NASA-CR-197784

Space, Telecommunications And Radioscience Laboratory

FINAL

March 30, 1995 IN-61-CR

(NASA-CR-197784) GRAYSCALE/RESOLUTION TRADE-OFF FOR TEXT: MODEL PREDICTIONS AND PSYCHOPHYSICAL RESULTS FOR LETTER CONFUSION AND LETTER DISCRIMINATION Final Report, Oct. 1992 - Mar. 1995 (Stanford Univ.) 10 p

N95-26930

Unclas

G3/61 0049075

Ms. Barbara Hastings
University Affairs Branch
NASA Ames Research Center
MS 241-1 Code ASC
Moffett Field, CA 94035

Dear Ms. Hastings:

This letter is the final report for the research entitled "Color Image Processing for Flat Panel Visual Display Modeling," under Cooperative Agreement No. NCC2-777. This contract began in October 1992 and ended in March 1995.

The research done for this grant was performed primarily by Dr. Ramin Samadani and Mr. Darrin Rath. Dr. Samadani was an associate investigator at Stanford University. He terminated employment with Stanford in autumn 1994. Darrin Rath is a graduate student at Stanford University. He will receive a Masters degree in electrical engineering in April 1995. Mr. Rath has accepted employment elsewhere and will leave Stanford when he graduates. Professor Allen M. Peterson was the principal investigator for this contract. He passed away suddenly in August 1994 and Professor G. Leonard Tyler became the new principal investigator.

The goal of the research was to provide computer tools that allow the simulation and evaluation of flat panel high-resolution display designs. The results of this research are being incorporated into the ViDEOS software simulation system, under the direction of Dr. James Larimer of NASA Ames Research Center. The ViDEOS software enables flat panel display designers to model various parameters of a display. The pixels of a typical display are modeled as a "sandwich" of layers that light passes through. These layers may include polarizers, glass, electrodes, and liquid crystal materials. The software calculates the transmissions of a spectrum of light (380 to 780 nm) as it passes through the sandwich. Voltage levels applied to the liquid crystal layers and viewing angles are taken into account. The calculated output spectrum is converted into CIE tristimulus values and the resulting color is displayed on a calibrated CRT screen.

Mr. Rath implemented the core set of graphical user interface routines used by the ViDEOS software. These routines, called "xobjects," are a hierarchy of C++ classes that are used as wrappers to the OpenLook library of routines. The entire ViDEOS interface uses xobjects as the base for implementing all graphics.

Dr. Samadani and Mr. Rath researched tiling methods. To model the display of an image on a flat panel display, the image is tiled so each pixel is typically composed of separate red, green, and blue subpixels. Also, there are often black areas surrounding the subpixels due to the opacity of the circuitry driving the display. The April 1993 status report for this grant by Dr. Samadani describes some theoretical aspects of this problem. Mr. Rath wrote Matlab and C++ algorithms to efficiently implement tiling. Mr. Rath's algorithm performs two dimensional filtering on an original sampled image. The original sampled image is treated as a 2D array of impulses of various magnitudes. These impulses are convolved with the tiling geometry of a single pixel to produce the tiled image.

STARLAB
Department of Electrical Engineering / EERA • Stanford University • Stanford, CA 94305-4055
Telex No. 348402 STANFRD STNU • FAX No. (415) 723-9251

Research on color digital halftoning was performed by Mr. Rath and Dr. Samadani. Existing halftoning methods were studied and algorithms were implemented that can be integrated into the ViDEOS system.

Mr. Rath also contributed technical and programming support to other researchers at NASA Ames Research Center. This especially included Dr. Jennifer Gille and Dr. James Larimer. Algorithms written by Mr. Rath were used to produce results for the paper "Grayscale/Resolution Trade-Off for Text: Model Predictions and Psychophysical Results for Letter Confusion and Letter Discrimination," by Dr. Jennifer Gille and others.

In summary, the research funded by this grant has contributed to the development of the ViDEOS modeling software and to research papers written by Dr. Ramin Samadani and Dr. Jennifer Gille. The ViDEOS software is available to US companies doing display research and development.

Yours Sincerely,

G. Leonard Tyler

G. Leonard Tyler
Professor, Electrical Engineering

Darrin Rath

Darrin Rath

GRAYSCALE/RESOLUTION TRADE-OFF FOR TEXT: Model predictions and psychophysical results for letter confusion and letter discrimination

Jennifer Gille,[1] *Russel Martin*[2]*, Jeffrey Lubin*[3]*, and James Larimer*[4]

(1) Western Aerospace Laboratories, Moffett Field, CA (2) Xerox PARC, Palo Alto, CA (3) David Sarnoff Research Laboratories, Princeton, NJ (4) NASA/Ames Research Center, Moffett Field, CA

1. Introduction

In a series of papers presented in 1994 at SPIE[1], SID[2] and IDRC[3], we examined the grayscale/resolution trade-off for natural images displayed on devices with discrete pixellation, such as AMLCDs. In the present paper we extend this study by examining the grayscale/resolution trade-off for text images on discrete-pixel displays.

