

Solomon, J. A., Watson, A. B. & Ahumada, A. (1994). Visibility of DCT basis functions: Effects of contrast masking. Proceedings, Data Compression Conference, Snowbird, Utah: IEEE Computer Society Press, 361-370.

Visibility of DCT basis functions: Effects of contrast masking.

Joshua A. Solomon Andrew B. Watson Albert Ahumada

NASA Ames Research Center
Moffett Field, CA 94035-1000
jsolomon@vision.arc.nasa.gov

INTRODUCTION

DCT image compression. The discrete cosine transform (DCT) is a standard method by which images may be compressed. Initially, the image is divided into 8x8-pixel blocks, each of which is then transformed into its DCT, which is an 8x8 matrix of coefficients. Each coefficient c_{ij} , specifies the magnitude of a DCT frequency (or basis function), indexed by i, j . For brevity, we adopt a vector notation for the frequencies, so $\mathbf{U} = \{i, j\}$ and $c_{\mathbf{U}} = c_{ij}$. Image

compression occurs when each coefficient block is quantized by dividing it, coefficient by coefficient, by a quantization matrix (QM), and rounding to the nearest integer. For any given block of any image, there will be an 8x8 matrix of remainders produced by this division. The distortion introduced by this method of image compression can be expressed as an error image, which is the inverse transform of each of these remainder matrices. The magnitude of the entries in the QM determine the image quality and bit-rate. Small entries yield little compression and little distortion; large entries yield much compression and much distortion. To effectively compress an image, one must determine the QM which yields the desired bit-rate/distortion trade-off.

The contrast sensitivity function. The human visual system has different sensitivities to different DCT basis functions. Peterson, Peng, Morgan and Pennebaker¹ displayed images of replicated DCT basis functions. For a given ambient luminance, Peterson *et al.* determined the magnitude of the basis function required for human detection. Iteration of this technique with different coefficients yielded a contrast sensitivity function over the range of DCT basis functions. This contrast sensitivity function suggests a QM which is, to some extent, consistent with the human visual system.

Display parameters. Recognizing that the human DCT contrast sensitivity function is itself a function of display parameters such as display luminance and pixel size, Ahumada *et al.*^{2,3} performed the measurements required to specify the contrast sensitivity function—and consequently, a more efficient QM—for any combination of these parameters.

Contrast masking. Watson⁴ 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 seminal work by Legge and Foley^{5,6}) for contrast masking: Given a DCT coefficient c_τ and a corresponding absolute threshold t_τ , the masked threshold m_τ will be

$$m_\tau = t_\tau \text{Max} \left[1, \left| \frac{c_\tau}{t_\tau} \right|^{w_\tau} \right], \quad (1)$$

where w_τ is an exponent that lies between 0 and 1. In the sequel, we will refer to this model as Model 0. In Model 0, 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 0 to account for between-coefficient contrast masking.

METHOD

Apparatus. Stimuli were displayed on a monochrome CRT with a background luminance of 40 cd m⁻². The (binocular) viewing distance was 97.4 cm. At this distance, the spatial extent of the screen was 16° by 16° of visual angle. There was no other source of illumination.

Stimuli. Each stimulus was the sum of a test basis function and a mask basis function. Every pixel of the resultant image was magnified by a factor of 2, both vertically and horizontally, and added to the background luminance of the monitor. Each stimulus subtended 0.25 degrees by 0.25 degrees.

Procedure. Psychophysical threshold estimates were obtained with a version of the Quest procedure⁷. Each trial consisted of two stimulus presentations, each marked by an auditory tone and modulated by a temporal Gaussian with 0.53 seconds between $\epsilon^{-\pi}$ points, and separated by 0.5 seconds. The observer identified the test interval by pressing one of two keys. A correct choice was identified by a low frequency tone; an incorrect choice by a high frequency tone. Each plotted symbol in the following figures represents a maximum likelihood estimate of threshold based on a session of 64 trials.

Masked thresholds m_{τ} for four test basis functions were measured as a function of masking contrast for three different masks. The tests, as identified by their DCT 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 $\{0,3\}$. Absolute threshold t_{τ} was also determined for each test. Theoretically, DCT coefficients can assume any real value. In the current study we use coefficients $c_{\mathbf{U}}$, such that $0 \leq c_{\mathbf{U}} \leq 1$. 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 ($\text{dB}[c_{\mathbf{U}}] = 20 \text{Log}_{10}[c_{\mathbf{U}}]$): -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.

An experimental run, lasting about 10 minutes, was devoted to a single test/mask combination and a single masking contrast. Test/mask/masking-contrast combinations were run in a randomized fashion. The first author was the only observer in these experiments.

