

Watson, A. B., Gale, A., Ahumada, A. J., Jr. & Solomon, J. (1994). DCT Basis Function Visibility: Effects of Viewing Distance and Contrast Masking. In B. E. Rogowitz (Ed.), *Human Vision, Visual Processing, and Digital Display IV* (pp. 99-108). Bellingham, WA: SPIE.

## **DCT BASIS FUNCTION VISIBILITY: EFFECTS OF VIEWING DISTANCE AND CONTRAST MASKING**

Andrew B. Watson    Joshua A. Solomon    Albert Ahumada

MS 262-2, NASA Ames Research Center, Moffett Field, CA 94035-1000  
beau@vision.arc.nasa.gov    al@vision.arc.nasa.gov    jsolomon@vision.arc.nasa.gov

Alan Gale

San Jose State University

### **ABSTRACT**

Several recent image compression standards rely upon the Discrete Cosine Transform (DCT). Models of DCT basis function visibility can be used to design quantization matrices for arbitrary viewing conditions and images. Here we report new results on the effects of viewing distance and contrast masking on basis function visibility. We measured contrast detection thresholds for DCT basis functions at viewing distances yielding 16, 32, and 64 pixels/degree. Our detection model has been elaborated to incorporate the observed effects. We have also measured detection thresholds for individual basis functions when superimposed upon another basis function of the same or a different frequency. We find considerable masking between nearby DCT frequencies. A model for these masking effects will also be presented.

### **1. INTRODUCTION**

The JPEG, MPEG, and CCITT H.261 image compression standards, and several proposed HDTV schemes employ the Discrete Cosine Transform (DCT) as a basic mechanism<sup>1, 2</sup>. Typically the DCT is applied to 8 by 8 pixel blocks, followed by uniform quantization of the DCT coefficient matrix. The quantization bin-widths for the various coefficients are specified by a quantization matrix (QM). The QM is not defined by the standards, but is supplied by the user and stored or transmitted with the compressed images.

The principle that should guide the design of a QM is that it provide optimum visual quality for a given bit rate. QM design thus depends upon the visibility of quantization errors at the various DCT frequencies. In recent papers<sup>3, 4</sup>, Peterson *et al.* have provided measurements of threshold amplitudes for DCT basis functions at one viewing distance and several mean luminances. Ahumada and Peterson<sup>5</sup> have devised a model that generalizes these measurements to other luminances and viewing distances, and Peterson *et al.*<sup>6</sup> have extended this model to deal with color images. From this model, a matrix can be computed which will insure that all quantization errors are below threshold. Watson<sup>7</sup> has shown how this model may be used to optimize the quantization matrix for an individual image.

## 2. EFFECTS OF DISPLAY RESOLUTION

Visual resolution of the display (in pixels/degree of visual angle) may be expected to have a strong effect upon the visibility of DCT basis functions, and we therefore collected data to document this effect and to validate and enhance the model.

### 2.1 Practical Pixel Sizes

Visual resolution of the display (in pixels/degree of visual angle) is determined by display resolution (in pixels/cm) and viewing distance (in cm), according to the formula

$$(\text{pixels/degree}) = (\text{pixels/cm}) / \cot^{-1}[\text{distance}]$$

In the viewing situations for which block-DCT compression is contemplated, there are limits to the practical range of visual resolutions. At the high end, display resolution will be wasted on spatial frequencies which are not visible to the human eye. The limit of human spatial resolution is about 60 cycles/degree. Nyquist sampling of this frequency would require 120 pixels/degree. This corresponds to 300 dpi printing viewed at a distance of about 23 inches. At the low end, the pixel raster becomes visible. In these experiments, we have examined three viewing distances, 16, 32, and 64 pixels/degree, that span a large part of the range of useful viewing distances.

### 2.2 Methods

Detection thresholds for single basis functions were measured by a two-alternative, forced-choice method. Each trial consisted of two time intervals, within one of which the stimulus appeared. The stimulus was a single DCT basis function, added to the uniform gray background that remained throughout the experiment. Background luminance was 40 cd m<sup>-2</sup>, and frame rate was 60 Hz. Observers viewed the display screen from distances of 48.7, 97.4, 194.8 cm. Display resolution was 37.65 pixels/cm. Images were magnified by two in each dimension, by pixel replication, to reduce monitor bandwidth limitations, resulting in magnified pixel sizes of 1/16, 1/32, and 1/64 of a degree, respectively at the three viewing distances (basis functions were 1/2, 1/4, and 1/8 degree in width). We describe these three viewing distances as yielding effective visual resolutions of 16, 32, and 64 (magnified) pixels/degree.

