Simulating Visual Pattern Detection and Brightness Perception
Based on Implicit Masking

Jian Yang[*]

Applied Vision Research and Consulting
Penfield, New York 14526

ABSTRACT

A quantitative model of implicit masking, with a front-end low-pass filter, a retinal local compressive nonlinearity described by a modified Naka-Rushton equation, a cortical representation of the image in the Fourier domain, and a frequency dependent compressive nonlinearity, was developed to simulate visual image processing. The model algorithm was used to estimate contrast sensitivity functions over 7 mean illuminance levels ranging from 0.0009 to 900 trolands, and fit to the contrast thresholds of 43 spatial patterns in the Modelfest study. The RMS errors between model estimations and experimental data in the literature were about 0.1 log unit. In addition, the same model was used to simulate the effects of simultaneous contrast, assimilation, and crispening. The model results matched the visual percepts qualitatively, showing the value of integrating the three diverse perceptual phenomena under a common theoretical framework.

Keywords: Human vision model, implicit masking, compressive nonlinearity, space domain, spatial frequency domain, pattern detection, brightness perception.

---

[*] jyang@visionrc.com; phone 1 585 410-5681; http://www.visionrc.com; Applied Vision Research and Consulting, 6 Royal Birkdale Ct, Penfield, NY 14526

1. INTRODUCTION

A human vision model would be attractive and extremely useful if it can simulate visual spatial perception and performance over a broad range of conditions. Vision models often aim at describing pattern detection and discrimination [1-3] or brightness perception [4, 5], but not both, due to the difficulty of simulating the complex behavior of the human visual system. In an effort to develop a general purpose vision model, the author of this paper proposed a framework of human visual image processing and demonstrated the capability of the model to describe visual performance such as grating detection and brightness perception [6]. This paper will further present a refined version of the visual image processing model and show more examples to investigate the usefulness of this approach.

In general, three major issues must be overcome to create a successful vision model. One issue is estimating the capacity of information captured by the visual system, which determines the degree of fine spatial structure that can be utilized by the visual system, which may be modeled by using a low-pass filter. The second issue, the central focus of this paper, is the modeling of nonlinear processes in the visual system, such as light adaptation and frequency masking. It is important to note that the effects of the nonlinear processes are local to each domain. For example, light adaptation describes the change of visual sensitivity with a background field, the effect of which is limited to a small spatial area [7, 8]. Frequency masking describes the effect of a background grating and occurs, if it does, only when the target and background contain similar frequencies [9]. This space or spatial frequency domain-specific effect makes it advantageous to transform the signals to the relevant domains to perform particular nonlinear operations. Moreover this transformation roughly mimics the transformations that are believed to occur in the human visual system. The third issue concerns information representation and decision making at a later stage.

In the endeavor of applying human vision detection models to engineering applications, several remarkable advances have been reported. Watson [10] proposed a so-called Cortex transform to simulate image-encoding mechanisms in the visual system, applying frequency filters similar to Gabor functions (i.e., a sinusoid multiplied by a Gaussian function) in terms of localization in the joint space and spatial-frequency domain. Later, Watson and Solomon [3] applied Gabor filters in their model to describe psychophysical data that was collected to understand the effects of spatial frequency masking and orientation masking. Peli [11, 12] considered the loss of information in visual processing, and boosted particular frequency bands of Gabor filters accordingly to obtain specific effects of image enhancements for visual impaired viewers. Based on the concept of the Cortex transform and other considerations, Daly [13] further developed a complete visual difference predictor to estimate visual performance for detecting the differences between two images. Lubin [14] also developed an impressive visual imaging model that attempts to model not only spatial, but also temporal aspects of human vision.

Most of the existing pattern detection models share at least one common feature. They incorporate the visual contrast sensitivity function (CSF) as a module within their models. These models either apply an empirical CSF as a front-end frequency filter [3, 11], or adjust the weighting factors of each Gabor filter based on the CSF values [14]. Therefore, obtaining an appropriate CSF is a critical step for these models. As the CSF plays such an important role in these models, it is worthwhile to review some CSF properties here.

Human visual CSF: A simple and widely used psychophysical test is the measurement of the contrast of sine-wave gratings that is just detectable against a uniform background. Such contrast threshold is reciprocal to contrast sensitivity [15]. Contrast values are calculated by using Michelson formula: $(L_{max}-L_{min})/(L_{max}+L_{min})$, where L_{max} and L_{min} are the peak and trough luminance of a

grating, respectively. As an example, Fig. 1 shows how the contrast threshold varies with spatial frequency and mean luminance, as reported by Van Nes and Bouman [16]. When the reciprocal of the contrast threshold value is expressed as a function of spatial frequency, the resulting function is referred to as the CSF. Under normal viewing conditions (i.e., photopic illumination level and slow temporal variations), the CSF has a band-pass shape, displaying attenuation at both low and high spatial frequencies [15-17]. To some extent, the CSF is similar to the MTF in optics, characterizing a system's response to different spatial frequencies. The behavior of the CSF is, however, much more complicated; it varies with the mean luminance, the temporal frequency, and the field size of the grating pattern.

      Although the CSF is an important model component, it is interesting to note that none of the mentioned image processing models tried to explain how and why CSF behaves differently in different conditions. One popular explanation of the CSF shape relies on retinal lateral inhibition [18]. In this theory, the visual responses are determined by retinal ganglion cells, which take light inputs from limited retinal areas. These areas are called receptive fields. They are circular in shape and each of them contains two distinct function zones: the center and surround. The inputs to the two zones tend to cancel each other, the so-called center-surround antagonism. Such spatial antagonism attenuates uniform signals, as well as low frequency signals. This might explain why the system as a whole is insensitive to low frequencies. However, I have not seen a coherent model emerging from this theory to offer a quantitative description of all the CSF curves simultaneously .

      In the literature, there are many descriptive models of the CSF [19-21]. These models can be useful in practical applications, but they provide little mechanistic insight into why the CSF should behave as it does, pertinent to how the images are processed in the visual system. In addition, the CSF represents the responses of the entire visual system to one type of stimuli, that is, sinusoidal gratings, and therefore, they are not a component of a visual image-processing model, as the visual system is not a linear system. The question becomes, can an image-processing model be built to simulate the behavior of the human visual system as shown in Fig. 1 when sine-wave gratings are used as inputs to the model.

