CONTRIBUTION TO IEEE STANDARDS SUBCOMMITTEE

Document Number:

DATE: August 5, 1997

COMMITTEE: G-2.1.6 Compression and Processing Subcommittee

TITLE: Sarnoff JND Vision Model

EDITOR: Dr. Jeffrey Lubin

SOURCE: Sarnoff Corporation

CONTACTS:

Dr. Jeffrey Lubin
Sarnoff Corporation
CN 5300
Princeton, NJ 08543-5300
(609)734-2678
(609)734-2662 (fax)
jlubin@sarnoff.com

David Fibush
Tektronix Inc., MS 50-353
P.O. Box 500
Beaverton, OR 97077
(503)627-6289
(503)627-1707 (fax)
davef@tv.tv.tek.com

DISTRIBUTION: G-2.1.6 Compression and Processing Subcommittee

ABSTRACT: This contribution describes the basic concepts of the Sarnoff JND Vision Model and comparisons with subjective rating data. The model is a method of predicting the perceptual ratings that human subjects will assign to a degraded color-image sequence relative to its nondegraded counterpart. The model takes in two image sequences and produces several difference estimates, including a single metric of perceptual differences between the sequences. These differences are quantified in units of the modeled human just-noticeable difference (JND). Sarnoff Corporation and Tektronix are cooperating on development of a measurement instrument using this model and expect to submit the operational algorithm for consideration by the ITU-T Joint Rapporteurs Group on subjective and objective audiovisual quality.

NOTICE

numerical value after more study. Sarnoff Corporation and Tektronix Inc., specifically reserve the right to add to, or amend, the statements contained herein.

Sarnoff Corporation

ANSI SUBMISSION: SARNOFF JND VISION MODEL

Prepared by: Visual Information Systems Research Group
Information Sciences Laboratory
Sarnoff Corporation
CN 5300
Princeton, NJ 08543-5300

July 29, 1997

TABLE OF CONTENTS

TABLE OF FIGURES

1. System Overview

The Sarnoff JND Vision Model is a method of predicting the perceptual ratings that human subjects will assign to a degraded color-image sequence relative to its nondegraded counterpart. The model takes in two image sequences and produces several difference estimates, including a single metric of perceptual differences between the sequences. These differences are quantified in units of the modeled human just-noticeable difference (JND). A version of the model that applies only to static, achromatic images is described by Lubin (1993, 1995).

The Sarnoff Vision Model can be useful in a general context (see Figure 1). An input video sequence passes through two different channels on the way to a human observer (not shown in the figure). One channel is uncorrupted (the reference channel), and the other distorts the image in some way (the channel under test). The distortion, a side effect of some measure taken for economy, can occur at an encoder prior to transmission, in the transmission channel itself, or in the decoding process. In Figure 1, the box called "system under test" refers schematically to any of these alternatives. Ordinarily, evaluation of the subjective quality of the test image relative to the reference sequence would involve the human observer and a real display device. This evaluation would be facilitated by replacing the display and observer by the JND model, which compares the test and reference sequences to produce a sequence of JND maps instead of the subjective comparison.

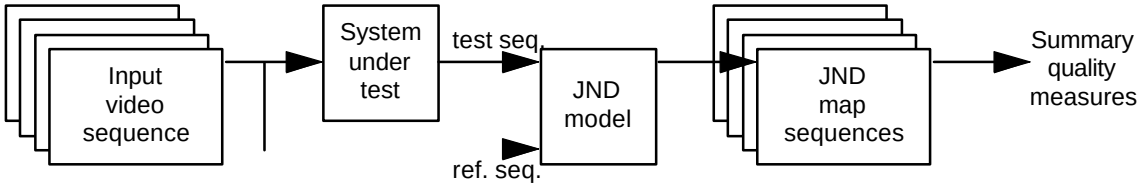

Figure 1. JND Model in System Evaluation

Figure 2 shows an overview of the algorithm. The inputs are two image sequences of arbitrary length. For each field of each input sequence, there are three data sets, labeled Y', C_b', and C_r' at the top of Figure 2 derived, e.g., from a

D1 tape. Y, C_b, C_r data are then transformed to R', G', and B' electron-gun voltages that give rise to the displayed pixel values. In the model, R', G', B' voltages undergo further processing to transform them to a luminance and two chromatic images that are passed to subsequent stages.

The purpose of the front-end processing is to transform video input signals to light outputs, and then to transform these light outputs to psychophysically defined quantities that separately characterize luma and chroma.

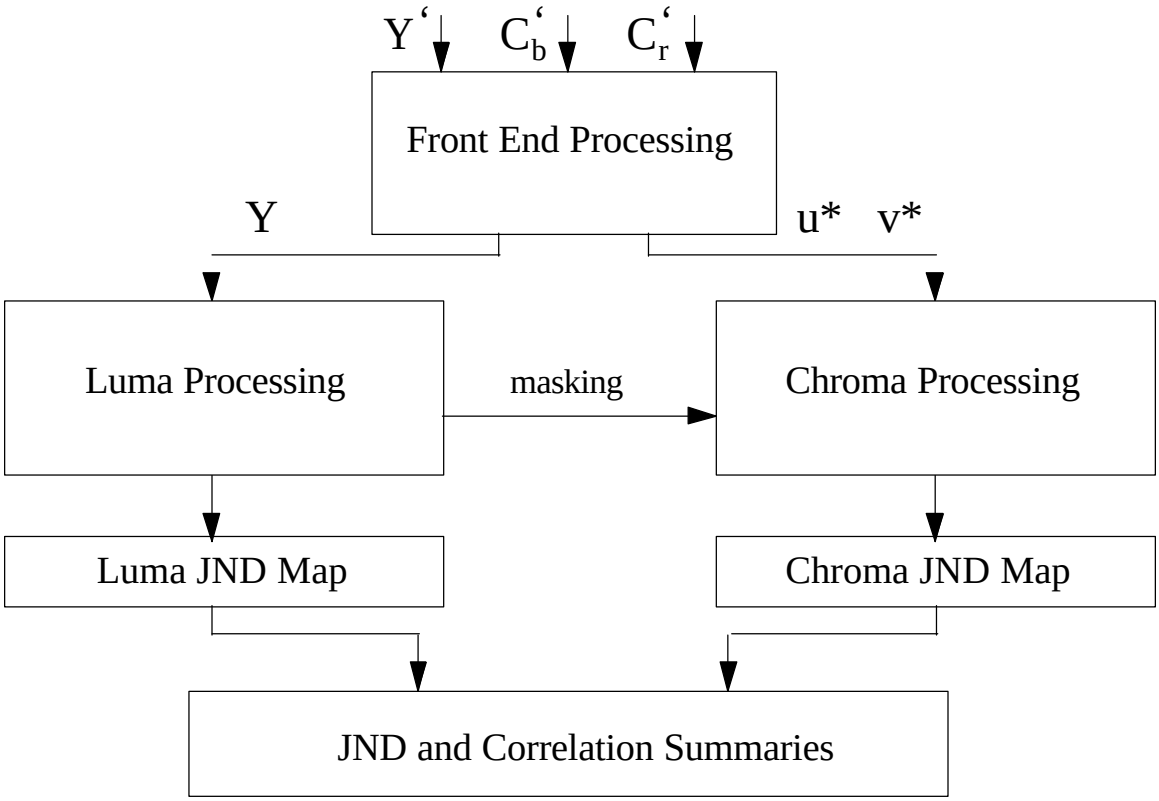

Figure 2. Sarnoff JND Vision Algorithm Flow Chart

A luma-processing stage accepts two images (test and reference) of luminances Y, expressed as fractions of the maximum luminance of the display. From these inputs, the luma-processing stage generates a luma JND map. This map is an image whose gray levels are proportional to the number of JNDs between the test and reference image at the corresponding pixel location.