During presentation, the luminance contrast of the stimulus was a Gaussian function of time, with a duration of 32 frames (0.53 sec) between  $e^{-\pi}$  points. The peak contrast on each trial was determined by an adaptive QUEST procedure<sup>8</sup>, which converged to the contrast yielding 82% correct. After completion of 64 trials, thresholds were estimated by fitting a Weibull psychometric function<sup>9</sup>. Thresholds are expressed as contrast (peak luminance, less mean luminance, divided by mean luminance), converted to decibel sensitivities ( $-20 \log_{10}[\text{threshold}]$ )

To reduce the burden of data collection, we measured thresholds for only 30 of the possible 64 basis functions, as indicated in Fig. 1. To the extent that thresholds change slowly as a function of DCT frequency, this sampling constrains our model sufficiently.

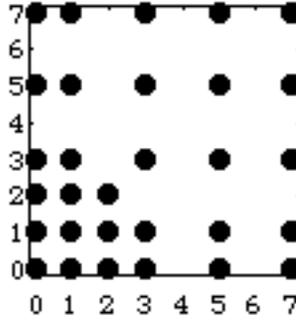


Figure 1. Subset of DCT frequencies used in the experiment.

To date, two data sets have been collected at the low resolution, five at the middle resolution, and one at the highest resolution, as shown in Table 1.

resolution (pixels/degree)	observer				
	abw	mjy	aig	sj	jas
16	0	30	0	30	0
32	7	30	60	30	30
64	0	30	0	2	0

Table 1. Thresholds collected for each observer and viewing distance.

### 2.3 Model of DCT Contrast Sensitivity

The model of DCT contrast sensitivity that we consider here is essentially that described by Peterson et al.<sup>6</sup> In that model, log sensitivity versus log frequency is a parabola, whose peak value, peak location, and width vary with mean luminance. In addition, sensitivity at oblique frequencies ( $\{u \neq 0, v \neq 0\}$ ) is reduced by a factor that is attributed to the orientation tuning of visual channels. The parameters of significance here are  $s_0$  (peak sensitivity),  $f_0$  (peak DCT frequency at high luminances), and  $k_0$  (inverse of the *latus rectum* of the parabola), and  $r$  (the orientation effect).

### 2.4 Results

Figures 2, 3, and 4 show decibel contrast sensitivities for the three viewing distances, along with curves showing the predictions of the best fitting version of the model. Within each figure, the three panels show data for horizontal frequencies  $\{u, 0\}$ , vertical frequencies  $\{0, v\}$ , 45 degree orientations  $\{u, v=u\}$ , and the remaining obliques  $\{u>0, 0<v \neq u\}$ , all plotted against the radial frequency  $f = \sqrt{u^2 + v^2}$ . In the case of the obliques, because there is no simple one-dimensional prediction to plot, we plot instead the actual sensitivity minus that predicted by the model. These plots, and the fits, do not include the thresholds at  $\{0,0\}$  (DC), which are reserved for a separate discussion. The data at 64 pixels/degree also omit 3 thresholds at very high frequencies which we suspect to be artifactual.