Figure 1. Contrast threshold versus spatial frequency, with mean retinal illuminance ranging from 0.0009 (top) to 900 (bottom) trolands in log steps. The data points are from Van Nes and Bouman [21] and the smooth curves are the fits with current model (see below).

Implicit Masking: In the effort to model the CSF, Yang and Makous [22-24] suggested that the DC component, that is, a component at 0 cycle per degree (cpd) and 0 Hz, in any visual stimulus has all the masking properties of any other Fourier component. The associated effect of the DC component in visual detection was called *implicit masking* [25]. The basic assumption here is that the energy of the DC component can spread to its neighboring frequencies, because of spatial inhomogeneities of

the visual system. When a target is superimposed on a background field of similar features, the required stimulus strength for detection, i.e., threshold strength, is generally increased. This is a nonlinear interaction. It follows that the DC component can reduce the visibility of the targets at low spatial frequencies as a consequence of the energy overlap, given such nonlinear interactions. This concept simplifies the explanation of CSF behavior considerably, as discussed in the following.

First, let us explore the roll-off of the CSF at the low spatial frequencies. Each of the frequency components spreads to a limited extent. The interaction between the target and the DC components should disappear when the spatial frequency of the stimulus is high enough. In this case, there is no effect of implicit masking. Therefore, the drop of contrast sensitivity because of implicit masking is restricted to low spatial frequencies.

Second, this assumption offers an explanation of the effect of luminance on the contrast sensitivity at low spatial frequencies: as mean luminance decreases, the component at zero frequency decreases too. When this happens, other factors such as noise can dominate, and thus the relative attenuation at low frequencies decreases.

Third, this assumption also offers an explanation of the dependence of the attenuation on temporal frequency [22]. The DC component of a grating is at zero temporal frequency and zero spatial frequency in a 2-D spatiotemporal frequency domain, so the effects of implicit masking apply only to very low temporal and spatial frequencies. Test gratings that are modulated at high temporal frequencies would be exempt from the effect of implicit masking, no matter what the spatial frequency of the grating.

Finally, the effect of field size on contrast sensitivity can be explained by the breadth of implicit masking. The extent of implicit masking is determined by the spread of the DC energy in the frequency domain. The larger the viewing field, the less the spread [26]. This explains why the peak sensitivity shifts to lower spatial frequency as field size increases, owing to the decreasing breadth of implicit masking. The exact amount of spread depends also on retinal inhomogeneities [26].
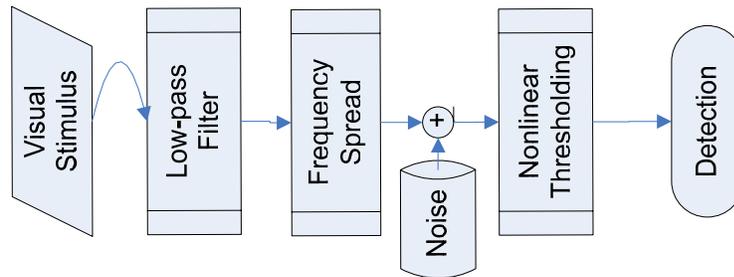
Figure 2. A three-stage model of CSF, based on implicit masking.

Based on the concept of implicit masking, Yang et al. [24] developed a quantitative model of the CSF. As schematized in Fig. 2, the form of visual processing is partitioned into three *functional* stages. The first stage represents a low-pass filter and it includes the effects of ocular optics, photoreceptors, and neural summation. The second stage represents a spread of grating energy to nearby frequencies. This stage represents frequency spreading caused by inhomogeneities in the stimulus, such as truncation of the field, and spatial inhomogeneities in the visual system, such as variation in the density of ganglion cells. The third stage, a nonlinear thresholding operation, is characterized by a nonlinear relationship between the required threshold amplitude and the background amplitude values. When the energy of the background field spreads to frequencies close to 0 cpd, the virtual masking amplitude at low frequencies increases and so does the threshold amplitude [24]. In this model, implicit masking is responsible for the CSF shape at low spatial

frequencies, and the low-pass filter determines the sensitivity roll-off at high spatial frequencies. In addition to the CSF shape, Fig. 1 shows that the overall contrast threshold reduces as the mean luminance level increases. It was found that the inclusion of a photon-like shot noise, as indicated in Fig. 2, provided a satisfactory account of the overall threshold changes [24]. The absolute shot noise increases, but the noise contrast reduces with mean luminance following a square-root law [27, 28].

In a further research, Yang and Stevenson [29] noticed that the interocular luminance masking affects low, but not high spatial frequencies, which suggests that the change of visual sensitivity at high spatial frequencies is determined by retinal processes, such as light adaptation, but not the luminance dependent noise.

So far the model is in an analytical form, taking parameter values, such as the frequency, the contrast, and the luminance of the stimulus as model inputs. It cannot, however, take stimulus profiles or images as the inputs. Later in this paper I will show how to extend such a model to perform visual image processing with incorporating implicit masking and compressive nonlinear processes.

Nonlinearity and Divisive Normalization: Nonlinear processes in vision have often been explained by a nonlinear transducer function [30, 31]. According to such a theory, threshold is inversely proportional to the derivative of the transducer function at any given pedestal amplitude [2, 32, 33]. Heeger [34, 35] suggested that the nonlinearity of the cells in striate cortex and related psychophysical data may be due to a normalization process. Foley [2] suggested that such normalization requires inhibitory inputs to the transducer function. However, specifying excitatory and inhibitory interactions among different stimulus components can be complicated in general cases. To deal with this difficulty, I use locally pooled signals in either the space domain or the spatial frequency domain to replace the signal in the denominator of the Naka-Rushton equation . Therefore, such modified compressive nonlinearity can display some features of divisive normalization.

