Chromatic Induction and Contrast Masking: similar models, different goals?

Sandra Jiménez[a], Xavier Otazu[b], Valero Laparra[a] and Jesús Malo[a]

[a]Image Processing Lab, Universitat de València, Spain;
[b]Computer Vision Center, Universitat Autònoma de Barcelona, Spain;

ABSTRACT

Normalization of signals coming from linear sensors is an ubiquitous mechanism of neural adaptation.[1] Local interaction between sensors tuned to a particular feature at certain spatial position and neighbor sensors explains a wide range of psychophysical facts including (1) masking of spatial patterns,[2] (2) non-linearities of motion sensors,[3] (3) adaptation of color perception,[4] (4) brightness and chromatic induction,[5,6] and (5) image quality assessment.[7]

Although the above models have formal and qualitative similarities, it does not necessarily mean that the mechanisms involved are pursuing the same statistical goal. For instance, in the case of chromatic mechanisms (disregarding spatial information), different parameters in the normalization give rise to optimal discrimination or adaptation,[8] and different non-linearities may give rise to error minimization or component independence.[9] In the case of spatial sensors (disregarding color information), a number of studies have pointed out the benefits of masking in statistical independence terms.[10–13] However, such statistical analysis has not been performed for spatio-chromatic induction models where chromatic perception depends on spatial configuration.

In this work we investigate whether successful spatio-chromatic induction models,[6] increase component independence similarly as previously reported for masking models.[12] Mutual information analysis suggests that seeking an efficient chromatic representation may explain the prevalence of induction effects in spatially simple images.

Keywords: Chromatic induction, contrast masking, efficient coding, mutual information

1. MODELS: FORMAL SIMILARITY

The particular Induction[5,6] and Masking[7,12] models considered here have the same linear+nonlinear scheme:

$$\mathbf{x} \xrightarrow{\mathbf{T}} \mathbf{w} \xrightarrow{\mathbf{R}} \mathbf{r}, \tag{1}$$

where the input image, $\mathbf{x}$, is first decomposed into achromatic and chromatically opponent channels, and each one is analyzed by a set of wavelet-like linear sensors, $\mathbf{w} = \mathbf{T}\mathbf{x}$, that provide a scale and orientation decomposition of the image.[2,14] Then, the second nonlinear stage of the models, $\mathbf{R}$, includes frequency selectivity and divisive interaction between spatially neighbor responses. Despite this general similarity, it is important to note that the particular parameters (e.g. frequency weights, excitation and inhibition exponents, and interaction neighborhoods) were obtained to fit different psychophysical datasets (e.g. induction and subjective distortion measures).

corresponding author: sandra.jimenez@uv.es
This work has been partially funded by projects CICYT TIN2012-38102-C03-01 and CICYT TIN2010-21771-C02-1

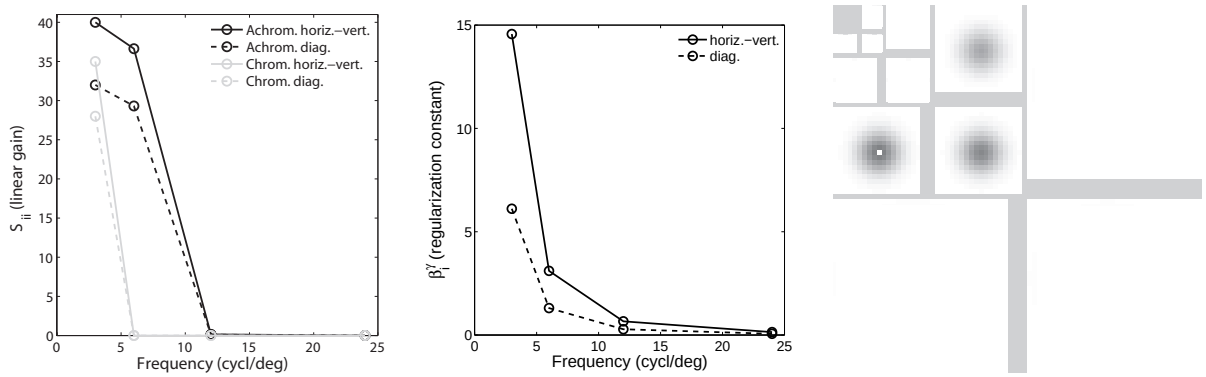

Figure 1. Parameters in the Divisive Normalization Masking model. The panels show S, the linear weights for the different scales, orientations and chromatic channels (left), the saturation constants, β (center), and an example of the interaction neighborhood in H (right). The best fit for the excitation and inhibition exponent was $\gamma = 1.7$.

