CRT Anti-glare Treatments, Image Quality, and Human Performance

by

Marc W. Hunter

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Industrial Engineering and Operations Research

APPROVED:

Harry L. Snyder, Chairman

Robert J. Beaton

Gregory J. Buhyoff

Albert M. Prestrude

Dennis L. Price

May, 1988

Blacksburg, Virginia

CRT Anti-glare Treatments, Image Quality, and Human Performance

by

Marc W. Hunter

Harry L. Snyder, Chairman

Industrial Engineering and Operations Research

(ABSTRACT)

This dissertation was a two-phase effort. Phase I investigated the physical image quality of 16 mesh, etched, and quarterwave antireflection filters for varying levels of filter transmissivity. Three levels of ambient lighting and two levels of monochrome CRT resolution were combined factorially with the filters. In addition, user measures of readability, legibility, and perceived image quality were obtained for these same filter and environmental conditions. Quantitative models were developed to predict the performance and subjective data based on signal and noise measures derived from the physical measurements. Phase II examined the effects of a wide range of filter transmissions and diffuse illuminance on measured image quality and the same user measures as in Phase I.

Phase I showed that while none of the glare filters yielded improved readability or legibility over a baseline condition, the etched and low transmission filters were notable for their degradation of human performance. Mesh and quarterwave filters were found to improve perceived image quality when a specular glare source was present. Modeling was minimally successful for the reading and legibility tasks, but yielded good fit models for perceived image quality. Phase II showed that when even extreme losses in display contrast occurred, users were capable of good reading and legibility performance. Perceived image quality was inversely related to illuminance level. Prediction of performance by image quality metrics was generally not too successful.

It was concluded that in office-type environments, mesh or quarterwave filters can be used to improve perceived image quality when specular glare sources are present, but that no anti-glare filters yielded enhanced short-term readability or legibility over a baseline. Etched filters were not

recommended. Measures of physical image quality proved to be good predictors of perceived image quality, but not of timed measures of readability or legibility. Under moderate lighting conditions, monochrome CRTs should be fitted with fairly high transmission filters as it was found the contrast enhancement offered by low transmission filters had negligible effects on performance. Finally, consistent and repeatable findings of degraded legibility for high luminance contrast levels (low illuminance) generated questions as to the existing standards regarding maximum contrast requirements for CRT use.

Acknowledgements

The author thanks and Texas Instruments, Inc. for the continued support of the graduate internship program that funded much of this effort. Acknowledgement is also give to and Optical Coatings Laboratories, Inc. for provision of all contrast enhancement, etched, and coated filters used in this research.

While many individuals contributed to this dissertation, a few deserve special mention. Committee members Drs. Beaton, Buhyoff, Prestrude, and Price are gratefully acknowledged for their insightful and valuable critiques of the study. assistance throughout the many tedious hours of data collection made the task almost enjoyable. was invaluable in his engineering support of this research, but is perhaps more appreciated for his encouragement and friendship. The author is indebted to Dr. Harry Snyder for his suppportive guidance and for his provision of an environment where individuals make their own mistakes, learn from these mistakes, and in the process become independent and capable researchers. The encouragement and perspective offered by tempered the sometimes self-centered and preoccupied behavior that accompanies an effort of this scope. Finally, the author's family is thanked for the patient and seemingly limitless support they have provided throughout this research.

Table of Contents

List of Illustrations

List of Tables

INTRODUCTION

Rationale

In 1981 the National Institute for Occupational Safety and Health (NIOSH) asked the National Academy of Sciences to critically examine the existing literature pertaining to visual issues encountered in occupational video viewing, to identify methodological problems, and to suggest areas of research needed to resolve these problems. With this in mind, the National Research Council's Committee on Vision established a panel of individuals with expertise ranging from ophthalmology and optometry to display and human factors engineering to investigate these issues. A 1983 report released by this panel (National Research Council, 1983) pointed out that one of the two major areas of research deficiency was the relationship of display characteristics to workplace conditions. Specifically, the report singled out research considering display design, workplace illumination, and the effective suppression of glare as requiring immediate and careful attention.

The panel suggested that glare filter effectiveness should be measured as a function of type and location of the glare source. Further, image quality measurements should be made to determine and quantify any degradation that may occur with the addition of the various filter technologies. Filter transmission characteristics under differing environmental conditions and their effects on visual task performance were also identified as important. The report went on to explain that one of the main goals behind this research should be "to cut through the morass of arbitrary, capricious, and often misleading claims made by some filter manufacturers." Clearly, this report sent out a strong signal to the scientific community that attention should be focused on these issues in a quantitative and objective manner. However, since publication of the NRC report, only a handful of relevant studies has been performed. Unfortunately, for the most part, even these investigations have lacked the level of scientific rigor and scope necessary to address adequately the issues or to generalize beyond their specific applications.

With the above in mind, the present investigation was designed to answer those research questions raised in the NRC report. In addition, the research extends into the areas of human performance prediction based on measures of image quality, as well as the optimization of glare filter characteristics such as antireflection and contrast enhancement. The displayed image quality under various anti-glare treatment and lighting conditions was quantified through microphotometric measurements and subsequent analyses. Upon completion of these measurements, extensive human performance studies were conducted using the same anti-glare and lighting conditions. Legibility, readability, and perceived image quality measures were used in an attempt to cover a range of human information processing complexity. Finally, although glare problems can exist in some form for nearly all electronic display technologies, the physical properties of the monochrome cathode ray tube (CRT) that leave it particularly susceptible to glare, and its widespread use in business, industry and the military singled it out as the logical host display.

FAMILIARIZATION WITH BASIC CONCEPTS

Types of Glare

The lighting research industry has made the distinction between visual discomfort glare and visual impairment or disability glare and it appears that most workers in this area adhere to the same approach (Brown, Dismukes, and Rinalducci, 1982; Cakir, Hart, and Stewart, 1980; National Research Council, 1983). Brown et al. (1982) define glare as a sensation produced by luminances within the visual field that are sufficiently greater than the luminance to which the eyes are adapted that causes annoyance, discomfort, or loss in visual performance or visibility. The magnitude of this sensation depends on factors such as size, position, and luminance of sources or reflecting surfaces as well as the luminance to which the eye is adapted.

The office environment rarely contains glare sources of sufficient intensity to be capable of producing visual discomfort (Cakir et al., 1980). Therefore, the present discussion will focus on the causes of disability glare, its effects on visual performance, and the anti-glare filters that have been designed to eliminate it. Disability glare is reflected or emitted light that reduces visual performance and visibility. Brown et al. (1982) point out that disability glare can be caused by light scattered within the eye that serves to reduce contrast at the retina or by specular reflections from glossy surfaces that produce a veil of light, thereby reducing contrast at the display. This loss of contrast at either the retina or the display has been shown to lead to decrements in selected visual tasks when contrast reduction is moderate and, when extreme, can completely mask the displayed information. These findings will be discussed in more detail in later sections.

Isensee and Bennett (1983) and Cakir et al. (1980) categorize discomfort and disability glare into either direct or reflected glare. Direct glare is produced by a luminous source being directed into the individual's eyes. Examples of direct glare sources can range from oncoming automobile headlights to sunlight entering through a window immediately behind the computer display at which the individual is working. Reflected glare occurs when a possible direct glare source is reflected by a specular or glossy surface and into the operator's eyes. While direct glare sources in the operating environment should certainly be avoided, anti-glare filters will have no effect on the loss of retinal contrast that occurs with their presence. Where anti-glare treatments are particularly useful is in dealing with reflected glare from either the untreated front surface of the CRT (about 4% reflectance) or the phosphor deposited surface (from 22% to 27% diffuse reflectance). Thus, the remainder of this dissertation will be limited to discussion of glare reflected from the front surface, the phosphor surface, or any of the glass interfaces between these two that are capable of causing some loss in visual performance or perceived image quality.

Effects of Reflected Glare

Reflection from the phosphor plane. Reflected glare from the CRT phosphor surface will ordinarily make the unfiltered, unexcited areas appear a diffuse, milky color of some luminance greater than zero. The magnitude of this luminance varies directly with the luminance of the light source and the distance of this source from the phosphor plane. When no light source is present (i.e., in a dark room), and the luminance of the unexcited phosphor will usually be very small and can be attributed to the "off" voltage of the electron beam producing some minimal phosphor excitation. In this dark environment, maximum luminance modulation (M) is found between the "on" and "off" areas of the screen, where M is defined as the difference between the maximum and minimum displayed luminance divided by the sum of the maximum and minimum luminances. This relationship is defined in the equation below as

$$M = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}, \qquad (1)$$

where M is the luminance modulation,
L_{max} is the maximum luminance, and
L_{min} is the minimum luminance.

Clearly, with zero background luminance present, M is unity. When a light source is introduced, however, and both the maximum and minimum luminance are increased, the luminance modulation can range from near zero under high illuminance levels to nearly unity when low illuminance is present.

Blackwell (1946) in his classic determination of contrast thresholds of the human visual system recorded responses to test stimuli, producing very stable threshold functions for various adaptation luminances and stimulus sizes. In general, he showed that increasing stimulus size decreased the contrast required for detection and that increasing the adaptation luminance decreased the required contrast. Blackwell's data are important in a basic research sense, but are not readily applicable to the problems of legibility or readability of electronically displayed information. In fact, many studies performed using CRTs, and interested in absolute or difference thresholds, may not be useful in prediction of legibility or reading performance (Gould, 1968).

Stocker (1964) attempted to build on Blackwell's data and to extend these threshold findings to electronic display applications. Stocker increased contrast to values well above those obtainable on CRTs and found that reading speed continued to increase. Howell and Kraft (1959), however, simulated CRT characters and found there to be little increase in character legibility when luminance modulation was increased from 0.86 to 0.95. Thus, Howell and Kraft recommended modulation values of 0.94 as desirable and 0.88 as acceptable if characters are greater than 16 arcminutes and are well focused. Snyder and Maddox (1978), using a high resolution storage CRT to simulate matrix addressed displays, found that legibility is reduced when modulation falls below 0.90 for non-contextual situations and below 0.75 for contextual.

Several studies have shown that modulation levels fall well below the acceptable range of 0.75 to 0.95 when only moderately high levels of diffuse ambient illumination are present and the CRT does not have a contrast enhancement filter attached (Hunter, Pigion, Bowers, and Snyder, 1987; Hunter, Reger, Farley, and Snyder, 1986; Morse, 1985). In fact, Hunter et al. (1987) found that modulation could fall to as low as 0.34 with only 530 lx of diffuse ambient illuminance and down to 0.12 when a 530 lx specular source was presented. Clearly, legibility at these levels should be severely degraded if the research cited above (Howell and Kraft, 1959; Snyder and Maddox, 1978) is accurate.

Front-surface reflection. While the loss of contrast at the phosphor surface is certainly a critical problem, the reflectance of glare sources by the face of the CRT may be equally important. Here, the 4% reflection at the air-to-glass interface makes most moderately intense objects in the user's environment easily detectable. Highly intense objects (i.e., light fixtures, unshaded windows, white shirts) can mask the displayed information completely due to the loss of contrast created by the "veiling" reflected image formed at the surface of the display. To compound this problem, conventional CRTs are designed so that their phosphor plane and faceplate are convex. This convexity serves to accept and reflect into the user's eyes a much greater glare source area than if the display were flat or slightly concave (flat panel display technologies and flat-tension shadow mask CRTs have this advantage).

Additional problems can arise with the presence of front surface reflections. These reflected images are formed at optical distances other than that from the user's eyes to the displayed information or phosphor plane. If the reflected image is clearly defined, highly patterned, or very luminous it has been theorized that the visual system will continually fluctuate in focus between the displayed information and the reflected image (Bauer, Bonacker, and Cavonius, 1981; Brown et al., 1982; Campbell and Durden, 1983; Snyder, 1984). It should be noted that performance decrements, visual fatigue or irritation, and short-term visual impairment due to this phenomenon have yet to be experimentally supported.

Some have suggested that a phototropic fixation response may occur if the reflection is intense, off to one side, and the user looks directly at the image, this leading to transient adaptation problems (DeBoer, 1977). Others have noted the possibility that the reflected images are capable of inducing binocular rivalry and binocular fusion problems (Reitmaier, 1979). Snyder (1984) summarizes the issue well when he writes "while such relationships have yet to be supported experimentally, there is no question that focused glare sources are at best annoying and distracting, and at worst may cause visual fatigue." (p. 299)

Contrast Enhancement Methods

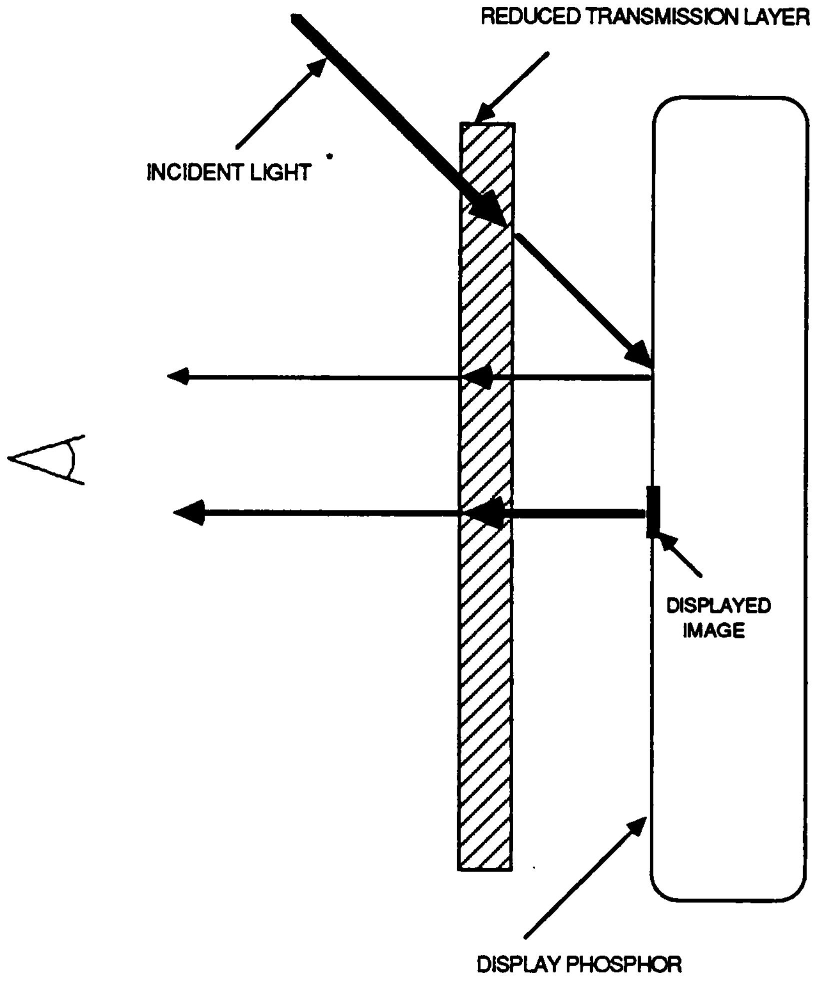
Undoubtedly the most effective and intuitively simple method to maintain high levels of displayed contrast is to reduce the ambient illumination present in the viewing environment. Snyder (1984) reports that ambient levels of about 200 lx are generally acceptable for the use of unfiltered CRTs, but that these levels are considered inadequate for viewing hard copy. Individuals with low vision or ocular opacities would be especially affected. Stammerjohn, Smith, and Cohen (1981) state that while the American National Standards Institute (ANSI, 1973) recommends levels of about 300-500 lx for workplaces with VDTs, the great majority of office environments with VDTs have values that range from 500-700 lx. Further, ANSI recommends ambient levels of about 750-1600 lx for environments where much of the work is being performed with hard copy. Recently, however, ANSI (1988) has revised these levels to 200-500 lx for average workplaces with CRTs, and to greater than 500 lx if poor quality hardcopy or high luminance displays are used. Still, if illumination is reduced to a level where unfiltered CRTs can be used, then it is clear that some source of local task lighting will be ordinarilly be required. If this solution is not feasible, a contrast enhancement treatment can be applied to the display terminal.