2. IMAGE PROCESSING BASED FRAMEWORK

The proposed model framework is based on the ideas of implicit masking, modified compressive nonlinear process, and other well-known properties of the visual system that have been used in many models. The model components are schematized in Fig. 3, and are elaborated in the following subsections.
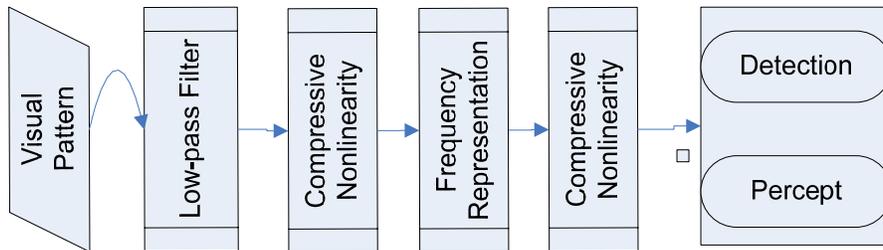
Figure 3. The schematized framework of visual image processing for pattern detection and brightness perception. The output of the last nonlinearity shows cortical information representation.

Low-pass Filtering: When the light modulated information of an image enters into human eyes, it passes through the optical lens of the eye and is captured by photoreceptors in the retina. One function of photoreceptors is to sample the continuous spatial variation of the image discretely. The cone signals are further processed through horizontal cells, bipolar cells, amacrine cells, and

ganglion cells with some re-sampling. From an image processing point of view, the effects of optical lens, sampling, and re-sampling in the retinal mosaic are low-pass filtering.

We estimate the front-end filter from psychophysical experiments. It has been shown that the visual behavior at high spatial frequencies follows an exponential curve [36]. Yang et al. [24] extrapolated this relationship to low spatial frequencies to describe the whole front-end filter with an exponential function of spatial frequency:

$$LPF(f) \ = Exp(-\alpha f), \tag{1}$$

where α is a parameter specifying the rate of attenuation for a specific viewing condition. Yang and Stevenson [37] modified the formula to account for the variation in α with the mean luminance of the image:

$$\alpha = \alpha_0 + \delta / L_0^{0.5}, \tag{2}$$

where α_0 and δ are two parameters and L_0 the mean luminance of the image.

Retinal Compressive Nonlinearity: In the retina, there are several major layers of cells, starting from photoreceptors including rods and three types of cones, to horizontal cells, bipolar cells, amacrine cells, and finally to ganglion cells where the information is transmitted out of the retina via optic nerve fibers to the central brain [38]. Retinal processes include a light adaptation, where the retina becomes less sensitive if continuously exposed to bright light. The adaptation effects are spatially localized [39, 40].

In the current model, the adaptation pools are assumed to be constrained by ganglion cells with an aperture window:

$$W_g(x, y) = \frac{1}{2\pi r_g^2} Exp(-\frac{x^2 + y^2}{2 r_g^2}), \tag{3}$$

where r_g is the standard deviation of the aperture. The adaptation signal at the level of ganglion cells I_g is the convolution of the low-passed input image I_c with the window function W_g. In this algorithm, the window profile is approximated as spatially invariant by considering only foveal vision. The retinal signal I_R is the output of a compressive nonlinearity. The form of this nonlinear function is assumed here to be the Naka-Rushton equation, which has been widely used in models of retinal light adaptation [41, 42]. One major difference here is that the adaptation signal I_g in the denominator is a pooled signal, which is similar to a divisive normalization process:

$$I_R = \frac{w_0(1 + I_0^n) I_c^n}{I_g^n + (I_0 w_0)^n}, \tag{4}$$

where n and I_0 are parameters that represent the exponent and the semi-saturation constant of the Naka-Rushton equation, respectively, and w_0 is a reference luminance value. In conditions where I_c and I_g are all equal to w_0, the retinal output signal is the same as the input signal strength.

Cortical Compressive Nonlinearity: Simple cells and complex cells in the visual striate cortex usually respond to stimuli of limited ranges in spatial frequency and orientation [43, 44]. To capture this frequency- and orientation-specific nonlinearity, one can transform the image I_R from a spatial domain to a frequency domain representation via a Fourier transform to $T(f_x, f_y)$, and divided by n_x and n_y to normalize the amplitude in the frequency domain. Here f_x and f_y are the spatial frequencies in x and y directions, respectively, and n_x by n_y is the number of image pixels.

These cells also exhibit nonlinear properties; their firing rate does not increase until the stimulation strength is above a threshold level and the firing rate saturates when the stimulation strength is very strong [44]. In the model calculation, the signal in the frequency domain passes through the same type of nonlinear compressive transform as it did in the retinal processing. Following the concept of frequency spread in implicit masking (see Fig 2), one major step here is to compute the frequency spreading that affect the masking signal in the denominator of the nonlinear formula. In this model, the signal strength in the masking pool, $T_m(f_x, f_y)$, is the convolution of the absolute signal amplitude $|T(f_x, f_y)|$ and an exponential window function:

$$W_c(f_x, f_y) = Exp[-(f_x^2 + f_y^2)^{0.5}/\sigma], \tag{5}$$

where σ correlates with the extent of the frequency spreading and the bandwidth of frequency channels. As the bandwidth of frequency channels increases with the spatial frequency [1], one should expect that the σ value increases with spatial frequency. To simplify the computation, however, this value is approximated as a fixed value in the current algorithm. Applying the same form of compressive nonlinearity as in the retina, the cortical signal in the frequency domain is expressed as:

$$T_c = sign(T) \frac{w_0(1 + T_0^v)|T|^v}{T_m^v + (T_0 w_0)^v}, \tag{6}$$

where v and T_0 are parameters that represent the exponent and the semi-saturation constant of the Naka-Rushton equation for the cortical nonlinear compression, respectively. The term T_m in the denominator includes the energy spread of the DC component (i.e., at 0 cpd) of the spatial pattern. This component is processed in the same way as other frequency maskers, if there are any, under Eq. 6. Thus, the concept of implicit masking is naturally implemented in the image processing framework. In summary, the major process in the cortex is modeled by a compressive nonlinearity applying to the spatial frequency and orientation components. The cortical image representation in the frequency domain is given by the function T_c. This function will be used to calculate visual responses for pattern detection and for estimating perceived brightness, as described in the following sections.

3. MODEL FITS TO PATTERN DETECTION DATA

As mentioned earlier, this paper focuses on the nonlinear parts of the visual process. In order to investigate whether the model estimates pattern visibility reasonably, a detection stage was added in the model to fit existing experimental data. A simple Minkowski summation was used to estimate the signal strength at a decision stage, although some other approach, such as linear summation within spatial frequency channels [45], or signal detection theory [46, 47], may ultimately turn out superior.