Halftoning in printing is an example of the grayscale/resolution trade-off. In printing, spatial resolution is sacrificed to produce grayscale. Another example of this trade-off is the inherent low-pass spatial filter of a CRT, caused by the point-spread function of the electron beam in the phosphor layer. On a CRT, sharp image edges are blurred by this inherent low-pass filtering, and the block noise created by spatial quantization is greatly reduced. A third example of this trade-off is text anti-aliasing, where grayscale is used to improve letter shape, size and location when rendered at a low spatial resolution.

There are additional implications for display system design from the grayscale/resolution trade-off. For example, reduced grayscale can reduce system costs by requiring less complexity in the framestore, allowing the use of lower cost drivers, potentially increasing data transfer rates in the image subsystem, and simplifying the manufacturing processes that are used to construct the active matrix for AMLCD (active-matrix liquid-crystal display) or AMTFEL (active-matrix thin-film electroluminescent) devices. Therefore, the study of these trade-offs is important for display designers and manufacturing and systems engineers who wish to create the highest performance, lowest cost device possible.

Our strategy for investigating this trade-off is to generate a set of simple test images, manipulate grayscale and resolution, predict discrimination performance using the ViDEOS$_{\text{Sarnoff}}$ Human Vision Model (see Larimer et al., 1994[4]), conduct an empirical study of discrimination using psychophysical procedures, and verify the computational results using the psychophysical results.

2. Grayscale/resolution trade-off

In our previous studies, the natural test images were zone-plate stimuli (radially symmetric spatial sine-wave chirp patterns) rendered at various combinations of pixel size and number of gray levels. Floyd & Steinberg error diffusion[5] was used to study the effects of dithering on the trade-off. Each zone plate was a downsampled rendering of a continuous ideal image. The zone plate images were chosen because they represent a wide range of spatial frequencies and orientations. The ideal image we used had a high contrast ratio, but was truncated at high spatial frequencies to fit on our screen. Truncation also eliminated aliasing that would be generated when the ideal image was rendered for display on our experimental apparatus.

Several effects of lower spatial and grayscale resolution were found to be exaggerated for discrete-pixel displays, such as liquid-crystal displays, where the fixed-pattern noise of pixellation is not reduced by an inherent spatial low-pass filter as is the case with CRTs. Jaggies, luminance banding, and simultaneous contrast effects (e.g. Mach Bands) are among the effects that are more visible on displays with spatially discrete pixels.[1,2]

Our results showed that when grayscale resolution was lowered via multi-level thresholding, i.e. simple quantization, there was little trade-off between grayscale and spatial resolution. For the display parameters used, we found that above about 130 dpi resolution, grayscale predominated in determining image quality. Below that point, spatial resolution predominated. With multi-level error diffusion, image quality was greatly improved in the region previously dominated by grayscale performance, and a true trade-off region was found. Dither, such as error diffusion, works by trading spatial resolution - the spatial noise that is introduced by the dither pattern - for grayscale. (See Ulichney, 1987[6]) When the introduced spatial noise is beyond the spatial resolution of the eye, it is hidden from view. This scheme is particularly effective when four or more levels of gray are used, because the contrast between adjacent pixels is never very high (except where image contrast is high), and the luminance of even small patches can be close to the desired value for that region; thus the spatial noise is hidden. The computational results from the $\text{ViDEOS}_{\text{Sarnoff}}$ model were strongly verified by the psychophysical results.[3]

There were two reasons to extend the study of the grayscale/resolution trade-off to text images. First, text is obviously an important type of image. Second, the information used in identifying letters is the high spatial-frequency content of the small, distinguishing features, information that was not tested with the zone-plate images.

3. Experiment 1 - Letter confusion

3.1 Computational study

We measured the legibility of a set of letters by (1) computing letter confusability using the $\text{ViDEOS}_{\text{Sarnoff}}$ model and (2) empirically measuring letter identification using psychophysical methods. We varied device spatial resolution (the equivalent of 288 dpi, 144 dpi, 72 dpi and 36 dpi), number of gray levels (8, 4 and 2), and "point" sizes (a subset of the equivalents of 12, 9, 6, 5, 4, 3 and 2 pts, depending on the device resolution). "Point" size in this study refers to the height of the letter in seventy-seconds of an inch, rather than to the height of an actual typeface. (See Rubinstein, 1988[7]) "Equivalent" in this study refers to the conditions of the empirical laboratory experiments, where the visual subtense of letters and simulated device pixels viewed at 6m corresponded exactly to the subtense of real letters and pixels viewed at 0.5m.