Similar processing, based on the CIE L*u*v* uniform-color space, occurs for each of the chroma images u* and v*. Outputs of u* and v* processing are

combined to produce the chroma JND map. Both chroma and luma processing are influenced by inputs from the luma channel called *masking*, which render perceived differences more or less visible depending on the structure of the luma images.

The chroma and luma JND maps are each available as output, together with a small number of summary measures derived from these maps. Whereas the single JND value output is useful to model an observer's overall rating of the distortions in a test sequence, the JND maps give a more detailed view of the location and severity of the artifacts.

It should be noted that two basic assumptions underlie the model presented here:

(a) Each pixel is square and subtends .03 degrees of viewing angle. This number was derived from a screen height of 480 pixels, and a viewing distance of four screen-heights (the closest viewing distance prescribed by the *Rec. 500* standard). When the model is compared with human perception at longer viewing distances than four screen heights, the model overestimates the human's sensitivity to spatial details. In the absence of hard constraints on viewing distance, we chose to make the model as sensitive as possible within the recommendations of the *Rec 500*. (See Section 3.1.2 & 3.2)

(b) The model applies to screen luminances of .01 to 100 ft-L (for which overall sensitivity was calibrated), but with greatest accuracy at about 20 ft-L (for which all spatiotemporal frequencies were calibrated). It is assumed that changing luminance incurs proportional sensitivity changes at all spatiotemporal frequencies, and this assumption is less important near 20 ft-L, where more calibration took place.

The processing shown in certain of the boxes in Figure 2 is described in more detail below, keyed to Figures 3, 4, and 5.

2. Algorithm Overview

2.1 Front End Processing

The stack of four fields labeled Y', C_b', C_r' at the top of Figure 3 indicates a set of four consecutive fields from either a test or reference image sequence. The first stage of processing transforms Y', C_b', C_r' data, to R', G', B' gun voltages. Currently, multiple transformations are included in the software.

The second stage of processing, applied to each R', G', B' image, is a point-nonlinearity. This stage models the transfer from R', G', B' gun voltages to model-intensities (R, G, B) of the display (fractions of maximum luminance). The nonlinearity also performs clipping at low luminances in each plane by the display.

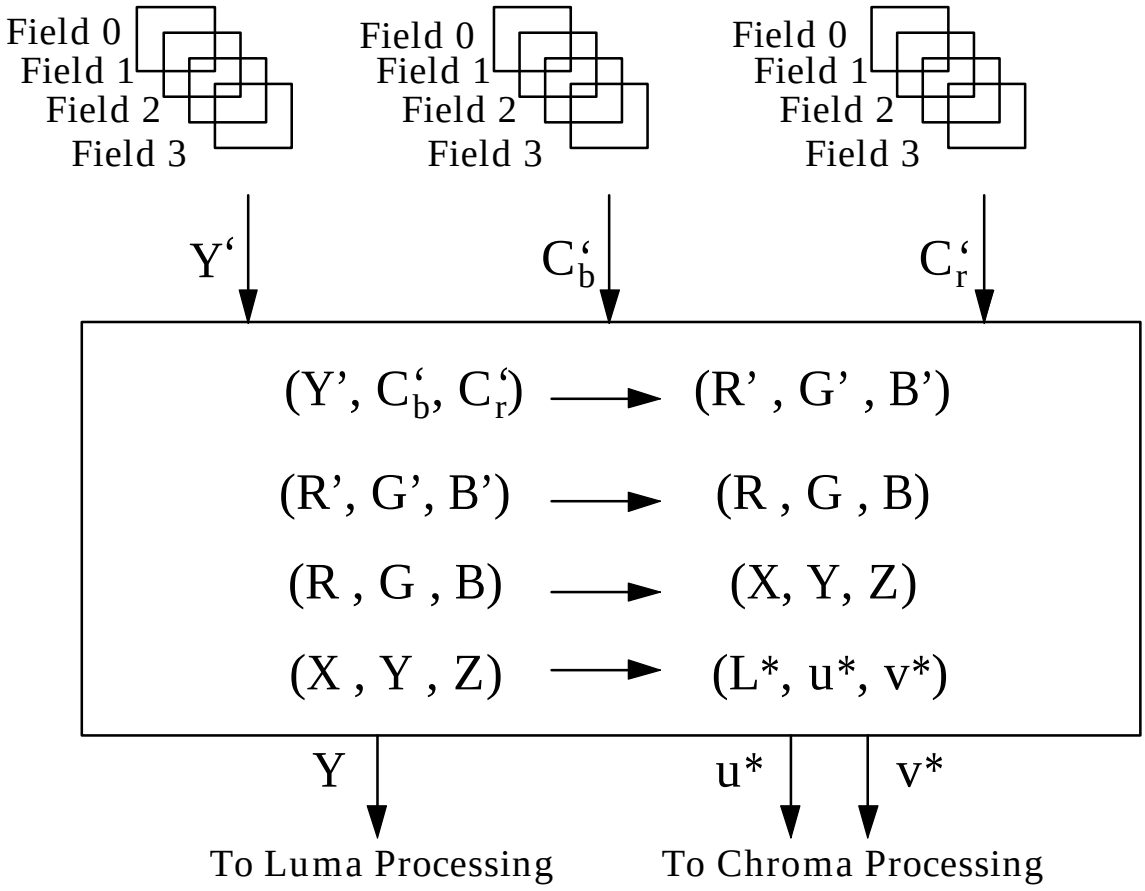

Figure 3. Front End Processing Overview

Following the nonlinearity, vertical electron-beam spot spread into interline locations is modeled by replacing the interline values in fields R, G, B by interpolated values from above and below. Then, the vector (R,G,B) at each pixel in the field is subjected to a linear transformation (which depends on the display phosphors) to CIE 1931 tristimulus coordinates (X, Y, Z). The luminance component Y of this vector is passed to luma processing.

To ensure (at each pixel) approximate perceptual uniformity of the color space to isoluminant color differences, we map the individual pixels into CIELUV, an international-standard uniform-color space (see Wyszecki and Stiles, 1982). The chroma components u^*, v^* of this space are passed to the chroma processing steps in the model. [1]

[1] The luminance channel L^* from CIELUV is not used in luma processing, but instead is replaced by a visual nonlinearity for which the vision model has been calibrated over a range of

2.2 Luma Processing

As shown in Figure 4, each luma value is first subjected to a compressive nonlinearity. Then, each luma field is filtered and down-sampled in a four-level Gaussian pyramid (Burt and Adelson, 1983), in order to model the psychophysically and physiologically observed decomposition of incoming visual signals into different spatial-frequency bands. After this decomposition, the bulk of subsequent processing by the model consists of similar operations (e.g., oriented filtering) performed at each pyramid level.

