1986

An investigation of resolution and addressability requirements for digital display systems used in word-processing and computer-aided-drafting applications

Stephen T. Knox
Portland State University

Let us know how access to this document benefits you.
Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds
Part of the Psychology Commons

Recommended Citation

10.15760/etd.5581

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

Title: An Investigation of Resolution and Addressability Requirements for Digital Display Systems Used in Word-Processing and Computer-Aided-Drafting Applications.

APPROVED BY MEMBERS OF THE THESIS COMMITTEE:

Gerald Murch, Ph.D., Chairman

Barry F. Anderson, Ph.D.

Robert J. Beaton, Ph.D.

James Paulson, Ph.D.

Two hardware factors contributing to the overall image quality of digital CRTs are display resolution and addressability. The relationship between these two factors and human performance was modeled by a metric of display quality, the Resolution Addressability Ratio (RAR), and investigated within the contexts of Word-Processing (WP) and
Computer-Aided-Drafting (CAD) tasks. The findings indicate a perceptual limit to MTF bandwidth improvements, and significant differences in display quality requirements between the two applications. The regression of image quality as a function of RAR metric values resulted in an $r^2 = 0.94$ for the WP task and an $r^2 = 0.79$ for the CAD task. These findings are discussed in terms of engineering guidelines for the design of CRT and flat-panel displays for applications which vary by the amount of the density of information in their typical displays.
AN INVESTIGATION OF
RESOLUTION AND
ADDRESSABILITY REQUIREMENTS
FOR DIGITAL DISPLAY SYSTEMS
USED IN WORD-PROCESSING AND
COMPUTER-AIDED-DRAFTING
APPLICATIONS

by

Stephen T. Knox

A thesis submitted in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE
in
PSYCHOLOGY

Portland State University
1986
TO THE OFFICE OF GRADUATE STUDIES AND RESEARCH:

The members of the committee approve the thesis of Stephen T. Knox presented October 15, 1986.

Gerald Murch, Ph.D., Chairman

Barry F. Anderson, Ph.D.

Robert J. Beaton, Ph.D.

James Paulson, Ph.D.

APPROVED:

Roger Jennings, Chair, Department of Psychology

Bernard Ross, Dean of Graduate Studies and Research
ACKNOWLEDGEMENTS

This research was sponsored by the Human Factors Research Laboratory, Tektronix, Inc., under the monitorship of Dr. Gerald M. Murch, Chief Scientist. Dr. Robert J. Beaton, Director of the Visual Interface Program, served as Senior Scientist.

The author is gratefully indebted to several individuals who contributed their time and talent to the conduct of this research. Dr. Robert J. Beaton skillfully directed this research and provided the author with patient instruction, encouragement and insight. Dr. Gerald Murch laid the theoretical foundation for the research conducted herein. Mr. Paul Glasser executed the design and development of the variable-resolution, variable-addressability monitor with enthusiasm, perseverance and engineering expertise. Finally, numerous technical problems were solved with considerable skill by Mr. Ban Van Nguyen.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>5</td>
</tr>
<tr>
<td>Display Resolution</td>
<td>5</td>
</tr>
<tr>
<td>Display Addressability</td>
<td>8</td>
</tr>
<tr>
<td>Relating Resolution and Addressability : The RAR</td>
<td>10</td>
</tr>
<tr>
<td>Application-Dependent Image Quality Criteria</td>
<td>13</td>
</tr>
<tr>
<td>METHOD</td>
<td>15</td>
</tr>
<tr>
<td>Subjects</td>
<td>15</td>
</tr>
<tr>
<td>Equipment</td>
<td>15</td>
</tr>
<tr>
<td>Procedure</td>
<td>18</td>
</tr>
<tr>
<td>RESULTS</td>
<td>23</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>41</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>47</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Luminance Profiles: Width / Separation</td>
<td>3</td>
</tr>
<tr>
<td>2. Hypothetical Modulation Transfer Function</td>
<td>6</td>
</tr>
<tr>
<td>3. Relationship of Line Spread Function to Modulation Transfer Function</td>
<td>7</td>
</tr>
<tr>
<td>4. Main Effect of Task: Accuracy Measure (Pre-test)</td>
<td>24</td>
</tr>
<tr>
<td>5. Main Effect of Task: Proficiency Measure (Pre-test)</td>
<td>25</td>
</tr>
<tr>
<td>6. Task by Group Interaction: Proficiency Measure (Pre-test)</td>
<td>26</td>
</tr>
<tr>
<td>7. Display Size Preference (Pre-test)</td>
<td>28</td>
</tr>
<tr>
<td>8. Main Effect of Task: Accuracy Measure</td>
<td>30</td>
</tr>
<tr>
<td>9. Main Effect of Task: Proficiency Measure</td>
<td>31</td>
</tr>
<tr>
<td>10. Main Effect of Task: Quality Ratings Measure</td>
<td>32</td>
</tr>
<tr>
<td>11. Main Effect of Resolution: Quality Ratings Measure</td>
<td>34</td>
</tr>
<tr>
<td>12. Main Effect of Addressability: Quality Ratings Measure</td>
<td>35</td>
</tr>
<tr>
<td>13. Addressability by Task Interaction: Quality Ratings Measure</td>
<td>36</td>
</tr>
<tr>
<td>14. Regression of Quality Ratings on RAR: Combined Applications</td>
<td>38</td>
</tr>
<tr>
<td>15. Regression of Quality Ratings on RAR: CAD Application</td>
<td>39</td>
</tr>
<tr>
<td>16. Regression of Quality Ratings on RAR: WP Application</td>
<td>40</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>12</td>
</tr>
<tr>
<td>II</td>
<td>44</td>
</tr>
<tr>
<td>Adjacent / Alternate Criteria Outcomes by RAR</td>
<td></td>
</tr>
<tr>
<td>Adjacent / Alternate Criteria Outcomes by Line Density</td>
<td></td>
</tr>
</tbody>
</table>
INTRODUCTION

Images presented on a digital display system are composed of discrete picture elements (pixels). A pixel represents a digitized sample of an image. In the "real world," images perceived by the human visual system are essentially continuous, with the scale of possible luminance values being virtually infinite. The digitization process samples an image into discrete luminance values obtained from a finite gray scale. A practical issue attending the digitization process is the introduction of image artifacts and anomalous detail in the displayed image that obscures the existing information (Beaton, 1984). Research directed toward producing "high-quality" images on digital displays has focused on characteristics of the individual pixel. Pixels are arranged in a rectangular matrix, with each $x,y$ location in the matrix being the pixel address. Two physical properties of this display matrix which affect image quality are: (1) the size or resolution of individual pixels, and (2) the density or addressability of pixels per unit area.