1.1 Divisive Normalization Masking Model

In this particular implementation of the above ideas,[7,12] the rows of the matrix $\mathbf{T}$ contain a simulation of the linear receptive fields of V1 neurons using an orthogonal 4-scales QMF wavelet transform.[15]

The second step consists of two main operations. First, a frequency dependent linear gain is applied according to the achromatic and chromatic Contrast Sensitivity Functions (CSFs),[16,17] through the application of a diagonal matrix $\mathbf{S}$ on the wavelet coefficients. Then, the weighted response of these sensors is non-linearly transformed according to the Divisive Normalization, $\mathbf{R}^{DN}$,[2,18] in which they are rectified and normalized by a pooling of the responses of the neighboring sensors in space and orientation:

$$r_i^{\mathrm{DN}} = \mathbf{R}(\mathbf{w})_i^{DN} = \left(\frac{|S_i \cdot w_i|^{\gamma-1}}{\beta_i^{\gamma} + \sum_{k=1}^{n} H_{ik} |S_k \cdot w_k|^{\gamma}} \right) \cdot w_i \tag{2}$$

where the subindex i refers to the scale, orientation and spatial position; β_i is a saturation constant; and H_{ik} is a kernel matrix that controls how the responses of neighboring linear sensors, k, affect the non-linear response of sensor i.[2]

The parameters of this model (Fig. 1) were fitted in order to match observers opinion in a natural images quality assessment experiment (see Laparra et al.[7] for details).

1.2 The Chromatic Induction Wavelet Model

This model[5,6] reproduces the brightness and color assimilation and contrast effects by applying a weighting function (referred to as Extended-CSF, α) to each scale, s, and orientation, o, of the achromatic and chromatic wavelet channels:

$$r_{s,o}^{\mathrm{IND}} = \mathbf{R}(\mathbf{w})_{s,o}^{\mathrm{IND}} = \alpha_{s,o}(\rho) \cdot w_{s,o} \tag{3}$$

It is important to note that the Extended-CSF, $\alpha_{s,o}(\rho)$, *is not* a traditional CSF-like linear weighting function, since its effects depend on ρ, which is the ratio of the energy of each wavelet coefficient (center) with regard to the energies of its spatial neighbors (surround), and hence it is image dependent.

From these internal responses, $r_{s,o}^{\mathrm{IND}}$, the authors propose to simulate the perceived image by inverting the wavelet transform:

$$\mathbf{x}_{\mathrm{perceived}}^{\mathrm{IND}} = \mathbf{T}^{-1}\mathbf{r}^{\mathrm{IND}} \tag{4}$$

In the implementation of the model, $\mathbf{T}$, is computed using a redundant biorthogonal wavelet transform.

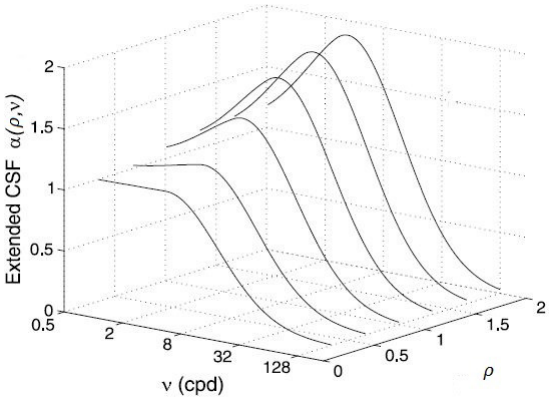

Figure 2. Illustration of the (achromatic) Extended-CSF, $\alpha_\nu(\rho)$, used in the Chromatic Induction model.

It is assumed that $\alpha_{s,o}(\rho)$ (figure 2) depends on the center-surround contrast energy ratio ρ and the spatial frequency ν of the scale s (more details in Otazu et al.[6]), specifically:

$$\alpha(\nu,\rho) = \begin{cases} \frac{\rho^2}{1+\rho^2}\exp\left(-\frac{\left(\log_2\frac{4}{\nu}\right)^2}{2\sigma_1^2}\right) + \exp\left(-\frac{\left(\log_2\frac{4}{\nu}\right)^2}{2\sigma_3^2}\right), & \nu \geq \nu_0 \\ \frac{\rho^2}{1+\rho^2}\exp\left(-\frac{\left(\log_2\frac{4}{\nu}\right)^2}{2\sigma_2^2}\right) + \exp\left(-\frac{\left(\log_2\frac{4}{\nu}\right)^2}{2\sigma_3^2}\right), & \nu_0 > \nu \geq \nu_0/4 \\ \frac{\rho^2}{1+\rho^2}\exp\left(-\frac{\left(\log_2\frac{4}{\nu}\right)^2}{2\sigma_2^2}\right) + 1, & \nu < \nu_0/4 \end{cases} \quad (5)$$

