Perceptual distortion measured with a gain control model of LGN response

Alexander Berardino, Valero Laparra, Johannes Ballé, Eero P. Simoncelli

Models for perceptual image distortion are generally designed to capture properties of human vision, such as the spatial frequency dependence of contrast sensitivity, and the masking effects of superimposed oriented patterns. However, the widely-used Structural Similarity (SSIM) index (Wang et al., 2004), is based on a unique construction devised to disregard changes in local luminance or contrast, while emphasizing changes in local structure. Since these properties are evident in the responses of early visual neurons, we wondered whether a more explicit model of physiological responses might provide a more suitable substrate for constructing a distortion measure. We built a functional model of early spatial visual processing (LGN-GC) that incorporates known properties of retina and LGN. Similar to the model of Mante et al., 2008, the model includes bandpass linear filtering, rectification, and local luminance and contrast gain controls. The distortion between an original and corrupted image is determined by passing each through the model, and measuring the Euclidean distance between the two response vectors. The model parameters (filter sizes and amplitudes) were fit to a database of human perceptual quality judgments (TID2008 - Ponomarenko et al., 2009). We find that the fitted parameters are consistent with measured physiological properties of LGN neurons in the macaque monkey. Moreover, LGN-GC outperforms SSIM, and performs comparably to multi-scale SSIM (MS-SSIM) at predicting perceptual distortions (despite its restriction to a single spatial scale), explaining more of the (cross-validated) variance in the human data (SSIM: 60%, MS-SSIM: 64%, LGN-GC: 67%). Finally, we performed a direct comparison of LGN-GC to SSIM by examining stimuli optimized to differentiate them (known as MAD competition; Wang et al., 2008). This comparison provides evidence that distance as measured by our physiologically-inspired model corresponds more closely with human perception than SSIM.

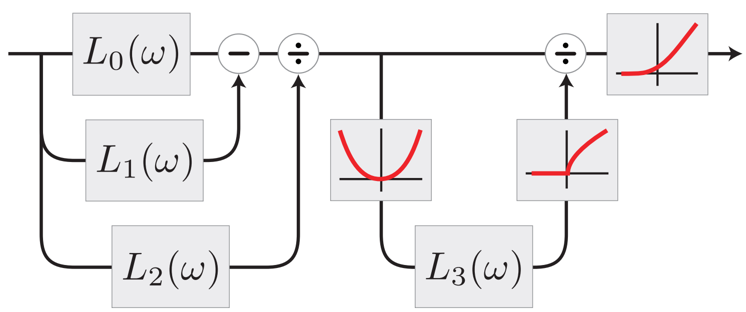

Figure 1: LGN-GC model architecture. The image is bandpass filtered ($L_0 * X - L_1 * X$) and is divided by the local average luminance ($L_2 * X$). These responses are then further normalized by a measure of local average contrast, computed by squaring, convolving with lowpass filter L_3, and taking the square root. A rectifying nonlinearity converts the resulting normalized responses into firing rates.

LGN model with local gain control(s)

The model (LGN-GC) is constructed using a cascade of static computational elements (Fig. 1), roughly based on the dynamic model of Mante et al., 2008 [1]. The two-dimensional image (X) is first convolved with a center-surround filter composed of a difference of Gaussians ($(L_0 - L_1) * X$). This filtered output is divided by a measure of the local luminance, computed by convolving with another lowpass filter, ($L_2 * X$):

$$Y = \frac{(L_0 - L_1) * X}{1 + L_2 * X}, \tag{1}$$

where L_0, L_1 and L_2 are bivariate symmetric Gaussians parameterized by their standard deviations, $\sigma_0, \sigma_1, \sigma_2$, constrained so that $\sigma_2 \geq \sigma_1$ and $\sigma_1 \geq 1.15\,\sigma_0$. These computations mimic operations carried out in the retina. In the second stage, the responses of the retinal stage are normalized by a measure of local contrast computed from their squared and blurred neighbors. Blurring is implemented with another Gaussian lowpass filter, L_3, with standard deviation $\sigma_3 \geq \sigma_0$. Although implemented as a single stage, the properties arising from this computation are first found in the retina, and are further enhanced in the LGN. Finally, the model response (firing rate) is obtained by applying a static nonlinearity:

$$R = \log\left(1 + \exp\left(\frac{Y}{1 + \sqrt{L_3 * Y^2}}\right)\right). \tag{2}$$

The LGN-GC model can be used as a distortion metric between two images by computing the Euclidean distance between the model's responses to each image. We optimized the parameters of the model to maximize the correlation of its predicted distances with the reported perceptual distances for half of the TID2008 image pairs [2]. Specifically, we fit the size of the center and surround filters, and the sizes and amplitudes of the filters used to gather the luminance and contrast normalization signals.