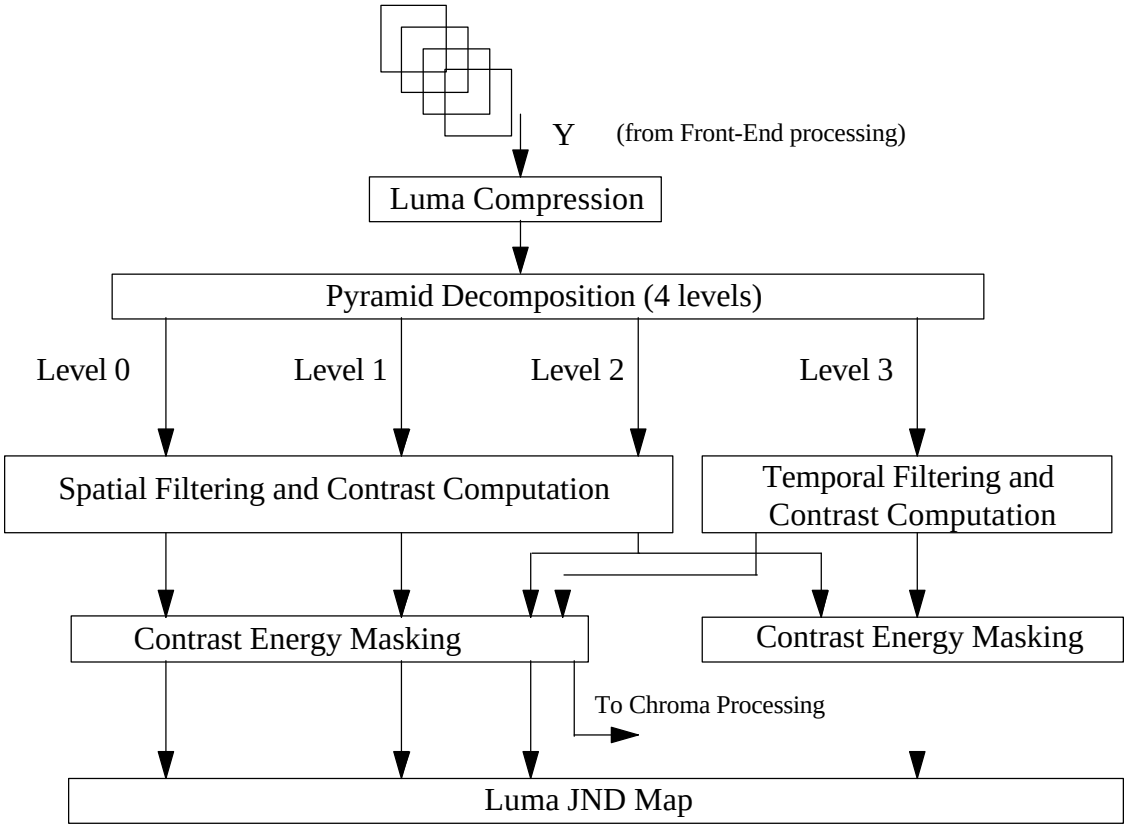

Figure 4. Luma Processing Overview

After this pyramid-making process, the lowest-resolution pyramid image is subjected to temporal filtering and contrast computation, and the other three levels are subjected to spatial filtering and contrast computation. In each case

luminance values. L* is used in chroma processing, however. to create a chroma metric that is approximately uniform and familiar to display engineers.

the contrast is a local difference of pixel values divided by a local sum, appropriately scaled. In the initial formulation of the model, this established the definition of 1 JND, which was passed on to subsequent stages of the model.[2] (Calibration iteratively revises the 1-JND interpretation at intermediate model stages. This is discussed in Section 3.1.) The absolute value of the contrast response is passed to the following stage, and the algebraic sign is preserved for reattachment just prior to image comparison (JND map computation).

The next stage (contrast masking) is a gain-setting operation in which each oriented contrast response is divided by a function of all the contrast responses. This combined attenuation of each response by other local responses is included to model visual "masking" effects such as the decrease in sensitivity to distortions in "busy" image areas (Nachmias and Sansbury, 1974). At this stage in the model, temporal structure (flicker) is made to mask spatial differences, and spatial structure is also made to mask temporal differences. Luma masking is also applied on the chroma side, as discussed below.

The masked contrast responses (together with the contrast signs) are used to produce the Luma JND map. This is done by:

- separating each image into positive and negative components (half-wave rectification)
- performing local pooling (averaging and downsampling, to model the local spatial summation observed in psychophysical experiments)
- evaluating the absolute image differences channel by channel
- up-sampling to the same resolution (which will be half the resolution of the original image due to the pooling stage).
- evaluating the sum over all channels, and also (in parallel) the maximum in all channels
- evaluating a linear combination of the channel-sum and the channel-maximum (an approximation to a Minkowski norm that was necessitated by limits on integer implementation)

[2] The association of a constant contrast with 1 JND is an implementation of what is known as Weber's law for vision.

2.3 Chroma Processing

Chroma processing parallels luma processing in several ways. Intra-image differences of chroma (u* and v*) of the CIELUV space are used to define the detection thresholds for the chroma model, in analogy to the way the Michelson contrast (and Weber's law) is used to define the detection threshold in the luminance model. Also, in analogy with the luminance model, the chromatic "contrasts" defined by u^* and v^* differences are subjected to a masking step. A transducer nonlinearity makes the discrimination of a contrast increment between one image and another depend on the contrast response that is common to both images.

Figure 5 shows that, as in luma processing, each chroma component u*, v* is subjected to pyramid decomposition. However, whereas luma processing needs only four pyramid levels, chroma processing is given seven levels. This captures the empirical fact that chromatic channels are sensitive to far lower spatial frequencies than luma channels (Mullen, 1985). Also, it takes into account the intuitive fact that color differences can be observed in large, uniform regions.

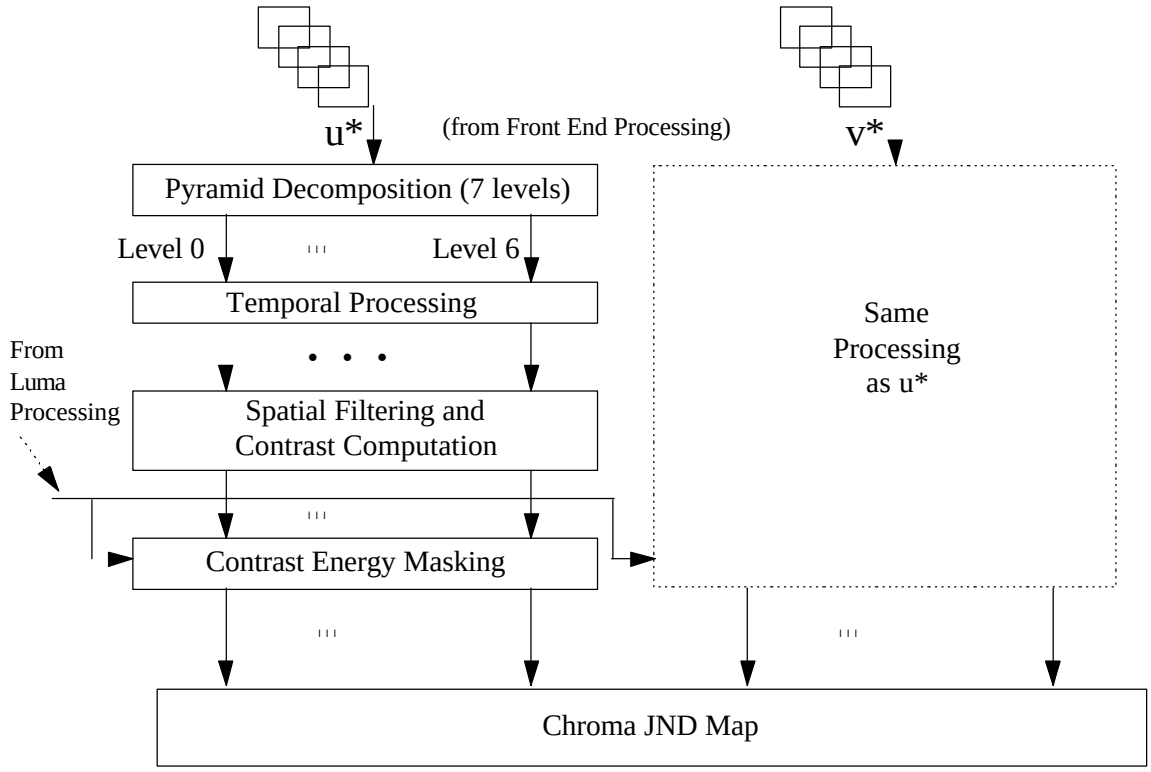

Figure 5. Chroma Processing Overview

To reflect the inherent insensitivity of the chroma channels to flicker, temporal processing is accomplished by averaging over four image fields.