Neutral density filters. A neutral density filter consists of a piece of glass, plastic, or gelatin material that has the property of absorbing equal amounts of light at each wavelength across the visible spectrum. Thus, the addition of the perfect neutral density filter will not alter the chromatic content of the displayed image. In reality, most inexpensive neutral density filters do not have a uniform transmittance spectrum, often tending to lose some of their absorption properties at either extreme of the visual spectrum. This minor nonuniformity is not usually critical except where color displays are being used and true chromatic rendition is important. The neutral density filter works as a contrast enhancer when placed in front of the phosphor plane of a CRT by causing any ambient light to pass through the absorption layer twice, while the displayed or emitted light will pass through the filter only once before reaching the operator's eye (Figure 1). Thus, by definition, any filter with reduced transmission will increase the contrast of the displayed information when ambient illumination is present. It also follows that filters with the least transmittance will increase contrast the greatest.

A problem with reduced transmission filters is that as their density is increased, the luminance of the displayed information is also reduced. If the CRT must serve double duty in both high and low ambient conditions, this reduction in displayed luminance may not be desirable. Another disadvantage that accompanies the use of neutral density filters is that unless they possess some antireflection treatment, they may actually increase the user's perception of the reflected image. For example, an untreated glass neutral density filter will still have a front surface with about 4% reflectance (the same as the CRT's faceplate), but because the contrast enhancement properties of the filter will reduce the background luminance of the CRT, any reflected image will be of increased contrast and may be more detectable.

Micromesh filters. Micromesh filters are composed of arrays of fine, black tubes that serve to not only enhance contrast because of their reduced transmissions, but to also scatter and diffuse reflections at the CRT's faceplate (Cakir et al., 1980). These filters can be designed so that their rows and columns of tubes can be oriented vertically and horizontally or diagonally. Reduced transmission is achieved through increasing the density or spatial frequency of the mesh (i.e., the higher the spatial frequency, the lower the transmission). While micromesh filters are able to enhance displayed contrast and reduce specular reflections simultaneously, they possess a number of undesirable characteristics as well.

First, because the mesh filters are composed of elements of some fixed spatial frequency, it is possible that interference or Moire patterns will be created when the filter is placed over the display.



*THICKNESS OF LIGHT RAY DENOTES INTENSITY

Figure 1. Contrast enhancement through the use of reduced transmission filters.

Particularly susceptible to this phenomenon are color CRTs that have a delta gun shadow mask with a fixed spatial frequency of its own. These Moire patterns have not been shown experimentally to degrade visual performance, but they can be annoying and distracting; certainly, they degrade perceived image quality.

Another disadvantage encountered with the use of micromesh filters is that viewing angles are drastically reduced, limiting the application when multiple operators or off-axis viewing are necessary. With increasing angles from normal, the mesh progressively occludes more and more of the displayed information due to the fact that the line of sight is tangential across the mesh rather than orthogonal to it. A final drawback with use of the mesh is the same as reported for the neutral density filter; increasing levels of contrast enhancement are achieved through increasing the spatial frequency of the mesh and thereby reducing the luminance of the displayed information.

Antireflection Treatments

Chemical or mechanical etching. One of the most frequently used methods for reducing the spatial frequency content and intensity of first surface reflections is the etching of the CRT's surface through caustic chemical application or mechanical abrasion. The surface of the faceplate is roughened so that incident light is diffused or scattered in a wide pattern. Thus, the sharpness or spatial frequency of the reflected image is defocused and the intensity of the image is reduced due to the scattering. Cakir et al. (1980) report that etched glass is normally reduced in reflectivity from four percent to about two percent or less. The degree of etch can be varied from fine treatments that still appear quite specular to harsh treatments where nearly all specular glare is eliminated.

Etches that are the most effective in reducing environmental reflection carry with them numerous trade-offs. First, while the etch is designed to scatter incident light, the roughness of the surface also diffuses the light being emitted from the display. Not only does this effect reduce the luminance of the displayed information, but it also tends to make it appear defocused. This defocusing will expand the width of the already-Gaussian spot found on the CRT and will make each spot physically larger (Hunter et al., 1986; Hunter et al., 1987). Clearly, if one of the goals of the display system is to maintain high image resolution, then the use of all but the most mild etches must be viewed with caution. Losses of nearly 30% in character luminance combined with the accompanying degradations in resolution have been found (Hunter et al., 1986). Additionally, increasing the thickness or distance between the phosphor plane and the CRT's front surface serve to only accentuate these problems (Sach, 1970). Uniformity across the CRT and replication of the desired degree of etch from one application to the next are problems with the technology that need to be resolved.

While the etched surface can be extremely effective in eliminating the specular content of the reflected image, it does so by merely redistributing the incident light. This redistribution can become problematic when high levels of incident light are scattered or spread out over the surface of the CRT and create a "veil" through which the displayed information must be viewed. Thus, while the etch will reduce the clarity and luminance of the specular reflection, it creates its own diffuse reflections that have undesirable effects on the operator's perception of the image.

Micromesh filters. Mesh filters reduce specular reflection through the same means that they enhance contrast. The fine, black tubes that make up the mesh serve to absorb and scatter a good deal of the incident specular reflection. Much of the incident light that does make it past the mesh and is reflected off the CRT's front surface will be absorbed or scattered by the tubes as it passes through the mesh the second time. Especially effective in reducing specular reflection are those filters composed of the higher spatial frequency meshes. However, the inherent loss of character luminance, the possible interference or Moire patterns, the dust and lint that collect on the filter, and the possiblity that the physical structure of the mesh may affect perceived image quality all must be weighed against the positive characteristics of the device.

Quarterwave coatings. Certainly, one of the most effective techniques available with which to reduce specular reflection while showing little if any loss in displayed resolution is that of applying a thin-film quarterwave coating. As its name implies, this coating consists of a film comprised of a single or multiple layers of transparent materials that have indices of refraction making the step from air to glass more gradual. As the incident light travels through the thin-film filter and is reflected off the glass substrate, the quarterwave thickness of the filter layer sets up an interference pattern at the air-to-filter interface. Effectively, the reflected light wave will interfere with the incoming incident light wave, the two cancelling each other and becoming imperceptable to the observer.

These thin films can be either sputter or vacuum deposited on glass substrates that are later bonded to the CRT with optical cement. This bonding is a permanent process that ordinarily will require replacement of the display should the filter be damaged to an extent that the image is badly degraded. Further, the deposition process is time consuming and requires complex and costly equipment. Thus, this production cost is passed down to the consumer, making quarterwave coatings the most expensive of the antireflection alternatives. A final nontrivial drawback is that the coatings must be cleaned regularly as they are quite susceptible to fingerprints and smudges that cause the thin film to lose its antireflection properties.

REVIEW OF RELEVANT LITERATURE

A review of the literature concerning the effects of glare and glare reduction techniques for electronic displays yielded a number of interesting observations. In general, most investigators agree that loss of contrast and the formation of specular reflections due to high ambient illumination should be avoided if at all possible. The same investigators were, however, unable to determine (as a group) the most effective methods by which to attain contrast enhancement and to reduce reflection. Most of the studies reported results that, while possibly meaningful for that specific situation, were not of sufficient scope to allow for generalizaion to other environments. In addition, a notable demarcation was evident in the goals and designs of the investigations. There was a clear split between those studies that approached the glare problem from the hardware, photometric measurement, and theoretic/analytic orientation and those that sought to address the issues of both subjective performance and objective human performance. Unfortunately, only a few studies were found to combine effectively the two approaches, thus characterizing the stimuli in a quantitative fashion so that their effects on human performance could later be determined. Equally as unsettling was the finding that proper control of important confounding variables was not evident, in a strict sense, yielding the majority of the results uninterpretable.

The following literature review describes the methodology employed by each of the studies and also discusses relevant findings. As noted above, the studies can be grouped primarily according to whether they investigated the hardware and photometric issues or the human performance aspects of glare and glare reduction treatments. Therefore, this same method of categorization is followed below.

Measurement and Analytic Studies

Beaton and Snyder (1984). Beaton and Snyder (1984) reported a study in which they measured 10 anti-glare filters' abilities to disrupt specular reflections, to pass the displayed image without degradation, and to transmit the displayed image without reducing the luminance. Both polished and etched CRTs were used as host displays for the anti-glare filters and measurements were made for a number of off-axis glare source angles.

The investigators collected modulation transfer functions (MTF) for each of the CRTs when the anti-glare filter was placed in front of it. The MTF concept is discussed in some detail later in this paper, but basically it is an image quality metric that indicates the amount of luminous power that is available at each of the spatial frequencies of interest. In general, display systems that are capable of transmitting power at high spatial frequencies are perceived by the observer as having sharper edges. Those that cannot transmit high frequencies appear blurred. Thus, glare filters that tend to degrade the displayed information through scattering or diffusion will reduce the high frequencies transmitted, this being reflected in the MTF. When averaged across angles of incidence, Beaton and Snyder found the circular polarizer and quarterwave coated neutral density filters to produce the least degradation in image quality, while the micromesh filters showed the greatest reduction. Additionally, due to its softening effect, the etched CRT was found to produce MTF values that were generally lower than those for the polished display.

The reflectance transfer function (RTF) is similar in concept to the MTF in that they are both spatial frequency based measures. The RTF, however, describes the surface's ability to diffuse or eliminate specular reflection rather than the image quality of the displayed information. Also, while larger values for MTF are considered desirable, surfaces that effectively reduce reflection through elimination of high spatial frequency information will result in small RTF values. As might be expected, Beaton and Snyder found that the etched CRT reduced reflection to a greater extent than did the polished, regardless of the anti-glare filter employed.

Finally, Beaton and Snyder formed a signal-to-noise ratio (SNR) by dividing the measure of the area under the displayed image quality (MTF) by the measure of the area under the reflected "noise" (RTF). Due to the etched CRT's ability to disrupt specular reflections, its SNR proved to be greater than the SNR for the polished CRT. The quarterwave coated neutral density filter was found to produce the largest SNR with the etched CRT, while the micromesh, green, and high neutral density filters performed well with the polished CRT.

Rancourt, Grenawalt, Hunter, and Snyder (1986). Rancourt, Grenawalt, Hunter, and Snyder (1986) investigated the effects that chemical etching, chemical etching plus a quarterwave filter, and a polished surface with a quarterwave filter would have in comparison to a polished-only CRT surface in reducing specular images and passing image quality. Using the MTF and RTF metrics previously employed in the anti-glare filter comparisons performed by Beaton and Snyder (1984), Rancourt et al. found substantial differences among the treatments. MTF measurements were recorded in dark room conditions, under overhead fluorescent lighting, and with a mildly collimated incandescent light that was varied from 20 to 50 degrees from normal. They found that the etched surface, regardless of whether it had a quarterwave filter, reduced the MTF by about 30% over the polished surface. Further, the addition of the quarterwave filter to either the etched or polished surfaces resulted in no loss in image quality under any lighting conditions. No effect of glare angle was found. The RTF measurements of antireflection capabilities indicated that the etched, etched with quarterwave, and polished with quarterwave were all very effective in reducing reflection when compared with the polished-only condition. Signal-to-noise ratios (SNR) formed by dividing the area under the MTF by the area under the RTF pointed out that the etched conditions produced the highest overall image quality and that including an antireflection coating increases the SNR over both bare etched and polished surfaces.

An aspect of the Rancourt et al. study that is perhaps more important than its findings is the attention to details that was evident in their design. Most notable was the use of a single, customized monochrome CRT on which all four of the treatment conditions were applied in close proximity to one another. The use of a single CRT reduces greatly the amount of variability in measurements that can occur between seemingly identical tubes. Great pains can be taken in "setting up" two CRTs to have identical outputs such as luminance, spot size, and modulation, but it is highly unlikely that they will ever be perfectly matched. Additionally, the custom CRT was designed so that all measurements could be performed within an area of only a few square centimeters; thus, spatial phosphor nonuniformity, luminance nonuniformity, and spot focusing problems could be minimized.

Stammerjohn, Smith, and Cohen (1981). Stammerjohn et al. (1981) evaluated VDT workstation designs at five different establishments in the San Francisco area and compared these designs with recommended design specifications. Workstation illumination and glare were among those measurements made at each of the sites. In addition, operators of the VDT workstations were surveyed to determine those factors that caused dissatisfaction, annoyance, or were the believed causes of health complaints. Stammerjohn et al. reported illumination levels ranging from 300 to over 1000 lx with 75% (39 of 52) of the workstations having illumination levels of 501-700 lx. Snyder (1984) recommended levels of about 200 lx when supplementary lighting is available, while Cakir et al. (1979) have proposed that ambient levels of 300-500 lx are appropriate for VDT viewing without supplementary lighting. More recently, the Human Factors Society (ANSI, 1988) has suggested that 200-500 lx is generally adequate. Clearly, the majority of workstations included in Stammerjohn et al.'s evaluation were above levels recommended as appropriate.

Stevenson, Fendly, and Wallner (1986). Stevenson, Fendly, and Wallner (1986) analytically investigated the use of the modulation transfer function (MTF) as a means by which to characterize CRT antireflection surfaces. Separate measurement systems for determining the transmitted MTF and the reflected MTF were described and sample measurements on both a chemically etched and mechanically abraded surface were reported. In addition, a simple diffuse reflectance apparatus was presented and was used to perform measurements on the above antireflection surfaces. The authors found mechanical abrasion and chemical etching to produce similar transmitted MTFs (i.e., similar in their degradation of the displayed image), but that mechanical abrasion was not as effective in reducing specular reflection. Further, due to its greater roughness or harshness, the mechanically abraded surface yielded a diffuse reflectance of about 2%, while the chemically etched surface produced values of about 0.4%.

Human Performance Studies

Baggen, Snyder, and Miller (1988). Baggen, Snyder, and Miller (1988) examined the effects of low illumination (10 lx), diffuse glare (1291 lx), and uniform specular glare (1291 lx) on the measured image quality, readability, and legibility of an electroluminiscent display (EL) that had been fitted with six different touch entry devices (TED). A variety of image quality signal and noise measurements were made to determine the optical quality of the TEDs and were later used in human performance modeling. Further, these metrics were used to form signal-to-noise (SNR) ratios that were also employed in both the hardware characterization and performance modeling. A number of performance measures were included to evaluate the touch input characteristics of the TEDs, but these are of little interest in the present context and will not be discussed here. The reader is referred to Baggen (1987) for a more complete treatment.

Briefly, Baggen et al. found large performance differences in reading speed to exist only under the specular glare condition. Here, a number of TEDs were found to degrade readability and this effect was accentuated when the display was viewed from an off-axis position. Modeling of these performance differences and those found for the legibility task was successful as R^2 values as high as 0.92 and 0.87 were reported, respectively. These findings are important in that they indicate the optical qualities of a display can be used to account for the variance found in human performance when using the device. However, it should be noted that many of the terms included in the multiple regression models carried with them signs that were contrary to the intended theoretical function of the term. For example, a number of SNRs were found to have a postive relationship with reading and search time, indicating that as the ratio of signal-to-noise increased (better image quality) so did the time required to read the displayed information or locate a target, respectively.

Bauer, Bonacker, and Cavonius (1981). Bauer et al. (1981) investigated the visibility of specular reflections for two levels of displayed character density (low and high) and two levels of display polarity (positive and negative). The authors′ pointed out that the detectability of reflections is a function of screen background luminance, screen reflectance, the relationship between the spatial frequency of the displayed information and that of the reflectance, as well as the luminance of the reflectance. Bauer et al. asked subjects to adjust the luminance of a checkerboard reflected image until it was perceived to have reached various degrees of annoyance or distraction for both positive and negative polarity screens. To provide maximum interference, the height of a single block of the checkerboard was matched to the height of the displayed characters. These characters were displayed in both high and low density formats, the former being considered by the authors to be more realistic.

The results of this investigation indicated that negative contrast video displays allowed the luminance of the checkerboard to be increased to a greater extent before the reflected image was perceived as being either "just visible" or "slightly distracting." This finding is not surprising as the increased background luminance associated with the negative contrast screen required that the luminance of the checkerboard also be increased to reach contrast values equivalent to the conventional, positive contrast screen. Further, while the average increase in luminance of the reflected checkerboard was only about 18% between the two polarities for the high character density conditions, nearly a 500% increase was found when low character densities were displayed. These differences led Bauer et al. to recommend that acceptable contrast levels can range from 2.2% to 4.4% modulation when negative video is employed to 6.5% to 17.0% when positive video is used. This apparent difference in sensitivity can be explained by Blackwell's (1946) data that showed the visual system to become more sensitive to changes in contrast as background luminance is increased.