RESULTS AND DISCUSSION

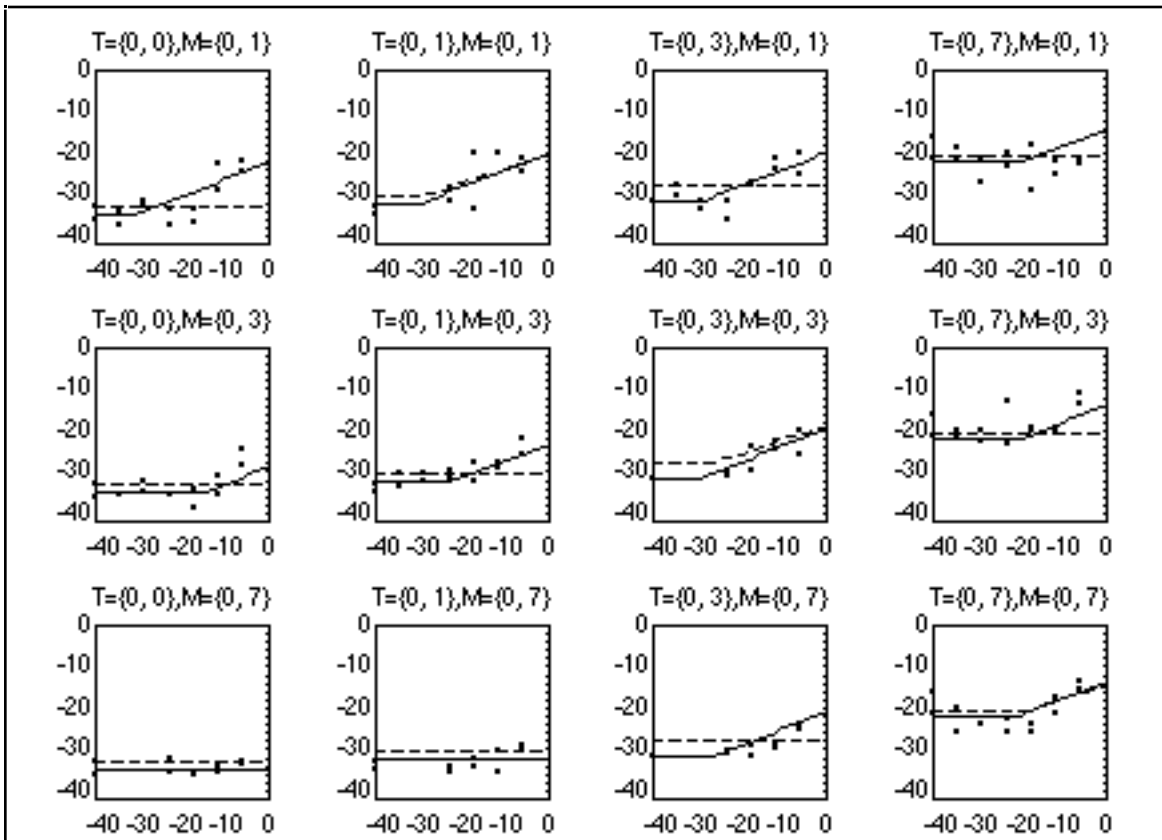


Fig. 1. Masked thresholds ($\text{dB}[m_T]$) for four test basis functions are plotted as a function of masking contrast ($\text{dB}[C_M]$) for three different masks. Unmasked thresholds ($\text{dB}[t_T]$) for the test basis functions are plotted on the ordinates. The dashed and solid lines are the predictions of Models 0 and 1, respectively, as described in the text.

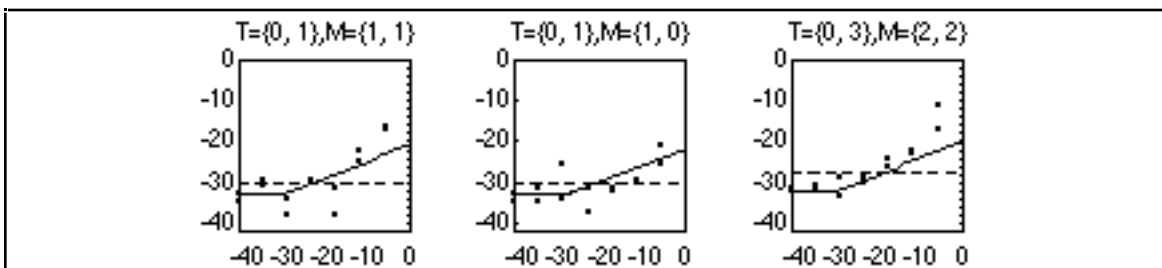


Fig. 2. Masked thresholds for test $\{0,1\}$ as a function of masking contrast for the masks $\{1,1\}$ and $\{1,0\}$, and for test $\{0,3\}$ as a function of masking contrast for the mask $\{2,2\}$.

The dipper effect. The results are plotted in Figs. 1 and 2. Data gathered with the $\{0,1\}/\{0,1\}$ test/mask combination at masking contrasts of -36 and -30 dB have been omitted from the graphs and further analysis. Similarly, we have omitted the $\{0,3\}/\{0,3\}$ data at -36 and -30 dB. 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^{5,8,9}. 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, exclusion of the "dipper data" will benefit our initial approximations.