The simple stimuli we chose were the block letters C, lowercase e, uppercase G, and O. Continuous ideal versions of these letters were defined by line segments and circular arcs, and then rendered at the desired spatial resolution. Figure 1 shows high-resolution versions of these letters.

Figure 1
High-resolution stimuli for the letter confusion study

Letters that were to be the equivalents of 1 inch (72 pts), black-and-white, at 288, 144, 72 and 36 dpi were calculated from the continuous ideal version. Each 72-point letter was filtered, using a simple triangle filter, and downsampled to create the test stimuli at the point sizes for that resolution. This filtering and downsampling technique was not intended to be the most effective method for producing a good, readable font. We selected one way for producing our letters; schemes have been developed that vary with type font and storage and rendering methods that are more or less optimal depending on the criterion applied for assessing optimality. (See Rubinstein, 1988[7])

The filtered, spatially downsampled letters were further manipulated by downsampling in grayscale from the original 256 levels. Grayscale bins with 8, 4 or 2 levels were created by linearly dividing the luminance range (0 to 86.6 cd/m^2) using 7, 3, or 1 thresholds. Each bin was assigned a value from 0 to 86.6 that represented a linear step, with 0 assigned to the lowest bin and 86.6 to the highest bin. This technique, as with the spatial downsampling, was not intended to produce optimum readability. In some cases, the filtering and spatial downsampling generated luminance values that, when downsampled in grayscale, did not use all available bins (for instance, a letter might use only 7 of the 8 possible gray levels.)

The simulation device for displaying the various letters was a BARCO Calibrator monitor with 72 dpi resolution. In order to simulate a discrete-pixel display on the CRT, all stimuli were pixel-replicated such that when they were viewed from a distance of 6 m, the size of their retinal image was the same as if they had been viewed on an actual discrete-pixel display at a distance of 0.5 m, normal viewing distance.

To summarize, then, there were four letters: C, e, G and O. There were 14 resolution/pointsize combinations: 288 dpi with 4, 3, and 2 point letters; 144 dpi with 12, 9, 6, 5 and 4 point letters; 72 dpi with 12, 9, 6, 5 and 4 point letters; and 36 dpi with 12 point letters. There were three grayscale values: 8, 4 and 2 levels. As it turned out, we had to create special versions of the 72 dpi, 4 and 5 point, 2 gray level letters, because the generation algorithm did not produce stimuli with appropriate letter features.

The computational study used letter discriminability to predict confusability. The study compared each of the four letters at a given resolution/pointsize/grayscale combination with each of the other letters at that combination, for a total of six comparisons at each resolution/pointsize/grayscale level. Each comparison measured how discriminable the pair of letters was, i.e. the inverse of how confusable they were.

The ViDEOS$_{\text{Sarnoff}}$ Human Vision Model used for the computational study has been described elsewhere.[8,9] Briefly, it follows the architecture of the visual system and uses classical psychophysical data for calibration. There are no free parameters to adjust when using the model; that is, the output measures discriminability without further scaling. The model has been verified in several different kinds of studies, both at and above threshold.

The input to the model is two images, and the output of the model is a map of their perceptual differences. Perceptual difference is measured in Just Noticeable Differences, or JNDs, where one JND represents a 70% chance of seeing a difference: a threshold difference. Multiple JNDs represent multiple independent looks, and therefore follow the binomial probability distribution. For example, 4 JNDs represents better than a 99% chance of seeing a difference. The output map, then, is an array of JNDs, where each point of the array corresponds to the same spatial location in each of the two images, and measures the perceptual difference between the images at that location.

The output JND map of the ViDEOS$_{\text{Sarnoff}}$ model can be visualized by displaying large values as bright points and small values as dim ones, as in Figure 2. Here a C and an O are compared using the model. The JND map is calculated and displayed. Not surprisingly, the C and the O are seen to differ most in the locations where the C has a gap and the O does not. Such a visualization of the JND map is an extremely valuable tool for exploring the effects of changing device and image parameters.