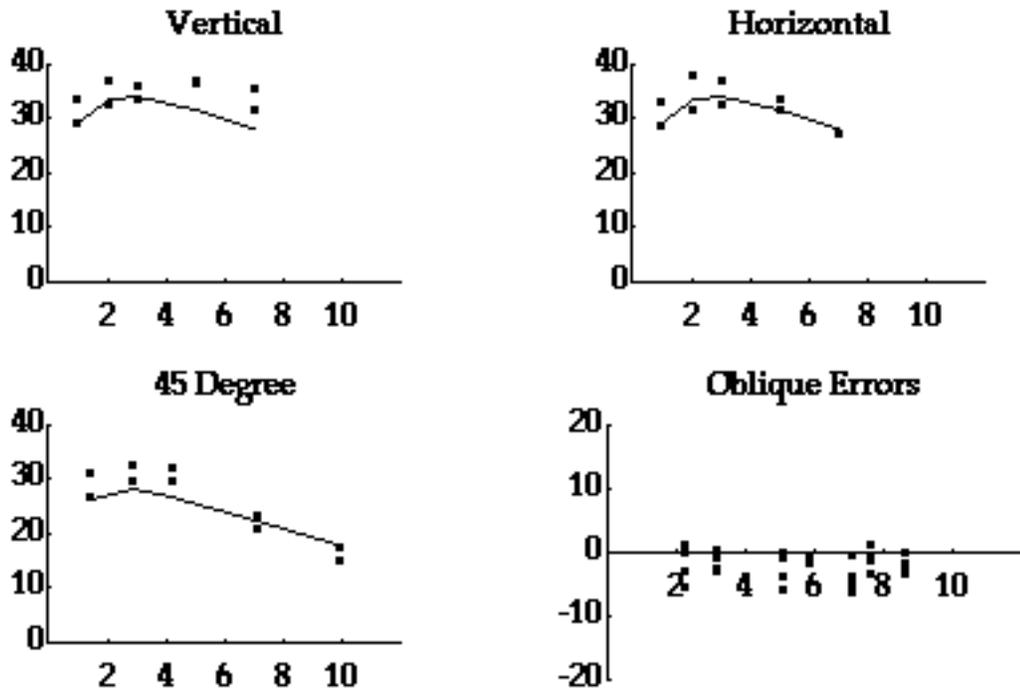


Figure 2. DCT basis function sensitivities at 16 pixels/degree.

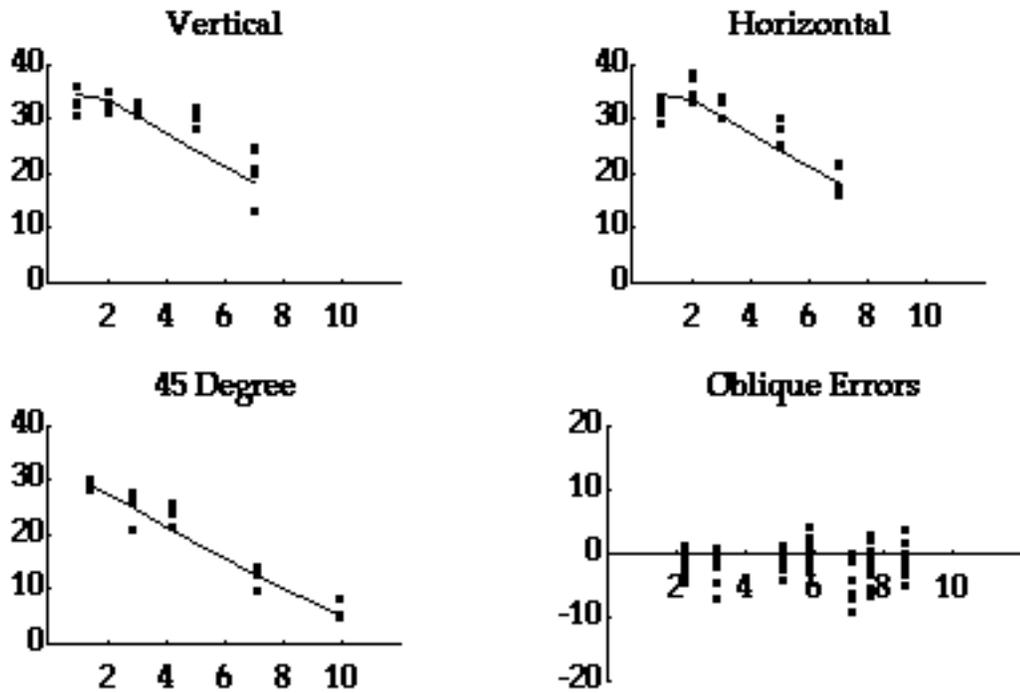


Figure 3. DCT basis function sensitivities at 32 pixels/degree.