The following examples shows model fits to two sets of experimental data on pattern detection performance. One set data is the contrast thresholds reported by Van Nes and Bouman [16] for detecting gratings at various mean luminance levels. The other set is from the Modelfest study with the contrast thresholds of 43 patterns at a mean luminance level of about 30 cd/m^2 [45, 48].

Pattern Detection Stage: Based on the block diagram (Fig. 3), a visual pattern passes through a low-pass filter, a retinal compressive nonlinearity, a frequency domain representation, and a cortical compressive nonlinearity to produce the cortical signal as described by T_c (see Eq. 6). In real

experiments, observers look for the target signal against a background field. To simulate this task in the computation, one can calculate the cortical visual response, T_{c_t}, in the spatial frequency domain in respect to the visual pattern, and, T_{c_r}, to the reference background field. The signal strength in the detection stage is assumed to equal to the Minkowski summation of the differences between T_{c_t} and T_{c_r} at every frequency components:

$$R = \{ \Delta f_x \Delta f_y \Sigma [(T_{c\_t} - T_{c\_r})^\beta] \}^{1/\beta}, \tag{7}$$

where Δf_x and Δf_y are the frequency intervals along x and y directions, respectively, and β is the exponent of the Minkowski summation over different frequency components. The response strength R is assumed to be a constant value R_t at a given threshold criterion.

Fits to Van Nes & Bouman Data: The Van Nes and Bouman [16] paper reported the contrast thresholds for detecting gratings with spatial frequencies in the range of 0.5 to 46 cpd, covering 7 mean illuminance levels in the range of 0.0009 to 900 trolands. The threshold values were measured using a method of limits, adjusting the contrast value to make the test grating just visible or just disappear to the observers. The major challenge for the computational model is to duplicate the thresholds, which change with luminance and spatial frequency as shown in Fig. 1. There are total of 102 data points corresponding to gratings of different spatial frequency and luminance combinations.

For each grating, the response strength R is determined by Eq. 7. The model estimated contrast threshold is the one that leads R being equal to a constant R_t value. Model parameters were optimized to minimize the root-mean-squared (RMS) error between the model estimates and the experimental data, both on a logarithmic scale:

$$E = \{\Sigma[log(C_i) - log(CE_i)]^2/n\}^{1/2}. \tag{8}$$

Here C_i is the model estimated contrast threshold, CE_i the contrast threshold reported by Van Nes and Bouman for the ith stimulus, and n is the number of data points. The summation is over all the102 gratings.

In model equations 1 to 7, there are 11 system parameters: α_0, δ, r_g, w_0, n, I_0, v, T_0, σ, β, and R_t. Each of the parameters is a positive real number; and some of them convey specific physical meaning about the visual system. These parameter values can be estimated by optimizing the fits between model predictions and experimental data. The quality of the fits was not sensitive to some parameter values when other parameters were optimized accordingly. These parameters, δ, r_g, w_0, and β, were thus set to 0.10 deg td$^{1/2}$, 0.9 min of arc, 100 cd/m^2, and 2.2, respectively, based on reasonable pilot data fits. The other 7 parameters were optimized to minimize the residual error as determined by Eq. 8. The contrast thresholds of the fits are plotted in smooth curves in Fig. 1, where the RMS error being 0.10 log unit.

Although there is no band-pass filter built in the model, the model output exhibits a band-pass behavior at high luminance levels. This result demonstrates the role of implicit masking. Furthermore, the model output captures the trend of the threshold variation with spatial frequency and luminance nicely.

Fits to the Modelfest Data: The above example shows that the model is adequate to capture visual performance on detecting the particular patterns, i.e., sinusoidal gratings. Now we examine how well this model deals with a variety of patterns. Modelfest was a collaboration between many laboratories to measure contrast thresholds of a broad range of patterns, including Gabor functions of varying aspect ratio, Bessel and Gaussian functions, lines, edges, checkerboard, natural scene, and

random noise, in order to provide a database for testing human vision models [45, 48]. There were 43 different monochromatic spatial patterns in the Modelfest test set. The field size was 2.13^o x 2.13^o and mean background luminance was about 30 cd/m^2. The contrast thresholds were determined using two-alternative-forced-choice (2AFC) with 84% correct responses.

The aim of developing a general purpose vision model will be one step closer if the model can produce contrast thresholds that are closely matched to the experimentally obtained results for all the stimuli, without varying the above determined model parameter values. To check this possibility, the luminance profile of each of the 43 visual stimuli was input to the model algorithm to calculate their contrast thresholds, which are shown in the dotted lines in Fig. 4. As a comparison, the circles show the mean experimental data over 16 observers. Clearly, the model underestimates the contrast thresholds in most of the cases. The model deviation in terms of RMS error is 0.22 log unit. Taking into the fact that the model parameters were obtained from a quite different experimental data set, the performance of the model is encouraging.

Two areas were identified that could contribute to the model deviations. One is on the low-pass filter. The Van Nes and Bouman study used Maxwellian view with optical apparatus, while the Modelfest study used direct view of video displays. Thus it is reasonable to have a greater α_0 value in the Modelfest study than that in the Van Nes and Bouman study. The second area is the decision making stage, as there were differences in the threshold measurements. This may require using different β and R_t values in the current model. Consequently, the solid lines in Fig. 4 show the model fits to the experimental data after optimizing the three parameters while the other 8 parameters were kept the same as in the previous case. The resulting RMS error is 0.11 log unit. The parameter value changed from 0.11 to 0.14 degree for α_0, 2.2 to 1.7 for β, and 0.36 to 0.53 for R_t.

The RMS error is larger than those reported by Watson and Ahumada [45], however the current model has the advantage in dealing with diverse data sets. As discussed earlier, this model can describe the luminance dependent CSFs. It can also explain brightness perception as shown in the next section.

Figure 4. Contrast thresholds of 43 Modelfest stimuli. The data points (circles) represent mean experimental results over 16 observers; the dotted lines represent model predications with a RMS error of 0.22 log unit; and the solid lines represent optimal model fits with a RMS error of 0.11 log unit.