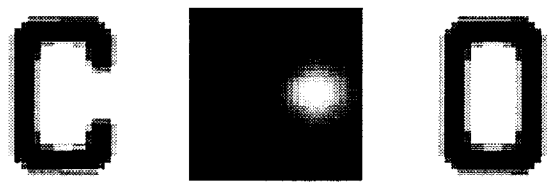

Figure 2
Letter stimuli and their JND map from the ViDEOS$_{\text{Sarnoff}}$ model
The map is a visualization of the perceptual discrimination between the two images. The spatial correspondence between the map and the images allows localization of perceptual differences. Brighter regions on the map indicate higher JND values and thus greater perceptual differences.

The output JND map can also be summarized by combining the JND values into a single measure. For example, finding the maximum JND value and its location in the map can be helpful in evaluating the visibility of small device-related artifacts. An average value such as the root-mean-square JND (RMS JND) would be more appropriate for looking at the general effects of, say, a compression algorithm on image quality.

The results of the computational study are shown in Figures 3a-3d. The results for each device resolution are plotted in the separate graphs. The ordinate is the summary discrimination measure: RMS JND for each output JND map. RMS JND is plotted against letter point size. The different grayscale values for each point size are separated slightly horizontally: left-to-right; they are 8, 4 and 2 levels, respectively. The six letter comparisons for each pointsize/grayscale combination are represented by different symbols. Since letter confusions will occur where letters are very similar and there are only small perceptual differences, poor legibility will be associated here with small RMS JND values. Threshold for confusability should be around 1 RMS JND.

Because the features that allow for letter discrimination are small and highly localized, maximum JND could have been used as the summary measure. However, since RMS JND was calculated from the map using the immediate neighborhood of the letter, it probably represents a better measure of discrimination.

The results show evidence of the following. Confusability increases as the letters are made smaller. There is no consistent effect of grayscale for 8, 4 and 2 levels of gray. Higher device resolution decreases confusability when letter size is held constant. Some letter pairs are consistently more confusable than others. Finally, very few of the realizable letters are predicted to be confusable. That is, RMS JND is rarely less than 1, and reading ordinarily represents something close to the limits of visibility, in the sense that a reader can tolerate much degredation up to a critical point where readability declines rapidly. (See Legge et al., 1985[8])