The relationship between perceived image quality and display resolution has received considerable attention from designers of cathode-ray tube (CRT) display devices. (Charman and Olin, 1965; Cohen, Gorog, and Carlson, 1975; Hufnagel, 1965; Schade, 1953; Snyder, 1973, 1974, 1976). The general approach taken by these workers has been to develop design guidelines that optimize the energy levels at which the spatial frequency components of displayed images are transmitted. These guidelines can be evaluated analytically by measuring size-related characteristics of individual pixels, which are derived from the width of the narrowest line that can be imaged on the display (Snyder, 1985). This conceptualization of resolution has been found to correlate highly
with subjective ratings of image quality and objective performance measures in specific task environments (Beaton, 1984; Task, 1979).

The relationship between image quality and display addressability is not so well specified. In a general sense, addressability refers to the number of rows of pixels times the number of columns of pixels over some unit area. These pixels may be addressed in a variety of ways, for example: (1) direct, with a hard connection to each pixel (small flat-panel displays); (2) matrix, with hard connections to each row and each column (larger flat-panel displays); and (3) scan, with separate amplifiers controlling horizontal deflection, vertical deflection and beam intensity (most CRTs) (Tannas, 1985). For many rastered CRT displays, addressability is determined first for the vertical dimension, which is the number of raster lines (rows of pixels) per unit distance. Once this is set, the horizontal addressability (columns of pixels) is configured so that it is proportional (pixels per unit distance) to the vertical addressability. Accordingly, this thesis defines addressability as the number of raster lines per inch on the display screen and defines resolution as the width (in inches) of a line.

Theoretical relations between resolution and addressability have been proposed (Lehrer, 1985; Murch & Beaton, 1986; Murch & Virgin, 1985). For example, consider the graphical relations in Figures 1a, 1b and 1c. The graphs plot the luminance profiles of adjacent raster lines. The dotted lines represent the resultant luminance output for each configuration. The line width (resolution) of each configuration is the same at 10 mils, but the addressability varies from 50 to 100 to 200 lines per inch. Since the resolution is the same for each configuration, the question to be asked is whether each configuration will result in the same image quality. The primary objective of this thesis is to investigate these relations empirically, and assess their effects upon image quality.
Figure 1a. Luminance profiles of several adjacent CRT raster lines. The line width (resolution) is 10 mils, and the lines per inch (addressability) is 100. The dotted line represents the resultant luminance output.

Figure 1b. Luminance profiles of several adjacent CRT raster lines. The line width (resolution) is 10 mils, and the lines per inch (addressability) is 50. The dotted line represents the resultant luminance output.

Figure 1c. Luminance profiles of several adjacent CRT raster lines. The line width (resolution) is 10 mils, and the lines per inch (addressability) is 200. The dotted line represents the resultant luminance output.
Additionally, this investigation will test the hypothesis that perceived image quality varies as a function of application-dependent task demands (Evans & Attaya, 1978; Goodenough, 1977; Greening, 1977; Hunt & Sera, 1978; Merrill & Metcalf, 1968; Peters, 1982;). That is, if there are constraining relations between resolution and addressability, do these relations vary across application environments? It has been proposed that the utility of displayed information, a function of the task demands of the application environment, will influence the perceived image quality of the information display (Bernath, Kretz & Wood, 1981; Lindenberg, 1976; Roetling, Trabka & Kinsky, 1968). Consequently, the empirical investigation conducted in this thesis will reference two specific application areas which, on the surface, appear to have widely divergent task demands: (1) Word-Processing (WP), where the primary task demand is the recognition and manipulation of text, and (2) Computer-Aided-Drafting (CAD), an application in which the placement of points in precise x,y coordinates is critical.

Subjects will perform standard WP or CAD tasks under various combinations of resolution and addressability. Human performance will be assessed through several dependent measures: (1) task accuracy, (2) task edit time, (3) mental workload and (4) a subjective rating of image quality. With these data, the functional relationships between the two display parameters and human performance within each application will be modeled with a metric of the relationship between line width and lines per inch, the Resolution Addressability Ratio, or RAR (Murch & Beaton, 1986; Murch & Virgin, 1985).
BACKGROUND

Display resolution

A fundamental measure of the resolution of a display system is the Modulation Transfer Function (MTF). The MTF (Figure 2) analytically describes system resolution by characterizing the attenuation of sine-wave patterns transmitted through the imaging system. When a sine-wave pattern of a given frequency is transmitted through the system components, the output modulation of the imaged pattern may not equal the modulation of the input signal.

The psychological importance of the output modulation is that it can be compared to the required modulation for the visual detection of a sine-wave pattern of a given spatial frequency. Higher spatial frequencies correspond to the perception of edges and fine details within an image, and require greater modulation to be seen (Campbell & Robson, 1968). As shown in Figure 2, these higher frequencies are the most severely attenuated in a display system with a smoothly decreasing MTF. Consequently, the MTF indexes the capacity of the display system to present fine spatial detail, and, thus, the “quality” of the displayed information. This conceptualization of resolution has been shown to correlate highly with subjective ratings of image quality and objective performance measures in specific task environments (Beaton, 1984; Task, 1979).

The MTF can be derived from the width of the narrowest line that can be imaged on the display. The luminance profile of this line is known as the line spread function. As shown in Figure 3, the line spread function and the MTF are inversely related to one
Figure 2. A hypothetical Modulation Transfer Function, for a digital imaging system. (C/D refers to cycles per degree of visual angle).
Figure 3. Relationship between the Line Spread Function and the Modulation Transfer Function. The Fourier transforms of the solid and dashed line spread functions yield the solid and dashed MTF's, respectively.
another through the Fourier transform (Dainty & Shaw, 1974; Gaskill, 1978). The "narrower" the line spread function, the "wider" the system MTF. A wide system MTF is desirable, because it means less attenuation of the high frequency information that relates to the perception of quality in displayed images. Since the MTF is derived from the width of a narrow line, system resolution may be operationally defined as line width.

**Display addressability**

For rastered displays, resolution is dependent on the design of the CRT. This is because the MTF is derived from the line width, and the line width is determined by characteristics of the phosphor surface and the electron beam. Addressability is a characteristic of the display controller which drives the CRT. Consequently, many CRT-based display systems have a fixed resolution, but may be driven at a variety of addressability levels. In general, addressability refers to the number of rows of pixels times the number of columns of pixels over some unit area. Since many display systems first determine addressability in the vertical dimension (number of rows of pixels), addressability may be operationally defined as the number of lines per inch.

Since the addressability of a display is under the control of the system design engineer, a question arises as to the proper relations between line width and lines per inch. For example, consider the graphical relations in Figures 1a, 1b and 1c. The graphs plot the luminance profiles of adjacent raster lines. The dotted lines represent the resultant luminance output for each configuration. The line width (resolution) of each configuration is the same 10 mils, but the addressability varies from 50 to 100 to 200 lines per inch. Although resolution is the same for each configuration, a question may be asked whether each configuration will result in the same level of image quality. Two image quality criteria have been proposed to theoretically answer this question (Murch & Bea-
The first of these, called the adjacent raster line criterion, states that individual raster lines should be perceived as a uniform field when all raster lines are activated. For example, Figure 1a will pass the adjacent criterion. The individual lines will be imperceptible when all lines are activated. Figure 1b, when all lines are activated, will be perceived as a series of alternating light and dark lines, because the spaces between the lines will be visible. Hence, it will fail the adjacent criterion. Finally, Figure 1c will also pass the adjacent criterion. The question becomes, what distinguishes the two configurations of 1a and 1c? Will they result in the same level of image quality?