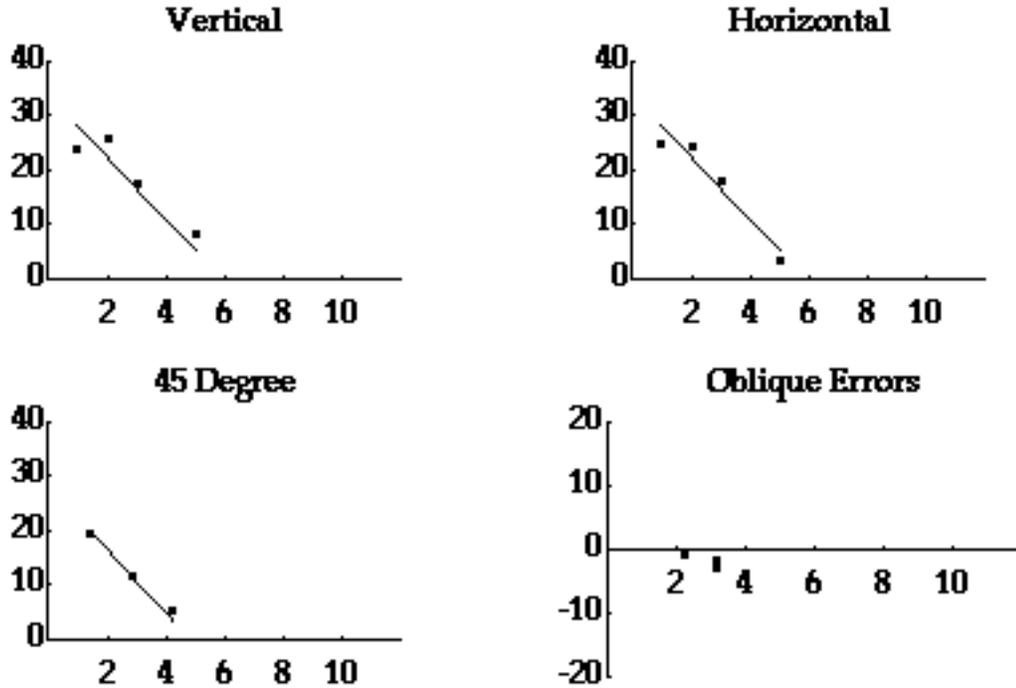


Figure 4. DCT basis function sensitivities at 64 pixels/degree.

The fits are reasonable, though there appear to be some systematic departures from the model. For reference, the RMS error of the raw data at the middle distance is 2.03 decibels, while the RMS error of the fit in Figs 2-4 is 2.94 decibels. The estimated parameters are shown in Table 2.

	pixels/degree		
	16	32	64
s0	51.1	56.17	29.84
f0	3.68		
k0	1.728		
r	0.5115		

Table 2. Estimated model parameters.

The parameters  $f_0$ ,  $k_0$ , and  $r$  (related to peak frequency, bandwidth, and orientation effects) are equated for all resolutions, while a separate value of  $s_0$  (peak contrast sensitivity) is estimated for each of the three resolutions. The behavior of this parameter is worth considering. Between 64 and 32 pixels/degree, it increases by a factor of 1.88. Between these two resolutions, the basis functions increase in size by a factor of two in each dimension. Thus if sensitivity increased linearly with area (as it should for very small targets<sup>10, 11, 12</sup>) we would expect an increase of a factor of 4. If sensitivity increased due only to spatial probability summation<sup>13, 14</sup>, we would expect a factor of about  $4^{1/4} = 1.414$ . Thus the obtained effect is nearer to that expected of probability summation. At the closest viewing distance, despite a further magnification by 2, the parameter  $s_0$  actual declines. While we would expect a smaller effect of size at the largest sizes, this decline is unexpected and may be due to 1) the relatively poor fit at this resolution, and 2) aspects of visual sensitivity which are not yet captured by the model.

## 2.5 DC Sensitivities

Figure 5 shows the sensitivities for DC basis functions at the three visual resolutions.

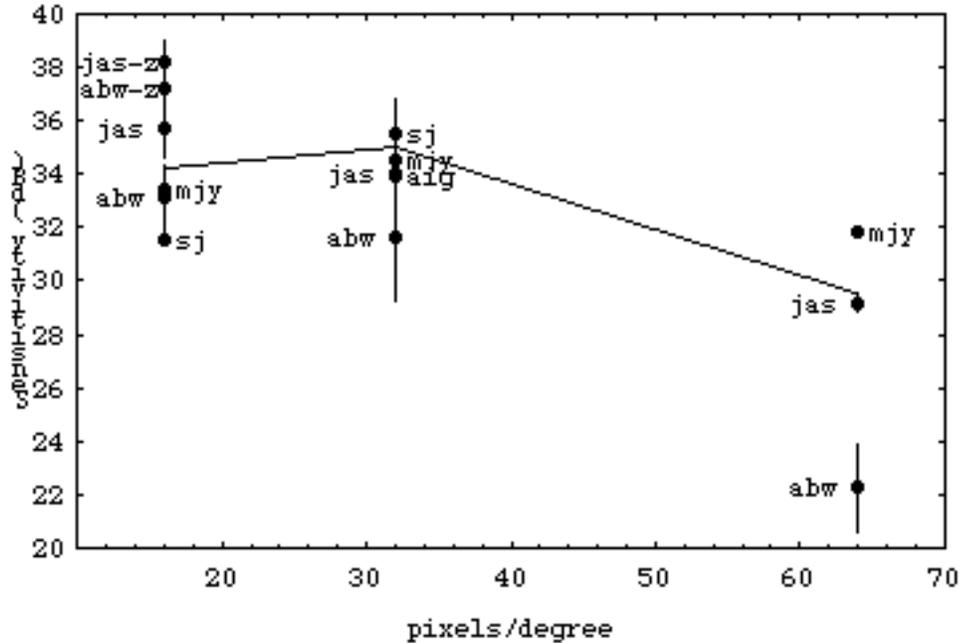


Figure 5. DC basis function sensitivities as a function of display visual resolution. Error bars of plus and minus one standard deviation are shown when multiple measurements were available. For clarity, points with error bars are labeled on the left, those without, on the right. The line indicates the parameter  $s_0$  from Table 2.