In this case the parameters were fitted to match the behavior of human observers in induction experiments. The experiment was performed by using classical induction stimuli (see Fig. 3 -left- for representative examples). The parameters obtained where $\nu_0 = 4$ cpd for the luminance channel and $\nu_0 = 2$ cpd for the red-green and yellow-blue chromatic channels. Both σ_2 and σ_3 were set to 1.25 and 2, respectively. To simulate the band-pass profile of the intensity channel CSF and the low-pass profile of chromatic channels CSFs, they set $\sigma_1 = 1.25$ for the luminance channel, and $\sigma_1 = 2$ for both the red-green and yellow-blue channels.

Equations 2 and 3 show the similarity between the models: the term in parenthesis in Eq. 2 could be interpreted as an Extended-CSF. Alternatively, the quotient between center-surround contrast ratios in Eq. 5 could be interpreted as a divisive normalization.

2. METHODS

In order to assess the coding efficiency of certain representation, one has to empirically assess the reduction in multi-information.[19] When the computation of the Jacobian of the transform is not straightforward, and the dimensionality is not preserved, reductions in mutual information between pairs of coefficients[20] is also a sensible approach.[12]

In this case, for a large set of images, we computed the responses in the two stages of the models: (1) after the wavelet transform, and (2) after the non-linear interaction stage. Then, we computed the statistical dependence mutual information (MI) between pairs of neighbor responses at each stage. A mechanism following a redundancy reduction principle should reduce the mutual information along the stages of the model.

We estimated the mutual information using 2D histograms since it has been shown that this straightforward approach is accurate enough for this class of heavy tailed signals.[12] Since MI is invariant under point-wise transforms,[20] our MI estimator first equalizes the marginal PDF of each coefficient to obtain uniform densities in the range [0, 1]. Then, the joint entropy is computed by using the 2D histogram and the Miller-Madow correction.[21] In our implementation, the total number of bins in the 2D histogram was set to be the square root

Figure 3. Representative images for two classes of stimuli. First and second: colorimetrically calibrated images where induction effects are substantial, assimilation effect in left side and contrast effect in right side. Third and fourth: examples of colorimetrically calibrated natural images.

of the number of available samples. In our case, the marginal entropies are zero due to the uniformization step. Therefore the MI is equal to minus the joint entropy.

We made this analysis for two classes of visual stimuli:

- Colorimetrically calibrated images where induction effects are substantial -we used the stimuli reported in Otazu et al.[6]-. These consisted of a set of two circularly symmetric patterns presented side by side and separated 7.6 deg of visual angle from the observer's viepoint on a dark background. We considered both *assimilation* and *contrast* stimuli. The left side stimulus consisted of a series of concentric rings alternating between two chromaticities with an extra ring of similar width. The right side was the reference stimulus (see Fig. 3, left, for representative examples). The spatial frequencies used in the *assimilation* and *contrast* stimuli were 0.81, 1.77, and 2.74 cpd, respectively. The stimulus rings were rendered using four sets of colored patterns. All the considered spatiochromatic configurations lead to 24 images to extract samples from. A total of $2 \cdot 10^5$ sample pairs were used in the MI computation.
- Colorimetrically calibrated natural images -from the Barcelona database[22]-. The database consists of 419 natural images where each colour plane corresponds to the human LMS cone activations. The images were chosen to represent five different visual environments and were taken under natural illumination at different times of the day. At the bottom-left corner of each picture there was a matte grey ball of approximately constant spectral reflectance and nearly Lambertian reflective properties, which allows to compute the illuminant (see Fig. 3, right, for representative examples). We did not consider the area corresponding to the grey ball in the MI computations. MI estimations were done using $3.2 \cdot 10^5$ samples.

3. RESULTS

This section shows the mutual information results for coefficients of the third scale after the wavelet transform and after the non-linear interaction, for the Chromatic Induction Model[6] (Figure 4, top panel) and for the Masking Model[7,12] (Figure 4, bottom panel). For useful reference, results for the achromatic channel and natural images of the Masking Model are consistent with the MI values previously reported on a different database.[7,12]

The different scale in the mutual information values in the Masking and the Induction models comes from the use of different wavelet transforms. Nevertheless, this is just a global scaling factor, while the relevant issue here is how the MI decreases or increases with the application of the nonlinear transform in each case. In the case of induction stimuli, the MI values are substantially larger than for natural images. This is not surprising since these images have a quite simple spatial structure leading to highly redundant values in the wavelet domain. Therefore, the MI values are large even though wavelet transforms are known to be efficient for natural images.