Beaton, Murch, and Knox (1985). Beaton, Murch, and Knox (1985) performed a series of three studies designed to investigate the effect of anti-glare filter selection on perceived image quality under dark, diffuse, and specular lighting conditions. Additionally, each of the studies used a different CRT to determine if the effects of these filters can be generalized across various CRT display technologies.

The first study was performed using a high resolution monochrome CRT as a host display to a neutral density spray filter, a rectangular oriented micromesh filter, a mechanically abraded glass filter, a quarterwave coated glass filter, and a quarterwave coated filter with a low neutral density layer. An additional condition consisting of a non-filtered, polished front surface was included as a baseline. Perceived image quality was determined through a series of paired comparisons that were presented under both dark and diffuse (900 lx) ambient conditions. Beaton et al. found that when viewed in the dark, the spray and etched filters were judged to significantly degrade the displayed image, while the quarterwave with the low neutral density layer was found to yield improved image quality over the baseline condition. The other four filters were not significantly different from the no-filter condition. When the diffuse glare source was present, the spray again produced a degraded image, while only the quarterwave with a neutral density layer enhanced the display quality. The authors suggested that the softening or blur caused by the etch and the luminance nonuniformity associated with the spray filter were the likely sources of image degradation.

A medium resolution color CRT was used in the second study to evaluate the effects of the same six filters used in the first experiment as well as a louvered filter, a green density filter, a diagonally oriented mesh filter, a glass circular polarizer, and a quarterwave coated filter with a high neutral density layer. All other physical characteristics of the study were identical to those used in the first study. Subjects were asked to rate each filter relative to the no-filter baseline condition on an 11-point scale ranging from extreme degradation to extreme improvement. When viewed in the dark, subjects found the quarterwave with a high density layer to mildly degrade the image, while the rectangular mesh, the etch, and the louvered filter moderately degraded the displayed information. Again, the spray proved to have the greatest negative effect. All other filters were not judged to be significantly different from the baseline. When viewed in the presence of the diffuse glare source, the spray was again found to degrade image quality, while the quarterwave coated filter and the diagonal mesh showed some improvement. The quarterwave filters with the low and high neutral density layers and the circular polarizer produced the greatest enhancement in judged image quality when compared to the no-filter condition. The authors concluded that the similarity of the results with those obtained from the first experiment suggested that the differences between filters could not be attributed to CRT type or bandwidth changes.

The third and final study used a high resolution monochrome CRT and only those six filters that had shown image quality enhancement in the previous two studies to investigate the effects of uniform and patterned specular reflection on perceived image quality. Because each filter had previously been presented under dark lighting conditions, only the two glare levels were used. For the uniform glare source, the circular polarizer degraded perceived image quality, while the quarterwave coating with either low or high neutral density layers was found to enhance the image. Explanation for this finding lies in the fact that the quarterwave coating reduced the first surface reflection and the neutral density layer served as a contrast enhancer. The circular polarizer, however, enhanced the contrast but was incapable of reducing the specular reflection; actually accentuating the reflection due to the reduced luminance of the background. Presentation of the patterned specular reflection produced nearly identical results, with the exception of the diagonal micromesh filter being rated as yielding slightly higher image quality than the no-filter condition. This finding is likely due to the fact that the mesh was able to effectively attenuate the high spatial frequency content of the patterned reflection. Beaton et al. point out that the poor showing of the circular polarizer and quarterwave filter with no density layer indicate that neither contrast enhancement or antireflection treatments alone are capable of improving image quality when a specular glare source is present.

A number of findings from this study have important implications for the present research. The first implication is that glare filter evaluations should incorporate a dark or low ambient condition as a lighting baseline. Second, both diffuse and specular glare sources should be employed to examine whether the filter is weak with respect to contrast enhancement or antireflection, respectively. Finally, humans are capable of discriminating among various lighting and filter conditions regardless of the psychophysical procedure used in data collection; these discriminations coincide well with expectations based on present knowledge of the glare filters and the human visual system.

Garcia and Wierwille (1985). Garcia and Wierwille (1985) recorded measures of reading response time and proportion of correct responses under diffuse (no-glare) and specular (glare) lighting conditions. A single CRT was used in this investigation, but no description was included as to whether antireflection or contrast enhancement treatments had been employed. Shielded and bare fluorescent tubes were used as the no-glare and glare light sources, respectively. Garcia and Wierwille also varied the difficulty of the content of the passage of textual material each subject was required to read under the two lighting conditions.

Basically, the subjects were instructed to take as long as they desired in reading the passage as their comprehension would be tested on reading completion. The results indicated that when specular lighting was present, the amount of time required to read easy passages was increased over the no-glare condition, but the time required to read difficult passages actually decreased. This result is certainly puzzling in that previous research would predict the introduction of an intense specular reflection to have either no effect or some negative effect on reading speed. Garcia and Wierwille suggested that one explanation for their findings could be that subjects were so disturbed by the intensity of the glare source that they hurried through the reading task in an effort to avoid further exposure to the condition. However, were this the case, a speed/accuracy trade-off would most likely surface when the proportion of correct responses was examined. No such trade-off occurred that would indicate the subjects rushed through the more difficult passages at the expense of more incorrect responses.

Habinek, Jacobsen, Miller, and Suther (1982). Habinek, Jacobsen, Miller, and Suther (1982) collected correct reading rate (CRR) and subjective preference data for three antireflection treatments under diffuse and specular lighting conditions. Specifically, three identical Sylvania CRTs were fitted with a micromesh filter, a quarterwave coated filter, and a no-filter polished faceplate, the last of which served as a baseline condition. A Clinton CRT with an etched front surface was employed as the fourth condition. As did Garcia and Wierwille (1985), Habinek et al. used diffuse fluorescent lighting for the no-glare condition and a bank of bare fluorescent tubes as the specular glare source. Positive and negative display polarities were combined factorially with the lighting and filter conditions.

Under diffuse lighting, no CRR differences were found among the three antireflection treatments and the baseline condition. Significant differences did exist, however, under diffuse lighting for the preference data, as the polished baseline treatment was judged to have "less blurry" characters. For specular lighting, the highest CRR was found with the micromesh filter, the etched surface, followed by the quarterwave filter, and finally the polished baseline condition. Preference data collected under specular lighting indicated that subjects found the micromesh filtered display the easiest to read, while the polished tube was judged to be significantly more difficult to read than the others. In general, the polished CRT was rated as inferior to the micromesh on a number of other criteria as well. The authors concluded that when a specular glare source is present in the viewing environment, the micromesh, etch, and quarterwave treatments are all preferable to a polished-only treatment. However, the three antireflection treatments did not differ significantly from one another; thus, the choice among these treatments should be based on economic, reliability, and maintenance criteria rather than on human factors considerations. These conclusions appear intuitively simple and logical; however, a number of methodological problems should be pointed out before they are accepted unquestioned.

First, as pointed out earlier, if at all possible, the same CRT should be used as the host display for all anti-glare treatments in order to control for resolution, luminance, and uniformity variations. The Habinek et al. study employed four different CRTs. Second, while the authors were apparently successful in statistically defending this point, the fact that the subjects were allowed to adjust the brightness controls of the CRT points out that additional differences in resolution, luminance, and contrast existed throughout the study. Third, an order effect may have been present in the data as subjects were always exposed to the diffuse lighting conditions prior to the specular glare. Fourth, as pointed out by Beaton et al. (1985), a dark or low ambient condition should be included to determine if the anti-glare treatments degrade image quality when no glare source is present. Finally, while the etched surface and the quarterwave coating are truly antireflective treatments, the micromesh filter is capable of reducing specular reflection and increasing contrast. Thus, the increased performance noted with use of the micromesh filter could be due to its contrast enhancing characteristics rather than its ability to reduce reflection.

Hunter, Pigion, Bowers, and Snyder (1987). Hunter et al. (1987) recently reported a study investigating the effect of chemical etching, chemical etching with a quarterwave coating, and a polished surface with a quarterwave coating on perceived image quality under dark, diffuse, and specular lighting conditions. A single medium-resolution, monochrome CRT was modified so that the three antireflection treatments could be presented simultaneously with a polished baseline condition. In addition, all conditions were centered on the CRT and were in close proximity with one another in order to minimize spatial nonuniformities. Importantly, to avoid perceived image quality differences due in some part to changes in displayed contrast levels, the illuminance of both the diffuse and specular glare sources was held constant at 530 lx. All treatment/lighting conditions were counterbalanced to control for order effects. To examine whether different psychophysical data collection methods yielded different results, magnitude estimation, an 9-point rating scale, and subjective ranking procedures were employed.

In general, Hunter et al. found the polished surface to yield higher perceived image quality than the etched surface under all lighting and filter conditions. Also, the quarterwave filter was found to have no effect on perceived image quality under dark or diffuse lighting, but nearly doubled the perceived quality of the image when a specular light source was present. The authors concluded that while the harshness of the etch reduced reflection, it did so at the expense of blurring the displayed information and therefore was consistently rated below the polished surface. Further, the finding that the quarterwave filter had no effect on perceived image quality under dark or diffuse lighting, but enhanced perceived quality under specular lighting conditions indicates that it is an effective antireflection treatment. The trends and magnitudes of the above results were nearly identical for each of the three dependent measures collected.

Notable is that Hunter et al. photometrically quantified the effects of the various antireflection and lighting conditions prior to the perceived image quality study. Using measurements similar to those reported by Beaton and Snyder (1984) and Rancourt et al. (1986), both the displayed image quality and the reflected luminance were captured in a single, simple MTF-based measurement. The measurement itself and the methodology employed to make the measurement will be discussed more fully later in this dissertation. The results of the MTF measurements directly reflect those of the perceived image quality study. That is, correlations of .90 and .83 with the MTF measures were obtained for magnitude estimation and verbal ratings, respectively.

McVey, Clauer, and Taylor (1984). McVey, Clauer, and Taylor (1984) investigated the effects that a number of anti-glare contrast enhancement filters have on subjective preference and typing performance for both positive and negative polarity displays. This study evaluated a concave plastic filter, a quarterwave coated filter, a micromesh filter, and a matte-surfaced filter. All were of roughly the same transmissivity. For the preference study, two monochrome CRTs were employed so that subjects could make paired comparisons among the various sets of filters presented to them. The typing performance study required the professional typists who served as subjects to key in alphabetic characters as quickly as possible. The number of keystrokes per second was recorded as the dependent measure for this task.

The results indicated that the quarterwave filter was the most preferred, followed by the concave. The difference between the two, however, was not significant. The micromesh and matte filters were the least preferred of the four and, again, no significant difference was found between them. These preference results remained constant regardless of the polarity of the display. Although clear preferences were evident within the group of filters tested, no statistically significant typing performance differences were found to exist among them. The negative display contrast did, however, produce keying rates that were slightly and significantly faster than the positive contrast polarity.

Based on the above findings, the authors concluded that while subjective preferences are developed among anti-glare filters, these preferences are not supported by objective performance differences. Unfortunately, methodological problems with the study again require the results and conclusions to be viewed with some caution. As in the Habinek et al. (1982) study, McVey et al. allowed subjects to adjust the display luminance at the beginning of each experimental session. While this procedure is acceptable when the display is to be used in a non-research fashion, adjustment of the brightness control changes a number of the parameters such as resolution, contrast, and luminance that are known to affect both perceived image quality and human performance. It is not likely that each of the subjects adjusted these parameters to equal extents; therefore, the results obtained by McVey et al. may be confounded and, in a strict sense, uninterpretable. Additionally, because two separate CRTs were used for the paired comparisons and no photometric measurements were reported that would support the CRTs' equivalence, it would be tenuous to assume that their outputs were identical. Thus, it is possible that even the preference results may have been biased.

Morse (1985). Morse (1985) reported an investigation in which subjective evaluations of brightness, sharpness, contrast, color, glare, and preference were obtained for a number of anti-glare filters. Photometric measurements of luminance, MTF, contrast, and glare were made in an attempt to determine the optical characteristics of the filters that may have produced subjective differences. Specifically, a quarterwave filter with a neutral density layer, a micromesh filter, a circular polarizing filter, a quarterwave filter with a blue density layer, an etched filter, and a no-filter polished surface were included in the evaluation. A specular glare source was used, but low ambient or dark conditions were not employed. Two monochrome CRTs served as host displays and subjects were allowed to adjust the brightness of each to a preferred setting for each of the six glare filters. While this procedure has been criticized in the studies described above, Morse took care to present each filter an equal number of times on each of the displays in order to minimize performance differences due only to CRT non-equivalence. For each pair of filters presented, subjects were asked to discriminate as to the brightness, sharpness, contrast, color, glare, and overall preference.

Photometric measurements were made of the luminance, contrast, glare, and area under the MTF for each of the filters when placed over the CRT. It should be noted that Morse did not directly measure the MTF of the display as done previously (Beaton and Snyder, 1984; Hunter, et al., 1987) but calculated the MTF based on a prediction equation that considers viewing distance, spot luminance, spot width, and average ambient illuminance. Thus, while separate MTFs were available for each of the six treatment conditions, the calculated MTF values could not account for reflected glare differences or changes in resolution and luminance that accompanied adjustment of the brightness controls. Although Morse expected to find high correlations between MTF and the preference ratings, the fact that the calculated MTF was unable to capture the changing conditions that occurred in the preference study makes this expectation somewhat unreasonable.

In general, the preference results indicated that the quarterwave filter with a neutral density layer, the quarterwave filter with a blue density layer, the micromesh filter, and the circular polarizing filter were all rated moderately high across all six criteria. The etched filter and no-filter conditions were rated significantly lower across most of the criteria. Note that all these findings were obtained with the specular glare source present and that no ratings were made in the dark to determine the degradation that may occur under more optimal viewing conditions. Calculated estimates found the quarterwave filters to yield the highest MTF values, followed by the circular polarizer, the micromesh, the polished surface, and finally the etched surface. In addition, the quarterwave filters, circular polarizer, and micromesh filters were all much more effective in reducing reflected luminance levels than were the etched and polished conditions. Due to the somewhat inappropriate application of the calculated MTF, only moderately high correlations among the subjective and objective measures were obtained.

Sach (1970). Sach (1970) varied both glass thickness and antireflection filter to determine their effects on subjects' abilities to detect targets on an Air Force standard resolution pattern. Sach also used resolution as a measure of the effectiveness of various selectively-absorbing and neutral density filters when evaluated under high ambient lighting conditions. A unity gain rear projection screen was employed to simulate the phosphor surface of a CRT and to display the resolution pattern. Two subjects were asked to report the target with the maximum number of lines per millimeter that they were able to resolve.

The glass thickness results led Sach to conclude that only optical coating antireflection treatments should be used if resolution is critical, as resolution was shown to seriously degrade for the etched surfaces when glass thicknesses approached those found on typical CRTs. The results of a second study investigating resolution changes as a function of both photometric and color contrast yielded interesting results. Here, Sach varied the saturation of a number of color filters to manipulate the level of color contrast. Not surprisingly, he found subjects' resolution to increase as the color differences between the target and background increased.

A third study investigated the effect that varying the transmission of a number of neutral density filters would have on the subjects' abilities to resolve. This study was conducted under high ambient illumination that effectively "washed out" the Air Force standard pattern when no contrast enhancement filter was present. Sach found that the introduction of any reduced transmission filter served to enhance photometric contrast, to decrease display luminance, and to improve the subjects' resolution scores over those obtained when no filter was present. The results indicate that as the density of the filter was further increased (also increasing the contrast), resolution was found to be degraded. Ordinarily, when contrast is increased, performance improvement would be expected. In this instance, however, Sach explained his findings through the reduced ability of the visual system to resolve high frequency detail at the low luminance levels produced by the reduced transmission filters. In addition, Sach noted that the filters also reduced the luminance of the background, making any diffuse first surface reflection (off the filter) more detectable, thus degrading performance.