Ahumada et al.<sup>5,6</sup> proposed as a working hypothesis that DC sensitivity is given by the peak sensitivity  $s_0$ . This prediction is given by the line drawn in Fig. 5. It captures some of the variation in the DC sensitivities, but further data will be needed to adequately test this model. The points in Fig. 5 at a resolution of 16 pixels/degree and labeled with the suffix "-z" were obtained by pixel-replication at the middle viewing distance, rather than use of the near distance. Their enhanced sensitivity suggests that viewing distance per se may have an effect, even when visual resolution is held constant. The substantial variability of DC thresholds at the highest resolution may be due to differences in accommodation between observers.

## 2.6 Discussion

We have examined the variation in visibility of single DCT basis functions as a function of display visual resolution. We have shown that the existing model<sup>5,6</sup> accommodates resolutions of 16, 32, and 64 pixels/degree, provided that one parameter, the peak sensitivity  $s_0$ , is allowed to vary. Variations in this parameter are to some extent consistent with spatial summation, although sensitivity is lower at the lowest resolution than summation would predict.

Practical DCT quantization matrices must take into account both the visibility of single basis functions, and the spatial pooling of artifacts from block to block. Elsewhere we have shown that to a first approximation this pooling is consistent with probability summation<sup>15</sup>. If we consider two images of equivalent size in degrees, but visual resolutions differing by a factor of two, then the sensitivity to individual artifacts would be lower by  $4^{1/4}$  in the higher resolution image due to the smaller block size in degrees, but higher by  $4^{1/4}$  in the same image due to the greater number of blocks. Thus the same matrix should be used with both. The point of this example is that the overall gain of the best quantization matrix must take into account both display resolution and image size.

### 3. EFFECTS OF CONTRAST MASKING

#### 3.1 Contrast masking

Watson<sup>7</sup> noted several image-dependent factors influencing the detectability of DCT basis functions and showed how to compute custom QMs for given images, in accord with these factors. One image-dependent factor influencing the detectability of DCT basis functions is contrast masking. Typically, sensitivity to quantization error, in a particular DCT coefficient, decreases with the magnitude of that coefficient. Watson's quantization scheme relies on the following model (based on work by Legge and Foley<sup>16, 17</sup>) for contrast masking: given a DCT coefficient  $c_T$  and a corresponding absolute threshold  $t_T$ , the masked threshold  $m_T$  will be

$$m_T = t_T \text{Max} \left[ 1, \left| c_T / t_T \right|^{w_T} \right], \quad (1)$$

where  $w_T$  is an exponent that lies between 0 and 1. In the sequel, we will refer to this model as Model 1. In Model 1, sensitivity to a particular coefficient's quantization error is independent of the magnitudes of all the other coefficients (except the DC). Here we present data which indicate that sensitivity to a particular coefficient's quantization error is affected by the magnitudes of other coefficients. We propose a revision of Model 1 to account for between-coefficient contrast masking.

#### 3.2 Methods

General methods were the same as in the earlier experiments (Section 2.2). Each stimulus was the sum of a test basis function and a mask basis function, added to the mean luminance of the display. The contrast of the mask remained constant throughout a block of 64 trials, while the contrast of the test was varied using the Quest procedure<sup>8</sup> to determine the threshold for the test in the presence of the mask. Effective visual resolution was 32 pixels/degree, so that each stimulus subtended 0.25 degrees by 0.25 degrees.