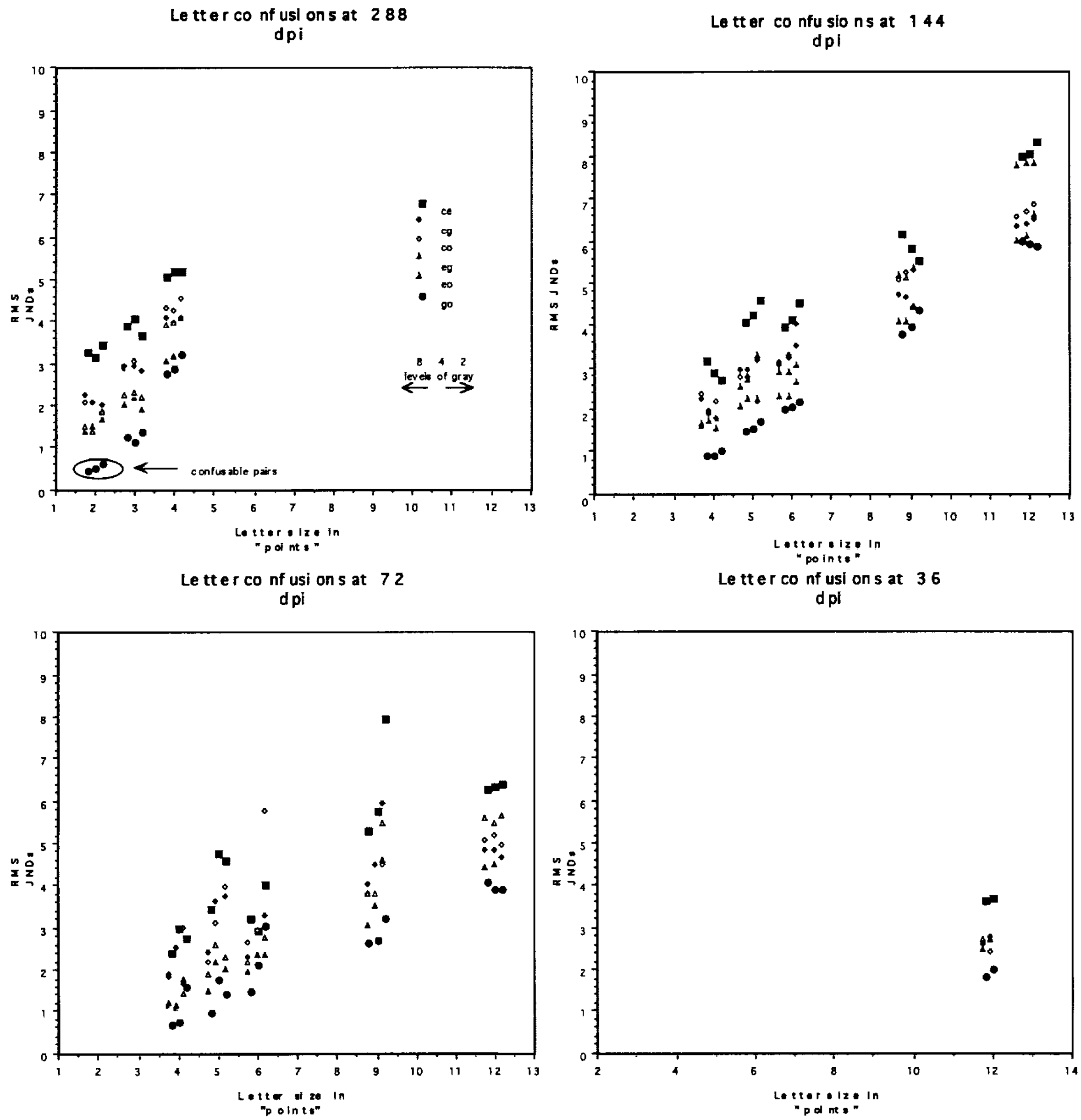

Figure 3
Letter confusion data

3.2 Empirical study

The empirical study used letter identification in a four-alternative forced-choice method-of-constant-stimuli procedure to measure letter confusability. The stimuli were the various letters described above. A set of letters of various types but with balanced numbers of C, e, G, and O was used in each experimental session.

The observer sat in a darkened room using a headrest to stabilize head position. The BARCO Calibrator monitor was viewed using a first-surface mirror such that the total viewing distance from the observer to the image of the monitor was 6m. During the session, 10 presentations of each of the letters in the set were shown in random order, under the control of a Macintosh Quadra computer. Observers were adapted to the image of the white screen at 86.6 cd/m^2.

On each trial, the observer would signal readiness, and the letter would be presented for one second on the white background. The observer would judge which of the four letters had been presented, and would indicate that judgment by pressing one of four keys on the computer keyboard. The computer tallied the responses of the observer by stimulus letter.

Since each stimulus letter was presented 10 times, and there were 4 alternative responses (one correct), the observer had to make an error on at least four trials in order to count the letter as not identifiable/below threshold for confusability. (This criterion was based on a probability of .0197 that an observer could be correct on 6 or more trials by chance.) Incorrect responses were then listed as the letter(s) with which the presentation letter was (were) confusable.

Very few of the letters tested were found to be confusable. Those that were are circled and indicated with an arrow in Figures 3a-3d. All confusions lay in the region predicted by the vision model.

We decided to further explore the grayscale/resolution trade-off, if any, for text by changing our empirical question. We asked what would happen if a letter were compared for discriminability with a high-resolution version of that same letter. The purpose of adding anti-aliasing to text is to create a letter that cannot be distinguished from, or that looks more like a high-resolution letter. The simple triangle filter we used to blur letters before downsampling is a type of anti-aliasing, and would allow us to further test the effects of grayscale at various spatial resolutions on image quality.

4. Experiment 2 - Letter discrimination

4.1 Computational study

We measured the quality, compared to high-resolution referents, of a set of letters by computing letter discrimination using the ViDEOS$_{Sarnoff}$ model and by empirically measuring letter discrimination using psychophysical methods. We varied device spatial resolution (the equivalent of 288 dpi, 144 dpi and 72 dpi), "point" size (the equivalent of 12, 9, 6 and 4 points), and grayscale (256, 8, 4 and 2 levels of gray). We used a single letter, the lowercase e of experiment one, shown enlarged in Figure 4 at various resolutions.

The letter stimuli were generated as follows. An e, the equivalent of one inch high on an 864 dpi device, was calculated from the continuous ideal version. For each of the stimuli used in the experiment, this highest resolution letter was filtered, with the simple triangle filter, downsampled to the appropriate device resolution and point size, and when necessary pixel-replicated to preserve point size when reduced-spatial-resolution media were simulated. Each was further downsampled in gray scale to the appropriate number of gray levels. Again, these letters represented images rendered on a discrete-pixel display.

The computational study compared each downsampled e of a particular resolution/pointsize/grayscale combination to a high-resolution referent at the same point size but at the equivalent of 864 dpi and 256 levels of gray. The comparison, using the ViDEOS$_{Sarnoff}$ model, measured their discriminability. The results are shown in Figure 5, where RMS JND for the comparison is plotted against grayscale bits ($\log_2$(number of graylevels).) The spatial resolution/point size combinations are plotted as separate lines.