From Fig. 4, one can see that the major contribution to the RMS error comes from stimuli #35 (a noise pattern) and #43 (a natural scene), where the model estimates are much lower than the experimental data as marked by the line segments (see Fig. 4). For the noise pattern, its spectra in the spatial frequency domain have random phases. Including a linear summation within narrow frequency channels can cancel some of the energies due to the phase differences, thus increasing the threshold estimate and potentially improving the fit. For the natural scene, energy cancellation can happen within linear channels too, due to the phase variations within the summation windows.

4. SIMULATING BRIGHTNESS PERCEPTION

The current model algorithm is designed to deliver visual information representation $\mathbf{T_c}$ at a cortical level (see Eq. 6). This information can be used to estimate pattern visibility as shown in the previous section. It is reasonable to believe that the cortical information presentation can be used to produce visual perception too when additional processes are included. In this section, I will show that the obtained cortical representation, after adding a fill-in process, can also be used to estimate the brightness perception of three well-know examples: simultaneous contrast, assimilation, and crispening.

Local Simultaneous Contrast: It is well known that the brightness of a visual target depends not only on the luminance of the target, but also on the local contrast of its edges in reference to the luminance of adjacent areas. Simultaneous contrast is often demonstrated by the brightness of a gray spot at different surrounding luminance levels (e.g., Ref. 49). Although the luminance level of the gray spot is fixed, the perceived brightness of the spot increases while the surrounding luminance decreases.
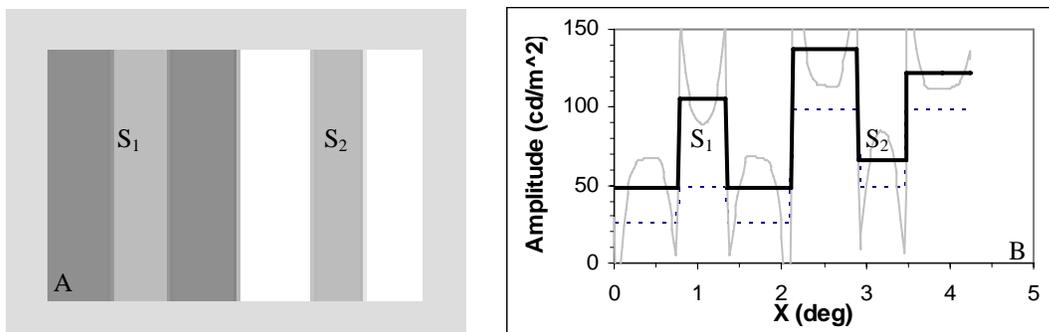
Figure 5. Panel A is a demonstrative pattern to show the effect of simultaneous contrast, where stripe S_1 looks brighter than S_2 while they have the same luminance, and panel B shows the luminance profile of the visual pattern (dotted lines), and the model simulation results of the brightness before (dim lines) and after (thick lines) a fill-in process.

For simplicity, the examples shown here are for one-dimensional patterns[1]. In the first example, the visual pattern with simultaneous contrast is demonstrated in Fig. 5A. Even though both of the stripes S_1 and S_2 have the same luminance level of about 50 cd/m^2 (see the dotted lines in the right panel of Fig. 5 for the corresponding luminance profile), stripe S_1 that is flanked by a lower luminance level of about 25 cd/m^2 looks brighter than stripe S_2 that is flanked by a higher luminance level of about 100 cd/m^2. This has been attributed to the effect of local contrast.

---

[1] note: The visual patterns in Figs. 5-7 are for the demonstrative purpose. The pattern luminance will not match the specified luminance profiles due to media limitation and the lack of standards to calibrate the printed or displayed images. Therefore, the perceived brightness by readers here may not reflect what it should be as in well controlled experiments .

In the model simulation of the perceived brightness, the luminance profile of the visual pattern is fed into the model algorithm as an input. Based on Eqs. 1 to 6, one can obtain the frequency domain representation, i.e., T_c, of the visual pattern. By performing an inverse FFT, one obtains the spatial representation of the pattern as shown by the dim lines in Fig. 5B. This spatial response contains overshoots near the edges. For estimating the brightness of each stripe, some investigators have suggested a fill-in process [50, 51] or an averaging process [4]. The thick lines in Fig. 5B are the average values of the dotted lines within each stripe after considering such a simple fill-in process. As the final simulation results (thick lines) show, the visual response to the left-side stripe is 105 that is larger than the response of 66 to the right-side stripe, in agreement with our percept in terms that S_1 is perceived brighter than S_2. As a clarification, this paper provides only qualitative comparison of the model prediction to actual visual percept; no efforts have been taken to attain an adequate match in numbers. The unit of brightness perception from the model does not provide a clear meaning yet, and the scale relies on the model parameter w_0, which was set to 100 cd/m^2 in the current model algorithm as mentioned earlier.
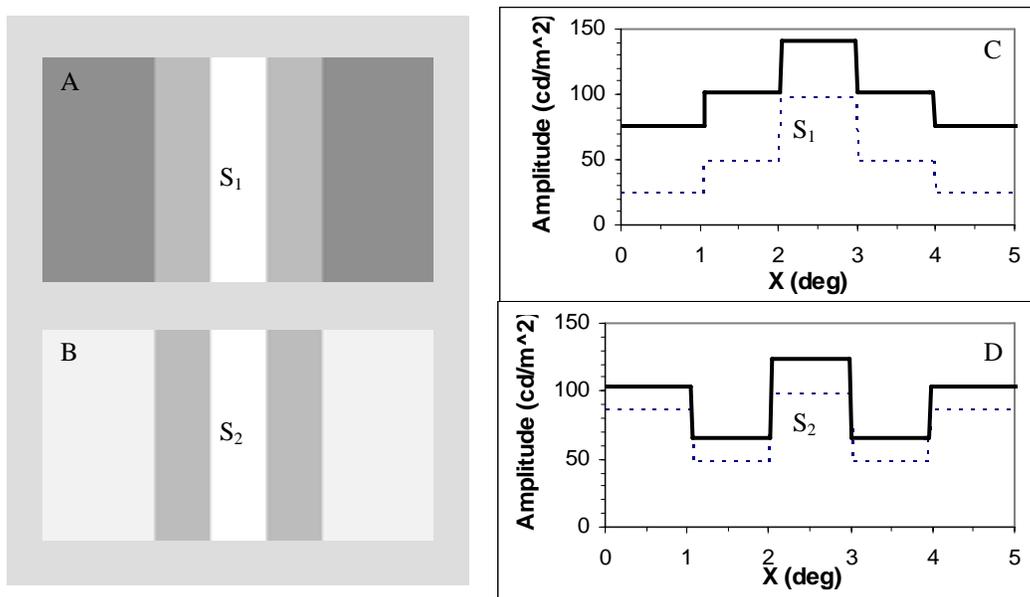
Figure 6. Left panel shows two patterns to demonstrate the effect of assimilation, where stripe S_1 looks brighter than S_2 while they have the same luminance value, and right panel shows the luminance profile of the patterns (dotted lines), and the corresponding model estimated brightness (thick lines) after a fill-in process.

