https://doi.org/10.2352/ISSN.2470-1173.2018.14.HVEI-506

A Dual Channel Spatial-Temporal Detection Model

Albert Ahumada; 1724 Luxton Street, Seaside, CA/USA Jihyun Yeonan-Kim; San Jose State University Foundation; NASA Ames Research Center, Moffett Field, CA/USA Andrew B. Watson; 56 Cleland Avenue, Los Gatos, CA/USA

Abstract

A dual channel spatial-temporal model for contrast detection is shown to improve the predictions for the spatial-temporal Modelfest data over those of the simplified Barten model.

Introduction

Carney, Mozaffari, Sun, Johnson, Shrivastava, Shen, and Ly [1] have recently reported contrast thresholds for spatial-temporal stimuli originally proposed by Carney, Klein, Beutter, Norcia, Chen, Tyler, Makous, Watson, Cropper, Popple, Robertson, Manahilov, Simpson, and Wenzel [2] to extend the spatial Modelfest data [3] into the temporal domain.

Barten[4] developed a model for the detection of arbitrary spatial temporal contrast stimuli based on an ideal observer limited only by photon noise, neural noise, and linear filters representing visual processing. When the mean luminance is high enough so that the photon noise can be ignored, Barten's model predicts that the threshold for detection depends only on the energy of the stimulus passed by a linear filter. This simplified Barten model is a spatial-temporal extension of the spatial energy detection model of Ahumada and Watson.[5] Ahumada, Watson and Yeonan-Kim[6] showed that a version of this spatial-temporal filter model provides a good account of the spatial-temporal contrast sensitivity data reported by Carney et al.[1]

This paper describes the simplified Barten model of Ahumada, Watson, and Yeonan-Kim[6] and a dual channel version of that model. The low spatial frequency P ("parvo") channel has parameters consistent with the model developed by Watson and Ahumada[4] for the original spatial Modelfest data [5]. The M ('magno") channel has higher temporal frequency information at a lower spatial resolution.

The Temporal Contrast Sensitivity Functions

Carney et al.[1] reported contrast thresholds for contrast stimuli that were the product of spatial and temporal functions

$$c = c_{X, Y, T}(x, y, t) = c_{X, Y}(x, y)c_T(t). \quad (1)$$

For their temporal contrast sensitivity functions (TCSFs) shown in Figure 1, the spatial functions are cosine phase horizontal Gabor functions with an envelope standard deviation $\sigma_{XY} = 0.5$ deg,

$$c_{X, Y}(x, y) = \exp(-0.5(x^2 + y^2)/ \sigma_{X, Y}^2) \times \cos(2 \pi f_Y y), f_Y = 0, 4, 11.3. \quad (2)$$

The temporal functions are sine phase Gabors with a Gaussian envelope $c_0(t)$ with standard deviation $\sigma_T = 0.25$ sec,

$$c_0(t) = \exp(-0.5 (t / \sigma_T)^2), \quad (3)$$

and

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

The temporal envelope itself will be referred to as the 0 Hz temporal waveform.

Figure 1 shows the TCSF data reported by Carney et al. [1] The TCSF curve for 0 cpd, the spatial Gaussian, is band-pass with a peak near 8 Hz, while the spatial Gabor TCSF for 4 cpd and 11.3 cpd are low-pass. Carney et al.[1] remarked that these two low-pass functions are "expected" to be parallel, but the difference between them appears to increase with temporal frequency.

