Spatial-Temporal Visible Contrast Energy Predictions of Detection Thresholds

Albert J. Ahumada, Andrew B. Watson, and Jihyun Yeonan-Kim

Ahumada and Watson (2013) used visible contrast energy to predict vision thresholds for visual images . For low contrast images, their local luminance based visible contrast image $v(x, y)$ can be computed from the contrast image $c(x, y)$ using an "optical" low pass filter $M_0(f_x, f_y)$ and an "inhibitory surround" low pass filter $M_1(f_x, f_y)$,

$$v(x, y) = FFT^{-1}(M_0(f_x, f_y) (1 - a M_1(f_x, f_y)) FFT(c(x, y)), \quad (1)$$

Barten (1994) added temporal low pass filters $H_0(f_t)$ and $H_1(f_t)$ to form a visible contrast "movie" $v(x, y, t)$ as

$$v(x, y, t) = FFT^{-1}(M_0(f_x, f_y) H_0(f_t) (1 - M_1(f_x, f_y) H_1(f_t)) FFT(c(x, y, t)), \quad (2)$$

If $v(x, y, t)$ is masked by white noise, the detection performance of an ideal observer is a function only of the noise level and the signal energy, $E_V = \int\int\int v(x, y, z)^2\, dx\, dy\, dt = \|v(x, y, z)\|$. Letting $V(f_x, f_y, f_t)$=FFT(v(x, y, t), the final inverse need not be computed since $E_V = \|V(f_x, f_y, f_t)\| = \int\int\int |V(f_x, f_y, f_t)|^2\, df_x\, df_y\, df_t$ if the FFT is appropriately normalized.

$$E_V = \|M_0(f_x, f_y) H_0(f_t) (1 - M_1(f_x, f_y) H_1(f_t)) C(f_x, f_y, f_t) \|, \quad (3)$$

where $C(f_x, f_y, f_t)$=FFT(c(x, y, t). When the image sequence is space-time separable, $C(f_x, f_y, f_t) = C_{XY}(f_x, f_y)\, C_T(f_t)$, and, dropping the f's for clarity,

$$E_V = \|(M_0 C_{XY}) (H_0 C_T) - (M_0 M_1 C_{XY}) (H_0 H_1 C_T)\|, \quad (4)$$

When $\|M_1 C_{XY}\| \ll 1$,the inhibitory response is negligible, and

$$E_V = \|M_0 C_{XY}\|\, \|H_0 C_T\|. \quad (5)$$

Using Gaussian spatial filters $M=\exp(-(f_x^2 + f_y^2)/f^2)$and Gamma temporal filters, $H=1/(1+i2\pi\tau f_t)^n$, we predicted the contrast energy threshold data of Carney et al. (2013) for

$c_{XY}(x,y) = \exp(-(x^2 + y^2)/(2\,(0.5\text{ deg})^2) \cos(2\pi f_Y y)$, $f_Y = 0, 4, 11.3$;

$c_T(t) = c_0(t) = \exp(-t^2/(2\,(0.25\text{ sec})^2)$ and

$c_T(t) = c_0(t) \sin(2\pi f_T t)$, $f_T = 1, 2, 4, 8, 15, 25$ Hz.

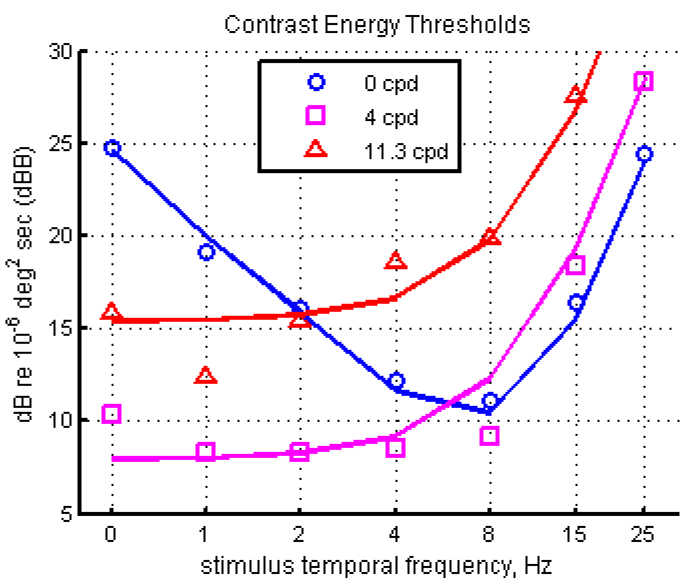

Figure 1. Data from Carney et al. (2013) [symbols] and model predictions [lines].

The model parameters were $[f_0, f_1]$ = [11.4, 0.88] cpd, $[\tau_0, \tau_1]$ = [12.5, 12.5] msec, $[n_0, n_1]$ = [3, 2]. The ideal observer noise spectral density estimate is 4.3 dBB. The RMS model fit is 1.6 dB with 20 - 6 df (n_1 was not allowed to vary). Additivity in log sensitivity is predicted to hold for f_Y = 4 and 11.3. The RMS error for 4 and 11.3 cpd additivity is 1.9 dB with 5 df.

References

Ahumada, A. J., Watson, A. B. (2013) Visible contrast energy metrics for detection and discrimination, Proc. SPIE 8651.

Barten, P. G. J. (1994) Spatio-temporal model for the contrast sensitivity of the human eye, in F. Engel, H. de Ritter, F. Blommaert, eds., Dynamic Properties of Vision VII, Eindhoven, Institute for Perception Research.

Carney, T., Mozaffari, S., Sun, S., Johnson, R., Shrivastava, S., Shen, P., Ly, E. (2013) Initial spatio-temporal domain expansion of the Modelfest database. HVEI XVIII, B. E. Rogowitz, T. N. Pappas, H. de Ridder, eds., SPIE Proc. Vol. 8651.

Acknowledgment

Supported by NASA Space Human Factors Research Project WBS 466199 and the NASA Engineering and Safety Center.