Sach's dependent measure could be criticized for its lack of generalization to CRT-based display systems. Additionally, very little mention is made of experimental controls that are normally associated with scientific research, and the use of only two subjects is not ordinarily acceptable. However, this study did introduce a number of concepts and approaches that are important to recognize. First, Sach clearly discriminated between filters designed to enhance contrast and those designed to reduce reflection. Second, the author manipulated the harshness level and type of the etching, recognizing that trade-offs exist between the etch's ability to reduce reflection and its effect on displayed resolution. Finally, Sach varied the transmission qualities of the neutral density filters employed in the third study, noting that while more dense filters will increase the contrast level of the displayed information, they do so at the expense of display luminance.

Conclusions

A number of important points can be identified from the above analyses and discussions of the photometric/measurement, analytic, and human performance research that has been conducted to date. These points serve to guide the design of the present experimental investigation, indicating areas that need further attention, as well as pointing out the pitfalls that previous researchers have not avoided.

1. Photometric and human performance measures should be collected to quantify both the stimulus and its resultant effect on the operator. These should be described as quantitatively as possible so that statistical analyses can be performed to allow the researcher to develop clear conclusions from the results.

2. Photometric measurements should consist not only of displayed luminance, contrast, and reflected luminance, but should also include some measure of resolution or image quality (preferably spatial-frequency based). Degradation of the displayed image that can be traced to introduction of a glare filter and measures of the amount of noise the filter may contribute or eliminate are also meaningful information.

3. Multiple human performance measures should be collected so that various complexity levels of human processing are represented. For example, alphanumeric search tasks, reading speed, and subjective ratings of image quality could be used to investigate the filter's effects on lower, intermediate, and higher levels of cognitive processing, respectively.

4. Any research performed in this area should recognize that clear design and intended functional differences exist between the antireflection and contrast enhancement filters. Direct comparisons between these two filter types are complicated by the fact that each is designed to perform optimally under specific environmental conditions. That is, antireflection treatments work best when specular

glare sources are present, while contrast enhancing treatments are most effective under diffuse ambient lighting.

5. The preceding point indicates the need to measure the performance of the glare filter under a number of different ambient lighting conditions. It is necessary that both specular and diffuse glare sources be used to determine fully the filter's strengths and weaknesses as it is likely that the environments will contain both. In addition, it is critical to include a dark or low ambient lighting condition that serves as a baseline by which the glare treatment can be evaluated under near-optimal viewing conditions.

6. Anti-glare research should always include an unfiltered, polished condition, the purpose of which is to serve as a baseline by which human performance or image quality decrements or enhancements can be judged. Failure to include such a condition leaves the researcher with the unanswered question of whether the glare treatment was more or less effective than no treatment at all.

7. If possible, a range or wide variety of each filter technology should be evaluated. The advantage of such an approach is that it yields increased generalization to other using environments as well as decreases the likelihood that the results obtained are dependent only on a specific filter. For example, the review of the available literature indicated the need for investigations into the effects of various levels and types of etch, levels of filter transmission, and the effects that different spatial frequency micromesh filters may have on performance.

8. Another important consideration is that all glare treatment evaluations, whether photometric or performance based, should be carried out on a single CRT. The intra-CRT variability is difficult enough to contend with without further complicating (and confounding) the issue by adding multiple host displays. The luminance, contrast, uniformity, and resolution characteristics are all singularly capable of affecting the results of the study and, unfortunately, are found to vary between displays, even those with identical specifications or model numbers.

9. No clear and consistent relationship between those photometric measures used to define glare filter image quality and measures of both subjective preference and human performance have been obtained. Thus, it is entirely possible that changes in image quality that are measurable with photometric equipment and seem to produce large numerical differences may not produce changes in human performance or preference.

10. Finally, the literature review reveals a clear need for research using proven experimental designs and methodologies. The failure to apply these basic scientific principles yielded results which were uninterpretable.

This research, then, contained two phases that were designed to address the above requirements. Phase I evaluated anti-glare treatments (etched, quarterwave, mesh, and polished filters) at varying densities, under dark, diffuse, and specular lighting conditions, and at two display resolutions (but using a single CRT). As per the points presented above, the filters were evaluated in terms of their effects on measured image quality, legibility and readability, and perceived image quality. Quantitative expressions were developed to model the relationship among the image quality measurements and the resultant changes in legibility, readability, and perceived image quality. Phase II examined, at a single display resolution, the interactive and singular effects that filter transmission (ranging from 11% to 92%) and diffuse illuminance (ranging from 0 lx to 2800 lx) would have on legibility, readability, and perceived image quality. Again, image quality measurements were used to predict the objective and subjective human performance changes.

IMAGE QUALITY METRICS

The preceding discussions have pointed out the need for precise quantitative descriptions of the effects that anti-glare treatments may have on displayed image quality. Recording the contrast, luminance, and reflected luminance yielded a good deal of basic information, but the inclusion of spatial-frequency based measures of the treatment's effect on displayed resolution and surface reflectance produced a more accurate and comprehensive description of the treatment. Beaton (1984) and Task (1979) experimentally evaluated a large number of the available image quality metrics and the reader is directed to these sources for more complete discussions.

Beaton and Snyder (1984) and Baggen et al. (1988) incorporated a number of image quality and noise measurements into the investigation of anti-glare filters and touch entry device overlays, respectively. The emphasis of the present research was to evaluate the effects of anti-glare treatments and not to compare among the metrics themselves. Thus, only a handful of the metrics were included, those being chosen because they were derived from strong theoretical frameworks, because they have been suggested or found to be good predictors of human performance, and because they are not overly complex in their measurement or calculation. The image quality measurements are (1) the area under the modulation transfer function for dark or diffuse lighting (MTF Area), and (2) the area bounded by the human contrast threshold and the MTF of the display (MTFA). The noise measures consisted of (1) the area under the reflectance transfer function (RTF Area), (2) the area under the Wiener spectrum (WS Area), (3) the area under the Wiener spectrum weighted by the human contrast sensitivity function (VWSA), and (4) values obtained through glossmeter measurement (GLOSS). Signal-to-noise ratios (SNRs) were formed by dividing each of the image quality values by a single noise metric or by some summative combination of noise metrics. In addition, pseudo SNRs were formed by a single MTF measurement that was obtained under glare conditions (GMTF area and GMTFA). Each of the above metrics is described in greater detail below.

Area under the Modulation Transfer Function (MTF Area)

MTF Area as it applies to electronic displays has been well described by a number of researchers (Beaton, 1984; Snyder, 1980; and Task, 1979), but for the present purposes, a more general discussion is appropriate. Basically, the MTF of a display is a describing function of the display system's spatial frequency response and can be determined empirically through microphotometric measurements of the response to various sinusoidal inputs, a single impulse, or some edge gradient. The former allows a direct mapping of the MTF, while the latter two require a Fourier transform of a line spread function (LSF) or differentiation and Fourier transform of the edge function, respectively, to produce the MTF.

Almost without exception, the amplitude of the output response will fall as the frequency of the input signal to the display system is increased. Figure 2 shows a typical MTF of a medium resolution monochrome CRT. It is evident that the CRT was able to display low frequency information with more signal strength than the higher frequency information, although the input signal to the system (in theory) remained at a modulation of unity. The importance of the high frequency fall-off is that it is this high frequency information that allows the display to appear "crisp" or sharply defined. Displays that produce MTFs that do not extend well into the upper spatial frequencies tend to appear blurry or defocused to the observer for reasons to be discussed in the next section on MTFA. In general, CRTs that are able to pass and display high frequency information will yield larger areas under the MTF and are considered superior to those with lower MTFs.

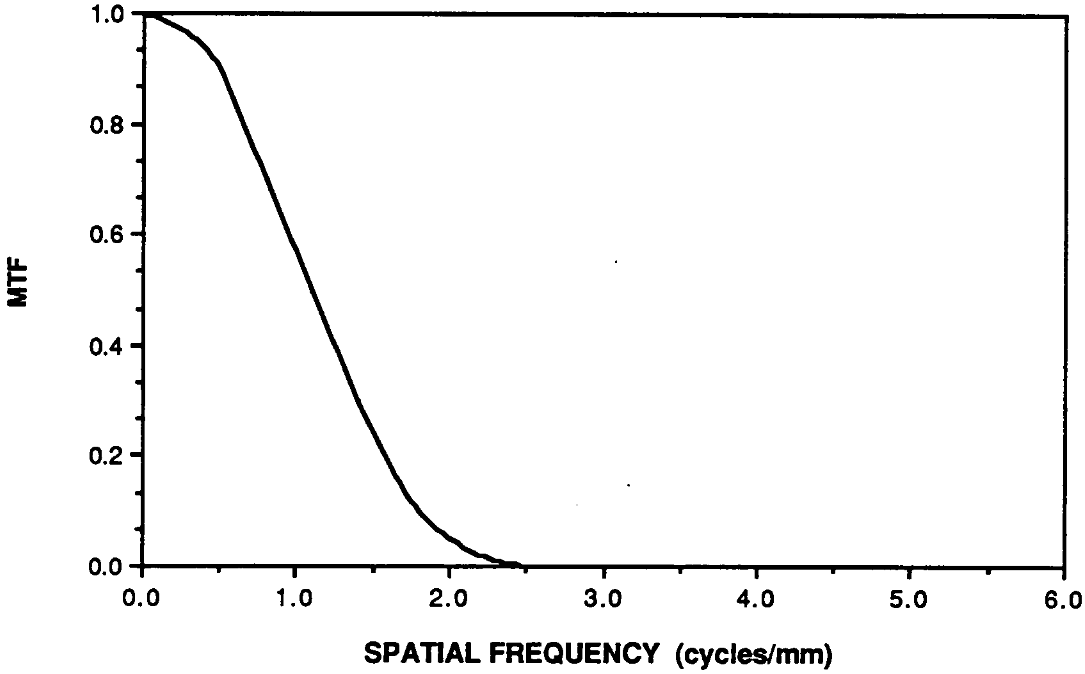


Figure 2. MTF of a non-filtered, medium resolution monochrome CRT measured in the dark.

While reduced MTFs can result from poor quality CRTs or imaging systems, the introduction of antireflection treatments that produce optical degradation of the displayed image will have much the same effect. Figure 3 shows the MTF of the same CRT display as employed in Figure 2, after a rather harsh chemical etch has been applied to its front surface. Note that the system is no longer able to display as high a spatial frequency. Further, the amplitude of the system's medium frequency response is also somewhat reduced. This reduction is caused by the diffraction of light and poor optical quality of the etched filter. Note that the MTFs of the individual display system components can be multiplied or cascaded onto one another to define the MTF of the overall system output. Figure 4 represents the effect that diffuse ambient lighting will have on the MTF of the display system. Basically, the introduction of the ambient lighting reduces the displayed modulation across the entire range of spatial frequencies, reducing the MTF Area and degrading the displayed image quality. Figures 3 and 4 serve to illustrate that measurements of the overall MTF of the display system will capture the reduction in image quality that can be attributed to poor optical quality filters as well as the ability of the filter to enhance contrast under ambient lighting. Mathematically, MTF Area is defined as

$$\text{MTF Area} = \int_0^{N_y} M(f)\, df\,, \qquad (2)$$

where $M(f)$ is the MTF at spatial frequency f in cycles/mm,
N_y is the Nyquist sampling limit, and

$$M(f) = \frac{Mo(f)}{Mi(f)}\,,$$

where $M_o(f)$ and $M_i(f)$ are output and input modulation at frequency f, respectively.

Modulation Transfer Function Area (MTFA)

While the MTF Area metric is capable of precisely defining the display system's ability to pass image quality, it does not consider the response of the human visual system to this displayed information. It has been suggested and shown experimentally that the initial components of the hu-

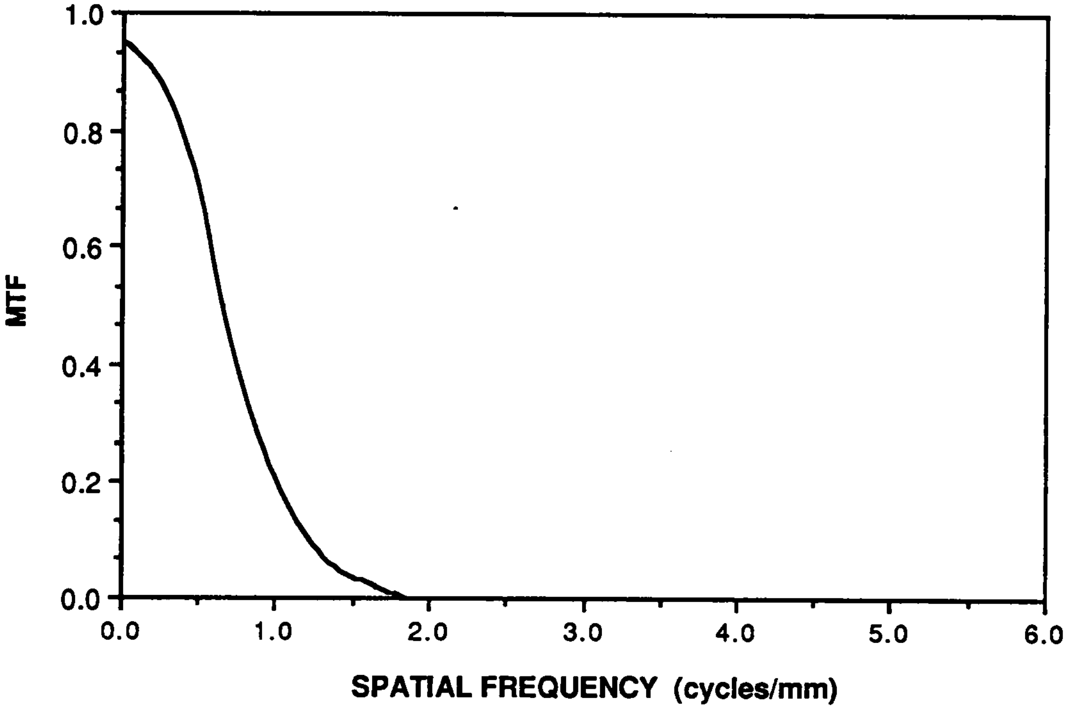


Figure 3. MTF of an etched, medium resolution monochrome CRT measured in the dark.

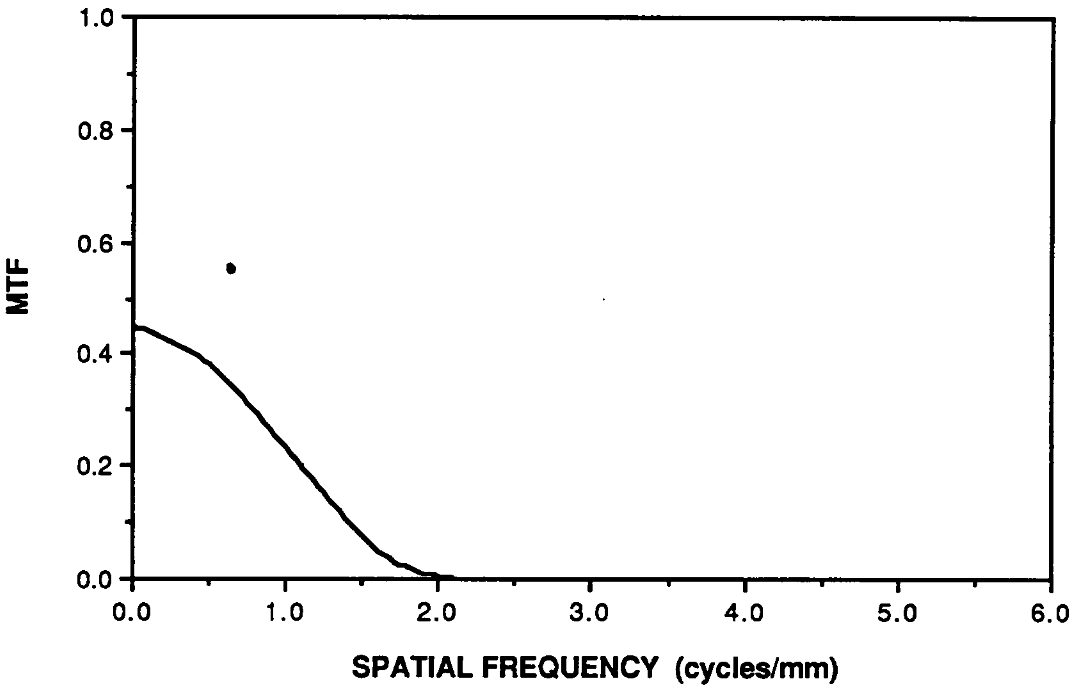


Figure 4. MTF of a non-filtered, medium resolution monochrome CRT measured under diffuse illuminance.

man visual system act as a Fourier analyzer, breaking incoming signals into their component frequencies for further processing (Campbell and Maffei, 1974; Campbell and Robson, 1968; Kelly, 1977). This characteristic allows the spatial frequency response of the human visual system to be determined empirically, this response being termed the contrast threshold function (CTF), the inverse of which is called the contrast sensitivity function (CSF). Experimentally determined CTFs have shown the visual system (at threshold) to be maximally sensitive to spatial frequencies between 2 and 6 cycles/deg and much less sensitive to higher or lower spatial frequencies.