Then, spatial filtering by a Laplacian kernel is performed in u* and v* . This operation produces a color difference in u*, v*, which (by definition of the uniform color space) is metrically connected to just-noticeable color differences. A value of 1 at this stage is taken to mean a single JND has been achieved, in analogy to the role of Weber's-law-based contrast in the luma channel. (As in the case of luma, the 1-JND chroma unit must undergo reinterpretation during calibration.)

This color difference value is weighted, absolute-valued, and passed (with the contrast algebraic sign) to the contrast-masking stage. The masking stage performs the same function as it did in the luma model. It is somewhat simpler, since it receives input only from the luma channels and from the chroma channel whose difference is being evaluated. Finally, the masked contrast responses are processed exactly as in the luma model (see last paragraph of Section 2.2).

2.4 JND Output Summaries

For each field in the video-sequence comparison, the luma and chroma JND maps are first reduced to single-number summaries, namely luma and chroma Picture Quality Ratings (PQRs). In each case, the reduction from map to number is done by histogramming the JND values for all pixels above a threshold, and then adopting the 90-percent value as the PQR value. Then, the luma and chroma PQR numbers are combined, again via a linear combination of a sum and a maximum, to produce the PQR estimate for the field being processed. A single performance measure for many fields of a video sequence is determined from the single-field PQRs by evaluating the 90th percentile of either:

- the actual histogram of single-field PQR values

or, for short sequences,

- a "faux histogram" fit to the actual histogram.

3. Luma Calibration and Prediction

Psychophysical data were used for two purposes, to calibrate the luma model (i.e., to determine values for certain model parameters), and to confirm the predictive value of the luma model once it was calibrated. In all cases, the stimuli were injected into the model as Y-value images immediately prior to the luma processing.

3.1 Calibration

The luma model was calibrated iteratively, using three sets of data. One data set was used to adjust the constants in the luma-compression step. The second data set was used to adjust the filtering and contrast of the mode (see Figure 4). The third set was used to adjust the masking-stage constants.

The details of the luma-compression adjustments are described in the subsections below.

3.1.1 Adjustment of luma-compression constants

The model predictions for sine-wave contrast sensitivity at various luminances were matched to contrast-sensitivity data for 8-cycles / deg sine waves presented by Van Nes et al. (1967) and reproduced by R. J. Farrell and J. M. Booth, *Design Handbook for Imagery Interpretation Equipment* (Boeing Aerospace Co., 1975, Report D180-19063-1) Fig. 3.2-33. To generate points on the model-based curve, a low-amplitude spatial sine wave at 8-cycles / deg was presented as a test image to the model, and the contrast threshold for 1 JND output was assessed. In each case the reference image implicitly had a uniform field with the same average luminance as the test field.

It should be noted that the Van Nes data expressed the light level in trolands of retinal illumination, not in external luminance (cd / m^2). However, because Van Nes et al. used an artificial 2-mm pupil, the conversion from trolands to cd / m^2 became a simple division by p.

The fit of spatial contrast sensitivity to data (see Figure 6 for final fit) was used to adjust two luma-compression constants. The solid line in Figure 6 represents the sensitivity predictions of the model.

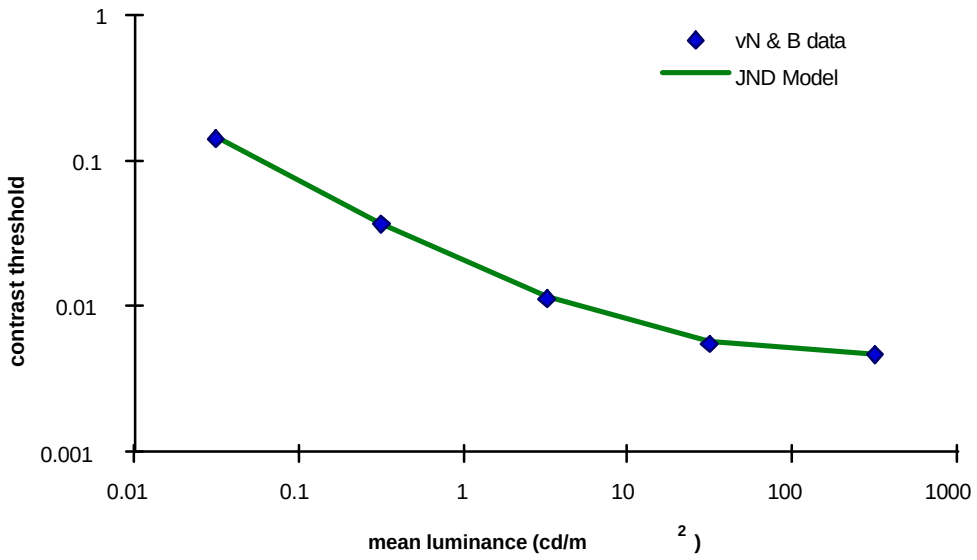

Figure 6. Luminance Contrast Sensitivity

It should be noted that the domain of the model fit was chosen to include the total range of luminances expected from the model display, or .01 to 100 ft-L. (The conversion from ft-L to cd / m^2 involves multiplying by 3.4).

3.1.2 Adjustment of contrast-normalization constants

The model predictions for spatial and temporal contrast sensitivities prior to masking were matched to contrast-sensitivity data for sine waves presented by Koenderink and Van Doorn (1979). To generate points on the model-based curve, a low-amplitude sine wave was presented as a test image to the model (either in space or in time), and the contrast threshold for 1 JND output was assessed. In each case the reference image implicitly had a uniform field with the same average luminance as the test field.

The fit of spatial contrast sensitivity to data (see Figure 7 for final fit) was used to adjust the contrast-pyramid sensitivity parameters. The dashed lines in Figure 7 represent the sensitivities of the separate pyramid channels that comprise the total sensitivity (solid line). It should be noted that the spatial model fit in Figure 7 was not extended beyond 15 cycles / deg, consistent with the viewing-distance constraint discussed in the System Overview: a viewing distance of four screen-heights. Similar adjustment of these constants could be performed to accommodate slightly different viewing distances; much greater viewing distances might require lower-resolution pyramid levels, and these could be easily incorporated at low computational expense.

