

Masking in color images

Albert J. Ahumada, Jr.^{*a}, William K. Krebs^{**b}

^aNASA Ames Research Center; ^bFAA

ABSTRACT

Masking of color targets was measured for fixed pattern noises made of all additive combinations of white/black, red/green, and blue/yellow noise. Results are compared with the predictions of a cone-contrast-based masking model with and without cross-channel masking. The model without cross-channel masking performed very well.

Keywords: color vision, target detection, image discrimination, vision models, image quality metrics, visual masking, contrast energy, color channels, color representations

1. INTRODUCTION

This study is part of a project to develop image discrimination models that can predict the detectability of targets in color images. Image discrimination models take as input two images and predict the number of just-noticeable differences (JNDs) between the images. If one image is a background image and the other image has the identical background plus a target image, the model predicts the detectability of the target in the background.

Previous work with luminance targets in complex imagery has shown that target detectability is strongly influenced by variations in background contrast.¹⁻³ A simple model for masking that was used with some success was to adjust the detectability prediction by a contrast gain control factor based on the RMS visible contrast.^{1,4-5} The theoretical goal here was to develop a color version of this simple model with three color channels and estimate parameters representing the masking effects within and between these channels. The experimental goal is to measure the masking of targets by fixed-pattern chromatic noise. These measurements were designed to simplify the calibration of the contrast masking components of simple color image discrimination models.

2. METHODS

2.1 Observers

The 16 observers were students at the Naval Postgraduate School and received course credit for their participation. They had passed standard tests for acuity and color vision.

2.2 Stimuli

The stimuli were presented using the red, green, and blue guns of a color CRT. The gun responses were made linear by lookup tables feeding 10 bit DACs. At full output, the individual guns had the luminance (Y) and CIE color chromaticities (x, y) shown in Table 1.

	Y	x	y
Red	18.5	0.616	0.337
Green	70.7	0.293	0.604
Blue	10.8	0.151	0.070

Table 1. Luminance (cd/m²) and chromaticity values for the three CRT guns.

* ahumada@mail.arc.nasa.gov; phone 1 650 604 6257; fax 1 650 604 0255; <http://vision.arc.nasa.gov/~al/ahumada.html>; NASA Ames Research Center, Mail Stop 262-2, Moffett Field CA 94035-1000; ** William.Krebs@faa.gov; phone 1 202 267 8758; fax 1 202 267 5968; FAA, 800 Independence Ave S.W., AAR-100 Attn: Dr Krebs, Washington D.C. 20591

The screen background was set to a neutral gray with equal contributions (128/255) from the three guns. Calibrations were done with a Minolta CS-100. The screen was viewed from a distance of 1 m, giving 52.7 pixels per degree of visual angle in the horizontal direction and 37.6 pixels per degree vertically. The screen was 19.4 by 13.6 deg (1024 by 512 pixels).

The fixed pattern noise images were 2.62 by 2.66 deg (138 by 100 pixels). They were generated as 69 by 50 pixel images and then pixel replicated. Three single color masks were generated from three samples of binary noise, using the contrast values of Table 1 for a high sample and the opposite contrasts for a low sample.

	Red	Green	Blue
White/Black	0.3000	0.3000	0.3000
Red/Cyan	0.4400	-0.0999	-0.0999
Yellow/Blue	0.0726	0.0726	-0.6000

Table 2. Gun contrasts for single color masks.

The Red/Cyan and the Yellow/Blue masks were physically isoluminant. Two color and three color masks were generated by summing the single color masks in the contrast domain. Three targets were used, a White/Black, a Red/Cyan, and a Yellow/Blue target. They were uniform horizontal rectangles 0.304 by 0.213 deg (16 by 8 pixels). The target contrasts were in the same RGB directions as the corresponding single color masks (Table 2). The target contrasts were arbitrarily defined as the largest of the gun contrasts. The targets set in a uniform background were mixed with the masks by interleaving frames at 120 frames per second, so the effective noise contrasts were halved. The masks were turned on and off abruptly for 0.5 sec, but the targets were linearly ramped on and off for 0.25 sec, resulting in effective contrast energy duration of 0.33 sec.