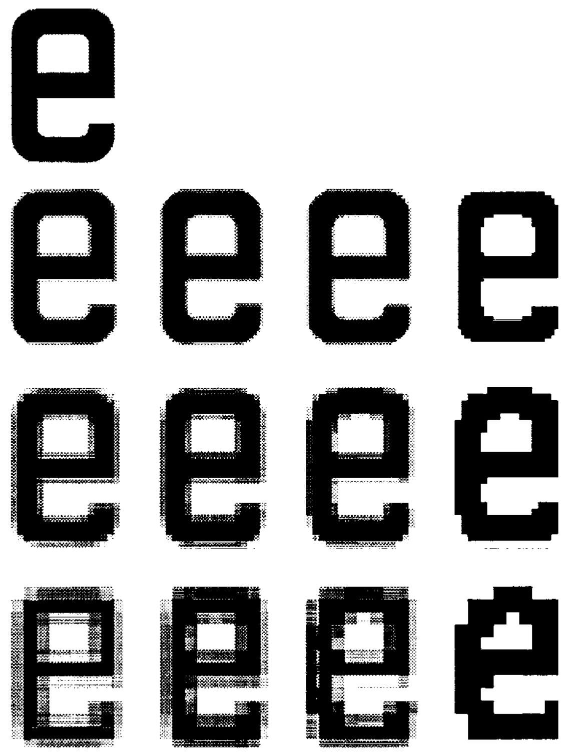

Figure 4
Letter discrimination stimuli
The high-resolution referent is in the upper right. Spatial resolution decreases moving down, and grayscale decreases moving to the right.

We interpret the results as follows. In this experiment, the lower-resolution letters that look like the high-resolution referents can be said to be good quality renderings of the image. That is, the lower the RMS JND value, the less discriminability from the referent and the better the quality. Letters with RMS JND values less than about one should lie below threshold for discriminability. From the figure, we see that the greatest determiner of discriminability is device spatial resolution: the higher the better. Point size interacts with device resolution: as the letters get smaller they get worse on the 72 dpi display, but letter size makes little difference on the 288 dpi display. In general, grayscale has little effect, although for the higher resolution displays there is evidence that there is an increase in quality when the number of gray levels is greater than two.

For the 72 dpi discrete-pixel display, there is evidence in the computational study for the non-intuitive result that increasing grayscale can lower the quality of the letters. The empirical study suggests reasons why this might be so.

