated a unified estimate for the input stimuli on the basis of the learned statistical regularities of natural scenes. In nature, these statistical regularities are dominated by the abundance of straight continuous bars and similar structures and therefore, these structures are much more probable, compared to other non-linear continuous structures, in the predictions made by the higher visual area (V2). These predictions remain uncorrected due to the absence of error carrying feed-forward connection in BS region at V1 and therefore, local optimum-estimate is achieved essentially by the top-down expectation.

In case of aligned bars on both sides of the BS region (the first case in figure 2(b)), the visual system received inputs corresponding to the presence bars on both sides. This is interpreted as the presence of a long straight continuous bar because in natural image statistics the presence of a broken bar or two bars side by side is less probable and this is reflected in the response of PE neurons announcing the prediction of a long bar across the BS resulting in the perception of a near perfect filling-in. In other cases, the model system received inputs corresponding to the presence of bars with different orientations. As long as the difference in orientation was small, the model system was able to interpret it as a continuous bar and filling-in occurred. With increasing angles, however, the higher area (V2) faces increased difficulties in interpreting two oriented bars as parts of a long continuous (may be non-linear) bar resulting in the deterioration of filling-in. Beyond a certain threshold, the system fails to predict anything more than what is visible and the filling-in ceases to happen.

To illustrate the varying filling-in with increasing angles, we have plotted average pixel value recorded at the centre of the BS region as a function of angle in figure 2(c). It shows that the value decreases till 60 degree and then increases a little. It may look like that filling-in never fails but varies with increasing angles. The problem lies in the interpretation of a threshold. In similar psychophysical experiments, participants report the occurrence or non-occurrence of filling-in depending on their perceptual experience. However, in our simulation study, there exists no such threshold and in fact, it is very difficult to define an equivalent one without further extension of the current model incorporating much more

details properties of the visual cortex. This is precisely why the filling-in never fails (figure 2(c)), though deteriorates, with increasing difference in orientation

In conclusion, we demonstrate that some aspects of oriented bar filling-in could be explained in the computational paradigm of HPC. We show that, in the absence of feed-forward connection due to the blind-spot, the top-down predictions (in accordance with learned statistical regularities of natural scene) mainly determines the nature of filling-in that corroborates well with experimental findings. These results suggest that the filling-in could be a manifestation of a hierarchical predictive coding principle.

References

[1] Ramachandran VS 1992 *Scientific American*. **266(5)** 86–91.
[2] Ramachandran VS, Gregory RL, et al 1991 *Nature*. **350(6320)** 699–702.
[3] Gerrits H, Timmerman G 1969 *Vision Research*. **9(3)** 439–442.
[4] Komatsu H 2006 *Nature reviews neuroscience*. **7(3)** 220–231.
[5] Júnior MF, Rosa M, Gattass R, Rocha-Miranda CE 1992 *PNAS*. **89(18)** 8547–8551.
[6] Komatsu H, Kinoshita M, Murakami I 2000 *The Journal of Neuroscience*. **20(24)** 9310–9319.
[7] Matsumoto M, Komatsu H 2005 *Journal of Neurophysiology*. **93(5)** 2374–2387.
[8] Neumann H, Pessoa L, Hansen T 2001 *Biological Cybernetics*. **85(5)** 355–369.
[9] Neumann H, Sepp W 1999 *Biological Cybernetics*. **81(5-6)** 425–444.
[10] Grossberg S, Mingolla E 1985 *Psychological review*. **92(2)** 173.
[11] Grossberg S, Kelly F 1999 *Vision Research*. **39(22)** 3796–3816.
[12] Grossberg S, Mingolla E, Ross WD 1997 *Trends in neurosciences*. **20(3)** 106–111.
[13] Satoh S, Usui S 2008 *Neural Networks*. **21(9)** 1261–1271.
[14] Raman R, Sarkar S 2015 *arXiv: 1507.01390* (q-bio.NC).
[15] Rao RP, Ballard DH 1999 *Nature neuroscience*. **2(1)** 79–87.
[16] Lee TS, Mumford D 2003 *JOSA A*. **20(7)** 1434–1448.
[17] Friston K 2005 *Phil. Trans. Roy. Soc. London B*. **360(1456)** 815–836.
[18] Hohwy J, Roepstorff A, Friston K 2008 *Cognition*. **108(3)** 687–701.
[19] Jehee JF, Ballard DH 2009 *PLoS computational biology*. **5(5)** e1000373.
[20] Jehee JF, Rothkopf C, Beck JM, Ballard DH 2006 *Journal of Physiology-Paris*. **100(1)** 125–132.
[21] Huang Y, Rao RP 2011 *Wiley Interdisciplinary Reviews: Cognitive Science*. **2(5)** 580–593.
[22] Clark A 2013 *Behavioral and Brain Sciences*. **36(03)** 181–204.
[23] Mumford D 1994 *Large-Scale Theories of the Cortex* (MIT Press). 125–152.
[24] Dayan P, Hinton GE, Neal RM, Zemel RS 1995 *Neural computation*. **7(5)** 889–904.
[25] Olshausen BA, et al 1996 *Nature*. **381(6583)** 607–609.
[26] Olshausen, BA, Field, DJ 1997 *Vision Research*. **37(23)** 3311–3325.
[27] Felleman DJ, Van Essen DC 1991 *Cerebral cortex*. **1(1)** 1–47.
[28] Araragi Y, Nakamizo S 2008 *Vision research*. **48(4)** 618–625.