2.3 Procedure

Thresholds were measured on the uniform gray background and on the seven fixed pattern noise backgrounds (Y, R, R+Y, W, W+Y, W+R, W+R+Y). Thresholds for the first 8 observers, group A, were estimated using a two-interval forced-choice (2IFC) QUEST procedure. Each block was run for at least 20 trials and until the 95% confidence interval width for the threshold was less than 2 dB. The thresholds were estimated by the QUEST procedure with the response error probability set to 0.01, the Weibul slope parameter set to 3.5, and the threshold probability correct assumed to be 0.75. A block was stopped when 30 trials were done and the estimated accuracy was 2 dB. The target and no-target intervals were each 0.5 sec in duration and separated by an interval of 0.1 sec. The next trial began 0.25 sec after the response to the previous trial. A feedback tone indicated an incorrect response. Horizontal and vertical lines outside the masking image region indicated the position of the possible target during the trial. Each observer ran one practice day and three test days. On each day the observer completed a warm-up session of 5 blocks of conditions chosen at random without replacement from the 14 conditions (2 targets by 7 maskers), and then ran 27 blocks (one block for each masking condition and two blocks of unmasked conditions) in a random order. The median threshold contrast was found from the three measurements in each condition and for each observer the masking effect of each masker was represented in decibels (20 times the logarithm to the base ten of the ratio of the masked threshold to the unmasked threshold).