The implications of this basic psychophysical research to the area of image quality are important. Figure 5 shows a CTF of the visual system plotted with the MTF of a CRT display system. Note that signals with modulation levels above the CTF are detectable to the observer while those below will not be perceptable. Thus, the area between the CTF and the MTF is especially interesting in that this area defines the displayed information that is visible to the human observer. This visible area is represented quantitatively by the MTFA metric (i.e., the area bounded by the CTF of the visual system and the MTF of the display system). MTFA is defined mathematically as

$$MTFA = \int_0^{Ny} [M(f) - T(f)]\, df, \qquad (3)$$

where T(w) is the contrast threshold function of the human observer in cycles/deg.

An important aspect of the MTFA is that it has been shown to correlate highly with a number of human performance tasks such as target detection, perceived image quality, and legibility/readability (Beaton, 1984; Snyder, 1973, 1974, 1984). Therefore, the predictive aspects of MTF-based image metrics should make them particularly appealing to human factors and display engineers interested in display systems with final applications to human user environments. Specifically, they were used in the present research to quantify the effects on image quality that various anti-glare treatments produce, as well as to predict perceived image quality and objective human performance changes that might accompany the introduction of these treatments.

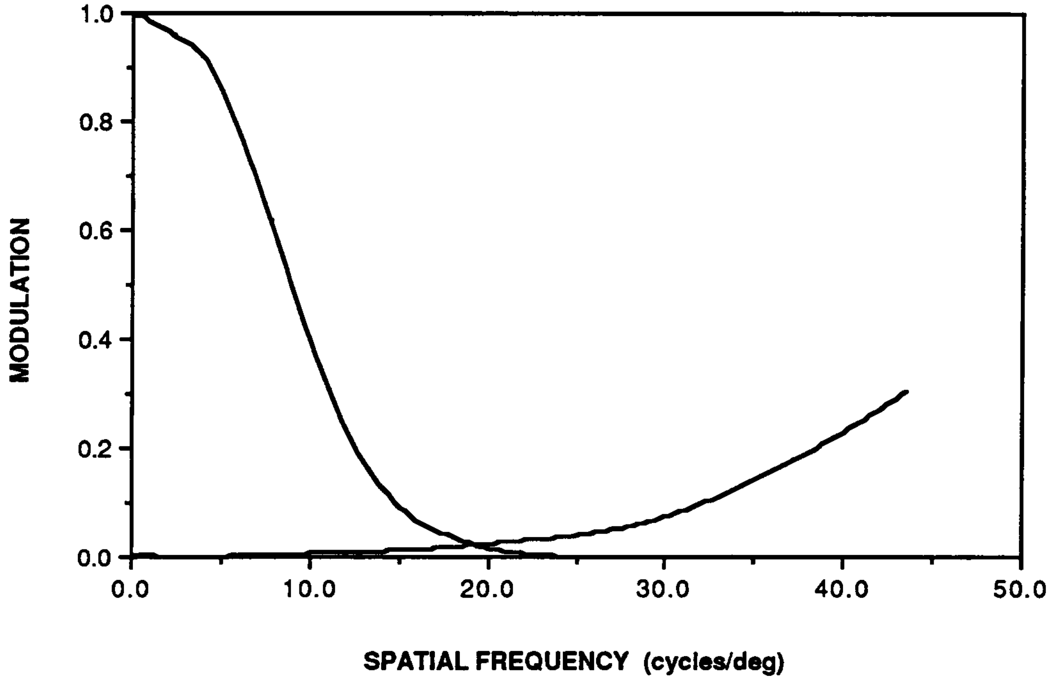


Figure 5. The area bounded by the MTF of the display and the CTF of the human observer.

Area under the Reflectance Transfer Function (RTF Area)

The area under the reflectance transfer function (RTF Area) has been used by a number of researchers to characterize the ability of the display's first surface to diffuse specular reflection (Beaton and Snyder, 1984; Hunter et al., 1986). As its name implies, the RTF Area is very similar in concept to the MTF metric discussed earlier. Simply, the RTF of a display surface is determined by Fourier transforming the photometrically collected line spread function of a reflected image of a very narrow slit, and then dividing this reflected MTF by the MTF of the slit used to produce the reflected image. In essence, the RTF is the output of the system (the reflected image after its interaction with the display's front surface) divided by the input to the system (the source of the reflected image), this ratio being conceptually equivalent to that which formed the MTF metric, and which defines any general transfer function.

While clearly the MTF Area and RTF Area metrics are closely related in terms of their measurement and calculation, their interpretations differ considerably. Remember that larger values of MTF area indicate the system's ability to display high spatial frequency information and that this ability is desirable as it produces sharp, clear images. Higher values of RTF Area also indicate the existence of high spatial frequency information, but here the information is being reflected from the front surface of the display and can be considered a source of noise which should be avoided. Intuitively, the more intense or sharply defined the reflected image, the more it is likely to degrade the displayed information by creating a loss of contrast and a source of distraction. Thus, the lower the RTF Area metric, the greater the ability of the surface to disrupt, diffuse, or reduce the intensity of the reflected information -- a desirable quality. The RTF Area is represented mathematically as

$$\text{RTF Area} = \int_0^{Ny} \frac{I(f)}{O(f)}\, df , \qquad (4)$$

where I is the amplitude of the Fourier transformed reflected image at spatial frequency f in cycles/mm, and

O is the amplitude of the Fourier transformed object slit at spatial frequency f in cycles/mm.

Area under the Weiner Spectrum (WS Area)

The area under the Weiner spectrum (WS Area) is another measure of the noise inherent in some displayed image or, for the present purpose, the noise that may be added to the displayed information through the introduction of some anti-glare treatment. As opposed to RTF Area which describes the front surface of the display, the Weiner spectrum quantifies the transmission fluctuations contributed by the filter. An excellent discussion of the Weiner spectrum as it has been used as a measure of density fluctuations for photographic processes can be found in Dainty and Shaw (1974) and can be generalized quite easily to electronic display applications. Baggen et al. (1988) employed the WS Area as a measure of the visual noise produced through the addition of various touch entry device overlays to displayed information on electroluminescent panels.

Basically, the WS Area is produced by collecting microphotometric measurements across a relatively large area of the display or some component of the display system that rests between the actual display surface (i.e., the phosphor layer of the CRT) and the observer. The output of this spatial luminance measurement will be found to vary around the mean luminance value of the host display or background light source and will often appear as a random process. The Fourier transform of these luminance-fluctuation-by-distance values produces the amount of power or amplitude available at each of the spatial frequencies of interest. Clearly, the larger the area under the Weiner spectrum, the more detectable the noise should appear and the greater the degradation in image quality. Note that the WS Area metric may be especially useful in capturing the noise added by fixed-element glare treatments such as the various spatial frequency micromesh filters. Mathematically, the WS Area is defined as

$$\text{WS Area} = \int_0^{Ny} w(f)\, df\,, \qquad (5)$$

where $w(f) = \Sigma\, f(x)\, e^{-j2\pi xf/N}$,

in which $f(x)$ is the deviation in luminance from a mean level,
f is spatial frequency,
N is the number of samples, and
limits on the summation are from $x=0$ to $N-1$.

Area under the Visually Weighted Weiner Spectrum (VWSA)

The WS Area metric does not take into account the relative sensitivity of the human visual system to different spatial frequencies, only quantifying the luminance fluctuations of the display system itself. A potential improvement on the WS Area measure is to weight the Wiener spectrum by the contrast sensitivity function (CSF) of the visual system. Thus, the area under the visually weighted Weiner spectrum (VWSA) could be a more meaningful predictor of human performance than the WS Area. Heavily weighted would be those spatial frequencies of the WS Area known to lie in the regions of maximum sensitivity of the CSF, while the contributions of WS Area in the frequencies to which the visual system is insensitive would be minimized through low weightings. Mathematically, the VWSA is defined as

$$\text{VWSA} = \int_0^{Ny} [w(f)\, T'(f)]\, df\,, \qquad (7)$$

where $T'(f)$ is the inverse of the contrast threshold function of the human observer, the CSF.

Glossmeter Values (GLOSS)

A final measurement of display noise that is very applicable to evaluations of antireflection treatments, but that cannot be truly considered a metric, is the glossmeter reading (GLOSS). Reported in arbitrary units, gloss values are attained from a simple device called a glossmeter, which consists of a directed light source used to create incident light and a photodetector used to measure the amount reflected from the surface. A glossmeter is calibrated to yield some maximum value when measuring the amount of reflected light obtained from a standard specular surface (i.e., polished black glass). Thus, less specular surfaces will, of course, yield smaller gloss values. The measurements of transmissive surfaces such as CRT faceplates or anti-glare filters are made with a highly absorbing, oil-soaked black felt placed behind the faceplate or filter to absorb any incident light that is not reflected from the front surface. Thus, the greater the ability of the surface to transmit rather than reflect incident light, the lower the value reported by the glossmeter. Clearly, lower values are preferable when considering the antireflection characteristics of the surface.

While not as theoretically attractive as RTF area, if GLOSS values are found to correlate highly with RTF Area and both are found to predict perceived image quality or human performance, then the ease with which the GLOSS value can be determined may make it a preferable alternative to RTF Area. To date, it does not appear that the relationship between GLOSS and RTF Area, or between GLOSS and human perception and performance has been investigated experimentally.

Signal-to-Noise Measures

Baggen (1988), Beaton (1984), Beaton and Snyder (1984), and Hunter et al. (1986) have each obtained signal-to-noise ratios (SNR) by placing the MTF Area or MTFA (signal) in the numerator and dividing by one or more of the various noise metrics in the denominator. The objective of such an approach is to capture both the ability of the display system to pass image quality as well as the level of noise present on the system in some single, overall measure. Thus, as per the above discussion, the MTF Area, MTFA, GMTF Area, and GMTFA, as well as log transforms of these metrics, were divided by RTF Area, WS Area, VWSA, and GLOSS values, as well as by meaningful combinations of these noise measures. An SNR would prove useful if it was more highly correlated with human task performance than either the signal or the noise when considered separately.

Glare MTF Area and Glare MTFA. A concept introduced by Hunter et al. (1987) was that of making MTF measurements that are capable of describing not only the signal of the display (as is done conventionally), but the amount of first surface reflectance as well. Simply, the approach taken was to make standard MTF Area measurements of the display under dark, diffuse, and specular lighting conditions. The advantage of this approach is that the effects of the anti-glare treatment on the ability of the display system to pass image quality will be described by the MTF Area values recorded under dark lighting, while the diffuse and specular conditions serve to test how well the treatment performs under less than optimal lighting. That is, desirable treatments should not degrade image quality under dark conditions and should be found to enhance contrast and/or reduce reflection under diffuse and specular lighting. Hunter et al. found this measure, termed the Glare MTF Area (GMTF Area), to be highly correlated with perceived image quality under these various lighting conditions.

It should be noted that neither GMTF area nor its visually weighted counterpart (GMTFA) is able to consider the spatial frequency content of the first surface reflection when a specular light source is present. Both are able, however, to measure the loss of contrast in the displayed information that accompanies surfaces that are unable to reduce the intensity of the reflected image. If found to be highly correlated with objective and subjective human performance, then GMTF area and GMTFA could be looked upon as pseudo SNRs that can be collected in a single measurement, merely by changing the ambient lighting present during the measurement.

HUMAN PERFORMANCE MEASURES

Detection/Search Task

A number of alternatives exist by which the observer's ability to detect or search out specific targets can be measured. Snyder and Maddox (1978) and Baggen (1988) asked subjects to search the display for randomly scattered alphanumeric characters and found the task to be fairly sensitive to changing image quality. Perhaps the most positive characteristic of the random search task is its similarity to real world tasks such as word processing or data entry. Due to the non-contextual nature in which the random characters are presented to the subject, the task truly becomes a measure of legibility. Typically, the data collected with such a task are quite variable due to between and within subject differences in search strategy, but the relative simplicity of the task and the fact that it has been used in previous image quality research make its inclusion worthwhile. Therefore, the present study incorporated a random alphabetic character search task as a measure of display legibility.

Tinker Reading Task

A task that requires the observer to employ a cognitively more advanced processing stage is that of reading for speed. A measure modified by Tinker (1958) that has previously been used in assessing the legibility of different typographies of printed text has been adapted and employed in numerous electronic display image quality evaluations (Abramson and Snyder, 1984; Baggen, 1988; Snyder and Maddox, 1978). Termed the Tinker Speed of Reading Test (Tinker, 1955), the measure used hundreds of short passages that were constructed to be very simple, removing reading comprehension as a factor in reading speed. The subject is required to read the passage as quickly as possible, pointing out the one word in the paragraph that is clearly out of context. Relatively high correlations have been obtained between reading speed and changes in image quality for various electronic displays. Thus, it seems logical that degradations or enhancements in the quality of the displayed information due to the introduction of anti-glare treatments would be reflected in corresponding reading speed variations.

Perceived Image Quality

The individual's perception of the displayed information could be considered an even higher level of processing complexity. Asking an observer's impression or rating of the quality of some displayed image requires that observer to cognitively consider a great many factors before making his/her response (i.e., previous experience with displays and what display aspects he deems most important). Many studies have investigated this issue, but those most appropriate for the present application have been reported by Beaton et al. (1985), Hunter et al. (1987). Both studies employed a number of different methods by which the perceived image quality of anti-glare treatments was measured. Psychophysical techniques such as paired comparisons, rating scales, magnitude estimation, and rank ordering all yielded fairly similar results, indicating that observers are very capable of differentiating among anti-glare treatments, regardless of the dependent measure employed.

The ease of data collection found with the use of rating scales certainly makes them the most attractive alternative for complex experimental designs such as the present research. Fortunately, a 9-point interval scale that ranges from "worst imaginable" to "best imaginable" has been shown by Beaton et al. (1985) and Hunter et al. (1987) to correspond well with other measures of perceived quality, and by Hunter et al. (1987) to be highly correlated with changes in GMTF Area (measured image quality). The use of such a scale allows data to be collected quickly and accurately with a minimum number of repeated presentations of the displayed image and anti-glare filter.

METHOD - PHASE I

The main objectives of Phase I were to determine the effects of various anti-glare (contrast enhancement, etched, mesh, and quarterwave) treatments on readability, legibility, and perceived image quality when viewed under dark, diffuse, and specular glare conditions. The spot size and character size of the information on the CRT display were varied to determine whether these filter and lighting effects would react differentially at low and high resolution levels. Another goal of Phase I was to quantify the relationship between the measured image quality of the display under the aforementioned filter, lighting, and resolution conditions and the observed changes in human performance. If successful, the models defining this relationship could be used by display designers and human factors engineers to predict future human performance from relatively simple display measurements.