Long Range Assimilation: The simultaneous contrast in the above example demonstrates the effect of local contrast on brightness perception. It has been shown in the literature that longer range interactions, other than local contrast, can also influence brightness perception as exampled by assimilation [52, 53]. Here, the perceived brightness is affected by the luminance level of non-adjacent background areas. The visual patterns on the left panel of Fig. 6 is a variant version of the bipartite field of Fig. 1 in Ref. 52. In this pattern, both stripes S_1 and S_2 have a same luminance of 97 cd/m^2, and their adjacent flanking stripes have a same luminance of 48 cd/m^2. The dotted lines of Fig. 6C and 6D show their luminance profiles. The percept of stripe S_1 being brighter than stripe S_2 cannot be explained by local contrast as there is no difference in local contrast. The only difference

between the two patterns is the luminance levels of the non-adjacent background fields, which are 25 cd/m^2 in A and 86 cd/m^2 in B. Such longer range effect was attributed to assimilation [52].

The model calculation follows the same way as described in the preceding example. Each of the luminance profiles of the patterns is fed into the model as an input to calculate its cortical representation. The simulated brightness following the fill-in process for pattern A is shown as the thick lines in Fig. 6C where stripe S_1 has a value of 141, and that for pattern B is shown as the thick lines in Fig. 6D where stripe S_2 has a value of 124. Therefore, the model predicts that stripe S_1 is perceived brighter than stripe S_2 by 17 units, which is consistent with our percepts in terms that stripe S_1 is likely perceived brighter than S_2.
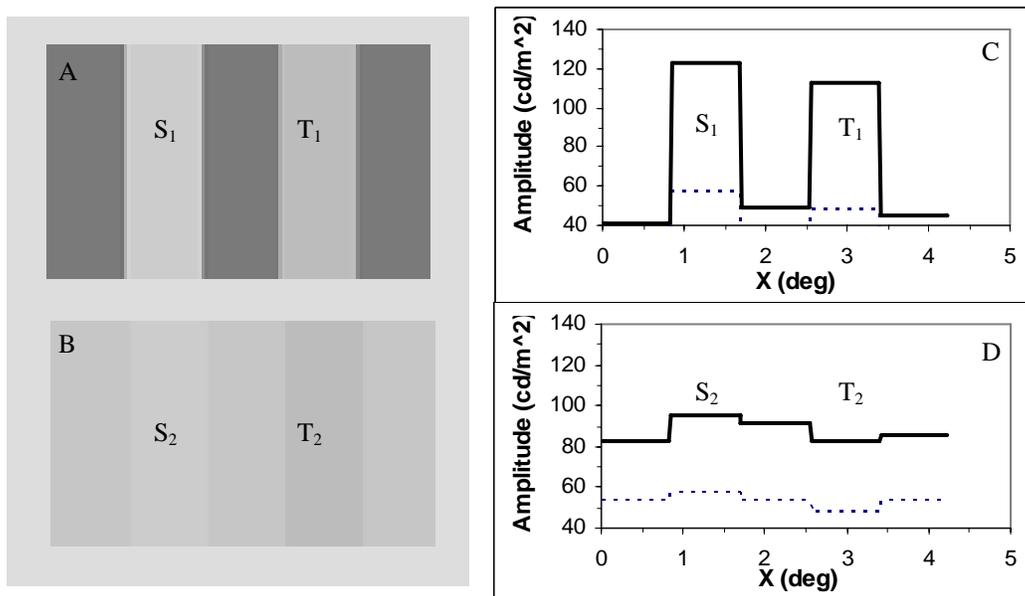
Figure 7. Stripes S_1 and S_2 have the same luminance of 57 cd/m^2; stripes T_1 and T_2 have the same luminance of 48 cd/m^2; and the background luminance is 17 cd/m^2 for pattern A and 54 cd/m^2 for pattern B. Model predicted brightness for stripes S_1 and T_1 are 123 and 112 (thick lines of panel C), with a difference of 11 units, and the predicted brightness for stripes S_2 and T_2 are 96 and 82 (thick lines of panel D), with a difference of 14 units.

Crispening Effect: Let's consider one more example here. It has been shown that the perceived brightness of a spot changes more rapidly with the luminance of the spot when its luminance is closer to the surrounding luminance [54]. Such crispening can also be demonstrated by seeing the effect of background luminance on the brightness difference of two spots (e.g., see Ref. 55). The perceived difference is the largest when the background luminance value is somewhere between the luminance values of the two spots. As illustrated in Fig. 7, the brightness difference between stripes S_1 and T_1 is barely detectable, while the difference between stripes S_2 and T_2 is easier to see, although S_1 and S_2 have the same luminance of 57 cd/m^2, and T_1 and T_2 have the same luminance of 48 cd/m^2. The dotted lines in Fig. 7C represent the luminance profile of Fig. 7A and the dotted lines of Fig. 7D represent the profile of Fig. 7B.

In the same way as in previous two examples, the luminance profile of each pattern is entered into the model algorithm to calculate its cortical representation, and then through a fill-in process. The thick lines of Fig. 7C represent the model predicted brightness for seeing pattern A,

and the thick lines of Fig. 7D represent the brightness for seeing pattern B. For a comparison, the model estimated brightness difference between S_1 and T_1, which is 11 units, is less than the difference between S_2 and T_2, which is 14 units. Thus, the model outputs are qualitatively consistent with the perceived brightness differences.