Results for the Chromatic Induction model show two distinctive trends: (1) for spatially simple images (in which induction is relevant), the induction nonlinearity substantially reduces mutual information (dashed versus solid lines); however (2) for natural stimuli, the induction nonlinearity gives rise to no reduction of redundancy with regard to the previous wavelet stage.

In the Masking model the situation reverses: while for natural images the efficiency gain is substantial in the achromatic case (as previously reported), and no clear gain is obtained in the chromatic channels; for spatially simple images the efficiency of the chromatic channels clearly decreases after the nonlinear stage.

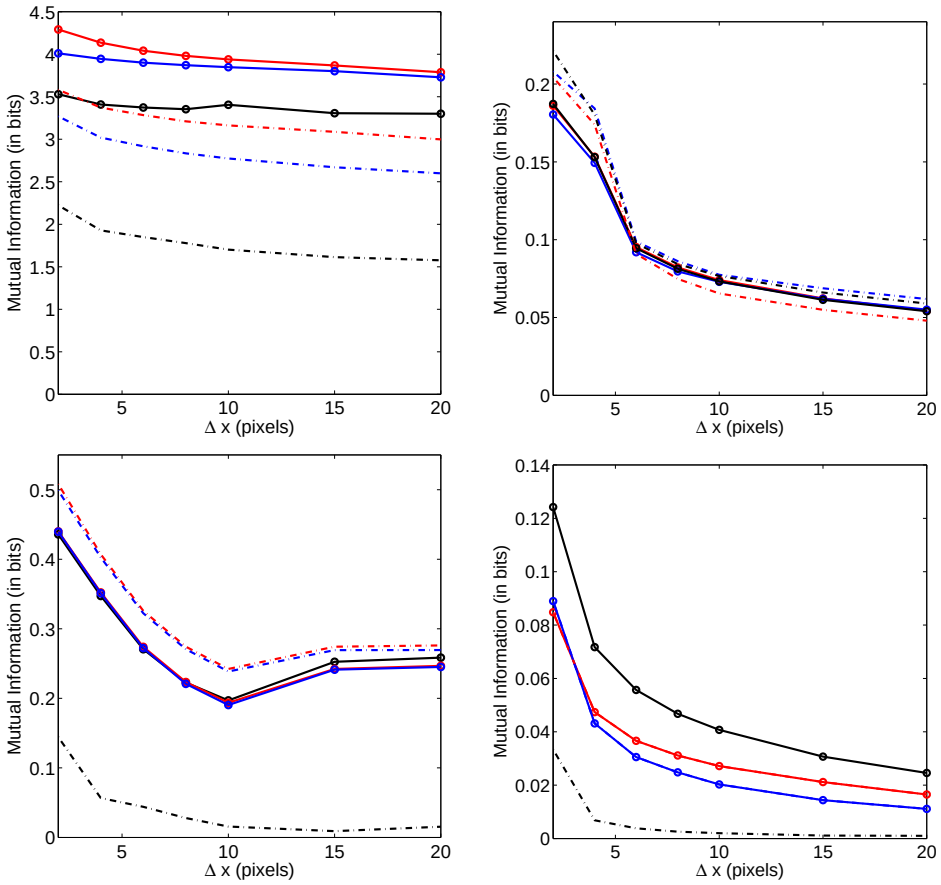

Figure 4. Assessment of coding efficiency for the Chromatic Induction Model (top row) and for the Masking Model (bottom row), for induction images (left column) and natural images (right column). Plots show the Mutual Information between pairs of coefficients (at different spatial distance Δx) of the third wavelet scale after the wavelet transform (solid lines) and after the non-linear interaction (dashed lines). Red, Blue and Black lines stand for the behavior at the Red-Green, Yellow-blue and Achromatic channels of the models.

4. DISCUSSION

Given the *formal* similarity of the considered brightness and chromatic induction model[5,6] with divisive normalization masking models[2,7,12] one could expect that chromatic induction also give rise to reductions in mutual information. However, this is not the case for natural images *with the current parameters of the induction model*. Interestingly, induction mechanisms seem to provide more efficient image representations for images in which induction effects are substantial. These facts can be interpreted in alternative ways:

- Efficient coding skeptics. The organization of mechanisms responsible for induction is not guided by a redundancy reduction principle since for the images that really matter (natural stimuli) they lead to no coding gain. In natural environments the advantage of induction (if any) is not in coding but somewhere else (e.g. increasing color contrast for better discrimination).