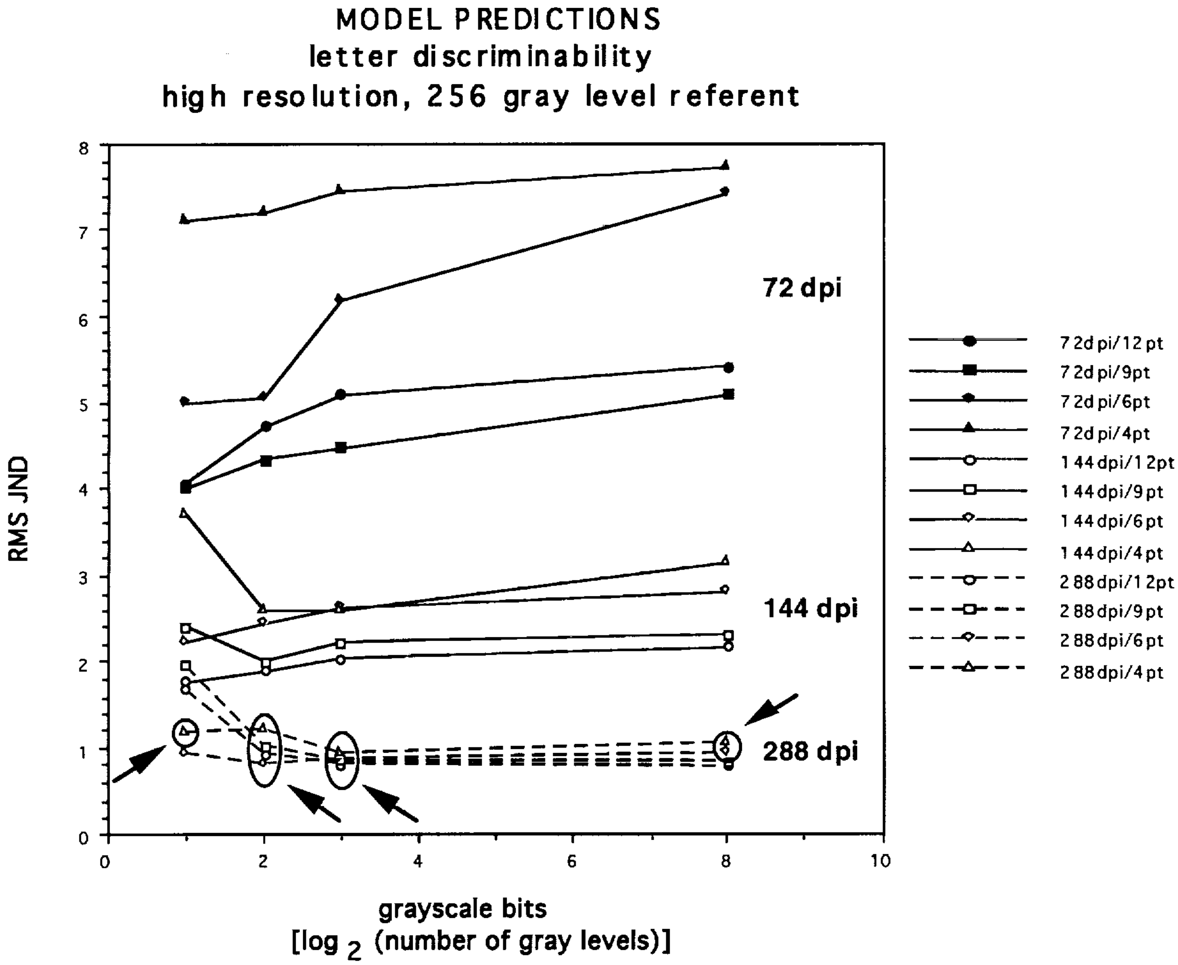

Figure 5
Letter discriminability results

4.2 Empirical study

The empirical study used letter discrimination in a modified four-alternative forced-choice method-of-constant-stimuli procedure to measure letter discrimination. The stimuli were the e's described above. On each trial, four letter e's were presented. Three of the e's were copies of the high-resolution referent, and one was the low-resolution test e. The observer's task was to indicate which of the letters was different from the others.

Most of the letters tested were discriminable from the high-resolution reference letter. Those that were not discriminable (correctly identified only at chance levels of performance) are circled and indicated with arrows in Figure 5. All of the letters that were not discriminable from the referent lay in the region predicted by the vision model. Therefore, the ViDEOS$_{Sarnoff}$ vision model is capable of predicting the effects of both reduced spatial resolution and reduced grayscale.

5. Discussion

The cues for discriminating the lower resolution test letter from the high-resolution referent were many. For lower resolution letters, jaggies were often visible. Lower resolution letters with more than two levels of gray could look fuzzy compared to the referent. Lower resolution letters with two levels of gray could have a shape that was distorted compared to the referent. As was found in the zone-plate studies, these discrimination cues were exaggerated for a discrete-pixel device compared to what might be expected on a CRT.

Figure 6 demonstrates, within the limitations of reproduction in this paper, the case where the inherent low-pass filter characteristic of a CRT can improve text quality for a small letter rendered on a lower-resolution device. The G on the left is a pixel-replicated version of a 6-point letter with 8 levels of gray on a 72 dpi discrete-pixel device, created by filtering and downsampling a high-resolution letter. The G on the right is this same image, blurred using the Adobe Photoshop Gaussian blur function, to simulate rendering on a CRT. Viewed from about 19 ft., these letters will have the appropriate visual subtense for comparison with actual letters viewed at a standard 0.5m.

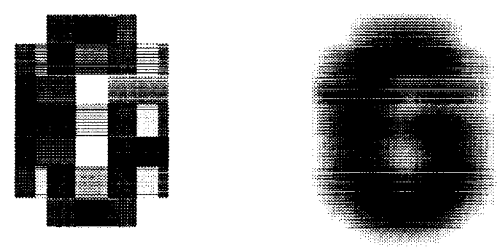

Figure 6
The effects of block noise *vs* blur

The block noise of the discrete pixels is visible and interferes with identification of the letter. As with the simultaneous contrast effects found for the zone-plate images, the edge-enhancing characteristics of the human visual system work against image quality in this case. When the edges are blurred, the underlying shape of the letter becomes more visible.

The counterintuitive result in the computational study, that adding grayscale might increase discriminability of a lower-resolution test letter from its high-resolution referent, can be explained in terms of the visibility of block noise. At 72 dpi, the block noise in the multiple gray level "anti-aliasing" regions is visible. This added visibility of the block noise makes the multiple grayscale letters at this dpi more discriminable from the high-resolution referent. At higher device resolutions, or dpi, this block noise is spatial-frequency shifted to higher frequencies where it is less visible and does not add to discriminability.