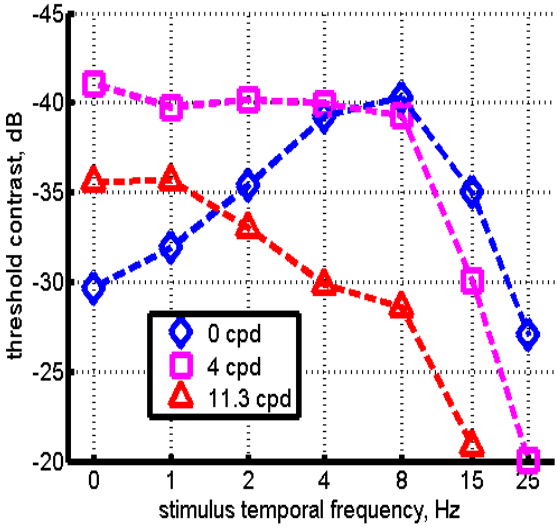

Figure 1. Temporal contrast sensitivity data from Carney et al. [1] Data are 20 times log to the base 10 of the contrast thresholds plotted on an inverted scale. Zero spatial and temporal frequencies indicate the Gaussian envelope alone, which had a standard deviation of 0.25 sec.

The contrast thresholds of Figure 1 are computed from the peak contrast of the envelopes. For sine phase Gabors this contrast level is not attained by the stimulus. The higher sensitivity in Figure 1 for the 0 Hz 4 cpd stimulus relative to

the 1 Hz 4 cpd stimulus is a result of this measurement strategy. Here the thresholds will be reported as contrast energy thresholds in dBB, dB re 10^{-6} deg^2 sec.[7, 8]

The single-channel simplified Barten model

The input for the model is a spatial-temporal contrast image sequence, c = c(x, y, t), represented in the frequency domain as

$$C = C(f_x, f_y, f_t) = FFT(c) \quad (5)$$

The visible contrast image sequence is computed from the input using two low pass spatial filters $S_0 = S_0(f_x, f_y)$ and $S_1 = S_1(f_x, f_y)$ and two low pass temporal filters $T_0 = T_0(f_t)$ and $T_1 = T_1(f_t)$. The visible contrast image in the frequency domain is given by

$$V = S_0\, T_0\, (1 - S_1\, T_1\,)\, C \quad (6)$$

S_0 and T_0 determine the high frequency limits of visibility while S_1 and T_1 are the integrators for the inhibitory surround. The detection threshold is be assumed to be inversely proportional to the visible contrast energy

$$E = \|V\|^2. \quad (7)$$

The spatial filters are Gaussian

$$Sj = \exp(-(f_x^2 + f_y^2)/fj^2), \; j = 0, 1 \quad (8)$$

and the temporal filters are Gamma

$$Tj = 1/(1+i2\pi\, \tau j\, f_t)^{nj}, \; i = 0, 1 \quad (9)$$

Figure 2 shows the contrast energy thresholds for these stimuli and the simplified Barten model predictions.

The least squares estimates of the model parameters are shown in Table 1. The RMS error corrected for the 6 parameters estimated is 1.50 dB. The contrast gain factor used to bring them into vertical alignment is 420 (52.4 dB).

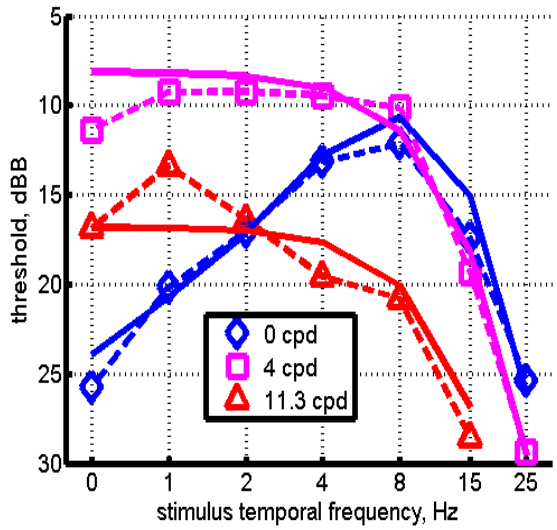

Figure 2. Temporal contrast energy sensitivity data from Carney et al. [1] [symbols and dashed lines] and the simplified Barten model predictions [solid lines].

Table 1: Parameter estimates for the simplified Barten model filters. Underlined parameter values were not estimated.