The three examples show that the current model can describe both the effects of local contrast and assimilation under a common theoretical framework. As a same algorithm and a same set of parameter values were used in each case, it is an encouraging evidence of showing the generality of the developed human vision model.

5. SUMMARY

Differing from most existing vision models, the current approach does not use CSF as the front-end filter in modeling visual image processing. Instead, the model simulates the CSF behavior at varying mean luminance by implementing implicit masking, using very basic components of visual image processing. They include a front-end low-pass filter, a nonlinear compressive process in the retina performed in the spatial domain, and a nonlinear compressive process in the Cortex performed in the frequency domain.

After including Minkowski summation in the decision stage, this model can describe the contrast thresholds obtained in two prominent and very different studies, namely the luminance dependent CSFs [16] and the Modelfest data [45, 48]. The residual RMS errors between the model and experimental data were about 0.1 log unit. It also suggests that further model improvement could be reached by applying more appropriate decision making roles such as adding linear frequency channels.

The same model can be used to identify the direction of visual illusion with respect to the change of perceived brightness in simultaneous contrast, assimilation, and crispening effect. While reports in the literature have shown that brightness perception can be simulated using the local-energy model of feature detection [56, 57], frequency channels [5, 58, 59], or natural scene statistics [60], the current approach relies on compressive nonlinear processes at both retina and visual cortex. Both Blakeslee et al. [59] and Dakin and Bex [60] use a frequency weight that increases with spatial frequency, in a way attenuating low frequency components. Similarly, the current model applies a concept of implicit masking to attenuate low frequency. The major differences here are that the amount of attenuation depends on the mean luminance level, and that frequency masking and spatially localized adaptation are included. It remains to see how important it is to apply these treatments in future studies. It is, nevertheless, encouraging to see the generality of the developed model, which integrates the three diverse perceptual phenomena under a common theoretical framework, in addition to its capability of estimating pattern visibility in a variety of conditions. In further studies, we need to concentrate on quantitative matches between the model predictions and experimental data on brightness perception.

6. ACKNOWLEDGEMENT

7. REFERENCES

1. H.R. Wilson, D.K. McFarlane, and G.C. Phillips, "Spatial frequency tuning of orientation selective units estimated by oblique masking," *Vis. Res.*, *23,* pp 873-882, 1983.
2. J.M. Foley, "Human luminance pattern-vision mechanisms: masking experiments require a new model," *J. Opt. Soc. Am. A*, *11,* pp 1710-1719, 1994.
3. A.B. Watson and J.A. Solomon, "A model of visual contrast gain control and pattern masking," *J. Opt. Soc. Am. A, 14,* pp 2378-2390, 1997.
4. E.G. Heinemann and S. Chase, "A quantitative model for simultaneous brightness induction," *Vis. Res., 35,* pp 2007-2020, 1995.
5. J. McCann, "Gestalt vision experiment from an image processing perspective," *Proc. 2001 PICS Conf.*, pp 9-14, 2001.
6. J. Yang, "Approaching a unified model of pattern detection and brightness perception," *Proc. SPIE*, *4662*, pp 84-95, 2002.
7. G.L. Fain and M.C. Cornwall, "Light and dark adaptation in vertebrate photoreceptors," In R. Shapley, and D.M-K Lam, *Contrast Sensitivity, MIT Press,* pp 3-32, 1993.
8. R. Shapley, E. Kaplan, and K. Purpura, "Contrast sensitivity and light adaptation in photoreceptors in the retinal network," In R. Shapley, R. and D.M-K. Lam, *Contrast Sensitivity, The MIT Press,* pp 103-116, 1993.
9. N.V.S. Graham, *Visual Pattern Analyzers*, Oxford University Press, 1989.
10. A.B. Watson, "Efficiency of a model human image code," *J. Opt. Soc. Am. A, 4,* pp 2401-2417, 1987.
11. E. Peli, "Contrast in complex images," *J. Opt. Soc. Am. A* 7, pp 2032-2040, 1990.
12. E. Peli, "Limitations of image enhancement for the visually impaired" *Optom. Vis. Sci.*, *69,* pp 15-24, 1992.
13. S. Daly, "The visible difference predictor: An algorithm for the assessment of image fidelity," *Human Vision, Visual processing, and Digital Display III, Proc. SPIE,* 1666, pp 2-15, 1992.
14. J. Lubin, "A visual discrimination model for imaging system design and evaluation," In Peli, E. (Ed). *Vision Models for Target Detection and Recognition*, World Science, pp 245-283, 1995.
15. O.H. Schade, "Optical and photoelectric analog of the eye," *J. Opt. Soc. Am., 46,* pp 721-739, 1956.
16. F. L. Van Nes and M.A. Bouman, "Spatial modulation transfer in the human eye," *J. Opt. Soc. A, 57*, pp 401-406, 1967.
17. F.W. Campbell and J.G. Robson, "Application of Fourier analysis to the visibility of gratings," *J. Physiol., London, 197*, pp 551-566, 1968.
18. B.A. Wandell, *Foundations of Vision,* Sinauer Associate, Inc. Publishers, 1995.
19. P.G.J. Barten, "Physical model for the contrast sensitivity of the human eye," *Human Vision, Visual Processing, and Digital Display IIII, Proc. SPIE*, 1666, pp 57-72, 1992.
20. P.G.J. Barten, *Contrast Sensitivity of the Human Eye and its Effects on Image Quality*, SPIE Optical Engineering Press, 1999.
21. J. Rovamo, J. Mustonen, and R. Näsänen, "Modelling contrast sensitivity as a function of retinal illuminance and grating area," *Vis. Res., 34*, pp 1301-1314, 1994.
22. J. Yang and W. Makous, "Spatiotemporal separability in contrast sensitivity," *Vis. Res., 34*, pp 2569-2576, 1994.
23. J. Yang and W. Makous, "Modeling pedestal experiments with amplitude instead of contrast," *Vis. Res.,* 35, pp 1979-1989, 1995.

24. J. Yang, X. Qi, and W. Makous, "Zero frequency masking and a model of contrast sensitivity," Vis. Res., 35, pp 1965-1978, 1995 .