Masked thresholds  $m_T$  for four test DCT frequencies were measured as a function of masking contrast for three different mask frequencies. The tests frequencies  $\mathbf{T}$  were  $\{0,0\}$ ,  $\{0,1\}$ ,  $\{0,3\}$  and  $\{0,7\}$ . These last three also served as the masks. Additionally,  $\{1,1\}$  and  $\{1,0\}$  were used to mask  $\{0,1\}$ ; and  $\{2,2\}$  was used to mask  $\{0,3\}$ . Un-masked threshold  $t_T$  was also determined for each test. Theoretically, DCT coefficients can assume any real value. In the current study we use coefficients  $c_U$ , such that  $0 \leq c_U \leq 1$ . A coefficient with value 1 fully utilizes the dynamic range of the display. For nearly every test/mask combination, six masking contrasts were used. Here we express these contrasts in decibels ( $\text{dB}[c_U] = 20 \log_{10}[c_U]$ ): -36, -30, -24, -18, -12 and -6. Because  $t_{07}$  is so high, when this basis function served to mask others, only the four greatest masking contrasts were used. Test and mask frequencies were fixed within a block of trials, and frequency combinations were run in a randomized fashion. The second author (jas) was the only observer in these experiments.

#### 3.3 Results and Discussion

The results are plotted in Figs. 6 and 7.

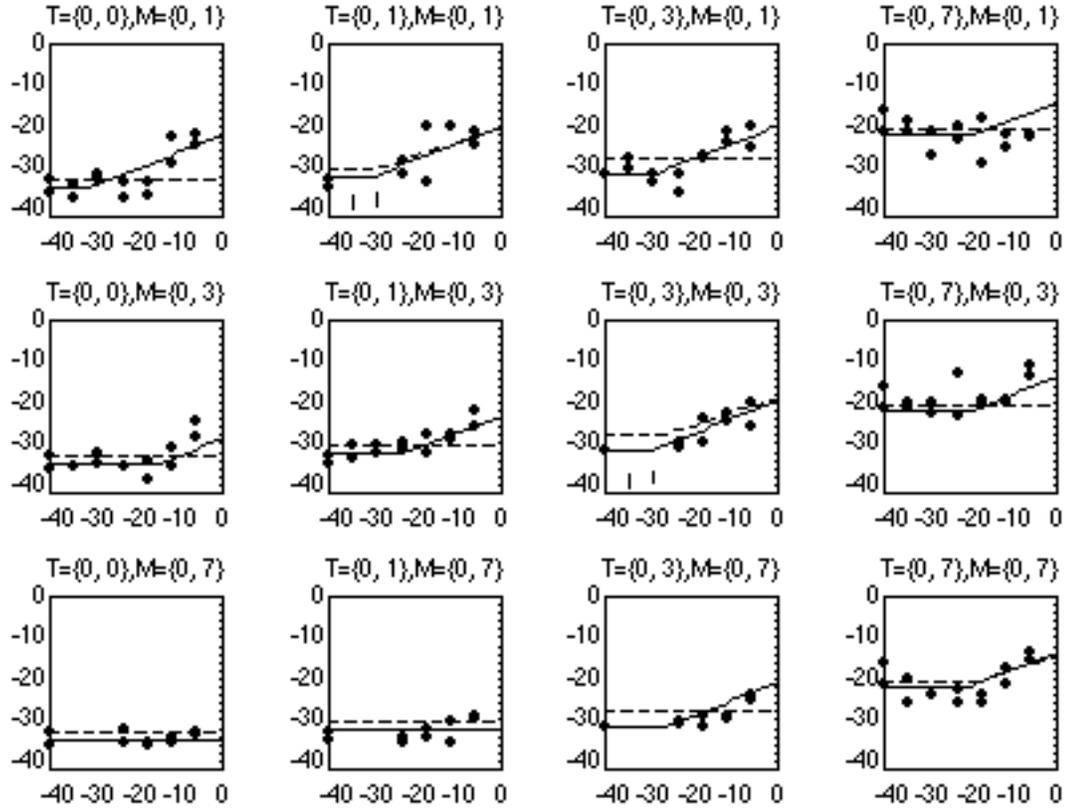


Fig. 6. Masked thresholds ( $\text{dB}[m_T]$ ) for four test basis functions are plotted as a function of masking contrast ( $\text{dB}[c_M]$ ) for three different masks. Unmasked thresholds ( $\text{dB}[t_T]$ ) for the test basis functions are plotted on the ordinates. The dashed and solid lines are the predictions of Models 1 and 2, respectively, as described in the text.

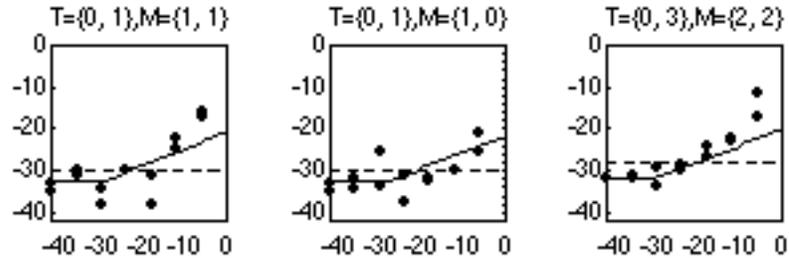


Fig. 7. Masked thresholds for test  $\{0,1\}$  as a function of masking contrast for the masks  $\{1,1\}$  and  $\{1,0\}$ , and for test  $\{0,3\}$  as a function of masking contrast for the mask  $\{2,2\}$ .