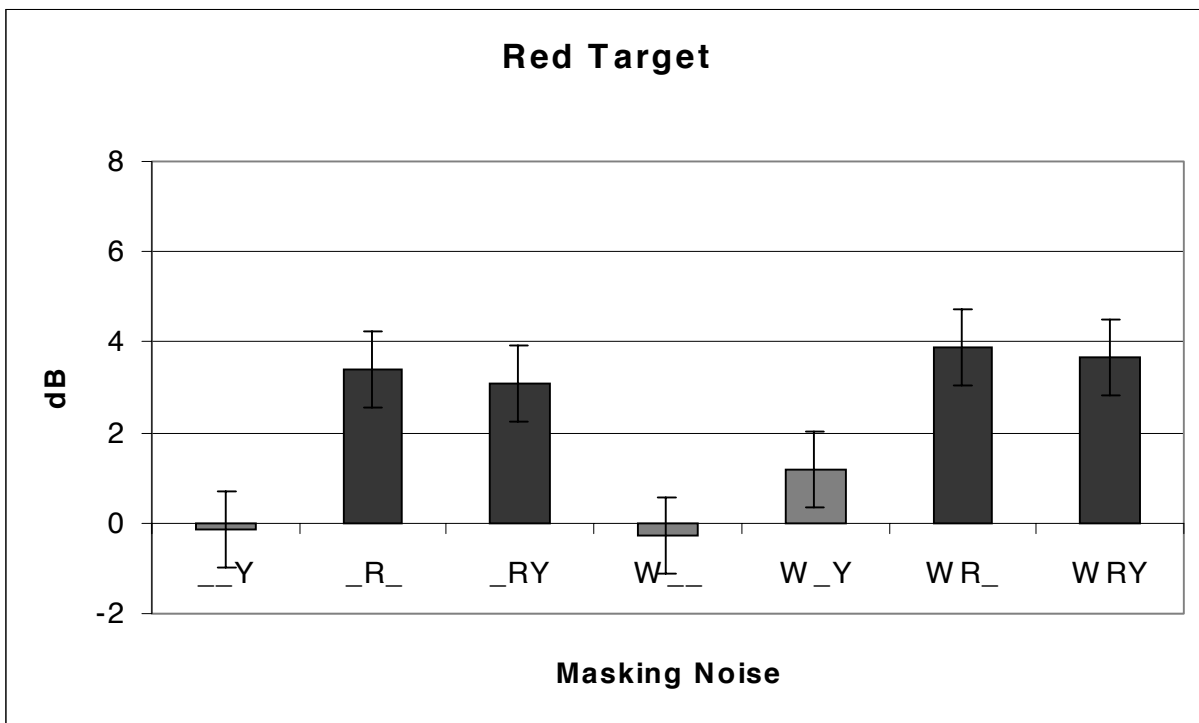
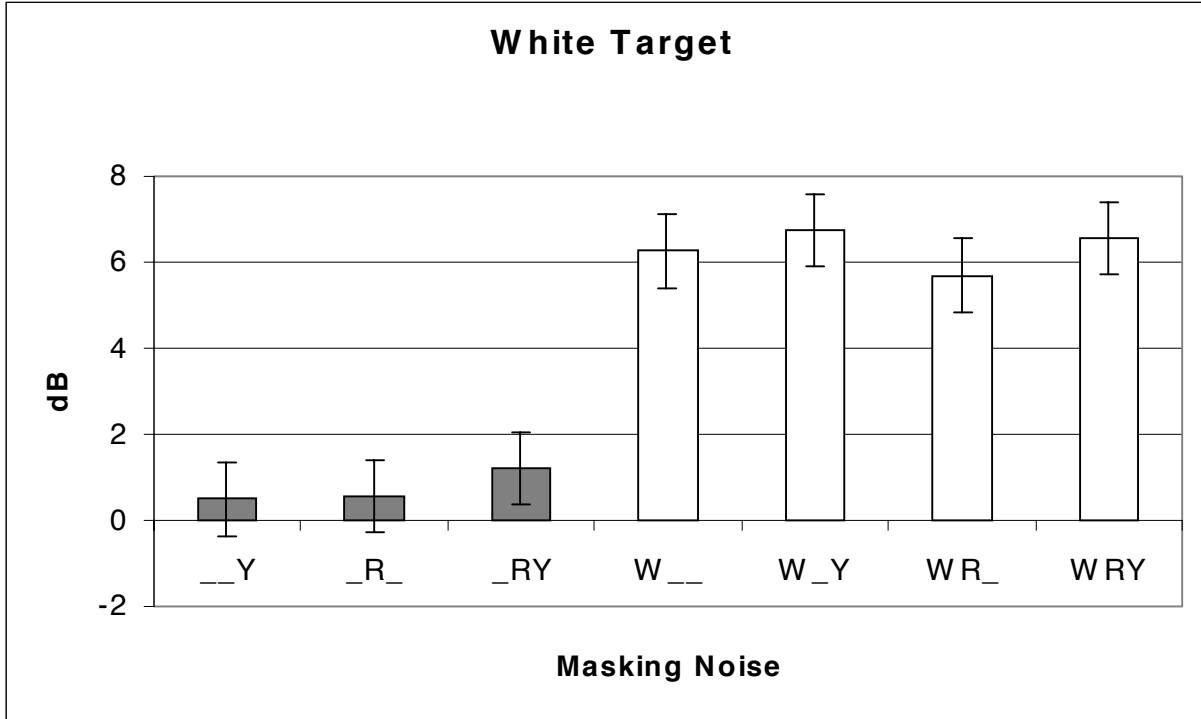
The procedure for the second group of 8 observers, group B, was the same as for group A except that the psychophysical procedure was four spatial interval forced choice. A black fixation cross marked the center of the image and the edge of the target was in one of the four quadrants to the left or right and up or down. Pilot work indicated that this method gave more reliable thresholds in a shorter time, so the minimum number of trials per block was reduced to 20.

3. RESULTS

As the pilot work indicated, group B did give more reliable thresholds in less time. Most blocks did stop in 20 trials (and only one temporal interval was used), while observers in group A averaged over 43 trials per block. Confidence intervals (95%) for the average thresholds on the uniform background in dB re 10^{-6} deg² sec of contrast energy) are shown in Table 3. The contrast for the colored targets is the largest of the three gun contrasts. The confidence intervals are based on the within-group variance, using the t distribution with 7 degrees of freedom. The thresholds are expected to be higher in experiment B because the 75% correct threshold occurs at a higher target level in the QUEST model when the guessing probability is reduced to 0.25 from 0.5.

Group	White	Red	Yellow
A	7.2 ± 1.9	22.0 ± 2.5	24.5 ± 3.2
B	10.6 ± 0.7	25.2 ± 1.5	24.5 ± 1.2

Table 3. Thresholds for targets on the uniform background in dBB.