Temporal	tau, msec	n
T_0	12.1	3
T_1	13.8	2
Spatial	f, cpd	
S_0	11.1	
S_1	1.12	

With the spatial frequency cutoffs so far apart, the model predicts that the inhibitory surround is negligible for both the 4 and 11.3 cpd stimuli. These two contrast energy temporal contrast energy sensitivity curves are predicted to be parallel with the shape of T_0, separated by 20 log (S_0(4 cpd)/S_0(11.3 cpd). The parameter estimate of f_0 = 11.4 predicts a difference that is too large at the low temporal frequencies and too small at the highs. The parameter value that best fits the difference at the two lowest temporal frequencies is f_0 = 14.2 cpd, close to the value of 15.2 cpd reported by Watson and Ahumada[8]. The difference between these two data functions at 4, 8, and 15 Hz corresponds to f_0 = 9.8 cpd. The dual channel model was developed to allow a larger difference between these functions at higher temporal frequencies by allowing a different spatial frequency response at the higher temporal frequencies.

Another feature of these two data curves is that they appear to be band-pass, while the assumed form of T0 is low-pass. A high-pass filter was added to the dual channel model to account for this effect.

The dual channel model

The dual channel model has two channels that are both Barten models. Each channel has its own excitatory temporal response

$$T_{0,P} = H\, P\, T_0, \quad (9)$$

and

$$T_{0,M} = H\, M\, T_0, \quad (10)$$

where M = 1 – P. Figure 3 shows the components of the excitatory temporal responses: H, P, M, and T_0. and Figure 4 shows the resulting combinations, $T_{0,P}$ and $T_{0,M}$. Each channel has its own excitatory spatial response, $S_{0,P}$ and $S_{0,M}$. Arbitrarily keeping the inhibitory integrators the same for both channels, the visible contrast images in the frequency domain are given by

$$V_P = S_{0,P}\, T_{0,P}\, (1 - S_1\, T_1\,)\, C, \quad (11)$$

and

$$V_M = S_{0,M}\, T_{0,M}\, (1 - S_1\, T_1\,)\, C. \quad (12)$$

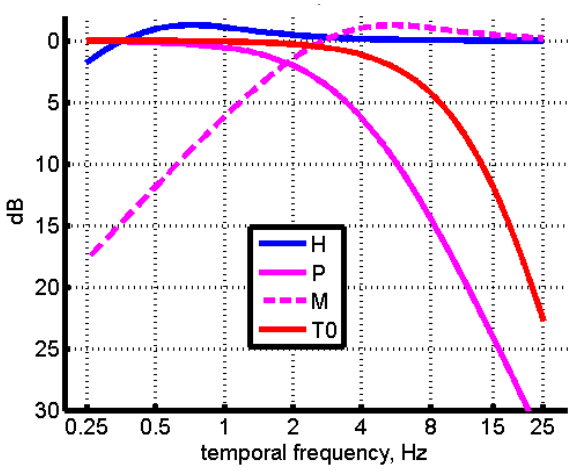

Figure 3. Amplitude responses of the components of the excitatory temporal filters. H is high pass filter with a low frequency cutoff that makes the overall response band pass. T_0 determines the high frequency falloff of the M channel. P and M = 1 – P segregate the two channels.

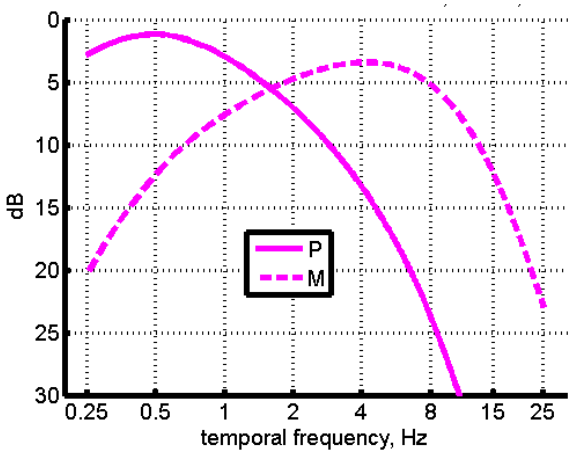

Figure 4. Amplitude responses combining the components of the excitatory temporal filters. The P channel excitatory temporal response is H P T_0. The M channel excitatory temporal response is H M T_0.