### 3.3.1 The dipper effect

Data gathered with the  $\{0,1\}/\{0,1\}$  test/mask combination at masking contrasts of -36 and -30 dB have been omitted from further analysis. Similarly, we have omitted the  $\{0,3\}/\{0,3\}$  data at -36 and -30 dB. These data appear as short vertical line segments in Fig. 6. Measured thresholds for these four viewing conditions fall well below their corresponding unmasked thresholds. These data demonstrate the "dipper effect," a well-documented phenomenon wherein a low contrast grating increases the detectability of a grating of the same frequency and phase<sup>16, 18, 19</sup>. These data have been omitted because it is not clear that the dipper effect comes into play for natural images. For images composed of more than one 8x8 pixel block, DCT basis functions can appear as gratings (uniform values) or noise (random values; with a quantifiable variance) or anything in between. The dipper effect would appear if both test and mask were gratings. However, there is no indication that it would

appear otherwise. The influence of a particular DCT coefficient on the detectability of quantization errors in natural images is similar in concept to the influence of a grating on the detectability of random visual noise. No dipper effect is expected in such a paradigm. Since we ultimately wish to model the detectability of quantization error in natural images, we believe that the exclusion of the "dipper data" will benefit our initial approximations.

### 3.3.2 Model 1

Model 1 was fit to the data. Model 1 does not include between-coefficient contrast masking. Consequently, for any given test basis function, its prediction for masked threshold is the same constant function of masking contrast for every mask having a non-zero coefficient at a different DCT index than the test. By setting all of the  $W_T$ s in Eq. 1. equal to a single parameter  $w$ , the total variance (on a log scale) from the model increased by less than 0.3%. Hereafter, when we refer to Model 1, we mean specifically: Given a test DCT basis function  $\mathbf{C}_T$ , its corresponding absolute threshold  $t_T$  and a mask DCT basis function  $\mathbf{C}_M$  the masked threshold  $m_T$  will be

$$m_T = \begin{cases} t_T \text{Max}\left[1, (c_M/t_T)^w\right] & \text{for } \mathbf{T} = \mathbf{M} \\ t_T & \text{otherwise} \end{cases}, \quad (2)$$

where  $0 \leq w \leq 1$ . Best fitting (method of least squares) values for  $w$  and  $t_T$ , as determined for Model 1, are given in Table 3. For comparison, we have also analyzed a Model 0 which predicts no contrast masking, i.e.  $m_T = t_T \forall \mathbf{T}$ . Best fitting values for  $t_T$ , as determined by Model 0 are also given in Table 3. Model 1 reflects the data for the viewing conditions in which the mask and target were identical more accurately than Model 0 does. However, it cannot reflect the between-coefficient masking evident by the increase in measured threshold with masking contrast for the other test/mask combinations.

### 3.3.3 Model 2

In order to reflect the between-coefficient masking, we propose the following revision of Model 1, referred to hereafter as Model 2. Given a test DCT basis function  $\mathbf{C}_T$ , its corresponding absolute threshold  $t_T$  and a mask DCT basis function  $\mathbf{C}_M$  the masked threshold  $m_T$  will be

$$m_T = t_T \text{Max}\left[1, \left(f[\mathbf{T}, \mathbf{M}] \frac{c_M}{t_T}\right)^w\right], \quad (3)$$

where  $w$  is an exponent that lies between 0 and 1 and  $f[\mathbf{T}, \mathbf{M}]$  is a positive, frequency-dependent scaling factor, that assumes a maximum value of 1 when  $\mathbf{T} = \mathbf{M}$ .  $f[\mathbf{T}, \mathbf{M}]$  may be described as a family of tuning functions. That is, for any test basis function  $\mathbf{C}_T$ ,  $f[\mathbf{T}, \mathbf{M}]$  reflects the sensitivity of  $\mathbf{C}_T$  detection to masks at different frequencies. We have chosen to specify these sensitivity functions with the following one-parameter rule:

$$f[\mathbf{T}, \mathbf{M}] = \exp\left[-\pi \|\mathbf{T} - \mathbf{M}\|^2 / \zeta_T^2\right], \quad (4)$$

where  $\zeta_T = \zeta \text{Max}\left[1, \|\mathbf{T}\|\right]$ . This is a radially symmetric Gaussian sensitivity function with a bandwidth that increases in proportion to frequency (except at DC). This is analogous to the spatial frequency channels that are believed to underlie the early stages of human visual processing.