Measurement Apparatus

Anti-glare filters. Sixteen anti-glare filters were evaluated in terms of their image quality during Phase I of this research. The filters and their pertinent characteristics are shown in Table 1. The Mesh-5.7 and Mesh-7.4 filters are commercially available and are produced by Sunflex (CD-90 and CD-44, respectively). The quarterwave (HEA) coatings on the POL-62%-QW and G65-62%-QW filters are also commercially available and were produced and applied by Optical Coatings Laboratories, Inc. All other filters were produced specifically for this research, although the transmission and front surface characteristics are not unusual and can be found in real-world applications. The harshness of chemical etch was varied, producing three levels of etch consisting

Table 1

The Sixteen Anti-glare Filters and Their Relevant Characteristics

Filter	Surface Treatment	Transmission (%)
POL-92%	polished	92
POL-62%	polished	62
POL-31%	polished	31
G65-92%	etched	92
G65-62%	etched	62
G65-31%	etched	31
G45-92%	etched	92
G45-62%	etched	62
G45-31%	etched	31
G25-92%	etched	92
G25-62%	etched	62
G25-31%	etched	31
POL-62%-QW	polished with AR coating	62
G65-62%-QW	etched with AR coating	62
MESH-5.7	5.7 lines/mm mesh	50
MESH-7.4	7.4 lines/mm mesh	37

of 25 gloss (heavy), 45 gloss (moderate), and 65 gloss (light). The filters with polished front surfaces can be considered the high-end of the etch scale and yielded gloss values of about 85. Each of the three etch levels and the polished filters also varied in neutral-density transmission with levels of 92%, 62%, and 31%. Two filters were included with thin-film anti-reflection coatings. One of these filters was polished prior to coating deposition, while the other had been etched to a 65 gloss; both were of 62% transmissivity. The final two filters were of the micromesh variety and contained fiber densities of about 5.7 lines/mm and 7.4 lines/mm, respectively.

The mix of filters was carefully selected to yield the maximum amount of generalizable information while minimizing the total number of filters included for measurement. For instance, the effect of contrast enhancement was examined by varying the transmissivity from 31% to 92% for the various polished and etched filters. The effect of surface etch harshness was studied by including panels ranging from 25 to 65 gloss and the polished filters with 85 gloss. Though no longer on a continuous scale, the coated and mesh filters were also of interest due to the different approaches the technologies take toward reduction of glare. The 62% transmissivity of the two coated filters was set to allow comparison of these filters with a similar control condition (the polished, 62% transmission) in order to isolate the effects of the coating. The mesh filters had no such control condition, but their transmission values fell within the range of transmissions examined for the polished filters.

Imaging system. The imaging system consisted of a 50-cm diagonal, monochrome P4 phosphor CRT, a high quality video signal generator, and an IBM-PC. The CRT was produced by Video Monitors Inc. (#E-M2400-155) and had been run previously with a full screen displayed for over 100 hr to promote phosphor burn-in and to minimize temporal luminance variation during the subsequent measurements and performance study. The monitor was run at a 1024 x 1024 pixel addressability and in a 60-Hz non-interlaced mode. The OPIX Imager, serving as the signal generator, was produced by Quantum Data and was capable of up to 200 MHz pixel rates with rise/fall times of about 1.8 ns. During the measurement phase of this research, the IBM-PC served as a terminal to load images into and control the output of the OPIX.

As mentioned earlier, one of the primary confounds with previous research has been the use of multiple host displays. Only one host CRT was used for the present research, thereby eliminating the inter-display variability and ensuring that the original stimulus behind the filter remained a constant. To allow manipulation of the various filters on a common CRT, a mounting apparatus was designed to press the filter to the face of the display. The apparatus consisted of a hinged aluminum frame, an over-center clamp, and a 10-cm by 20-cm viewing window. A three-sided, 11-cm x 21-cm rubber seal was applied to the face of the CRT and was open at the top. When the filter was placed in the mounting frame and clamped to the CRT, the 0.4-cm deep cavity existing between the filter and the CRT face was filled with glycerin. This eliminated the air-to-glass interface and correspondingly eliminated the reflections that accompany interfaces of materials with different refractive indices. The glycerin was chosen because it is optically transparent and its index of refraction of 1.427 closely approximates that of glass, reducing reflection to near zero.

Measurement system. A diagram of the microphotometric measurement system is shown in Figure 6. Briefly, the system was composed of a telemicroscope; photomultiplier tube (PMT) with photopic correction filter; an x,y,z stage; a granite optical table; an intelligent radiometer; and an IBM-PC. EG&G Gamma Scientific produced the telemicroscope (GS2110), the photomultiplier tube (D46-A), and the radiometer (GS4100). An air-cushioned, granite optical table, produced by Technical Manufacturing Corporation (#24-6413), served to isolate the display and microscope from high frequency vibration. Ealing Instruments produced the x,y,z stage that provided precise control over the positioning of the microscope. With the exception of the Wiener spectrum measurements, a 1.0X objective lens and a 0.010-mm x 3.0-mm scanning aperture slit were used for all spatial luminance measurements.

Diffuse and specular glare sources. Two, 122-cm, four-lamp, fluorescent light fixtures served as glare sources during the MTF-base image quality measurements. The lamps themselves were General Electric WattMiser II, 34-watt with a color temperature of 4200 deg K and maximum single-lamp outputs of about 2750 lumens. One of the fixtures provided a diffuse source of illumination similar in intensity to levels found in typical office environments (about 650 lx). The other fixture produced a specular reflection on the face of the CRT of sufficient luminance to simulate an environment with open windows or poorly located lighting fixtures.

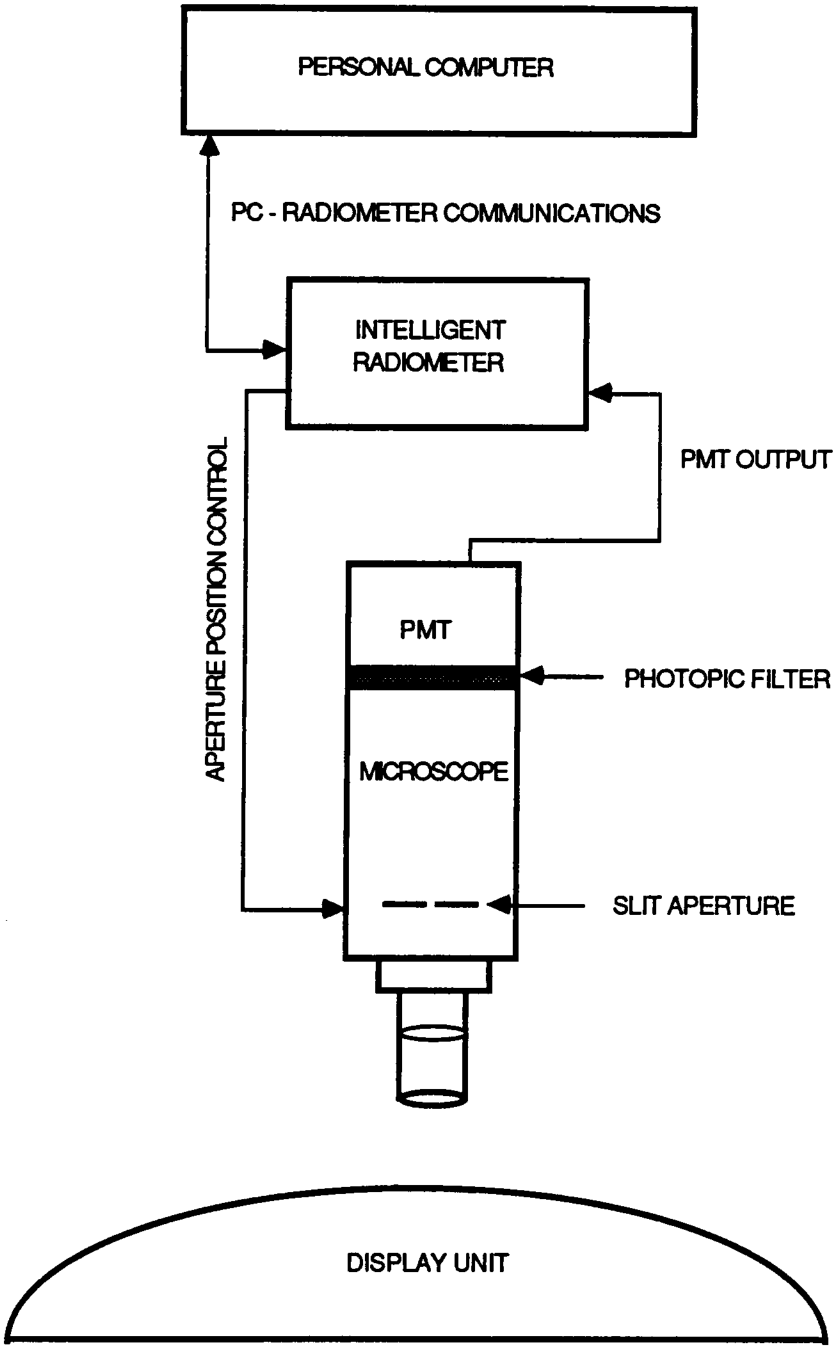


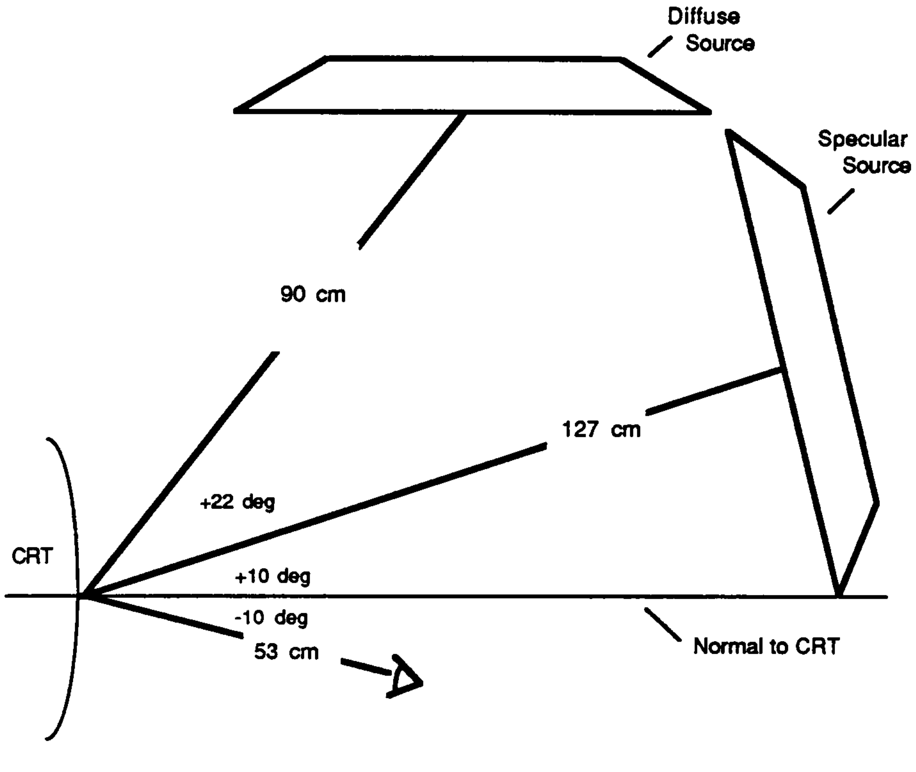
Figure 6. Microphotometric measurement system used in all spatial-luminance determinations.

It is important to note that the illuminance at the face of the CRT was 650 lx for the specular source as well as the diffuse. Therefore, measurement and performance differences between the two conditions could be attributed to the diffuse and specular nature of the glare and not to differences in the total illuminance incident on the display. The positioning of the glare sources in relation to the display is shown in Figure 7. A black curtain was lowered in front of the specular source during diffuse measurements in order to eliminate all specular reflections. To minimize luminance nonuniformities, a 63-mm thick, white Acrylite FF (color code 3.0) acrylic sheet was placed directly in front of the sources. Luminance of the diffuse and specular sources was about 2220 cd/m^2 and 1520 cd/m^2, while the x,y chromaticity coordinates were .39,.41 and .40,.42, respectively.

RTF glare source. A General Electric, tungsten DYT bulb, enclosed in an aluminum housing and powered by a stable DC power supply, served as the glare source for the RTF measurements. The output of the bulb was passed through a 0.100- mm x 8.0-mm slit aperture and the image of this slit was reflected off the face of the glare filter. A diffuser was placed between the DYT and the object slit to eliminate "hot spots" created by the filaments of the 53,000 cd/m^2 source. The bulb, housing, diffuser, and slit were all positioned by lens holders and clamped in place on an optical bench to assure stability throughout the measurements.

Measurement Procedure

MTF measurements. As described earlier, the MTF of a display system can be derived from the line spread function (LSF) of the display and the width, shape, and relative amplitude of this LSF determine the values of the resultant MTF. This study manipulated the width of the LSF while holding the shape and amplitude constant, thereby varying the resolution of the display. In the high resolution case, a single pixel wide vertical line was displayed at the center of the CRT and the DC focus grid voltage of the monitor was adjusted so as to yield the smallest pixel width the system was capable of producing. The low resolution case consisted of a two pixel wide vertical line also displayed at the center of the CRT and the focus grid voltage set so as to produce a Gaussian-shaped LSF of about twice the width of the high resolution line. The monitor was operated in the 1024 x 1024 addressability mode, and the video signal of the OPIX Imager was set at



SIDE ELEVATION VIEW

Figure 7. Relationship among the observer, the CRT, and the diffuse and specular glare sources for Phase I.

the maximum output (255 bits). The peak luminance of the line for both the low and high resolution conditions was set at 42 cd/m^2 when measured in the dark and with no glare filter present. Half-amplitude spot widths of the low and high resolution lines under these measurement conditions were 0.380 and 0.800 mm, respectively.

All LSF measurements were made with the fluorescent glare sources positioned as shown in Figure 7 and the CRT and telemicroscope mounted to the optical table. After about a 2-hr warm-up for the glare sources and CRT, the measurement system was calibrated and three consecutive LSF scans were made for each of the 96 (2 Resolution x 3 Lighting x 16 Filter) experimental conditions. To minimize variability in the spot width, all low resolution measurements were completed before the high resolution ones were begun, thereby requiring only one change in display format and focus voltage. All data were initially stored on the PC and later transferred to the university's IBM-4341 mainframe computer for the MTF calculations and plotting.

RTF measurements. For each of the RTF measurements, the glare filter was mounted to the CRT and clamped into place just as was done on the MTF measurements. However, because the RTF is a characterization of the first-surface reflection qualities of the filter, it was necessary to ensure that the light passing through the first surface of the filter was not reflected back into the microphotometer after striking the phosphor of the CRT. To avoid this potential problem, black felt soaked in glycerin was placed between the back surface of the filter and the front surface of the CRT faceplate, this felt serving to absorb any light passing through the filter. Three consecutive scans of the slit's reflected image were made for each of the 16 filters. For all measurements, the slit was positioned at a horizontal angle of incidence of 20 deg, while the microphotometer was set at the corresponding 20 deg angle of reflection. Measurements of the mesh filter were made with the mesh placed over a polished, 92% transmissivity filter.

Wiener spectrum measurements. Wiener spectrum measurements determined the degree of luminance fluctuation or nonuniformity introduced into the display system by a filter. As per Dainty and Shaw (1973), the required parameters of the measurement can be determined analytically for the specific application. A measurement step size of 0.067 mm, slit length of 10.00 mm, and measurement distance of 2.0 mm were specified. The Wiener measurements were made by placing a luminance standard (Hoffman Engineering LS-65) behind the glare filter and recording the luminance as a function of aperture position over the 2.0 mm scanning distance. The luminance of this standard was set at 50 cd/m^2, but this value is not critical as calculation of the Wiener spectrum works with the delta luminance values around the DC or mean luminance. Measurement of the mesh filters was performed with the mesh mated to a polished, 92% transmissivity glass filter. Ten consecutive scans were made of each filter evaluated.

Gloss measurements. Gloss measurements were made with a Pacific Scientific, 60-deg Glossgard II glossmeter. The device consisted of a collimated tungsten light source set at a 60-deg angle from normal and a silicon photodector at the reflected 60-deg angle. The meter was calibrated to a polished, black glass (high gloss) and a matte white ceramic (medium gloss) standard prior to each measurement session. Similar to the RTF measurements, it was necessary to eliminate reflections from the back surface of the glare filter so that any changes in the gloss value among filters could be attributed solely to the front surface reflectivity. Therefore, black felt soaked in glycerin was applied to the back surface of each filter during measurement to absorb the incident light not reflected by the front surface. Five repeated gloss measures were made for each of the 16 filters by placing the meter on the front, convex side of the filter and recording the gloss value from the meter's digital display. Again, both mesh filters were placed over polished, 92% filters for these measurements.