25. W. L Makous, "Fourier models and the loci of adaptation," J. Opt. Soc. Am. A, 14, , pp 2323-2345, 1997.

26. J. Yang and W. Makous, "Implicit masking constrained by spatial inhomogeneities," Vis. Res., 37, pp1917-1927, 1997

27. J. Krauskopf and A. Reeves, "Measurement of the effect of photon noise on detection," Vis. Res., 20, pp 193-196., 1980.

28. A. Reeves, S. Wu, and J. Schirillo, "The effect of photon noise on the detection of white flashes," Vis. Res., 38, pp 2639-2647, 1997.

29. J. Yang and S.B. Stevenson, "Post-retinal processing of background luminance," Vis. Res., 39, pp 4045-4051, 1999.

30. J. Nachmias and R.V. Sansbury, "Grating contrast: discrimination may be better than detection," *Vis. Res., 14*, pp 1039-1042, 1974.

31. J.M. Foley and G.E. Legge, "Contrast detection and near-threshold discrimination in human vision," *Vis. Res., 21*, pp 1041-1053, 1981.

32. G.E. Legge and J.M. Foley, "Contrast masking in human vision," *J. Opt. Soc. Am.*, *70,* pp 1458-1471, 1980.

33. J. Ross and H.D. Speed, "Contrast adaptation and contrast masking in human vision," *Proc. R. Soc. London, Ser. B, 246*, pp 61-69, 1991.

34. D.J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual Neurosci., 9,* pp 181-197, 1992.

35. D.J. Heeger, "The representation of visual stimuli in primary visual cortex," *Curr. Dir. Psychol. Sci., 3,* pp 159-163, 1994.

36. F.W. Campbell, J.J. Kulikowski, and J. Levinson, "The effect of orientation on the visual resolution of gratings," *J. Physiol. London, 187,* pp 427-436, 1966.

37. J. Yang and S.B. Stevenson, "Effect of background components on spatial frequency masking," *J. Opt. Soc. Am. A, 15,* pp 1027-1035, 1998.

38. R.W. Rodieck, *The Vertebrate Retina,* W. H. Freeman and Company.

39. D. I. A. MacLeod, D. R. Williams, and W. Makous, "A visual nonlinearity fed by single cones," *Vis. Res.,* 32, pp 347-363, 1992.

40. S. He and D. I. A. MacLeod, "Contrast-modulation flicker: dynamics and spatial resolution of the light adaptation process," *Vis. Res.,* 38, pp 985-1000, 1998.

41. R.M. Boynton and D.N. Whitten, "Visual adaptation in monkey cones: recordings of late receptor potentials," *Science*, *170,* pp 1423-1425, 1970.

42. J.E. Dowling, *The retina: an approachable part of the brain*. The Belknap Press of Harvard University Press, 1987.

43. R. Shapley and P. Lennie, "Spatial frequency analysis in the visual system," *Annu. Rev. Neurosci., 8*, pp 547-583, 1985.

44. R.L. De Valois and K.K. De Valois, *Spatial Vision*. New York: Oxford University Press, 1988.

45. A. B. Watson and A. J. Ahumada, "A standard model for foveal detection of spatial contrast," *J. of Vision, 5*, pp 717-740, 2005.

46. W. S. Geisler, "Sequential ideal-observer analysis of visual discriminations," *Psychological Review 96*, pp. 267–314, 1989.

47. M. P. Eckstein, C.K. Abbey, and F.O. Bochud, "A practical guide to model observers for visual detection in synthetic and natural noise images," in J. Beutel, H. L. Kundel, and R. L. Van

Metter (Eds.) *The Handbook of Medical Imaging: Volume 1, Progress in Medical Physics and Psychophysics,* SPIE press, pp. 593-628, 2000.

48. T. Carney, S. A. Klein, C. W. Tyler, A. D. Silverstein, B. Beutter, D. Levi, A. B. Watson, A. J. Reeves, A. M. Norcia, C. C. Chen, W. Makous, and M. P. Eckstein, "The development of an image/threshold database for designing and testing human vision models," *Proc. SPIE, 3644*, pp 542-551, 1999.

49. E. Hering, Outlines of a Theory of *the Light Sense*. Harvard University Press. Cambridge, Massachusetts, 1964.

50. L.E. Arend and R. Goldstein, "Lightness models, gradient illusions, and curl," *Percept. Psychophys.*, *42,* pp 65-80, 1987.

51. S. Grossberg and D. Todorovic, "Neural dynamics of 1-D and 2-D brightness perception: a unified model of classical and recent phenomena," *Percept. Psychophys.*, *43,* pp 241-277, 1988.

52. R. Shapley and R. C. Reid, "Contrast and assimilation in the perception of brightness," *Proc. Natl. Acad. Sci. USA, 82*, pp 5983-5986, 1985.

53. S. S. Shimozaki, M. P. Eckstein, and C. K. Abbey, "Spatial profiles of local and nonlocal effects upon contrast detection/discrimination from classification images," *J. of Vision, 5*, pp 45-57, 2005.

54. H. Takasaki, "Lightness change of grays induced by change in reflectance of gray background," *J. Opt. Soc. Am. 56,* pp 504-509, 1966.

55. M.D. Fairchild, *Color Appearance Models*. Reading, MA, Addison-Wesley, 1998.

56. M. C. Morrone and D. C. Burr, "Feature detection in human vision: a phase-dependent energy mode," Proc. R. Soc. Lond B 235, pp 221-245, 1988.

57. M. C. Morrone, D. C. Burr, and J. Ross, "Illusory brightness step in the Chevreul illusion," Vision Res. 34, pp 1267-1574,1994.

58. J. A. McArthur and B. Moulden, "A two-dimensional model of brightness perception based on spatial filtering consistent with retinal processing," *Vision Res. 39*, pp 1199-219, 1999.

59. B. Blakeslee, W. Pasieka, and M. E. McCourt, "Oriented multiscale spatial filtering and contrast normalization: a parsimonious model of brightness induction in a continuum of stimuli including White, Howe and simultaneous brightness contrast," *Vision Res. 45*, pp 607-615, 2005.

60. S. C. Dakin and P. J. Bex, "Natural image statistics mediate brightness 'filling in'," Proc. R. Soc. Lond. B, 270, pp 2341-2348, 2003.