Best fitting (method of least squares) values for  $\zeta$ ,  $w$  and  $t_T$ , as determined for Model 2, are also given in Table 3. The average variance (squared rms error on a decibel scale) from Models 0, 1 and 2 is also provided in Table 3. The best fitting predictions of Model 2 are also drawn as solid lines in Figs. 6 and 7.

Parameter	Model 0	Model 1	Model 2
-----------	---------	---------	---------

$\text{dB}[t_{00}]$	-32.9	-32.9	-35.1
$\text{dB}[t_{01}]$	-29.2	-30.2	-32.6
$\text{dB}[t_{03}]$	-27.4	-27.8	-31.9
$\text{dB}[t_{07}]$	-20.5	-20.9	-22.1
$w$	n/a	0.324	0.396
$\zeta$	n/a	n/a	5.5
Average variance from model	18.5	16.5	8.95

Table 3. Residual variance from Models 0, 1 and 2.

### 3.4 Conclusions

With the addition of a single parameter ( $\zeta$ ), our Model 1 captures 46% more of the variance in our data than does Model 0. Incorporating this modification into the current method for computing DCT quantization matrices will yield more efficient image compression. The estimated value of  $\zeta$  indicates a rather broad bandwidth for the masking effect. This may be due in part to the rather broad bandwidth of the basis functions themselves.

## 4. ACKNOWLEDGMENTS

We thank Mark Young for extensive assistance and Heidi Peterson for useful discussions. This work was supported by NASA RTOPs 506-59-65 and 505-64-53.

## 5. REFERENCES

1. W.B. Pennebaker and J.L. Mitchell, JPEG Still image data compression standard, Van Nostrand Reinhold, New York (1993).
2. G. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, 34(4), 30-44 (1991).
3. H.A. Peterson, "DCT basis function visibility in RGB space," (1992).
4. H.A. Peterson, H. Peng, J.H. Morgan and W.B. Pennebaker, "Quantization of color image components in the DCT domain," (1991).
5. A.J. Ahumada Jr. and H.A. Peterson, "Luminance-Model-Based DCT Quantization for Color Image Compression," (1992).
6. H. Peterson, A. Ahumada and A. Watson, "An Improved Detection Model for DCT Coefficient Quantization," (1993).
7. A.B. Watson, "DCT quantization matrices visually optimized for individual images," (1993).
8. A.B. Watson and D.G. Pelli, "QUEST: A Bayesian adaptive psychometric method," *Perception and Psychophysics*, 33(2), 113-120 (1983).
9. A.B. Watson, "Probability summation over time," *Vision Research*, 19, 515-522 (1979).
10. C. Noorlander, M.J.G. Heuts and J.J. Koenderink, "Influence of the target size on the detection threshold for luminance and chromaticity contrast," *Journal of the Optical Society of America*, 70(9), 1116-1121 (1980).
11. C.H. Graham, R.H. Brown and F.A. Mote, "The relation of size of stimulus and intensity in the human eye: I. Intensity thresholds for white light," *J. Exp. Psychol.*, 24, 555-573 (1939).
12. H.B. Barlow, "Temporal and spatial summation in human vision at different background intensities," *Journal of Physiology*, 141, 337-350 (1958).
13. N. Graham, J.G. Robson and J. Nachmias, "Grating summation in fovea and periphery," *Vision Research*, 18, 815-825 (1978).
14. J.G. Robson and N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," *Vision Research*, 21, 409-418 (1981).

15. H.A. Peterson, A.J. Ahumada Jr. and A.B. Watson, "The Visibility of DCT Quantization Noise," SID Digest of Technical Papers, XXIV, 942-945 (1993).
16. G.E. Legge and J.M. Foley, "Contrast masking in human vision," Journal of the Optical Society of America, 70(12), 1458-1471 (1980).
17. G.E. Legge, "A power law for contrast discrimination," Vision Research, 21, 457-467 (1981).
18. C.F. Stromeyer III and S. Klein, "Spatial frequency channels in human vision as asymmetric (edge) mechanisms," Vision Research, 14, 1409- 1420 (1974).
19. J. Nachmias and R. Sansbury, "Grating contrast: discrimination may be better than detection," Vision Research, 14 , 1039-1042 (1974 ).