Human Performance Apparatus

The human performance phase of this research used the same fluorescent glare sources, filters, and imaging and display system as the measurement phase. The geometric relationships among the two glare sources and the CRT were not changed, nor was the 650 lx illuminance present at the face of the display from both sources. A black facade was fixed to the filter mounting apparatus and served to limit the subject's view of the filter and display to the 10-cm wide x 20-cm high region in the center of the CRT. All displayed information was presented in this area. Subjects were seated in a hydraulic, height-adjustable office chair that had been fitted with a headrest. Eye height from the floor (118 cm) was held constant for all subjects, as was the distance from the eye to the center of the display (53 cm) when resting against the headrest. In this position, the viewing angle (eye to display center) was 10 deg above horizontal.

Glare sources were controlled by remote switches at the experimenter workstation, while timing and presentation of the reading, search, and rating trials were controlled by software run on the PC. Subjects' response times and the accuracy of their responses for each trial were recorded on the PC. A 7 x 9 Huddleston font was used for all tasks, and at the 53 cm viewing distance, the characters subtended 16.1 and 32.2 arcminutes vertically for the high and low resolution characters, respectively. Figure 8 shows the difference in structure between a character displayed at high resolution and the same character at low resolution. Note that the small, high resolution pixels are "grouped" to make up a single low resolution pixel. The DC grid voltage of the CRT was varied to yield the minimum displayable spot size (best focus) in the high resolution case (half-amplitude width of about 0.380 mm). For low resolution, the grid voltage was varied to produce a defocused spot with a half-amplitude width of approximately twice that of the high resolution spot (about 0.800 mm). Center-to-center dot spacing was 0.270 mm for high resolution and 0.540 for low resolution.

Human Performance Procedure

Subjects. Eight subjects (four female) participated in the research and were paid for their time. Each was screened for 20/20 near and 20/22 distant vision, normal lateral and vertical phoria, and for color vision abnormalities using a Bausch & Lomb Orthorater. In addition, near and distant contrast sensitivity was measured using a test system produced by Vistech Consultants Inc. Refracted subjects were accepted only if the contact lenses or glasses worn were not photo-sensitive or tinted. The age of the participants ranged from 18 to 21, with a mean age of about 20 yr. In addition, each subject was required to speak English as a native language. All were obtained from the university community and as a group averaged about four hours per week of computer use.

Experimental design. The experimental design for the human performance study is shown in Figure 9. Each of the eight subjects was exposed to all 96 of the possible experimental condi-

a)

b)

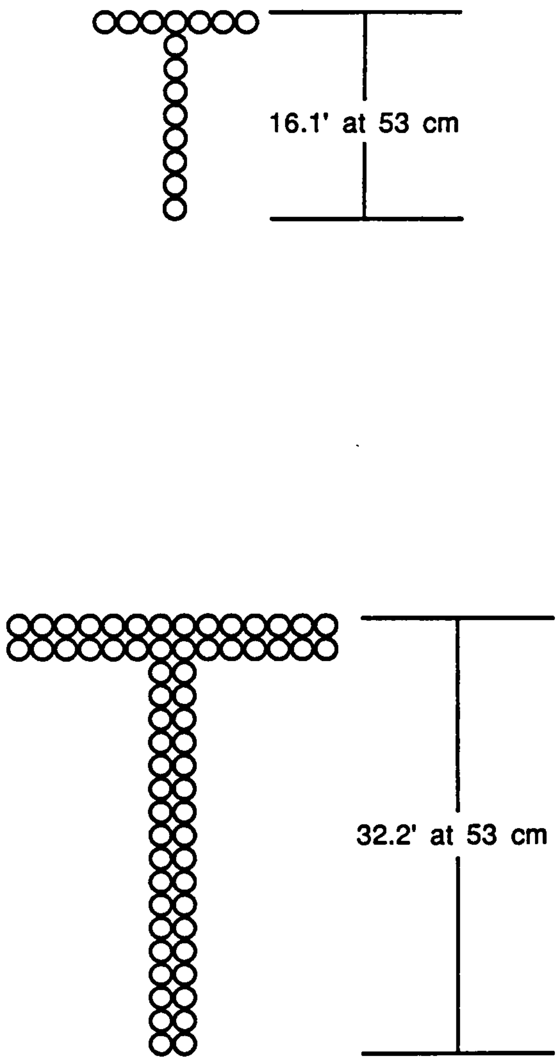


Figure 8. Illustration of a) the high resolution character format and b) the low resolution character format.

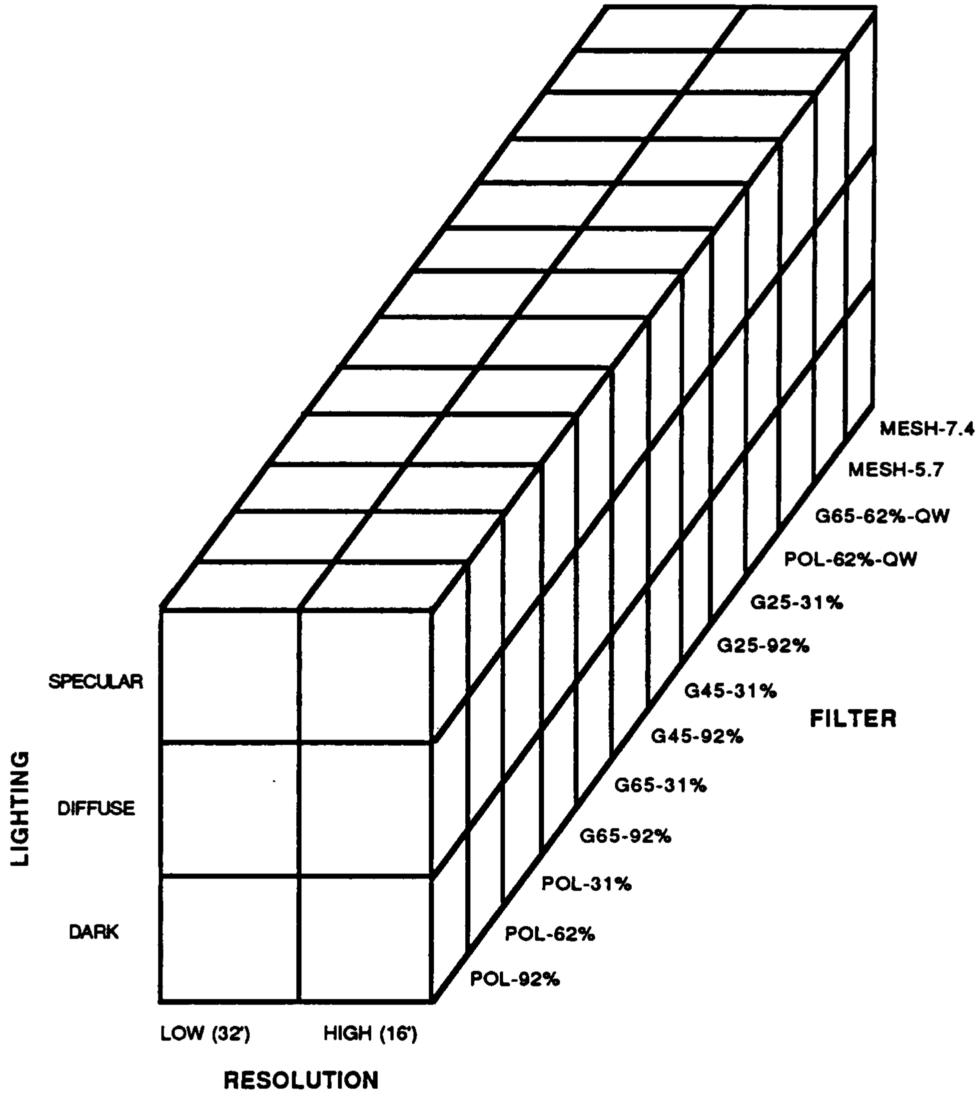


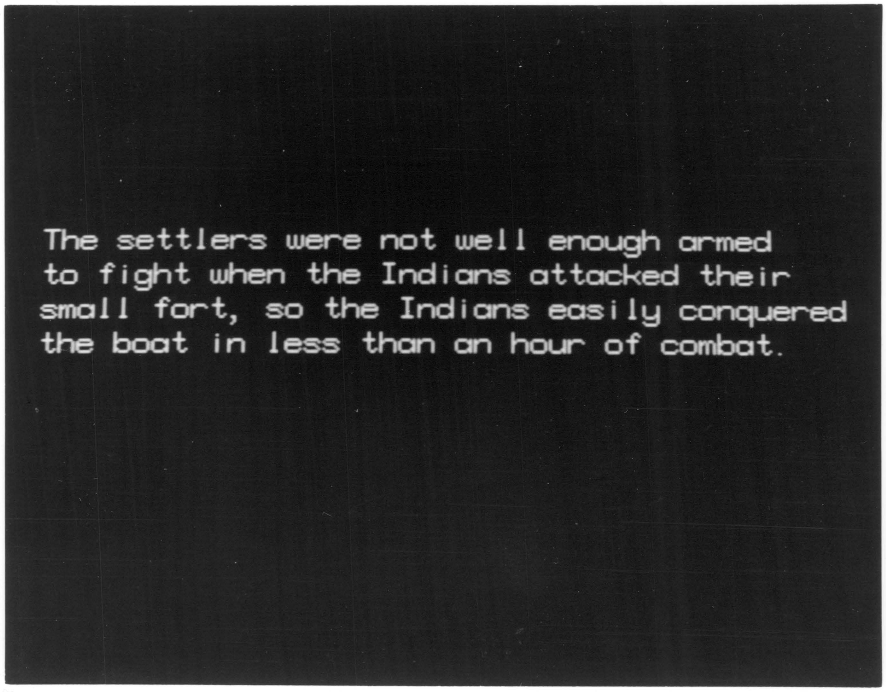
Figure 9. Experimental design for Phase I human performance evaluations.

tions. A 16 x 16 Latin-square was developed in order to fully counterbalance for filter presentation order. Of the 16 available presentation orders, 8 were chosen at random to be used in the study. Resolution and lighting order were also balanced across the 16 filters. In addition, the reading task was performed first for eight of the filters and the search task first for the other eight, thereby controlling the influence that the dependent measures might have on one another. However, the image quality ratings for each filter were collected prior to both of the timed tasks, so as to minimize the effect of the subject's objective performance on his or her perceptions of the quality of the image. Each subject was presented two filters per day for eight consecutive days, with data collection for each filter requiring about 50 min. After completion of data collection for the first filter of the day, a 15-min rest period was allowed while the experimenter mounted a new filter to the CRT.

Reading, search, and rating procedures. As mentioned earlier, the reading task incorporated passages from the Tinker Speed of Reading Test (Tinker, 1955). Each of these simple passages was two sentences long, and contained one word in the second sentence that was clearly out of context. After an auditory "ready" cue, the subject depressed and held down the mouse button (starting the clock and displaying the passage simultaneously) until the out-of-context word was located. At this time, the subject released the button (stopping the clock) and reported the target word to the experimenter. On button release, the display was blanked with a pattern of alternating on and off one-pixel-wide vertical lines. This pattern was set to the space-averaged luminance of the reading passage (about 8 cd/m^2) and was used to ensure that the passage was not still readable due to either phosphor persistence or afterimages present at the retina of the subject. All trials were subject-initiated and contained intertrial intervals of about 10.0 s. Photographs of example high and low resolution reading passages are shown in Figure 10.

The search task required the subject to search for a target character (either upper or lower case) from among 52 (26 upper and 26 lower case) randomly positioned alphabetic characters. As for the reading task, after an auditory "ready" cue, the subject depressed the mouse button to display the passage and released the button when the target character was found. The blanking pattern was displayed upon button release. Photographs of example low and high resolution search patterns are shown in Figure 11. To check the accuracy of the trial, the subject was required to report

a)



b)

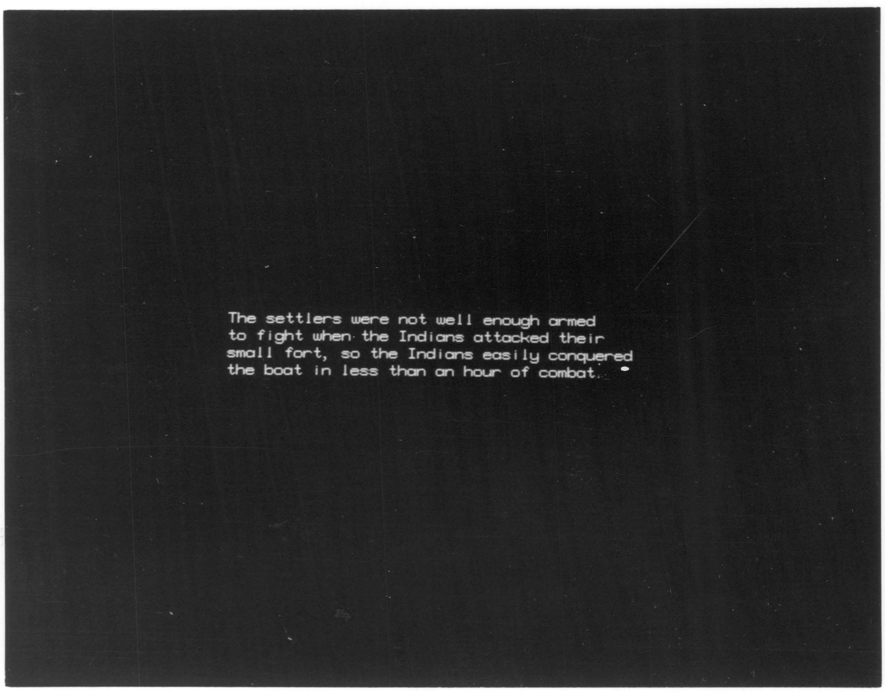


Figure 10. Photographs of Tinker reading passages under a) low resolution and b) high resolution display formats.

a)

b)