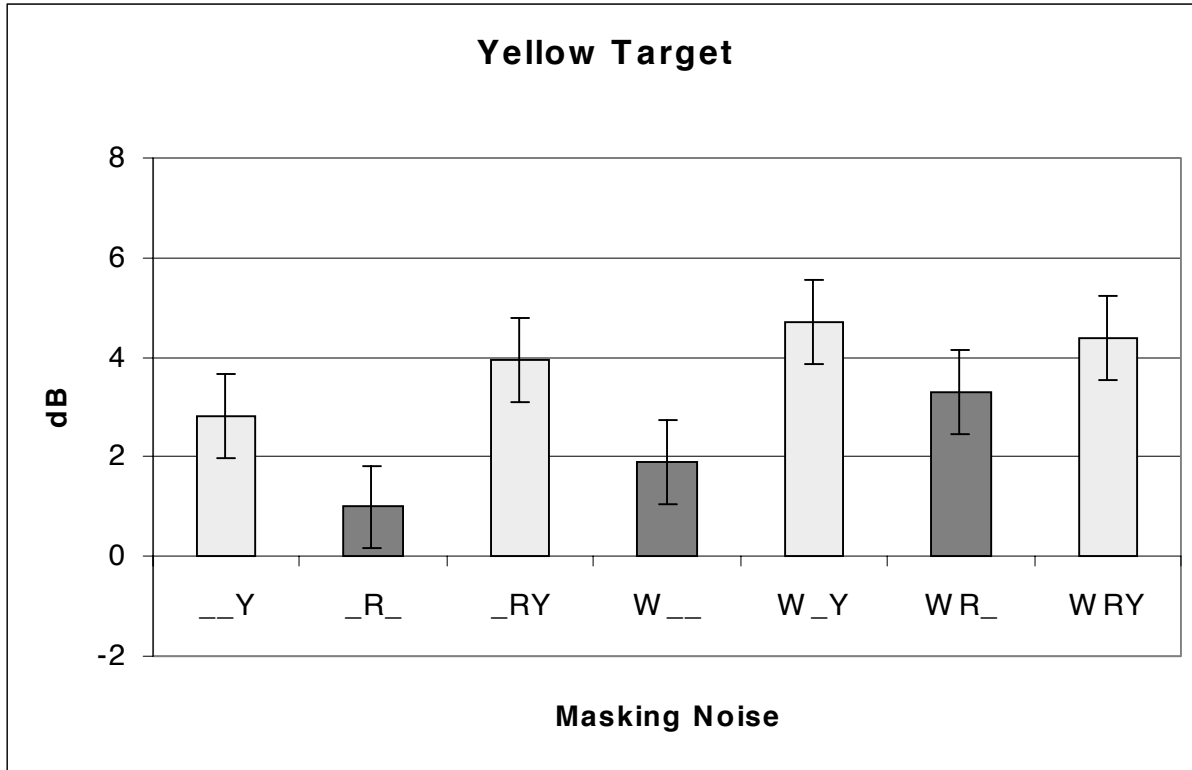


Figure 1. Threshold increases in dB for the 7 maskers on the 3 targets for the pooled data of groups A and B. The error bar heights, 95% confidence intervals for these increases, are ± 0.8 dB.

Table 4 shows for the three targets the ratios in dB of the mean masked thresholds to those on the uniform background. 95% confidence limits for these ratios, based on the variance of the masking scores over observers averaged over maskers, are ± 1.4 dB and ± 0.8 dB, for groups A and B, respectively). The pooled data is a weighted average giving more weight to the group B data by virtue of its higher reliability. The pooled data is plotted in Figure 1. Confidence intervals for the pooled estimates are ± 0.8 dB. The masking is dominated by the masker of the same color as the target, although some cross masking occurs, especially for the yellow target.

group	Target	--Y	-R-	-RY	W--	W-Y	WR-	WRY
A	W	1.1	1.0	1.7	5.3	4.8	4.2	4.8
	R	-0.2	2.7	3.5	1.1	2.3	4.6	4.2
	Y	1.0	2.0	2.8	2.7	3.8	3.8	3.2
B	W	0.2	0.3	1.0	6.7	7.7	6.4	7.4
	R	-0.1	3.7	2.9	-0.9	0.6	3.5	3.4
	Y	3.7	0.4	4.4	1.5	5.2	3.1	4.9
Pooled	W	0.5	0.5	1.2	6.3	6.7	5.7	6.5
	R	-0.1	3.4	3.1	-0.3	1.2	3.9	3.7
	Y	2.8	0.9	3.9	1.9	4.7	3.3	4.4

Table 4. Mean masking results in dB by group, target, and masker.