The answer is found in the second criterion, called the alternate raster line criterion, which states that individual lines should be visible when alternate lines are made inactive. While the adjacent criterion is intended to eliminate image "noise" due to visible dark lines associated with perceptible raster structures, the alternate criterion theoretically optimizes the energy levels (modulation) of high spatial frequency components. In addition, optimizing the modulation of the high frequency components optimizes the modulation of lower frequencies as well (Beaton, 1984). The logic underlying this concept is that the sine-wave pattern generated by the alternate criterion is the highest spatial frequency pattern which the CRT is capable of transmitting. The greater the modulation of this frequency, the greater the modulation of all other frequencies for a display system with a smoothly decreasing MTF.

In the example, when alternate lines of Figure 1c are made inactive, it will fail the alternate criterion, while Figure 1a will pass. Consequently, although each configuration has the same level of resolution, Figure 1b will be perceptually "noisy" and Figure 1c will have a less than optimized MTF. Highest image quality should be obtained from the configuration in Figure 1a.
Relating resolution and addressability: The RAR

A quality metric intended to model the adjacent and alternate criteria is the Resolution Addressability Ratio (RAR) (Murch & Beaton, 1986; Murch & Virgin, 1985). The RAR is expressed as

\[
RAR = \frac{\text{line width}}{\text{line separation}} \tag{1}
\]

where line width (in inches) is given by the full-width of the line spread function at one-half maximum intensity, and line separation is the conversion of lines per inch to inches per line. The RAR values for the three example configurations of Figures 1a, 1b and 1c are 1.0, 0.5 and 2.0 respectively. In an engineering setting, the RAR of a given display is used as a predictor variable in a least-squares regression equation expressed as

\[
M = \frac{2}{\pi} \exp[3.6(RAR) - 7.0(RAR)^2 + (RAR)^3], \tag{2}
\]

where M denotes modulation. This equation computes the modulation of the raster structures associated with the adjacent and alternate criteria. Because RAR is usually calculated in the adjacent case, one-half its value is used when computing the modulation of the alternate case.

Additionally, the addressability level determines the spatial frequency of the raster structures, and can be computed by

\[
\omega = \left[\frac{\pi}{180} \left(\frac{D}{S}\right)\right], \tag{3}
\]

where \(\omega\) is the spatial frequency, D is the viewing distance of the operator from the display screen and S is the line separation from equation 1. This is the spatial frequency of the adjacent case, and is equal to twice the frequency of the alternate case.
The computed modulation values are then compared to the modulation required to
detect periodic gratings of a given spatial frequency (Snyder, 1980), given by
\[ M = b_0 \exp[b_1(\omega)+b_2(\omega^2)+b_3(\omega^4)], \] (4)
in which \( M \) denotes modulation, and the values for the regression coefficients \( b_0, b_1, b_2 \) and \( b_3 \) are \( 1.7062 \times 10^{-3}, 201.6188 \times 10^{-3}, -2.3161 \times 10^{-3} \) and \( 0.2000 \times 10^{-6} \), respectively.

Through the application of equations 1, 2, 3 and 4, the adjacent and alternate image quality
criteria can be evaluated at any level of RAR.

The developers of the RAR have provided display design engineers with a set of
equations modeling the outcome of the adjacent and alternate criteria for any combina-
tion of line width and lines per inch. For example, consider the modeled relations
presented in Table I for the factorial combinations of four levels of resolution (line
widths of 5, 10, 15 and 20 mils) and four levels of addressability (100, 120, 150 and 200
lines per inch). The columns of Table I give the RAR value, the addressability level, the
modulation required for the visibility of the raster structure, the computed modulation
from the RAR equations, and the result of comparing the computed to the required
modulation for the given criterion. The RAR modeling outcomes are that example
displays with an RAR < 0.75 fail the adjacent criterion, and that displays with an RAR
> 2.0 fail the alternate criterion.

Because optimizing the modulation of the alternate criterion optimizes the display
system MTF, it is probably the more critical of the two criteria. Thus, attention should
be directed toward the column presenting the computed modulation of the alternate cri-
terion for each level of RAR. Note that as RAR increases, the modulation of the alternate
criterion decreases. Consequently, for the range of resolution and addressability lev-
els in the example, image quality should decrease as a function of RAR. These example
levels of resolution and addressability will be used in the experiment.
TABLE I

PREDICTED OUTCOMES FOR ADJACENT AND ALTERNATE CRITERIA

TABLE ORDERED ON RAR (LINE WIDTH / LINE SEPARATION)

<table>
<thead>
<tr>
<th>RAR</th>
<th>Line Density</th>
<th>adjacent modulation</th>
<th>alternate modulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>required</td>
<td>computed</td>
</tr>
<tr>
<td>0.50</td>
<td>100</td>
<td>0.20</td>
<td>0.76</td>
</tr>
<tr>
<td>0.60</td>
<td>120</td>
<td>0.33</td>
<td>0.60</td>
</tr>
<tr>
<td>0.75</td>
<td>150</td>
<td>0.78</td>
<td>0.28</td>
</tr>
<tr>
<td>1.00</td>
<td>200</td>
<td>1.00</td>
<td>0.06</td>
</tr>
<tr>
<td>1.00</td>
<td>100</td>
<td>0.20</td>
<td>0.06</td>
</tr>
<tr>
<td>1.20</td>
<td>120</td>
<td>0.33</td>
<td>0.01</td>
</tr>
<tr>
<td>1.50</td>
<td>150</td>
<td>0.78</td>
<td>0.01</td>
</tr>
<tr>
<td>1.50</td>
<td>100</td>
<td>0.20</td>
<td>0.01</td>
</tr>
<tr>
<td>1.80</td>
<td>120</td>
<td>0.33</td>
<td>0.01</td>
</tr>
<tr>
<td>2.00</td>
<td>100</td>
<td>0.20</td>
<td>0.01</td>
</tr>
<tr>
<td>2.00</td>
<td>200</td>
<td>1.00</td>
<td>0.01</td>
</tr>
<tr>
<td>2.25</td>
<td>150</td>
<td>0.78</td>
<td>0.01</td>
</tr>
<tr>
<td>2.40</td>
<td>120</td>
<td>0.33</td>
<td>0.01</td>
</tr>
<tr>
<td>3.00</td>
<td>150</td>
<td>0.78</td>
<td>0.01</td>
</tr>
<tr>
<td>3.00</td>
<td>200</td>
<td>1.00</td>
<td>0.01</td>
</tr>
<tr>
<td>4.00</td>
<td>200</td>
<td>1.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Application-dependent differences in perceived image quality