The temporal response low pass filters 1 – H, P, T_0, and T_1 are all assumed to be Gamma filters and the spatial low-pass filters $S_{0,P}$, $S_{0,M}$, and S_1 are assumed to be Gaussians. The detection threshold is be assumed to be inversely proportional to the total visible contrast energy in both channels

$$E = k\,||V_P||^2 + ||V_M||^2, \qquad (13)$$

where k is the relative weight of the P channel energy with respect to the M channel energy. Figure 5 shows the predictions of the dual channel model for the temporal contrast energy function data. The model captures both the band-pass nature of the 4 and 11.3 cpd curves and the increased spacing between these curves at the higher temporal frequencies.

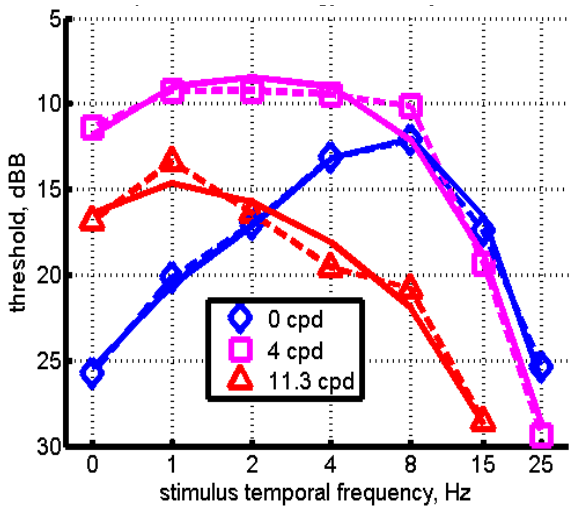

Figure 3. Temporal contrast energy sensitivity data from Carney et al. [1] [symbols and dashed lines] and dual channel model predictions [solid lines].

Table 2 shows the least squares parameter estimates.

Table 2: Parameter estimates for the dual channel filters. Underlined parameter values were not estimated.

Temporal	τ, msec	n
1 – H	310	2
P	53.8	2
T_0	9	4
T_1	11	2
Spatial	f, cpd	
$S_{0,P}$	15.6	
$S_{0,M}$	9.9	
S_1	0.81	

Ten parameters were estimated by least squares, 8 parameters whose values are shown in Table 2, a k value of 0.60 and a contrast gain factor of 382 (51.6 dB). The n parameters of the L, B, and T_1 filters were arbitrarily set to 2. The RMS fit corrected for the 10 estimated parameters was 1.1144 dB.

We evaluate the significance of this improvement using the F test for whether the more particular (nested) hypotheses H_1 fits as well as the more general hypothesis H_0,

$$F(df_1\text{-}df_0, df_0) = (SS_1 - SS_0)/(df_1 - df_0)) / (SS_0/df_0),$$

where SSj is the sum of the squared errors of prediction for hypothesis j and dfj is the number of data points minus the number of parameters estimated.
The corrected RMS error for hypothesis j is given by

$e_j = \sqrt{(SS_j / df_j)}$,

so

$F(df_1\text{-}df_0, df_0) = ((df_1 e_1 2 – df0 e_0 2)/(df_1 – df_0)) / e_0 2$,

Let hypothesis B be the simplified Barten model and hypothesis A be the simplified Barten model with the adaptation filter H. The corrected RMS errors are $e_B = 1.5031$ dB and $e_A = 1.3582$ dB. The degrees of freedom are $df_B = 14$ and $df_A = 13$, so $F(1,13) = 4.15$, $p = 0.06$. Let hypothesis C be the channel model with $e_C = 1.1144$ dB. If we compare the channel model with the Barten model with adaptation we obtain an $F(3,10) = 3.10$, $p = 0.11$. Finally, comparing the channel model with the Barten model, we get $F(4,10) = 3.87$, $p = 0.04$. We conclude that the channel model is a significant improvement over the Barten model and that it is likely that both the adaptation filter and the channels would contribute significantly to the improvement if more data were available.