3.1 Additivity Analysis

An additivity analysis was performed on the pooled data to see whether the effects of adding the masking noises were additive in the contrast energy domain. Table 5 shows the additive energy components that minimize the squared error in the dB domain. For example, to obtain the additive energy prediction of mask W-Y on target Y in dB we compute $10 \cdot \log_{10}(1+0.675+1.032)$. The RMS error corrected for the 9 estimated parameters is only 0.45 dB. A more restrictive model that assumes masking only by the same component has a corrected RMS error of 1.1 dB, which is also small, but the F

test for this more restrictive hypothesis assuming the additive hypothesis gives an F ratio of 15.9 with 6 and 12 degrees of freedom, which is significant at $p < 0.01$.

Mask\Target	W	R	Y
W	3.125	0.100	0.675
R	0.118	1.171	0.297
Y	0.173	0.051	1.032

Table 5. Additive contributions in the contrast energy domain of each mask component upon each target.

These additive masking components summarize the results of the experiment. Each mask component is seen to mask the target in its own direction. The only strong cross masking results are the masking of the Y target by the W and R mask components. A more general additivity analysis was also performed, allowing the additivity to occur after an arbitrary exponent, rather than the fixed energy exponent of 2. The estimated best exponent was 1.86 and the nested F test for whether the value of 2 was worse gave an F of 0.08, so the data cannot reject 2 as the best exponent, supporting the additivity of masking in the contrast energy domain.

4. A COLOR MASKING MODEL

Since the colors of the maskers were not chosen to be aligned or orthogonal to directions of theoretical color mechanisms, the data cannot be interpreted directly as indicating whether color mechanisms mask each other or not. Models of the mechanisms and their interactions are needed to answer this question. Since the goal here is to find models that engineers can use to optimize displays, we restrict ourselves to color channels based on readily available measurements (CIE XYZ).

4.1 The single channel masking model

The single-channel masking model takes as input a background luminance image Y_0 and a target plus background image Y_t . They are then converted to contrast images by subtracting and dividing them by the mean luminance of the background image. The contrast images are filtered by a luminance contrast sensitivity filter and the unmasked target detectability is simply the norm of the difference of the filtered contrast images. The effect of the background contrast energy on the detectability is accounted for by a multiplicative factor $(1 + (a c)^2)^{-0.5}$. If c is the RMS contrast energy after filtering with a filter with unity peak gain, a typical value for a would be 20, corresponding to a masking contrast threshold of 5%. A straightforward extension of this model would be to add two more color channels as has been done before^{6, 7} and estimate three masking coefficients for each of the three channels. Several iterations using various directions and contrast normalizations were tried with this data. The model below gave the best fit of all models that we happened to try.

4.2 A cone contrast channel model. The model begins by the conversion of RGB images to CIE XYZ images. Next the XYZ images are converted to cone catch images LMS (long, middle, and short wavelength sensitive cones, also called red, green, and blue cones) by the transformation from Boynton (p. 404)⁸:

$$\begin{aligned} L &= 0.157 X + 0.542 Y - 0.0332 Z \\ M &= -0.157 X + 0.458 Y + 0.0332 Z \\ S &= 0.161 Z \end{aligned}$$

The cone images are then converted to cone contrast images based on the average level for that cone.

$$\begin{aligned} L_C &= (L - L_0) / L_0, \\ M_C &= (M - M_0) / M_0, \\ S_C &= (S - S_0) / S_0 \end{aligned}$$

Luminance Y_C and color-opponent O_C channels are then built from the L and M cone contrast images.

$$\begin{aligned} Y_C &= L_C + M_C \\ O_C &= L_C - M_C \end{aligned}$$

These YOS contrast images are then contrast sensitivity filtered. The Y_C channel is band pass filtered by a difference-of-Gaussians filter. The wide band Gaussian has a center spread (1/e amplitude reduction) in the spatial domain of 2.0 arc min.