The proposal that the display system characteristics for optimal image quality vary with specific task or application demands has received a great deal of mention in the literature. Evans and Attaya (1978) suggest that image quality studies should be conducted within the context of specific application areas, using subjects who are highly trained and experienced with the normal demands of that application. Hunt and Sera (1978) propose that an application-dependent dimension affecting perceived image quality is whether the display is utilized in a performance or non-performance environment. Peters (1982) notes that optimal system performance may not be guaranteed by an optimum MTF, since the judged quality of the final system output depends on how individuals use that information. Hence, the utility of an image quality metric must be weighted by a consideration of the application requirements (Lindenberg, 1976). In other words, it appears likely that system design guidelines which receive support from empirical modeling of the information transfer process may interact with criteria for the usefulness of that information as demanded by the application. (Roetling et al., 1968; Bernath et al. 1981).

Systematic investigations of image quality requirements across applications are notably lacking in the literature. In order to explore these proposed differences, two application areas will be tested: (1) Word-Processing (WP) and (2) Computer-Aided-Drafting (CAD). It is hypothesized that subjects in a CAD application, where the exact placement and selection of coordinate locations on the display is critical, should require a higher level of image quality (a function of addressability and resolution) than subjects in a WP application, where the primary task demands involve the manipulation of text and graphic material. The test of this hypothesis will be made by comparing the functional trends between task performance and various levels of display resolution and display
addressability across these two applications.

According to the literature on MTF-based measures of resolution, task performance and subjective ratings of image quality should decrease with increasing line width. According to the RAR modeling outcomes represented in Table I, there should be a strong linear relationship between image quality and RAR. If the two applications referenced in this study demand different levels of image quality, then a main effect of task and significant effects of task by resolution and task by addressability should be anticipated.
METHOD

Subjects

Twenty volunteer subjects were solicited from Tektronix, Inc. to participate in this study. Ten female subjects were selected to participate in the WP task, while ten male subjects participated in the CAD task. Each subject participated in three separate experimental sessions. The first session consisted of pre-testing and training, while the second and third sessions were used to collect actual trial data. No subject participated in more than one session per day, and no subject took longer than five days to complete all three sessions.

In the WP group, all subjects were professional secretaries. Seven reported having attained a moderate level of word processing skill, while the remaining three were self-reported experts. These subjects ranged in age from 26 to 45 years. Subjects in the CAD group were engineers, four of whom reported expert skill levels with computer-aided-drafting packages, while the remaining six reported moderate skill levels. This group of subjects ranged in age from 22 to 60 years. All subjects were solicited from specific application areas in order to conduct the experimental investigation with well-trained participants.

Equipment

A custom-designed monitor was used in the experiment (Tektronix, Prototype GMA 201). The monitor was driven by a computer-graphics workstation (Tektronix, 4115). Resolution and addressability of the digital display system were varied through pro-
grammable circuits. The programmable focus and raster size circuits were calibrated by a direct-view microscope (SKS, Model 25X-SD, with an 0.002 inch reticule), and were controlled during the experimental sessions by a micro-computer (Digital Equipment Corporation, PDP 11/24).

Resolution was varied by manipulating the focus voltage levels with an 8-bit DAC. The voltage manipulations changed the size of the electron beam and, thus, the width of a narrow line. However, it is unlikely that line widths departing from the factory setting maintained the fidelity of other CRT factors contributing to display resolution. The line width of a CRT achieved at the factory may be considered to consist of the unaberrated line width increased by contributions from various aberrations (e.g. magnification, lens aberrations, thermal effects, space charge ...etc) (Lehrer, 1985). Consequently, manipulations of one of these factors, such as focus voltage, may have had a twofold effect: (1) the width of a raster line varied from 5 to 20 mils, and (2) the fidelity of the line width decreased with increasing line width as the edges of the wider lines became more diffuse and defocused. Hence, it is possible that resolution may have been confounded with focus. To eliminate this possible confound, the alternative would have been to build four identical monitors which varied only in their line width. Since no two CRTs are completely identical, this alternative was virtually impossible.

Addressability was varied by manipulating the deflection voltages of the vertical and horizontal amplifiers given by two separate 12-bit DACs. One consequence of the addressability manipulation was the resulting change in image size. Thus, as the active raster area varied vertically from 14 (cm) to 28 (cm), the vertical size of a capital "L" varied from 3.0 (mm) to 6.0 (mm). In a sense, addressability was inextricably confounded with image size in this experiment. In order to avoid the confound, image size could have been held constant by resampling the image (e.g. changing the dot-matrix size of the
display character generator). However, this correction for image size would have introduced a more serious confound, that of changing the structure of the characters. For example, to maintain the vertical size of the capital "L" at 3.0 (mm), the number of dot-matrix rows would have had to vary from 8 (for an addressability of 200 lines per inch) to 4 (for an addressability of 100 lines per inch). It is known that any decrease in the number of dot-matrix rows or columns significantly reduces the image quality of dot-matrix characters (Maddox, 1980).

The experimental workstation consisted of a table with the monochrome monitor mounted at one end, leaving space at the other end for a work area. Each subject was seated in an adjustable-height chair at a "normal" viewing distance (22 inches from the display). The chair height was adjusted so that the subjects' line of sight was perpendicular to the center of the display screen.

Subjects in the WP group used the EDT text-editor (Digital Equipment Corporation), operating in "keypad mode" within an RSX-11M environment. Subjects in the CAD group used the Teknicad drafting package (Tektronix, Inc.), running locally under the control of the graphics workstation.

The experimental manipulations of resolution and addressability were under the control of a laboratory computer (Digital Equipment Corporation, PDP 11/24) with all control software written in FORTRAN-77.
Procedure

The experimental tasks performed by the subjects were developed from a model of a standard task performed by experienced operators within each application. The model resulted from data obtained during an informal survey, and the standard task was defined by a set of discrete operations categorized as activities of reading, function selection, data entry, or cursor positioning. From this taxonomy, multiple renditions of the standard task were developed for each application. Hence, for each level of resolution and addressability, the subject performed the same task; however, the task conditions varied across trials.

Pre-test

Before the experimental trials began, each subject participated in a pre-test session. The pre-test consisted of two parts. In the first part, subjects performed a simulated paper-and-pencil editing task. Subjects were presented with either a text or line drawing stimulus developed in the same manner as the experimental stimuli. Each pre-test stimulus was partitioned into two halves. The top half contained the "correct" version of text or line drawing, while the bottom half contained an "altered" version. The text had been altered by substituting random words with the word that followed it in the dictionary, or by substituting a random letter in a word to the letter that followed it in the alphabet. The line drawings had been altered by deleting random lines. The number of alterations varied from 2, 4, 6 or 8.