Correlation with perceptual distortion databases

We compared model-predicted perceptual distances to human quality ratings from several different databases, including cross-validated comparisons to a holdout set of the TID2008 database used for fitting (Fig. 2). We found that LGN-GC achieved a remarkable correlation with perceptual distance reports, consistently outperforming the *de facto* standard (SSIM) and performing comparably to the multiscale version (MS-SSIM), even though it is handicapped by operating at a single scale. Additional cascaded stages that mimic properties found in later visual areas (orientation tuning, multiple scales), as well as replacement of the Euclidean norm by a metric more appropriate for spiking responses, may further improve performance.

	TID 2008		CSIQ		TID 2013	
	ρ_p	ρ_s	ρ_p	ρ_s	ρ_p	ρ_s
SSIM	0.74	0.77	0.79	0.85	0.78	0.82
MS-SSIM	0.79	0.84	0.77	0.89	0.80	**0.87**
LGN-GC	**0.82**	**0.85**	**0.88**	**0.93**	**0.85**	0.85

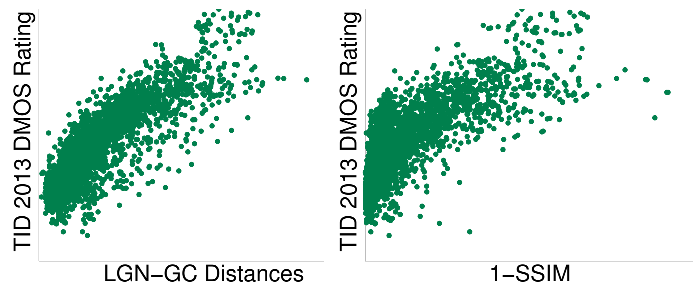

Figure 2: Left: Correlations (Pearson -ρ_p- and Spearman -ρ_s-) for different databases. Right: predictions of the LGN-GC model (left) and SSIM (right) against the differential mean opinion scores from TID2013.

MAD competition between LGN-GC and SSIM

Finally, we compared the perceptual relevance of distances reported by LGN-GC and SSIM using the "MAD competition" methodology [3]. Visual comparisons suggest that distortions captured by LGN-GC more closely match human perception than those of SSIM (Fig. 3).

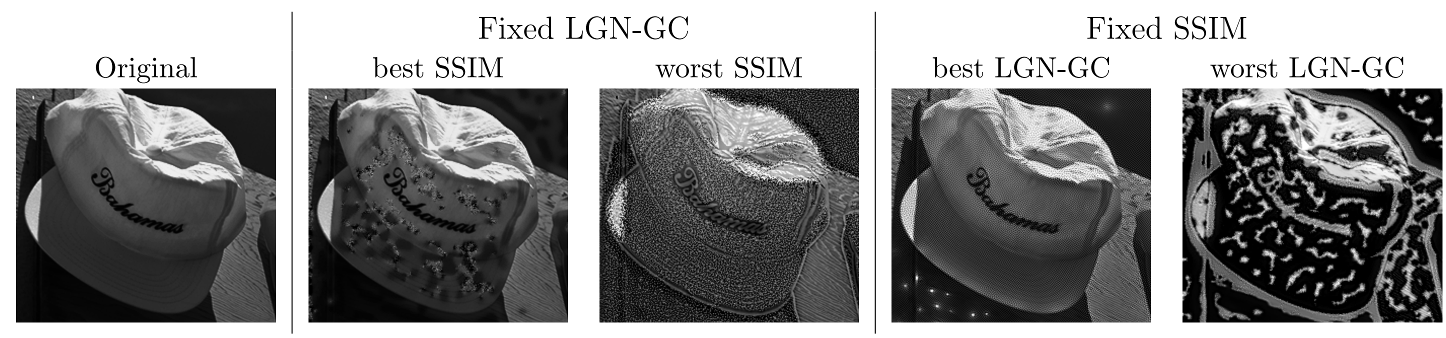

Figure 3: Comparison of SSIM and LGN-GC using MAD competition. Left: Original image. Middle: Maximally and minimally distorted images according to SSIM, with equal distortions according to LGN-GC. Right: Maximally and minimally distorted images according to LGN-GC, with equal distortions according to SSIM. Middle two images exhibit approximately equal perceptual quality (in agreement with LGN-GC), whereas the right two are noticeably different (in disagreement with SSIM).

References

[1] Valerio Mante, Vincent Bonin, and Matteo Carandini. Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli. *Neuron*, 58(4):625–638, May 2008.

[2] N Ponomarenko, V Lukin, and A Zelensky. TID2008 – a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10:30–45, 2009.

[3] Zhou Wang and Eero P Simoncelli. Maximum differentiation (MAD) competition: a methodology for comparing computational models of perceptual quantities. *Journal of Vision*, 8(12):8.1–13, 2008.