This corresponds to a low pass $1/e$ spatial frequency amplitude cutoff of 16.9 cycles per deg. The surround spread was set to 8 times that of the center and its DC response was set to 0.685 times that of the center. This filter was then normalized to have unit maximum gain in the spatial frequency domain and then scaled by a luminance channel contrast sensitivity factor. The contrast sensitivity factor is defined to be unity if a one degree square area of unit filtered contrast is at threshold (gives a d' of unity). The contrast sensitivity filters for O_C and S_C were low pass Gaussian filters with unity gain at DC and spatial frequency cutoffs of 5.46 and 2.73 cycles per deg, respectively. The chromatic channel filters were also multiplied by similarly defined contrast sensitivity factors. Conveniently, the white, red, and yellow targets were predicted to be detected primarily by the Y, O, and S channels, respectively, so the sensitivity parameters of the model do not play a role in the masking.

The RMS contrast sensitivity filtered channel contrast of the single color masks are shown in Table 6.

Mask\Channel	Y	O	S
W	0.2228	0.0000	0.0240
R	0.0072	0.0110	0.0065
Y	0.0031	0.0025	0.0293

Table 6. RMS values in each channel for each of the single color masks.

The RMS value of each channel for a combination mask is obtained by summing these values these values in the energy domain (squaring the values, summing them, and then taking the square root). Note that the contrast sensitivity filtering of the O and S channels sharply reduces the RMS contrast in the masks in these channels. Note also that the W noise generates almost as much variation in the S channel as does the Y noise.

In our simplified model of masking, the effect of the background masker is represented by a contrast-gain-type masking coefficient. Masking coefficients M are computed from the RMS value of the filtered no-target images in each channel. For the three channels (Y, O, S), these coefficients are

$$M_Y = (1 + (a_{YY} c_Y)^2 + (a_{YO} c_O)^2 + (a_{YS} c_S)^2)^{-0.5},$$

$$M_O = (1 + (a_{OY} c_Y)^2 + (a_{OO} c_O)^2 + (a_{OS} c_S)^2)^{-0.5},$$

$$M_S = (1 + (a_{SY} c_Y)^2 + (a_{SO} c_O)^2 + (a_{SS} c_S)^2)^{-0.5}.$$

The c values are the RMS contrast of the filtered mask-only images and the a values give the contribution of each channel to the masking of the other. If the nonlinear effects of the masking noises are ignored (here they are less than 0.5 dB) these equations can be used to estimate the masking directly. For the values given in Table 7, the RMS prediction error corrected for the 9 estimated parameters was 0.45 dB, essentially the same as the unrestricted additive model (which would allow negative contributions).

From\To	Y	O	S
Y	7.8	1.4	0.0
O	29.5	99.4	43.1
S	13.9	0.0	33.9

Table 7. Masking coefficients a_{JI} from channel I to channel J for the pooled masking data.

The model with masking restricted to only the same channel has a very good fit of 0.56 dB and the estimated within-channel coefficients are (8.2, 101.6, 35.6) for (Y, O, S). The F test for this more restrictive hypothesis assuming the general masking hypothesis gives an F ratio of 2.59 with 6 and 12 degrees of freedom, with significance $0.1 < p < 0.05$.

4.3 Results with Other Models

The best results of the other models tested was obtained with weightings of the cone contrast channels suggested by Eskew.⁹ His suggested channel directions based on a number of masking and detection studies give a heavier weighting to the L cone in the luminance channel and make the S cone channel a blue/yellow opponent. The 9 parameter version is only slightly worse (parameter corrected RMS error fit = 0.59 dB), but the blue yellow opponent was blind to the white noise and thus the

no-cross-masking version did a poor job of predicting the masking of the yellow signal by the white noise (0.94 dB). A model with the three channels being just the three individual cone contrast images with cross masking was much worse (1.7 dB).

5. DISCUSSION

Although the model presented can make predictions for arbitrary color images, it will not always perform well. The simple luminance discrimination model can be regarded as a useful approximation to more sophisticated models with oriented band pass filtering.¹⁰ The simple model can only be successful when there is substantial overlap in the spatial frequency content of the background and the target. The small value of $a_{YY} = 7.8$ (< 20) is probably the result of the large amount of background energy in the high spatial frequencies that had little effect on the target detectability.