At the start of each pre-test trial, subjects were allowed to review the "correct" version of the stimulus with the "altered" version hidden from view. When the subject indicated "READY," the experimenter presented the "altered" version. The task was to identify all alterations to the "correct" version by either circling the altered portions of text or
replacing the missing lines. The subject was able to reference the "correct" version throughout the trial. The trial ended when the subject indicated "FINISHED." Upon completion of the editing task, the subject was asked to estimate the amount of time to the nearest second that elapsed between "BEGIN" and "FINISHED." Each subject received eight trials, a factorial combination of two types of stimulus with four levels of alterations. The order of presentation was uniquely randomized for each subject.

Three performance measures were recorded: (1) accuracy (proportion of alterations identified), (2) edit time (time-to-complete the trial divided by the number of alterations identified), and (3) mental workload (ratio of actual-to-estimated time to complete the trial). These three performance measures were used to establish baseline data for possible group differences in editing speed and accuracy. In addition, these performance measures provided data on possible differences between the two experimental tasks in terms of their relative difficulty, mental workload and time to complete.

In the second part of the pre-test, display size preferences were examined using the method of pair comparisons (Guilford, 1954). The displayed images used for the stimuli were photographs of either text or line drawings. The photographs were developed at four different sizes, which matched the experimental display sizes resulting from the manipulation of display addressability. Subjects were presented with pairs of a single image type in all possible size combinations. The order of presentation was randomized uniquely for each subject. Subjects were instructed that the photographs represented displays of varying size, and the task was to indicate their preferred display size. The results from this pre-test were used to establish a baseline for display size preferences independent of manipulations of line width and line separation.
Word-Processing task

Following the pre-test, subjects in the WP group received instruction and practice on a limited subset of commands to operate the EDT editor. This command set allowed the subjects to move the cursor in four directions (up, down, forward and backward), to vary cursor movement speed (character, word or line at a time), and to delete either a character or a word from the displayed text. The subjects made text entry into the display by directly typing in the text.

The standard WP task involved editing a document on the display. On each experimental trial, the subject was presented with a hard-copy source document that served as the "correct" version of the stimulus, while the displayed soft-copy served as the "altered" version. Unlike the pre-test, where alterations were identified, the subject used the editor to correct the displayed version. Each trial began when the experimenter indicated "BEGIN" and ended when the subject indicated "FINISHED."

Immediately following each trial, the subject estimated to the nearest second the time taken to complete the trial. Next, the subject was asked to rate the "quality" of the display, on a scale of 0 (worst imaginable) to 8 (best imaginable). Each integer step on the rating scale was associated with a word describing display quality; that is, "awful" (1), "poor" (2), "marginal" (3), "passable" (4), "OK" (5), "good" (6) and "excellent" (7). These descriptors of display quality have been found to share equal psychological distance (Jones & Marks, 1985).

Computer-Aided-Drafting task

Following the pre-test session, subjects in the CAD group received instruction and practice on a command subset of the Teknicad (Tektronix) drafting package. The command subset allowed the subjects to draw and delete lines and arcs from the screen.
workspace. The standard CAD task involved the correction of line drawings presented on the display. Subjects in this group received the same number of practice sessions as subjects in the WP group.

On each experimental trial, the subject was presented with a hard-copy source document of the line-drawing that served as the "correct" version. The soft-copy displayed drawing was the "altered" version, and the subject used the Teknicad package to edit the displayed version so that it corresponded to the hard-copy source document. The trial began and ended in the identical fashion used for the WP group. A time estimation and "quality" rating followed each trial.

On each trial in the experiment, four measures of human performance were collected: (1) accuracy (proportion of alterations edited), (2) edit time (time-per-edit, defined as the time to complete trial divided by number of alterations edited), (3) workload (ratio of actual-to-estimated time to complete trial) and (4) a subjective assessment of image quality. Of these four dependent measures, only time estimations (workload) have not been previously reported as an index of imaging system performance. However, the literature does report time-estimations to be reliable and non-intrusive predictors of mental workload (Casali & Weirwille, 1983). It was hypothesized that mental workload would increase as image quality decreased.

Experimental display factors

Four levels of resolution (line widths of 5, 10, 15, and 20 mils) and four levels of addressability (100, 120, 150 and 200 lines per inch) were used in this experiment. For the purpose of computing RAR values, the conversion of line density (lines per inch) to line separation (inches per line) resulted in peak-to-peak separations of 10, 8.34, 6.67 and 5 mils. This resulted in twelve unique RAR values (0.50, 0.60, 0.75, 1.00, 1.20, 1.50, 1.80,
2.00, 2.25, 2.40, 3.00, and 4.00). Each subject received a total of 16 trials in the experiment, consisting of the factorial combination of four levels of addressability and four levels of resolution. The presentation order of the resolution and addressability levels were randomized uniquely for each subject.
RESULTS

Pre-test data

The data from the first part of the pre-test were analyzed with a three factor (group, task, number-of-alterations), mixed factor (i.e. subjects nested within group) analysis of variance procedure for each of the accuracy, edit time and workload dependent measures.

The accuracy measure was computed by dividing the number-of-edits by the number-of-alterations in each trial condition. As shown in Figure 4, a main effect of task was obtained \( F(1,18) = 19.71, p = 0.0003 \), indicating it was more difficult to be accurate on the CAD task. No other effects were significant at \( \alpha = 0.05 \).

The edit time measure was defined as the time per trial divided by the number of edits per trial. As depicted in Figure 5, the main effect of task was significant \( F(1,18) = 100.08, p < 0.0001 \). Hence, subjects in the CAD task took less time-per-edit than subjects in the WP task. A significant task by group interaction also was obtained with the edit time measure, as shown in Figure 6 \( F(1,18) = 47.53, p < 0.0001 \). A post hoc Newman-Keuls Multiple Range Test indicated that all four means comprising this interaction were significantly different from each other \{ Cdiff \( 2 \) (18 d.f.) = 1.53 \( p < 0.05 \), \{ Cdiff \( 3 \) (18 d.f.) = 1.82 \( p < 0.05 \), \{ Cdiff \( 4 \) (18 d.f.) = 2.00 \( p < 0.05 \). The trends reveal that while both groups took less time-per-edit when editing the line drawings (i.e. CAD task) as compared to text (i.e. WP task), the secretaries were more consistent performers across the two tasks, because their edit time varied less between the two tasks.
Figure 4. Main effect of task, averaged over 10 subjects, on the accuracy measure for the pre-test session.
Figure 5. Main effect of task, averaged over 10 subjects, on the edit time measure for the pre-test session.
Figure 6. Task by group interaction, averaged over 10 subjects, on the edit time measure for the pre-test session.
No other effects were found to be statistically significant at \( \alpha = 0.05 \).

Mental workload was defined as the ratio of actual and estimated time-to-complete each trial. No effects were found from the ANOVA on the workload measure.