Model 0. Model 0 was fit to the data. Model 0 does not predict 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_τ 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 0, we mean specifically: Given a test DCT basis function c_τ , its corresponding absolute threshold t_τ and a mask DCT basis

function c_M the masked threshold m_T will be

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

where w is an exponent that lies between 0 and 1. Best fitting (method of least squares) values for w and the t_T s, as determined for Model 0, are given in Table 1. Model 0 adequately reflects the data for the viewing conditions in which the mask and target were identical. However, it cannot reflect the between-coefficient masking evident by the increase in measured threshold with masking contrast for the other test/mask combinations.

Model 1. In order to reflect the between-coefficient masking, we propose the following revision of Watson's original model for contrast masking. It will be referred to hereafter as Model 1. Given a test DCT basis function c_T , its corresponding absolute threshold t_T and a mask DCT basis function c_M the masked threshold m_T will be

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

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

$$f(T, M) = e^{-\pi \|T-M\|^2 / \sigma_T^2}, \quad (4)$$

where $\sigma_T = \sigma \text{Max} [1, \|T\|]$. Best fitting (method of least squares) values for σ , w and the t_T s, as determined for Model 1, are also given in Table 1.

Model 0 vs Model 1. The average variance (squared rms error on a decibel scale) from Models 0 and 1 is also provided in Table 1.

Table 1

Parameter	Model 0	Model 1
dB[t_{00}]	-32.9	-35.1
dB[t_{01}]	-30.2	-32.6
dB[t_{03}]	-27.8	-31.9
dB[t_{07}]	-20.9	-22.1
w	0.324	0.396
σ	n/a	5.50
Average variance from model	16.5	8.95

CONCLUSION

With the addition of a single parameter (σ), our model for contrast masking captures 46% more of the variance in our data than Watson's original model. Incorporating this modification into the current method for computing DCT quantization matrices will yield more efficient image compression.

ACKNOWLEDGEMENTS

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

REFERENCES

¹H. A. Peterson, H. Peng, J. H. Morgan and W. B. Pennebaker. "Quantization of color image components in the DCT domain," in Human Vision, Visual

Processing, and Digital Display II, B. E. Rogowitz, ed. (Proceedings of the SPIE, 1991).

²A. J. Ahumada Jr. and H. A. Peterson. "Luminance-Model-Based DCT Quantization for Color Image Compression," in Human Vision, Visual Processing, and Digital Display III, B. E. Rogowitz, ed. (Proceedings of the SPIE, 1992).

³H. A. Peterson, A. J. Ahumada Jr. and A. B. Watson. "An improved detection model for DCT coefficient quantization," (Proceedings of the SPIE, 1993).

⁴A. B. Watson. "DCT quantization matrices visually optimized for individual images," in Human Vision, Visual Processing, and Digital Display IV, B. E. Rogowitz, ed. (SPIE, Bellingham, WA, 1993).

⁵G. E. Legge and J. M. Foley. "Contrast masking in human vision," *Journal of the Optical Society of America*, 70, 1458-1471 (1980).

⁶G. E. Legge. "A power law for contrast discrimination," *Vision Research*, 21 457-467 (1981).

⁷A. B. Watson and D. G. Pelli. "QUEST: A Bayesian adaptive psychometric method," *Perception and Psychophysics*, 33, 113-120 (1983).

⁸J. Nachmias and R. V. Sansbury. "Grating contrast: discrimination may be better than detection," *Vision Research*, 14, 1039-1042 (1974).

⁹C. F. Stromeyer, III and S. Klein. "Spatial frequency channels in human vision as asymmetric (edge) mechanisms," *Vision Research*, 14, 1409- 1420 (1974).

REFERENCES

- ¹H. A. Peterson, H. Peng, J. H. Morgan and W. B. Pennebaker. "Quantization of color image components in the DCT domain," in Human Vision, Visual Processing, and Digital Display II, B. E. Rogowitz, ed. (Proceedings of the SPIE, 1991).
- ²A. J. Ahumada Jr. and H. A. Peterson. "Luminance-Model-Based DCT Quantization for Color Image Compression," in Human Vision, Visual Processing, and Digital Display III, B. E. Rogowitz, ed. (Proceedings of the SPIE, 1992).
- ³H. A. Peterson, A. J. Ahumada Jr. and A. B. Watson. "An improved detection model for DCT coefficient quantization," (Proceedings of the SPIE, 1993).
- ⁴A. B. Watson. "DCT quantization matrices visually optimized for individual images," in Human Vision, Visual Processing, and Digital Display IV, B. E. Rogowitz, ed. (SPIE, Bellingham, WA, 1993).
- ⁵G. E. Legge and J. M. Foley. "Contrast masking in human vision," *Journal of the Optical Society of America*, 70, 1458-1471 (1980).
- ⁶G. E. Legge. "A power law for contrast discrimination," *Vision Research*, 21 457-467 (1981).

⁷A. B. Watson and D. G. Pelli. "QUEST: A Bayesian adaptive psychometric method," *Perception and Psychophysics*, 33, 113-120 (1983).

⁸J. Nachmias and R. V. Sansbury. "Grating contrast: discrimination may be better than detection," *Vision Research*, 14, 1039-1042 (1974).

⁹C. F. Stromeyer, III and S. Klein. "Spatial frequency channels in human vision as asymmetric (edge) mechanisms," *Vision Research*, 14, 1409- 1420 (1974).