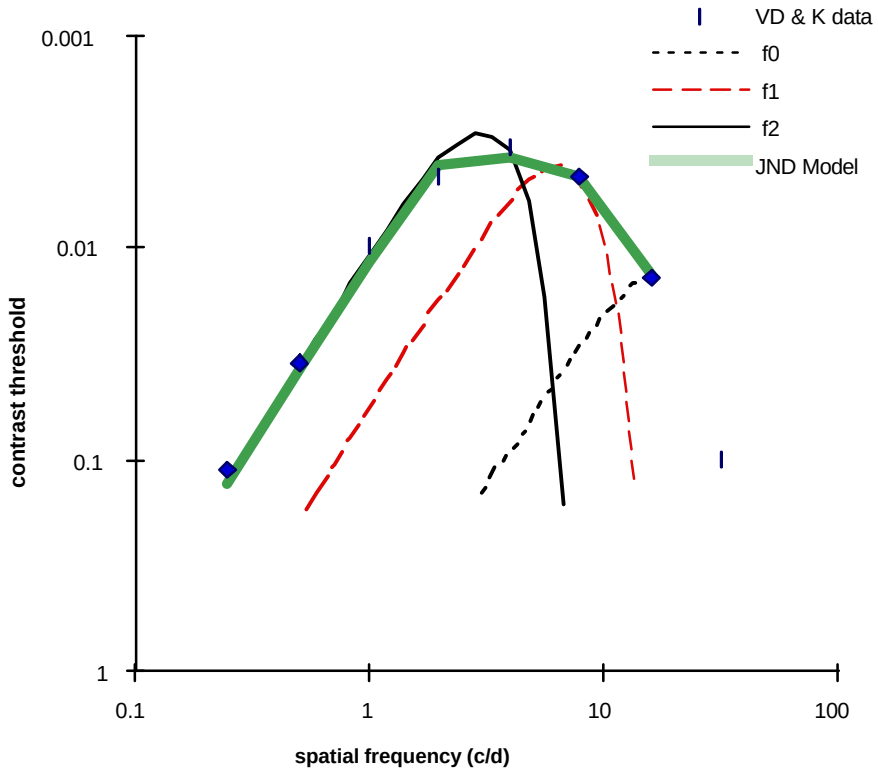

Figure 7. Luma Spatial Sensitivity

The fit of temporal contrast-sensitivity to data (see Figure 8 for final fits) was used to adjust the temporal filter-tap parameters, as well as the contrast-pyramid sensitivity parameter. The method used to fit these parameters is analogous to the spatial-contrast calibration. The lowest-spatial- frequency data of Van Doorn and Koenderink at various temporal frequencies were matched against the sensitivities computed for spatially uniform temporal sine waves. In each case, the vision-model field rate sampled the temporal sine wave at 50 and 60 Hz.

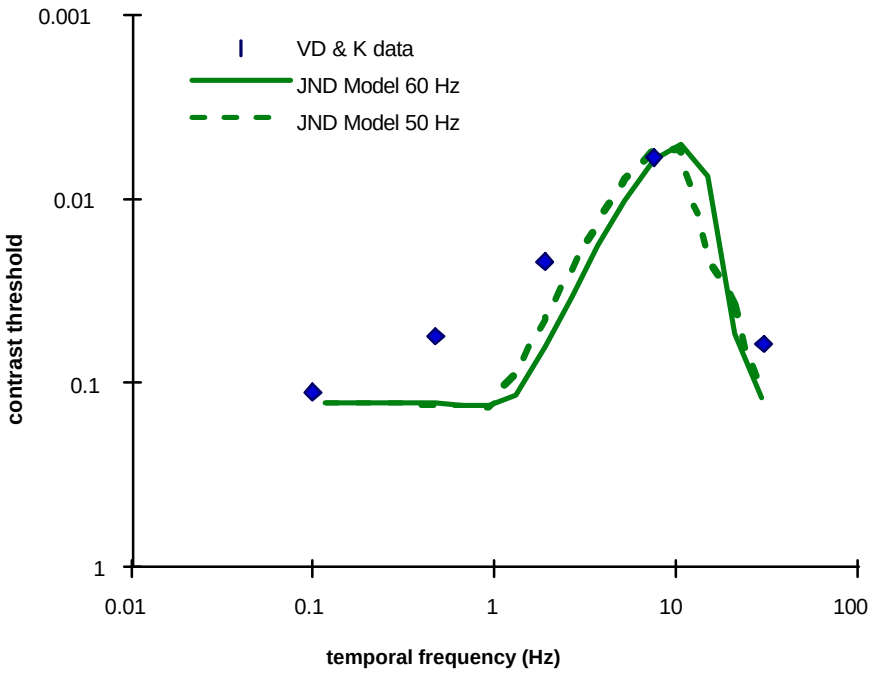

Figure 8. Luma Temporal Sensitivity

3.1.3 Adjustment of masking constants

The masking-parameter values were fit by comparing predictions for masked contrast discrimination with data acquired by Carlson and Cohen (1978). The results of the final-fit comparison appear in Figure 9. From the Carlson-Cohen study, a single observer's data was chosen subject to the criteria of being representative and also of having sufficient data points. In this case, the model stimulus consisted of a spatial sine wave of given pedestal contrast in both test and reference fields, and additionally a contrast increment of the test-field sine wave. The contrast-increment necessary to achieve 1 JND was determined from the model for each contrast-pedestal value, and then plotted in Figure 9.

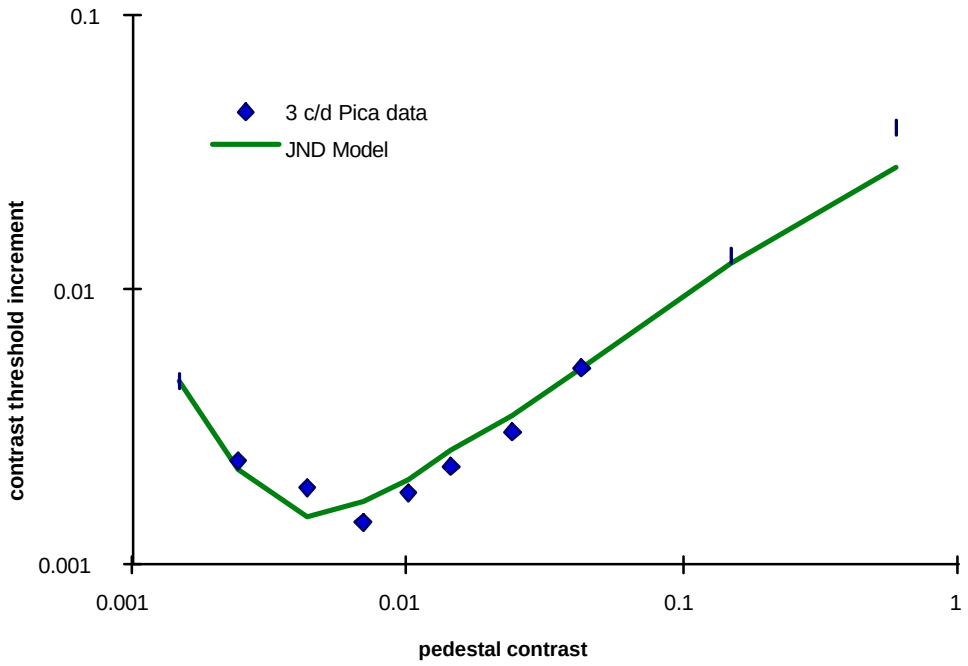

Figure 9. Luma Contrast Discrimination

3.2 Prediction

After model calibration, model predictions were compared with detection and discrimination data from stimuli that were not sine waves. This was done in order to check the transferability of the sine-wave results to more general stimuli. It will be seen from Figures 10, 11, and 12 that the predictions were not applied to patterns with nominal spatial frequencies above 10 cycles / deg. Such patterns would have had appreciable energies at spatial frequencies above 15 cycles / deg, and would have aliased with the pixel sampling rate (30 samples per degree - see System Overview).

In the first study (Figure 10), low-contrast disks in the test field were detected against a uniform reference field. The experimental data are from Blackwell and Blackwell (1971). In running the model for this particular study, it was necessary to replace the spatial summary measure with a maximum. Otherwise the JND result was sensitive to the size of the background of the disk (i.e., to image size).