The facilitation and masking among Gabor images of the same shape and different color has been measured and modeled¹¹ with parameters similar to those of the Watson and Solomon model¹⁰ and with a cross masking coefficient array like that of Table 5. Unfortunately, the directions of the channels were determined principally by the facilitation data and the model did not take into account the facilitation that is not specific to color direction.

The simple model ignores the problems with inhomogeneous backgrounds. Each cone contrast denominator really should be represented by a low pass filtered image and each masking contrast should really be a low pass contrast energy image.^{12, 13} Although color vision researchers commonly use cone contrast as the basic stimulus for color vision, it can lead to problems when the images have little short wavelength energy. Boynton proposed a solution to this problem based on the addition of a "dark light" constant to the contrast denominator,⁸ but it is not clear whether there is an appropriate fixed value.

Despite the many problems, the cone contrast color channel model described above can be used to assess the visibility of targets in color displays. The version without cross masking should do well, even if cross masking exists, in images where the color contrast tends to be low.

ACKNOWLEDGEMENTS

This work was supported in part by NASA RTOP 548-51-12.

REFERENCES

1. A. M. Rohaly, A. J. Ahumada, Jr., A. B. Watson (1997) Object Detection in natural backgrounds predicted by discrimination performance and models, *Vision Res.* **37**, pp. 3225-3235.
2. M. P. Eckstein, A. B. Watson, A. J. Ahumada, Jr. (1997) Visual signal detection in structured backgrounds. II. Effects of contrast gain control, background variations, and white noise, *J. Opt. Soc. Amer. A* **14**, pp. 2406-2419.
3. M. P. Eckstein, A. J. Ahumada, Jr., A. B. Watson (1997) Image discrimination models predict signal detection in natural medical image backgrounds, in Human Vision, Visual Processing, and Digital Display VIII, ed. B.E. Rogowitz and T.N. Pappas, Proceedings Volume 3016, pp. 44-56, SPIE, Bellingham, WA.
4. A. J. Ahumada, Jr. (1996) Simplified vision models for image quality assessment, in Society for Information Display International Symposium Digest of Technical Papers, ed J. Morreale, Volume 27, pp. 397-400, Society for Information Display, Santa Ana, CA.
5. A. J. Ahumada, Jr., B. L. Beard (1997) Image discrimination models predict detection in fixed but not random noise, *J. Opt. Soc. Amer. A* **14**, pp. 2471-2476.
6. H. A. Peterson, A. J. Ahumada, Jr., A. B. Watson (1993) An improved detection model for DCT coefficient quantization, in Human Vision, Visual Processing, and Digital Display IV, ed. B.E. Rogowitz and J. Allebach, Proceedings Volume 1913, pp. 191-201, SPIE, Bellingham, WA.
7. A. B. Watson, J.Q. Hu, J.F. McGowan III, J.B. Mulligan (1999) Design and performance of a digital video quality metric, in Human Vision and Electronic Imaging IV, ed. B.E. Rogowitz and T.N. Pappas, Proceedings Volume 3644, pp. paper 17, SPIE.
8. R. M. Boynton (1992) Human Color Vision. Optical Society of America.
9. R. T. Eskew, Jr., J. S. McLellan, F. Giulianni (1999) Chromatic detection and discrimination. In K. Gegenfurtner, L.T. Sharpe, eds., Color vision: from molecular genetics to perception. Cambridge: Cambridge U. Press, Ch. 18, pp. 345-358.
10. A. B. Watson, J. Solomon (1997) A model of contrast gain control and pattern masking, *J. Opt. Soc. Amer. A* **14**, pp. 2379-2391.

11. C. Chen, J. M. Foley, D. H. Brainard (2000) Detection of chromoluminance patterns on chromoluminance pedestals, *Vision Res.* **40**, pp. 773-803.
12. A. J. Ahumada, Jr., B. L. Beard (1998) A simple vision model for inhomogeneous image quality assessment, in Society for Information Display Digest of Technical Papers, ed. J. Morreale, vol. 29, Paper 40.1, Santa Ana, CA.
13. K. Brunnström, R. Eriksson, A. J. Ahumada, Jr. (1999) Spatio-temporal discrimination model predicting IR target detection, in Human Vision and Electronic Imaging III, ed. B.E. Rogowitz and T.N. Pappas, Proceedings Volume 3644, pp. 403-410, SPIE, San Jose, CA.