- Efficient coding lovers. Masking and Induction models differ mainly in the values of their parameters (for example, the saturation non-linearity is smoother in the masking case). One could think that the same spatio-chromatic sensors change depending on the environment to attain the same efficiency goal. When dealing with complex images the sensors behave according to masking-like nonlinearities since they give rise to better coding efficiency for those images. However, when facing spatially simple images their nonlinearity changes to the one used in the induction model to obtain better coding performance. Induction

would be just a by product of this change of behavior. This would explain why induction is more apparent in spatially simple images.

Further work may clarify which is the correct interpretation. In this regard we will explore modifications of the model parameters (both in the masking and the induction models) to see whether the reproduction of chromatic induction psychophysics is compatible with coding gains in natural environments.

REFERENCES

[1] Carandini, M. and Heeger, D. J., "Normalization as a canonical neural computation.," *Nature reviews. Neuroscience* **13**, 51–62 (Jan. 2012).

[2] Watson, A. and Solomon, J., "A model of visual contrast gain control and pattern masking," *JOSA A* **14**, 2379–2391 (1997).

[3] Simoncelli, E. P. and Heeger, D. J., "A model of neuronal responses in visual area MT," *Vision Research* **38**(5), 743–761 (1998).

[4] Hillis, J. M. and Brainard, D., "Do common mechanisms of adaptation mediate color discrimination and appearance? Uniform backgrounds," *JOSA A* **22**(10), 2090–2106 (2005).

[5] Otazu, X., Vanrell, M., and Parraga, A., "Mutiresolution wavelet framework models brigthness induction effects," *Vision Research* **48**(5), 733–751 (2008).

[6] Otazu, X., Parraga, A., and Vanrell, M., "Toward a unified chromatic induction model," *Journal of Vision* **10**(12) (2010).

[7] Laparra, V., Muñoz Marí, J., and Malo, J., "Divisive normalization image quality metric revisited," *JOSA A* **27**(4), 852–864 (2010).

[8] Abrams, A., Hillis, J., and Brainard, D., "The relation between color discrimination and color constancy: When is optimal adaptation task dependent?," *Neural Computation* **19**(10), 2610–2637 (2007).

[9] Laparra, V., Jimenez, S., Camps, G., and Malo, J., "Nonlinearities and adaptation of color vision from sequential principal curves analysis," *Neural Computation* **24**(10), 2751–2788 (2012).

[10] Schwartz, O. and Simoncelli, E., "Natural signal statistics and sensory gain control," *Nat. Neurosci.* **4**(8), 819–825 (2001).

[11] Malo, J. and Gutiérrez, J., "V1 non-linear properties emerge from local-to-global non-linear ICA," *Network: Computation in Neural Systems* **17**, 85–102 (2006).

[12] Malo, J. and Laparra, V., "Psychophysically tuned divisive normalization approximately factorizes the PDF of natural images," *Neural Computation* **22**(12), 3179–3206 (2010).

[13] Lyu, S., "Dependency Reduction with Divisive Normalization: Justification and Effectiveness," *Neural Computation* **23**(11), 2942–2973 (2011).

[14] J.Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision Research* **29**, 847–856 (1980).

[15] E.Simoncelli and E.Adelson, [*Subband Image Coding*], MA: Kluwer Academic Publishers, Norwell (1999).

[16] Mullen, K. T., "The CSF of human colour vision to red-green and yellow-blue chromatic gratings," *J. Physiol.* **359**, 381–400 (1985).

[17] J.Malo, Pons, A., A.Felipe, and Artigas, J., "Characterization of human visual system threshold performance by a weighting function in the gabor domain," *Journal of Modern Optics* **44**, 127–148 (1997).

[18] Heeger, D., "Normalization of cell responses in cat striate cortex," *Journal of Modern Optics Vis. Neurosci.* **9**, 181–198 (1992).

[19] Bethge, M., "Factorial coding of natural images: how effective are linear models in removing higher-order dependencies?," *Journal of the Optical Society of America A* **23**(6), 1253–1268 (2006).

[20] Cover, T. M. and Tomas, J. A., [*Elements of Information Theory*], John Wiley & Sons, New York (1991).

[21] Miller, G., "Note on the bias of information estimates," *Information Theory in Psychology* **II-b**, 95–100 (1955).

[22] Parraga, C., Vazquez, J., and Vanrell, M., "A new cone activation-based natural image dataset," *Perception (Suppl.)* **36**, 180 (2009).