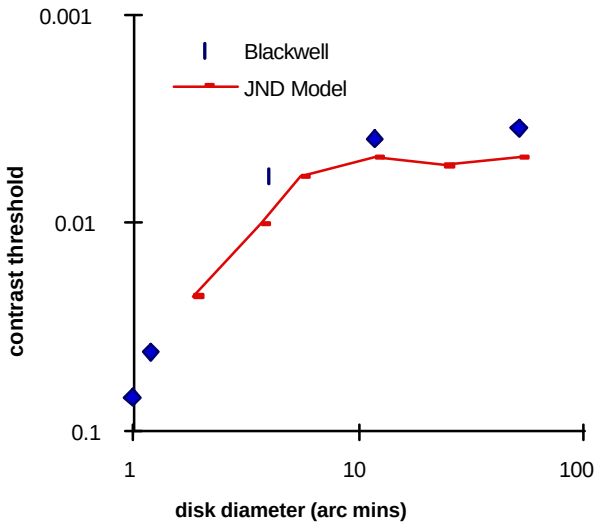

Figure 10. Disk Detection

In the second study (Figure 11), the detection of a low-amplitude checkerboard, the data was acquired in an unpublished study at Sarnoff. The third study (data from Carlson and Cohen, 1980) was somewhat different from the first two. A blurred edge given by erf(ax) was presented in the reference image, and discrimination was attempted against an edge given by erf(a'x) in the test image. Here, x is retinal distance in visual degrees, $a = \pi f / [\ln(2)]^{0.5}$, $a' = \pi (f + \Delta f) / [\ln(2)]^{0.5}$, and f is in cycles / deg. Here, Δf is the change in f required for one JND. The plot in Figure 12 is $\Delta f / f$ versus f.

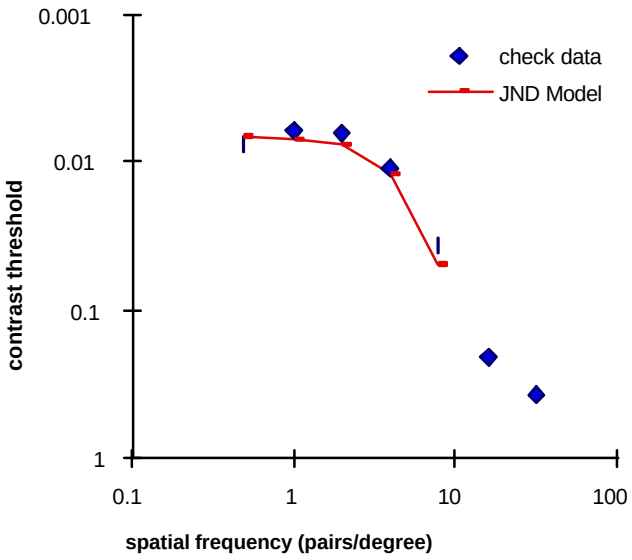

Figure 11. Checkerboard Detection

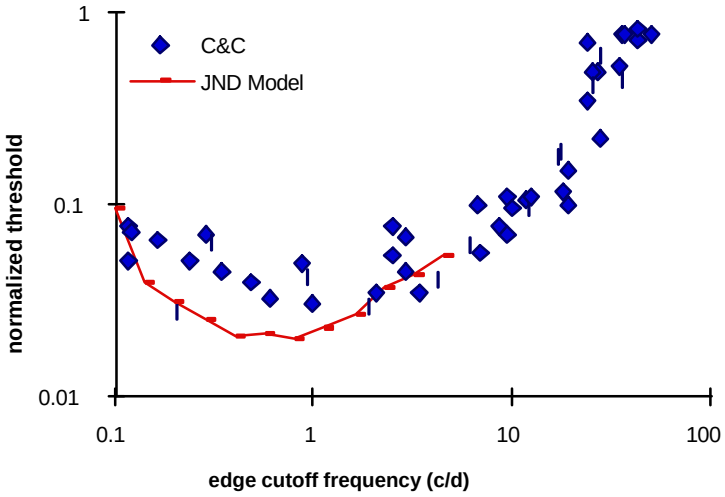

Figure 12. Edge Sharpness Discrimination

It can be seen that the model predictions are well fit to the data, for the range of spatial frequencies characteristic of the display at the four-screen-height viewing distance.

4. Chroma Calibration

As in luma-parameter calibration, psychophysical data were used to calibrate chroma parameters (i.e., to adjust their values for best model fits). In all cases, the stimuli were four equal fields, injected into the model as images in CIE X, Y, and Z just prior to conversion to CIELUV.

4.1 Adjustment of contrast-normalization constants

The model predictions for chromatic contrast sensitivities prior to masking were matched to contrast-sensitivity data presented by Mullen (1985). The test sequences used were four equal fields, each with a horizontally varying spatial sine-wave grating injected as (X, Y, Z) values. The data used for calibration were from Mullen's Figure 6, corresponding to which each test image was a red-green isoluminous sine-wave. At pixel i, the test-image sine wave had tristimulus values given by

$$X(i) = (Y_o / 2)\,\{(x_r / y_r + x_g / y_g) + \cos(2\pi f a i)\,\Delta m(x_r / y_r - x_g / y_g)\}$$

$$Y(i) = Y_o \ , \qquad (1)$$

$$Z(i) = (Yo / 2)\,\{(zr / yr + zg / yg) + \cos(2\pi f a i)\,\Delta m(zr / yr - zg / yg)\}.$$

Here Δm is the threshold incremental discrimination contrast, $(x_r, y_r) = (.636, .364)$ is the chromaticity of the red interference filter (at 602 nm) , $(x_g, y_g) = (.122, .823)$ is the chromaticity of the green interference filter (at 526 nm), $z_r = 1 - x_r - y_r$, $z_g = 1 - x_g - y_g$, and $a = .03$ deg / pixel . The reference-image is a uniform field represented by Equation (1) but with $\Delta m = 0$. For purposes of the model, it is sufficient to set $Y_o = 1$.

To generate points on the model-based curve, the above stimulus was presented at various values of f, and the contrast threshold Δm for 1 JND output was assessed. The fit of modeled chromatic-contrast sensitivity to data (see Figure 13 for final fit) was used to adjust the constrast-sensitivity parameters in the model.

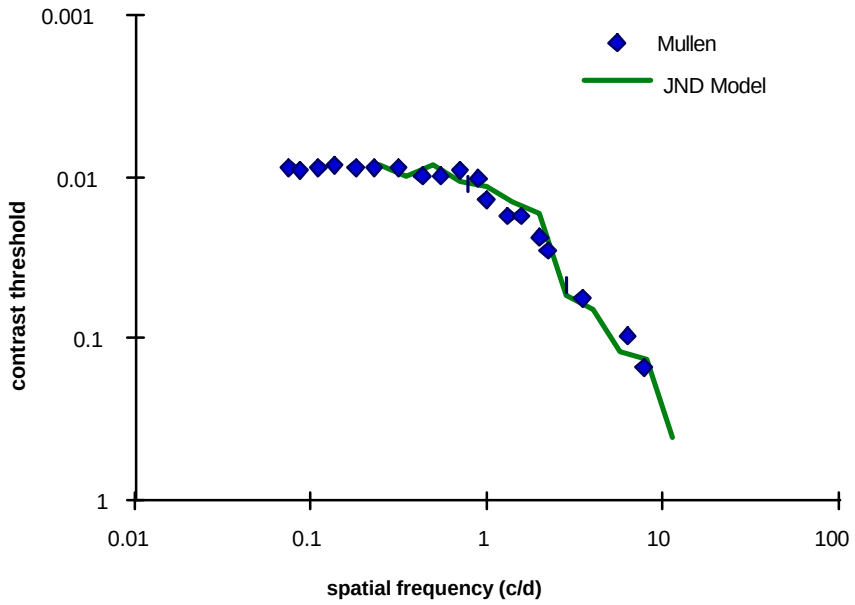

Figure 13. Chroma Spatial Sensitivity

4.2 Adjustment of masking constants

The model predictions for chroma masking were matched to data presented by Switkes. et al. (1988). The test sequences used were four equal fields, each with a horizontally varying spatial sine-wave grating injected as (X, Y, Z) values. To correspond with Figure 4 of that work (chroma masking of chroma), at pixel i, the test-image sine wave had tristimulus values given by

$$X(i) = (Y_o / 2)\,\{(x_r / y_r + x_g / y_g) + \cos(\ 2\pi\, f\, a\, i)\,[(m + \Delta m)(x_r / y_r - x_g / y_g)]\}$$

$$Y(i) = Y_o \qquad (2)$$

$$Z(i) = (Y_o / 2)\,\{(z_r / y_r + z_g / y_g) + \cos(\ 2\pi\, f\, a\, i)\,[(m + \Delta m)\,(z_r / y_r - z_g / y_g)]\},$$

where Δm is the threshold incremental discrimination contrast, $(x_r, y_r) = (.580, .362)$ is the chromaticity of the red phosphor, $(x_g, y_g) = (.301, .589)$ is the chromaticity of the green phosphor, $z_r = 1 - x_r - y_r$, $z_g = 1 - x_g - y_g$, and $f\, a = 2$ c / deg * .03 deg / pixel = .06. The reference-image sine wave is the same as the test-image sine wave but with $\Delta m = 0$. For purposes of the model, it is sufficient to set $Y_o = 1$.