The size-preference data obtained in the second part of the pre-test session were analyzed according to the Pair-Comparisons Procedure outlined by Guilford (1954). The preference data were transformed into psychological scale values sharing a common zero point and a standard deviation of 2.0. Figure 7 plots the scale values as a function of subject group and size. The four display sizes of these photographic stimuli matched the size variations resulting from the experimental manipulations of addressability. The smallest display size (14 cm X 18 cm) was achieved at an addressability of 200 lines per inch, and the largest display size (28 cm X 35 cm) was achieved at an addressability of 100 lines per inch. The trends in the data indicate that subjects in both groups preferred the mid-range sizes over the smallest and largest sizes. Guilford’s procedure does not provide a statistical test to determine if these apparent differences are significant. However, it is possible to estimate the standard error of the mean for this distribution of values. The number of raw scores comprising each scale value was 4, one score from each subject group by stimulus type (line drawing image versus text image) condition. Since the common standard deviation was 2.0, the common standard error was 1.0 (the standard deviation divided by the square root of 4). Using this as a rule of thumb measure of significance, it is apparent that the engineers had no clear display size preference, but the secretaries clearly disliked the smallest size and showed greatest preference for the middle sizes.
Figure 7. Display size by subject group interaction, averaged over 10 subjects, on the size preference scale values for the pre-test session.
Experimental data

The experimental data obtained from the four measures (accuracy, edit time, workload and subjective ratings) were analyzed separately, using a three factor (resolution, addressability, and task), mixed factor (i.e. subjects nested within task) analysis of variance procedure.

The accuracy measure was computed as in the pre-test, and a main effect of task (Figure 8) was obtained \( F (1,18) = 10.76, p = 0.0042 \). As in the pre-test results (Figure 4), the WP task produced fewer errors than the CAD task, suggesting it was more difficult to be accurate on the CAD task. No other significant effects were obtained with the accuracy measure.

The edit time measure, computed in the same manner as the pre-test above was significant \( F (1,18) = 5.37, p = 0.0324 \), as shown in Figure 9. However, unlike the pre-test results, where subjects took less time-per-edit in the CAD task (Figure 5), subjects took less time-per-edit in the WP task. No other significant effects were obtained form the ANOVA on the edit time measure.

Similar to the pre-test analyses, mental workload was computed as the ratio of the subjects' actual to estimated time-to-complete each trial. No significant effects were observed with this dependent measure.

With the subjective image quality rating measure, main effects of resolution \( F (3,54) = 68.95, p < 0.0001 \), addressability \( F (3,54) = 19.48, p < 0.0001 \), and task \( F (1,18) = 9.85, p = 0.0057 \) were significant.

As shown in Figure 10, the task main effects suggests implied that, over all levels of resolution and addressability, the display was perceived as having less quality for subjects in the WP task than subjects in the CAD task.
Figures 2. Main effect of task, averaged over 10 subjects, on the accuracy measure for the experimental trials.
Figure 9. Main effect of task, averaged over 10 subjects, on the edit time measure for the experimental trials.
Figure 10. Main effect of task, averaged over 10 subjects, on the quality rating measure for the experimental trials.
The main effect of resolution (Figure 11) reveals that quality ratings decreased as line width increased. The post hoc Newman-Keuls pairwise comparison indicated that the effect of resolution reached asymptote at 10 mils \( \text{Cdiff}_2 \) (54 d.f.) = 0.44, \( p < 0.05 \), \( \text{Cdiff}_3 \) (54 d.f.) = 0.54, \( p < 0.05 \), and \( \text{Cdiff}_4 \) (54 d.f.) = 0.59, \( p < 0.05 \).

Post hoc comparisons on the main effect of addressability (Figure 12) found that only at 200 lines per inch were the subjective ratings different \( \text{Cdiff}_2 \) (54 d.f.) = 0.40, \( p < 0.05 \), \( \text{Cdiff}_3 \) (54 d.f.) = 0.49, \( p < 0.05 \), and \( \text{Cdiff}_4 \) = 0.54, \( p < 0.05 \).

A significant interaction of task by addressability was obtained (Figure 13) with the subjective ratings measure \( F (3,54) = 11.97, p < 0.0001 \). Post hoc comparisons indicated that the source of the interaction was the subjective ratings given in the WP task to those displays at 200 lines per inch. This point differed from all others, and all others did not differ from each other \( \text{Cdiff}_2 \) (54 d.f.) = 0.57, \( p < 0.05 \), \( \text{Cdiff}_3 \) (54 d.f.) = 0.69, \( p < 0.05 \), \( \text{Cdiff}_4 \) (54 d.f.) = 0.76, \( p < 0.05 \), \( \text{Cdiff}_5 \) (54 d.f.) = 0.80, \( p < 0.05 \), \( \text{Cdiff}_6 \) (54 d.f.) = 0.84, \( p < 0.05 \), \( \text{Cdiff}_7 \) (54 d.f.) = 0.87, \( p < 0.05 \), \( \text{Cdiff}_8 \) (54 d.f.) = 0.90, \( p < 0.05 \).
Figure 11. Main effect of resolution, averaged over 20 subjects, on the quality rating measure for the experimental trials. Vertical bars represent the standard error.
Figure 12. Main effect of addressability, averaged over 20 subjects, on the quality rating measure for the experimental trials. Vertical bars represent the standard error.
Figure 13. Addressability by task interaction, averaged over 10 subjects, on the quality rating measure for the experimental trials. Vertical bars represent the standard error.
RAR analysis

From the factorial combinations of four levels of resolution and four levels of addressability, 16 levels of RAR were achieved. A linear regression of the mean quality ratings as a function of RAR (Figure 14) resulted in $r^2 = 0.78$. Because of the significant task by addressability interaction, the data was partitioned by task, giving separate regressions of mean quality ratings for each application. In the CAD application (Figure 15), the obtained $r^2$ was 0.79, while $r^2 = 0.94$ was obtained in the WP application (Figure 16). It is clear that RAR accounts for a sizable portion of the variance for both sets of subjects, and particularly so for the WP application.