Discussion

At the meeting the question arose as to how this model could be extended. One dimension is stimulus spatial extent and eccentricity. The simple approach of Watson and Ahumada[9] is to consider that the main effect of fewer samples in the periphery is to increase the neural noise level , which can be approximated by assuming the noise is constant and reducing the signal level appropriately. Extension to color could be done by adding opponent color channels as suggested by Wuerger, Watson, and Ahumada. [10] These additional opponent color channels should not need an M channel.

References

1. T. Carney, S. Mozaffari, S. Sun, R. Johnson, S. Shrivastava, P. Shen, E. Ly, "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, 2013.

2. T. Carney, S. Klein, B. Beutter, A. Norcia, C. Chen, C. W. Tyler, W. Makous, A. Watson, S. J. Cropper, A. V. Popple, K. Robertson, V. Manahilov, B. Simpson, K. Wenzel, "Extending the Modelfest image/threshold database into the spatio-temporal domain," Proc. SPIE 4662, 138-148, 2002.

3. Modelfest, http://visionscience.com/data/modelfest/index.html, accessed 9/8/2017.

4. P. G. J. Barten, 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, 1994 .

5. A. J. Ahumada,A. B. Watson, " Visible contrast energy metrics for detection and discrimination, "Proc. SPIE 8651, Human Vision and Electronic Imaging XVIII, 86510D (March 14, 2013); doi:10.1117/12.2009383; **http://dx.doi.org/10.1117/12.2009383**

6. A. Ahumada, A. B. Watson, J. Yeonan-Kim , "Spatial-Temporal Visible Contrast Energy Predictions of Detection Thresholds," in Computational and Mathematical Models in Vision Conference, St. Pete's Beach, FL, 2017.

7. A. B. Watson, J. Robson, H. Barlow, "What does the eye see best," Nature, vol. 302, pp. 419-422 , 1983.

8. A. B. Watson, A. Ahumada, "A standard spatial observer for contrast detection," J. Vis., vol. 5, pp.717-740, 2005.

9. A. B. Watson, A. Ahumada, "Letter identification and the neural image classifier," J. Vis., vol. 15, no. 2, pp.1-26, 2015.

10. S. Wuerger, A.J. Ahumada, A. B. Watson, "Toward a standard observer for spatio-chromatic detection," Proc. SPIE 4662, 159-172, 2002.

Author Biographies

Al Ahumada received his BA in mathematics from Stanford University (1961) and his PhD in experimental psychology from UCLA (1967). He was an assistant professor of psychology at UC Irvine (1967-75) and a researcher at Stanford (1975-80). He was a research psychologist at NASA Ames Research Center (1980-2017). His work centers on vision modeling.

Jihyun Yeonan-Kim completed B.A. in English Literature (2007) and M.S. in Cognitive Sciences (2009) at Yonsei University, and M.S. in Electrical Engineering (2013) and Ph.D. in Psychological Sciences (2013) at Purdue University. She is currently a research associate at San Jose State University Foundation at NASA Ames Research Center. She received traditional training in psychophysics and neuroscience and now develops the biophysical models of retinal and early cortical circuitry for both theoretic and applicational purposes.

Andrew Watson received his PhD in Psychology from University of Pennsylvania (1976) and did postdoctoral work at University of Cambridge, UK. From 1982 to 2016 he worked at NASA Ames Research Center in California. He is the founder and current Editor-in-Chief of the Journal of Vision. He is a Fellow of ARVO, OSA, and SID, and is on the organizing committee of HVEI.