To generate points on the model-based curve, the above stimulus was presented at various values of mask contrast m, and the contrast threshold Δ m for 1 JND output was assessed. The fit of modeled chromatic-contrast sensitivity to data (see Figure 14 for final fit) was used to adjust the chromatic transducer parameters in the model.

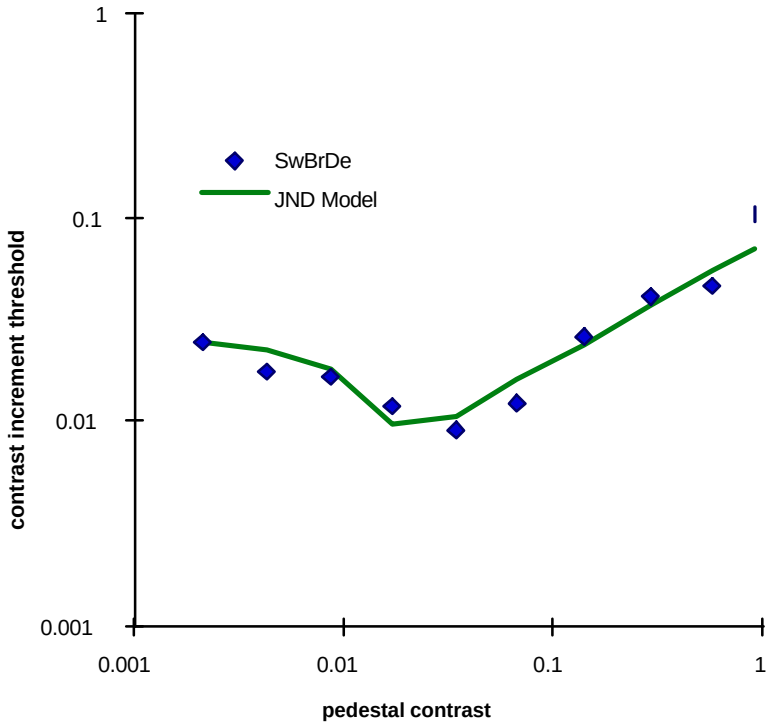

Figure 14. Chroma Contrast Discrimination

5. Comparisons with Rating Data

Four image sequences, each with various degrees of distortion, were used to compare the Sarnoff Vision Model with DSCQS rating data collected by an ISO / IEC group on high bit-rate (20 to 50 mbits / sec) MPEG-2 sequences that had been subjected to different numbers of compression and reconstruction, both with and without intervening spatial and / or temporal shifts. The results are plotted in Figure 15, and reveal a correlation 0.92 between the model and the data. For each of the sequences, the Vision Model processed 30 fields.

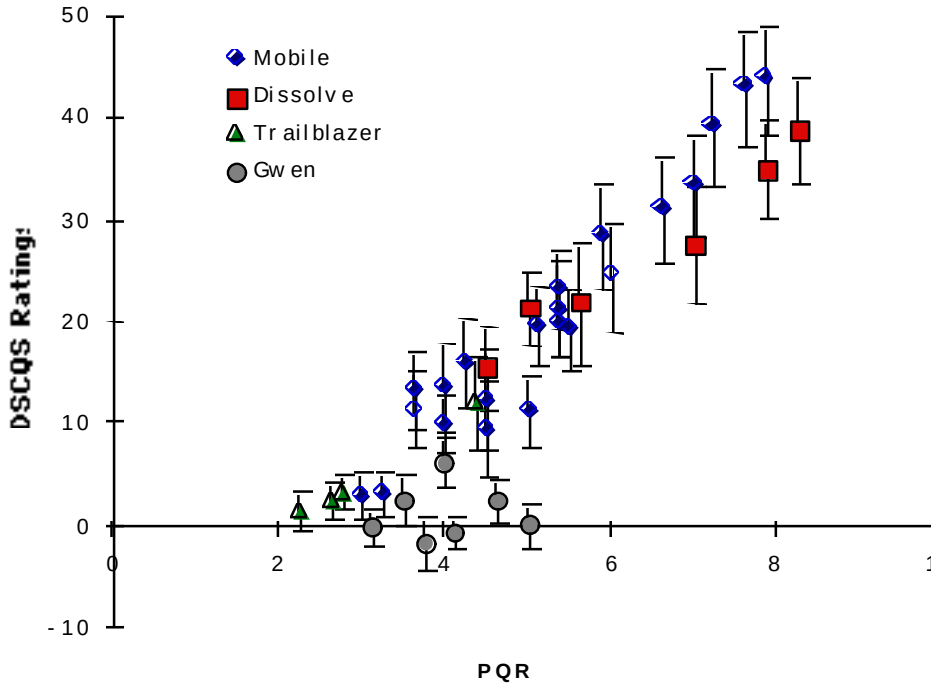

Figure 15. MPEG-2 Rating Predictions, 30 Fields Per Sequence.

In addition to these MPEG rating predictions, we have run the model on some lower bit rate MPEG-2 sequences (2 to 7 mbits / sec), and have rerun the model on some JPEG rating data first reported in Lubin, 1995. For the MPEG-2 sequences, data were collected at Sarnoff under a different rating paradigm. In this paradigm, subjects were shown, in each trial, five-second clips of the reference and then the test sequence, and were asked to rate the perceptibility of distortions using the following five point scale:

0. not perceptible
1. barely perceptible
2. mildly perceptible
3. clearly perceptible
4. extremely perceptible

Each data point is an average of at least 80 separate trials (20 subjects x 4 replications per subject). Data were collected for five different sequences (as shown in the legend in Figure 16) at three bit-rates each, using a standard TM5 encoder. Figure 16 shows the correlations of the JND Vision Model with these data. For comparison purposes, Figure 17 shows the correlation of an MSE measure.

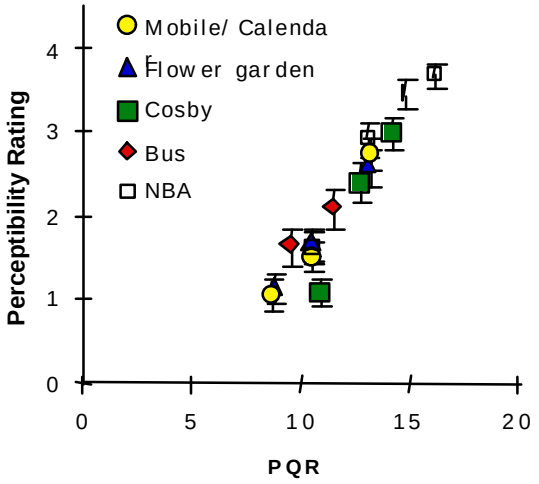

Figure 16. MPEG-2 low bit-rate rating predictions

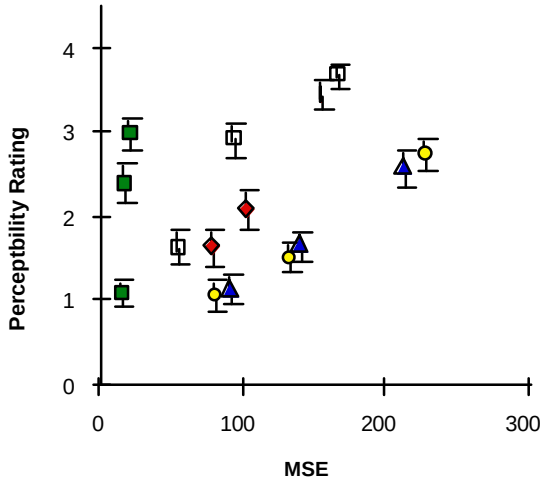

Figure 17. MSE predictions on low bit-rate MPEG-2 data.