Of the sixteen experimental RAR values, only twelve were unique. Thus, at RAR values of 1.0, 1.5, 2.0 and 3.0, there were two data points, each having different levels of resolution and addressability. To gain more insight into a possible application difference, t-tests were performed for each application on the two displays at the duplicated RAR values, using an overall combined $\alpha = 0.0315$ for the eight separate tests. The purpose in conducting these pairwise comparisons was to investigate statistical evidence in support of an hypothesis that RAR accounted for more of the variance in the WP application than in the CAD application. If RAR alone accounts for more variance in the WP application, then fewer significant t-tests might result. Accordingly, in the CAD application, the two displays at RAR = 1.0 were not different, but the two displays at each RAR of 1.5, 2.0 and 3.0 were significantly different from each other, { $T$ (19 d.f.) = 18.58, $p = 0.002$}, { $T$ (19 d.f.) = 37.55, $p = 0.0002$} and { $T$ (19 d.f.) = 9.53, $p = 0.013$}, respectively. For the WP application, only the two displays at RAR = 1.0 were significantly different from each other { $T$ (19 d.f.) = 8.68, $p = 0.0163$}, while the two displays at the remaining RAR values were not different.
Figure 14. Regression of quality ratings on RAR, averaged over 20 subjects in the combined Word-Processing and Computer-Aided-Drafting applications. The number over each data point is the line width in mils.
Figure 15. Regression of quality ratings on RAR, averaged over 10 subjects in the Computer-Aided-Drafting application. The number over each data point is the line width in mils.
Figure 16. Regression of quality ratings on RAR, averaged over 10 subjects in the Word-Processing application. The number over each data point is the line width in mils.
DISCUSSION

Pre-test results

The task by group interaction shown in Figure 6 provides insight into the editing skills of the two groups of subjects. Overall, secretaries perform faster than engineers when editing text, whereas engineers are faster than secretaries when editing line drawings. Secretaries appear to be more consistent performers across the two tasks, which might be attributed to greater editing skills. However, the lack of a significant main effect of group with the edit time measure implies that neither group was faster than the other. In addition, the failure to obtain a significant main effect of group with the accuracy measure suggests secretaries and engineers are about equal in their accuracy when editing either text or line drawings. Thus, it can reasonably be concluded there were no major differences in editing skills between the two groups of subjects.

The significant main effects of task obtained with the edit time and accuracy measures indicate the experimental tasks differ in the degree of accuracy and speed with which they can be accomplished. The main effect of task on the pre-test accuracy measure (Figure 5) suggests it was more difficult to be accurate on the CAD task, while the main effect of task on the pre-test edit time measure indicates it took a longer period of time to make a single edit on the WP task (Figure 9). The failure to detect any pre-test workload difference between the two tasks suggests that there were no appreciable differences in workload demands between the two tasks.
The implications to be drawn from part one of the pre-test are clear. There were no major group differences in editing skills. The tasks were not equal in their degree of difficulty and the time it took to perform them, but these differences did not bring about differences in workload.

Experimental results

Since the tasks differed in their degree of difficulty and the time it took to complete them, it is likely that the main effects of task found on the experimental accuracy (Figure 8) and edit time (Figure 9) measures were not due to manipulations of resolution and addressability. In both the pre-test and the experiment, subjects were more accurate on the WP task. However, the trend for edit time reverses itself from pre-test to experimental results. Subjects took less time on the CAD task when editing was done with paper and pencil, as in the pre-test (Figure 5). But during the experimental trials, subjects took less time to edit the WP task (Figure 9). A reasonable explanation for the apparent reversal is the difference in additional editing time imposed by the text and drafting editors. That is, the EDT text editor allowed cursor movement, text deletion and text entry all within a single mode, while the Teknicad drafting package required the draw-a-line, draw-an-arc or delete functions to be read into RAM prior to execution.

Of the four dependent measures recorded during the experimental trials, the subjective rating of image quality was the most sensitive to the effects of task, resolution and addressability. The main effect of resolution (Figure 11) is consistent with previous studies (Snyder, 1985), and can be quantitatively predicted by MTF-based measures of image quality (Beaton, 1984; Task, 1979). The subjective rating asymptote at 10 mils suggests a perceptual "bound" for MTF bandwidth improvements, since no appreciable effect on image quality was observed for the smallest line width. It may be the case that
the increased bandwidth associated with the smallest line width is beyond the spatial frequency response of the human visual system (Snyder, 1980).

The main effect of addressability (Figure 12) is consistent with the theoretical position that very high line density levels which produce RAR values much greater than 1.0 may fail to pass the alternate criterion, constituting a less than optimized MTF (Murch & Beaton, 1986). The post hoc analysis found a significant decrease in the image quality ratings at an addressability of 200 lines per inch. Of the 4 displays at this addressability level, 3 fail the alternate criterion. Table II lists the predicted outcomes of the adjacent and alternate criteria for each RAR level. The list is ordered by line density, and the results can be compared to the addressability main effect in Figure 12, where image quality ratings are equal for displays of 100, 120 and 150 lines per inch, but show a decrease at 200 lines per inch. Three out of the four displays at 200 lines per inch fail the alternate criterion, while this occurs only 2 out of 4 times for displays at 150 lines per inch, 1 out of four times for displays at 120 lines per inch, and 0 times for displays at 100 lines per inch.

The main effect of task found on subjective image quality ratings supports an hypothesis that image quality criteria are governed by application demands (Peters, 1982; Bernath, Kretz & Wood, 1981; Evans & Attaya, 1978; Lindenberg, 1976; Hunt & Sera, 1968; Roetling, Trabka & Kinsly, 1968). Over all levels of resolution and addressability, the displayed images were judged as having less quality for subjects in the WP task. This result implies that image quality is more "critical" in a word processing application, which is opposite of the trend anticipated by the experimental hypothesis. Since the exact placement of a one-pixel wide line is critical to drafting, it was hypothesized that decreases in image quality should affect that application more than word-processing, where the primary task demand is the recognition and manipulation of text.
TABLE II

PREDICTED OUTCOMES FOR ADJACENT AND ALTERNATE CRITERIA

TABLE ORDERED ON LINE DENSITY (LINES / INCH)