In Figure 7, two of the test stimuli for the 72 dpi/12 point comparisons, 72dpi/12pt/8 levels and 72dpi/12pt/2 levels of gray, are rendered along with the 12pt high-resolution referent and the respective JND maps generated in the vision model. (Viewed from 2.25m or about 7.5ft, these letters will have the appropriate visual subtense for comparison with actual letters viewed at a standard 0.5m.) The JND maps clearly show the following. For the 8-gray-level stimulus, it is the block noise of the gray regions that makes the letter discriminable from the referent, and this noise is present over the entire letter. For the 2-

gray-level stimulus, it is the more localized jaggies of the curved portions of the letter that predominate. One of the strengths of the ViDEOS$_{\text{Sarnoff}}$ vision model as an engineering tool is the ability to examine in detail where differences lie and to establish the sources of discriminability.

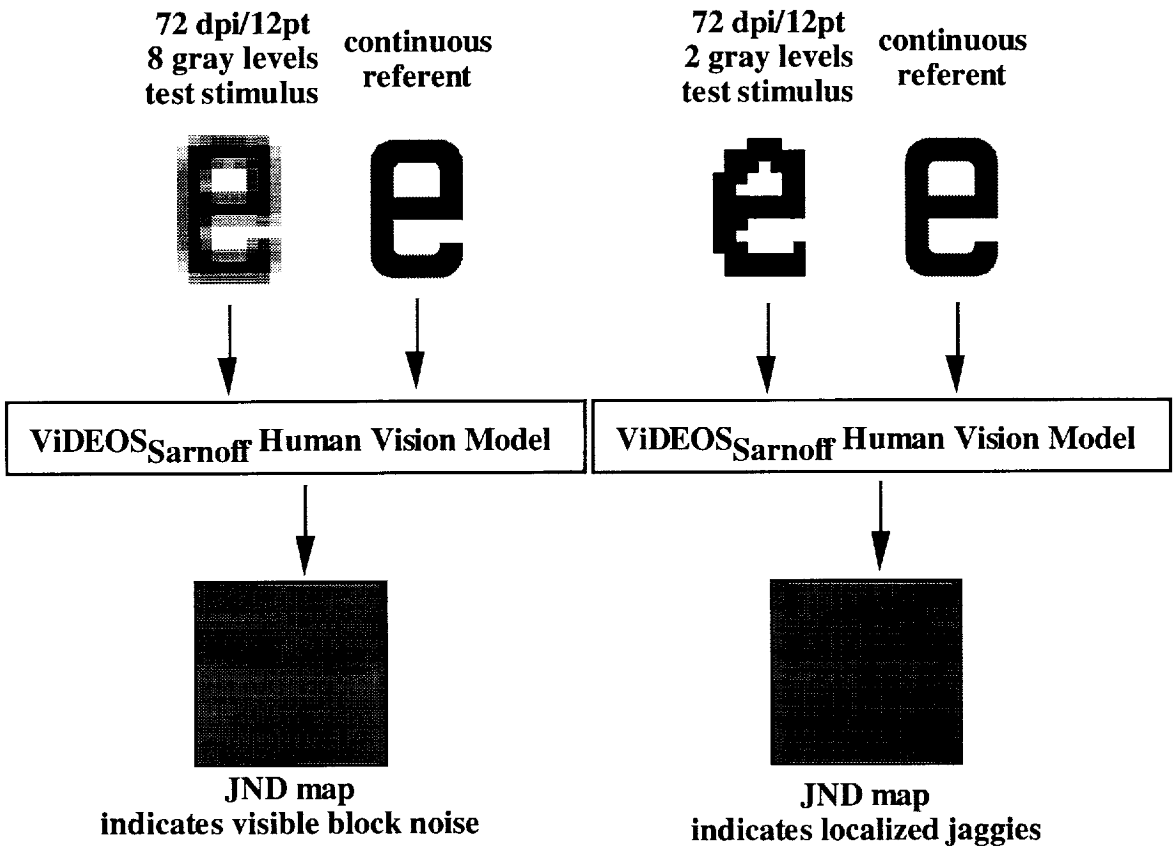

Figure 7
JND maps used to explore sources of perceptual differences

The results confirm that the ViDEOS$_{\text{Sarnoff}}$ vision model is a useful tool for analysing image quality on spatially discrete display devices. In future work, we will further explore the counterintuitive finding that at some device resolutions, adding grayscale degrades performance. We will investigate a masking mechanism explanation of this result.

6. Acknowledgments

This work was funded by the ARPA High-Definition Systems Program and a NASA research grant awarded to Western Aerospace Laboratories under NCC 2-788 (Dr. James Larimer, Technical Monitor).

We wish to thank Darrin Rath for his technical support on this project.