Note that, although MSE shows a monotonic relationship between quality and bit-rate for each image sequence, it does not adequately capture the differences in this relationship from one sequence to the next. On the other hand, the JND Vision Model produces quite good correlation.

For the JPEG task, observers were shown four different scenes (p1, p2, p3, and p4) each compressed at 11 different JPEG levels. Observers were then asked to rate the quality of each resulting still image on a 100-point scale (100 being best).

As shown in Figure 18, the model does a good job predicting the rating data, with excellent clustering across image types and a strong linear correlation over the entire rating range (.94). Even better correlation (0.97) results when one omits the four points above 15 JNDs, for which some saturation at the low end of the rating scale has evidently occurred.

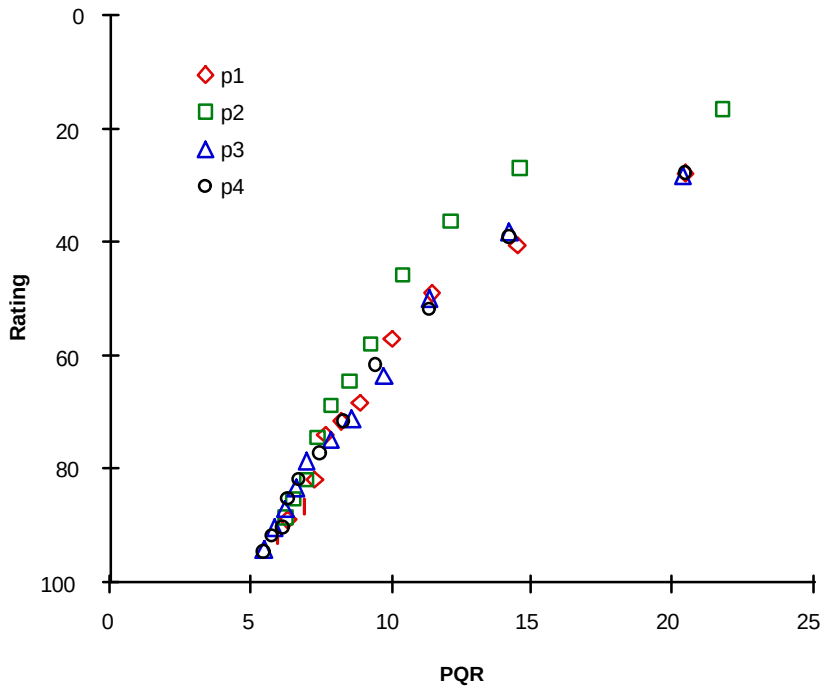

Figure 18. Predictions of Final Model on JPEG rating data

On the other hand, as shown in Figure 19, correlation among ratings and predictions based on the root mean-squared error between the original and compressed images are not nearly as good (.81). Here, the predictions do not track well across image types, even though a monotonic relation between rating and predicted value is observed within each image.

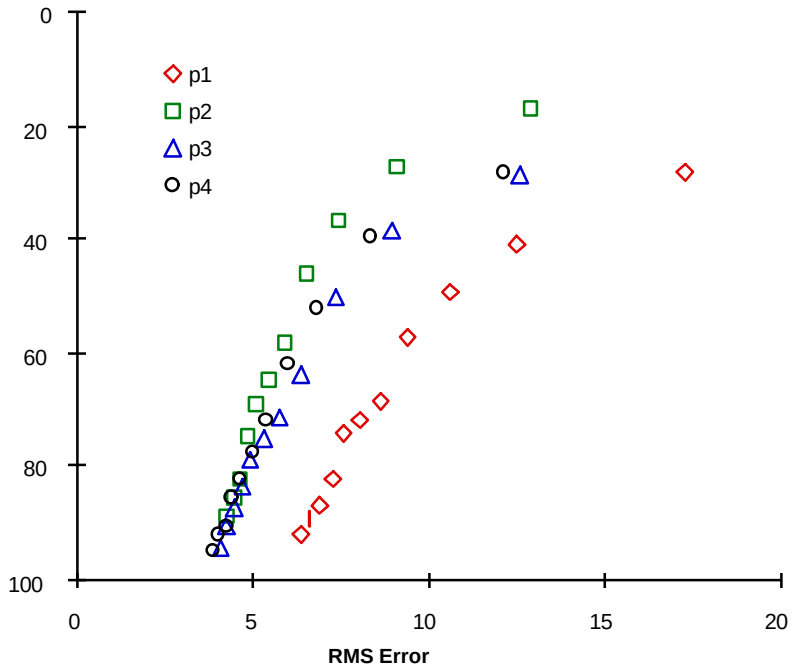

Figure 19. RMS error predictions on JPEG rating data

6. Conclusions

As described in the above sections, the Sarnoff JND Vision Model is based on known physiological mechanisms of vision, and is carefully calibrated according to basic psychophysical data. This methodology provides it with a great deal of robustness in predicting visibility of distortions, not only in MPEG-2 video, but in an extremely broad range of current and future video applications. The excellent predictions of the model across the range of data described in this document should provide confidence that this model is a useful tool for image quality metering applications.

7. References

O. M. Blackwell and H. R. Blackwell, "Visual performance data for 156 normal observers of various ages," *J. Illum. Engr. Soc.* **61**, 3-13 (1971).

P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications,* **COM-31**, 532-540 (1983).

C. R. Carlson and R. W. Cohen, "Visibility of Displayed Information", Report prepared for Office of Naval Research by RCA Laboratories, Princeton, NJ, July 1978. Contract No. N00014-74-C-0184.

C. R. Carlson and R. W. Cohen, "A simple psychophysical model for predicting the visibility of displayed information. *Proc. Soc. Inform. Display* **21**, 229-245 (1980).

C. H. Graham (Ed.), *Vision and Visual Perception*.. Wiley, 1966.

J. J. Koenderink and A. J. van Doorn, "Spatiotemporal contrast detection threshold surface is bimodal," *Optics Letters* **4**, 32-34 (1979).

J. Lubin, The use of psychophysical data and models in the analysis of display system performance. In A. B. Watson (ed.), *Digital Images and Human Vision*, MIT Press, 1993, pp. 163-178.

J. Lubin, A visual system discrimination model for imaging system design and evaluation, to be published in E. Peli (ed.),*Visual Models for Target Detection and Recognition*, World Scientific Publishers, 1995.

K. T. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings," *J. Physiol.* **359**, 381-400, 1985).

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

E. Switkes, A. Bradley, and K. De Valois, "Contrast dependence and mechanisms of masking interactions among chromatic and luminance gratings,"*J . Opt. Soc. Am.* **A5**, 1149-1162 (1988).

F. L. van Nes, J. J. Koenderink, H. Nas, and M. A. Bouman, "Spatiotemporal modulation transfer in the human eye," *J. Opt. Soc. Am.* **57**, 1082-1088 (1967).

G. Wyszecki and W. S. Stiles, *Color Science*, 2nd ed., Wiley, 1982.