<table>
<thead>
<tr>
<th>Line Density</th>
<th>RAR</th>
<th>required</th>
<th>computed</th>
<th>pass/fail</th>
<th>required</th>
<th>computed</th>
<th>pass/fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>1.00</td>
<td>1.00</td>
<td>0.06</td>
<td>pass</td>
<td>0.20</td>
<td>0.76</td>
<td>pass</td>
</tr>
<tr>
<td>200</td>
<td>2.00</td>
<td>1.00</td>
<td>0.01</td>
<td>pass</td>
<td>0.20</td>
<td>0.06</td>
<td>fail</td>
</tr>
<tr>
<td>200</td>
<td>3.00</td>
<td>1.00</td>
<td>0.01</td>
<td>pass</td>
<td>0.20</td>
<td>0.01</td>
<td>fail</td>
</tr>
<tr>
<td>200</td>
<td>4.00</td>
<td>1.00</td>
<td>0.01</td>
<td>pass</td>
<td>0.20</td>
<td>0.01</td>
<td>fail</td>
</tr>
<tr>
<td>150</td>
<td>0.75</td>
<td>0.78</td>
<td>0.28</td>
<td>pass</td>
<td>0.09</td>
<td>0.97</td>
<td>pass</td>
</tr>
<tr>
<td>150</td>
<td>1.50</td>
<td>0.78</td>
<td>0.01</td>
<td>pass</td>
<td>0.09</td>
<td>0.28</td>
<td>pass</td>
</tr>
<tr>
<td>150</td>
<td>2.25</td>
<td>0.78</td>
<td>0.01</td>
<td>pass</td>
<td>0.09</td>
<td>0.02</td>
<td>fail</td>
</tr>
<tr>
<td>150</td>
<td>3.00</td>
<td>0.78</td>
<td>0.01</td>
<td>pass</td>
<td>0.09</td>
<td>0.01</td>
<td>fail</td>
</tr>
<tr>
<td>120</td>
<td>0.60</td>
<td>0.33</td>
<td>0.55</td>
<td>fail</td>
<td>0.05</td>
<td>1.00</td>
<td>pass</td>
</tr>
<tr>
<td>120</td>
<td>1.20</td>
<td>0.33</td>
<td>0.01</td>
<td>pass</td>
<td>0.05</td>
<td>0.55</td>
<td>pass</td>
</tr>
<tr>
<td>120</td>
<td>1.80</td>
<td>0.33</td>
<td>0.01</td>
<td>pass</td>
<td>0.05</td>
<td>0.12</td>
<td>pass</td>
</tr>
<tr>
<td>120</td>
<td>2.40</td>
<td>0.33</td>
<td>0.01</td>
<td>pass</td>
<td>0.05</td>
<td>0.01</td>
<td>fail</td>
</tr>
<tr>
<td>100</td>
<td>0.50</td>
<td>0.20</td>
<td>0.76</td>
<td>fail</td>
<td>0.04</td>
<td>1.00</td>
<td>pass</td>
</tr>
<tr>
<td>100</td>
<td>1.00</td>
<td>0.20</td>
<td>0.06</td>
<td>pass</td>
<td>0.04</td>
<td>0.76</td>
<td>pass</td>
</tr>
<tr>
<td>100</td>
<td>1.50</td>
<td>0.20</td>
<td>0.01</td>
<td>pass</td>
<td>0.04</td>
<td>0.28</td>
<td>pass</td>
</tr>
<tr>
<td>100</td>
<td>2.00</td>
<td>0.20</td>
<td>0.01</td>
<td>pass</td>
<td>0.04</td>
<td>0.06</td>
<td>pass</td>
</tr>
</tbody>
</table>
A possible explanation for the result that image quality may be more critical for text than for line drawings can be deduced from the addressability by task interaction (Figure 13). The post hoc analysis found that the quality rating given by subjects in the WP task for displays at 200 lines per inch differed significantly from all other data points, none of which differed from each other. One account of this interaction is that addressability affected image quality more for images of greater active pixel density. That is, the text images may have contained more active pixels per unit distance than the line drawing images. A narrow line imaged by the character generator of the GMA-201 monitor, such as the vertical or horizontal portion of a capital "L", is two pixels wide, while a narrow line drawn with the Teknicad drafting package is one pixel wide. This difference in the luminance profile of single lines becomes critical when the width of the line is greater than the distance between lines, such as occurs with a line separation of 5 mils (200 lines per inch) and line widths of 10, 15 and 20 mils. In these cases, adjacent pixels and lines overwrite each other, reducing the high frequency content of the image. Thus, at an addressability of 200 lines per inch, the high frequency content of the one-pixel wide line drawings may have been unaffected, while the high frequency content of the two-pixel wide text may have become obscured, making it difficult to read.

RAR results

The a priori hypothesis of Table I was that image quality should decrease as a function of RAR, and that there should be a strong linear relationship between the two variables for the range of resolution and addressability levels used in the experiment. The results are that RAR accounts for 79% of the observed variance in the data. These results support the theoretical position of optimizing display image quality by optimizing the modulation of the alternate criterion. Also, when the data is partitioned by application, RAR appears to account for more of the variance in the WP application (94%)
than in the CAD application (77%). Additional evidence for this position was found in the t-tests of the two displays at the duplicated RAR values. For the CAD application, 3 of 4 t-tests found a significant difference, but only 1 of 4 were significantly different in the WP application.

In summary, the RAR analyses suggest that the theory underlying the display metric receives empirical validation. Optimum relations between resolution and addressability can be analytically determined by modeling the relationship through the adjacent and alternate criteria. The RAR metric has also appeared to detect an application difference in display image quality requirements. As seen in Figure 15, resolution is critical to image quality in the CAD application, but the effect of addressability is minimal across the four resolution levels. This is not the case for the WP application in Figure 16, where image quality decreases at the higher addressability levels for each level of resolution. This trend suggests that while resolution is important in the WP application, it is constrained by addressability.

The suggested RAR finding of application-specific image quality requirements is consistent with the significant effects of task and the task by addressability interaction. It also offers additional support for an hypothesis that applications with typical information displays of high active pixel density, such as word-processing, will be critically affected by addressability as well as resolution.
CONCLUSIONS

The four dependent measures taken in this experiment can be conceptualized as assessing image quality in terms of productivity (accuracy and edit time measures), mental workload (time-estimations) and esthetic judgements (subjective ratings). The failure of the productivity measures to detect significant effects due to experimental manipulations of image quality implies that task performance is not an issue over the range of resolution and addressability levels used. CRT technology may have advanced to a state where displays are capable of delivering image quality at a level sufficient for the accurate and efficient performance of word-processing and computer-aided-drafting tasks. Additionally, any difficulties the early visual system may have in encoding stimulus energy associated with poor image quality does not appear to cause increased mental workload demands on the human information processing system. However, people are capable of perceiving differences in image quality. Some displays will image better than others. This study replicates previous findings that the greater the bandwidth of the display MTF, the higher the image quality. Addressability becomes critical at excessive line densities that attenuate the MTF associated with a given resolution level.

There is an apparent difference in image quality criteria between the two applications tested, and this difference significantly impacts display addressability. The human factors literature suggests that criterion differences result from application-dependent demands made on the visual system (Roetling et.al., 1968; Bernath et.al., 1981). The suggestion is that some variable or set of variables common to applications requiring a display device influence the perceived image quality. Consequently, if those variables
could be found and their relationship to perceived image quality understood, then engineering efforts could be directed toward application specific displays. The hypothesis to be advanced here is that these variables might best be understood by characterizing the prototypical images of a given application, as opposed to the proto-typical tasks or proto-typical operators. As discussed in the addressability by task interaction, the images of the two applications in this study may have differed in their active pixel density. Thus, application-dependent image quality criteria may vary with such factors as the active pixel density of the proto-typical image associated with a given application.

The data trends of Figures 15 and 16 imply that in applications heavily dependent on the display of alphanumerics, such as word-processing and business accounting programs, display addressability will have a major influence upon the perception of image quality. As the density of active pixels is likely to be high in the typical information displays of these applications, it is critical not to overaddress. Conversely, in applications where the active pixel density may be less, such as drafting, technical data analysis and instrument displays, resolution will influence perceived image quality more than display addressability. Hence, it is possible to underaddress in these applications without experiencing severe image quality decrements.
REFERENCES