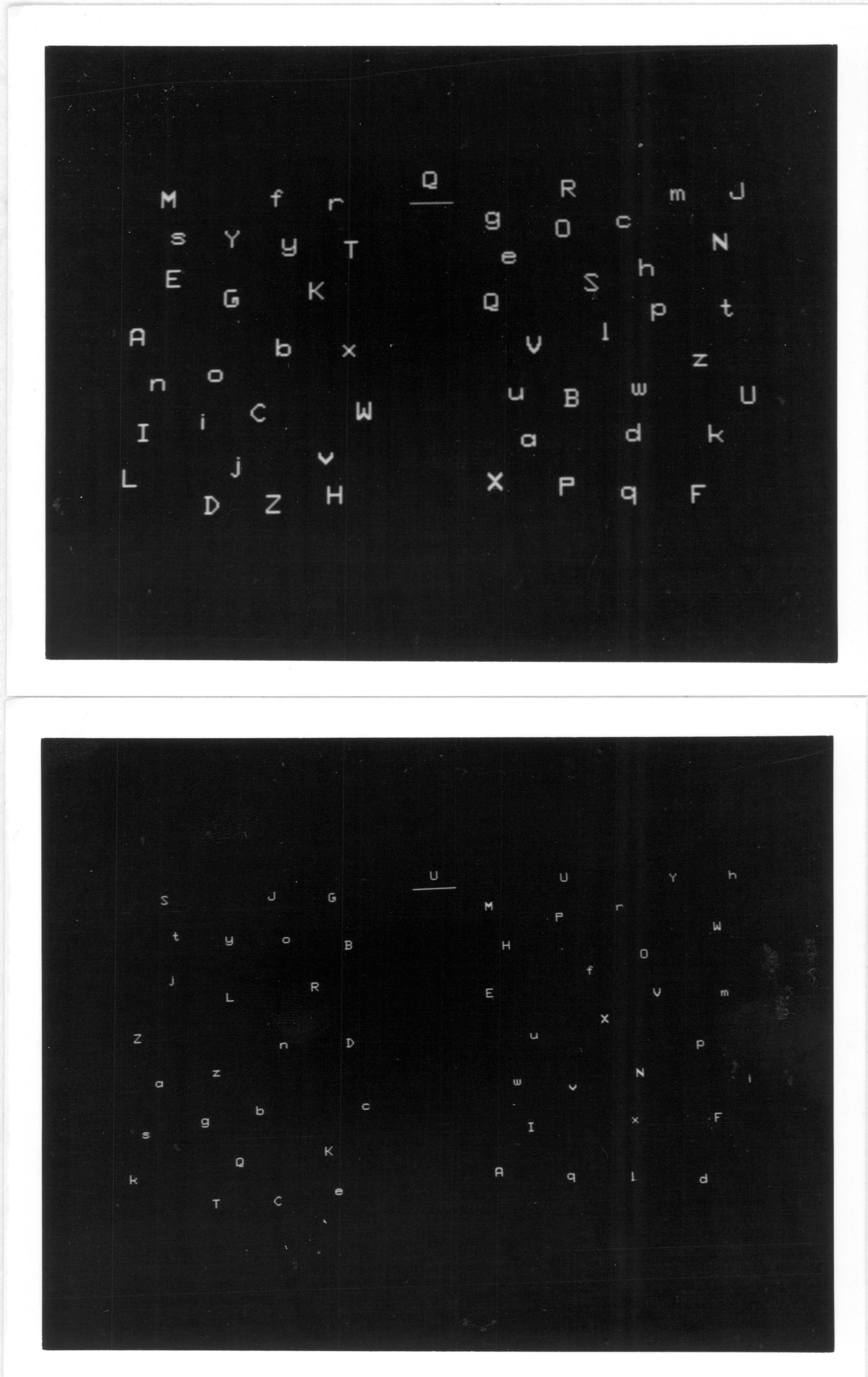
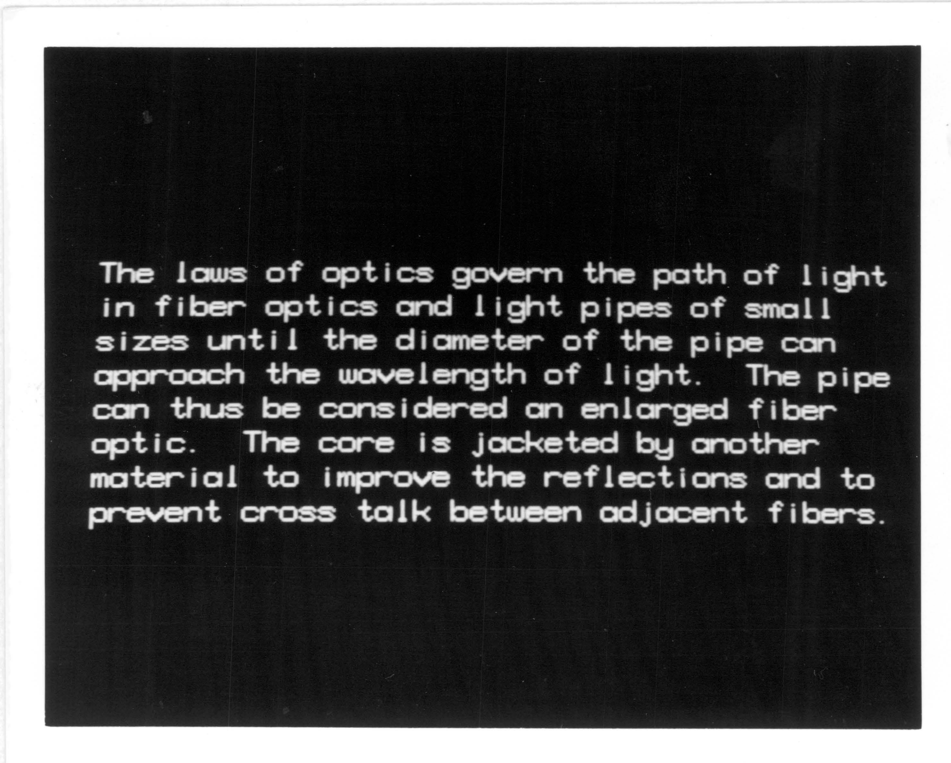


Figure 11. Photographs of alphabetic search patterns under a) low resolution and b) high resolution display formats.

verbally whether the character was in the left or right half of the pattern. The target character was always displayed in the upper-center of the pattern and was underlined. Again, all trials were subject initiated and were separated by 30.0 s intertrial intervals.

The image quality rating task required the subject to view an eight-line passage of text and to report verbally the perceived image quality of that passage. The passage was experimenter initiated, and required no input from the subject other than their verbal response. The subject chose from among a printed list of nine image quality descriptors ranging from "best imaginable" to "worst imaginable" and reported only the descriptor. Numerals were not associated with the descriptors until the statistical analyses. Photographs of the low and high resolution passage used for all rating trials are shown in Figure 12.

a)



b)

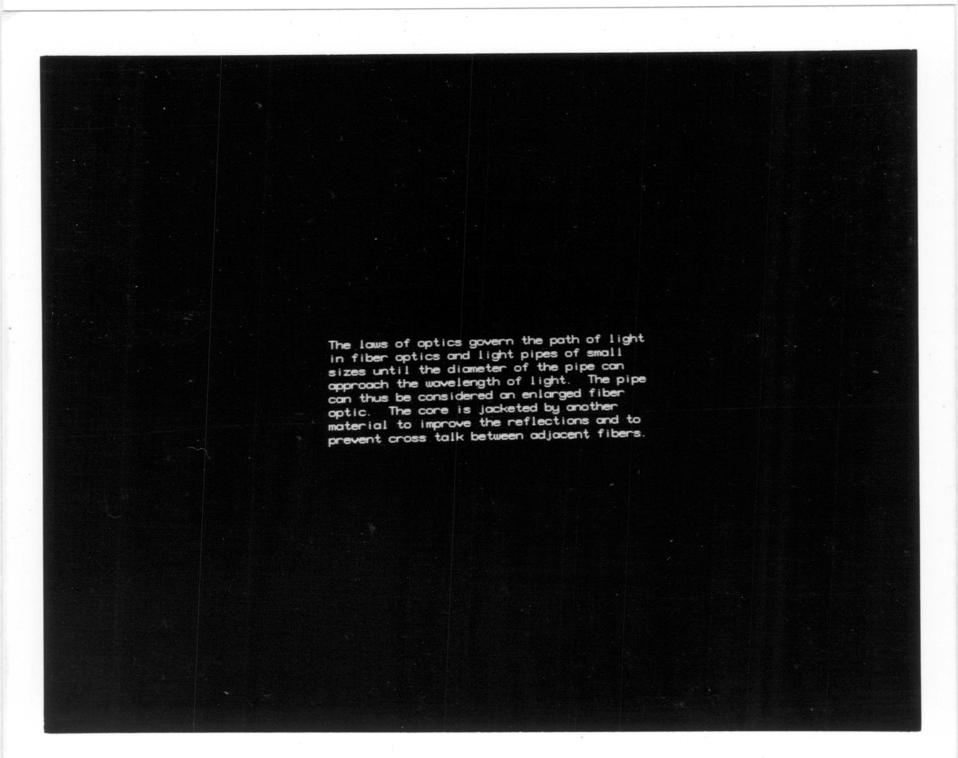


Figure 12. Photographs of image quality rating passages under a) low resolution and b) high resolution display formats.

RESULTS AND DISCUSSION - PHASE I

Image Quality Measurements

MTF results. Given that the only difference between MTFA and GMTFA (and MTF Area and GMTF Area) is that the latter is collected under glare conditions and is scaled to this glare, the metrics are then equivalent when collected under dark conditions. Remember that MTFA and MTF Area determinations are made only in the dark and do not directly reflect the ability of the display to minimize the effects of glare. Due to this equivalency and to simplify this analysis, only the GMTFA results will be discussed.

The results of the GMTFA measurements made under dark, diffuse, and specular lighting are shown in Figures 13, 14, and 15, respectively. The values used to construct these figures, as well as the results of the MTFA, MTF Area, and GMTF Area measurements are listed in Table 2. Plots of the line spread functions and resultant MTFs for each of the 16 filters and for the unfiltered, bare CRT under dark conditions are included in Appendix A. Note that all values reported in Table 2 are the means of the three, repeated-measurements made for each experimental condition. At the resolution settings used in this research, the metrics that consider only the quality of the display system (MTF Area and GMTF Area) and those that consider the display/observer system (MTFA and GMTFA), differ only by a scale factor from cycles/mm to cycles/deg and a small, additive constant that defines the area of the MTF below the CTF of the human observer (see Figure 5). For practical purposes, the metrics are directly proportional to one another and need not be discussed individually.

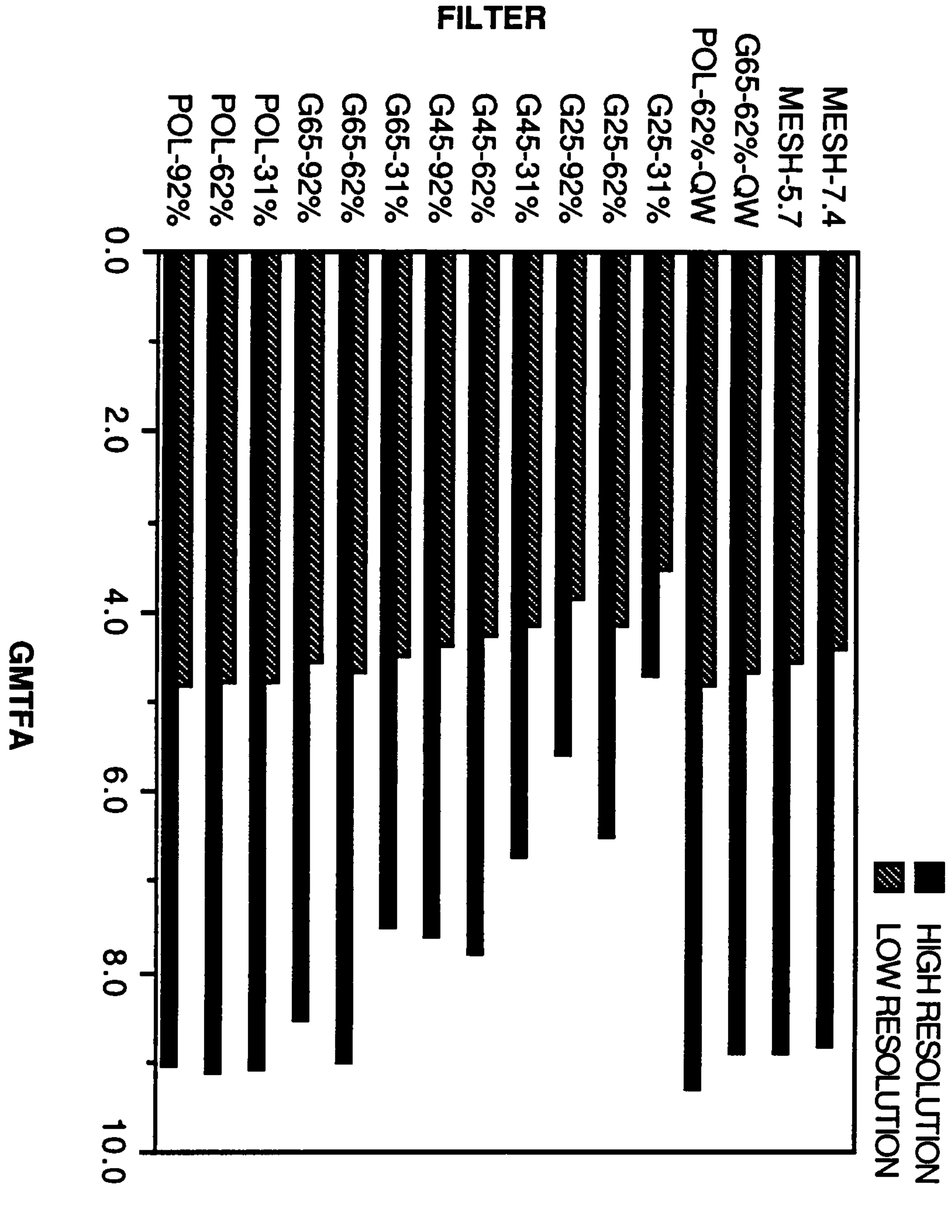


Figure 13. GMTFA as a function of filter and display resolution under dark lighting conditions.

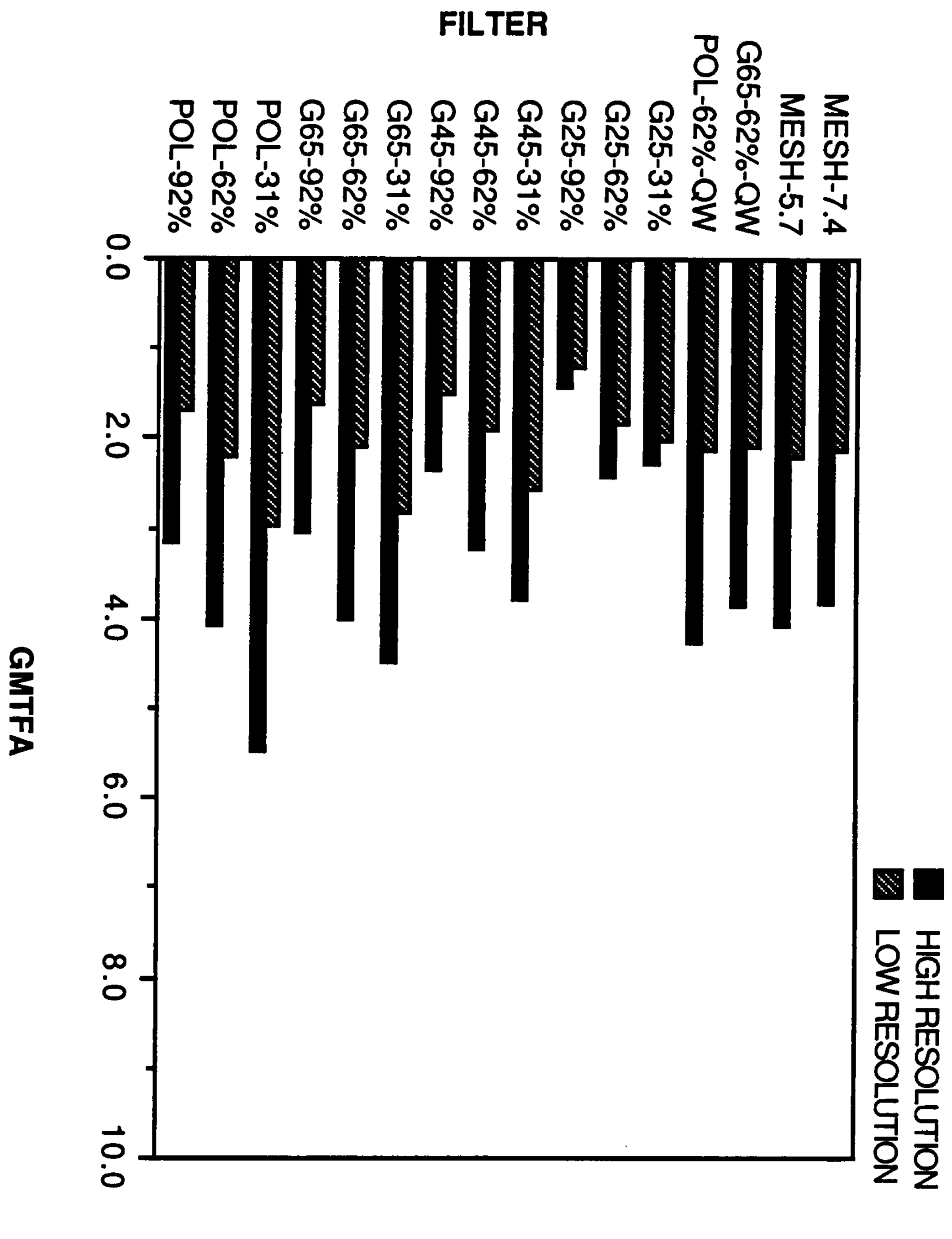


Figure 14. GMTFA as a function of filter and display resolution under diffuse lighting conditions.

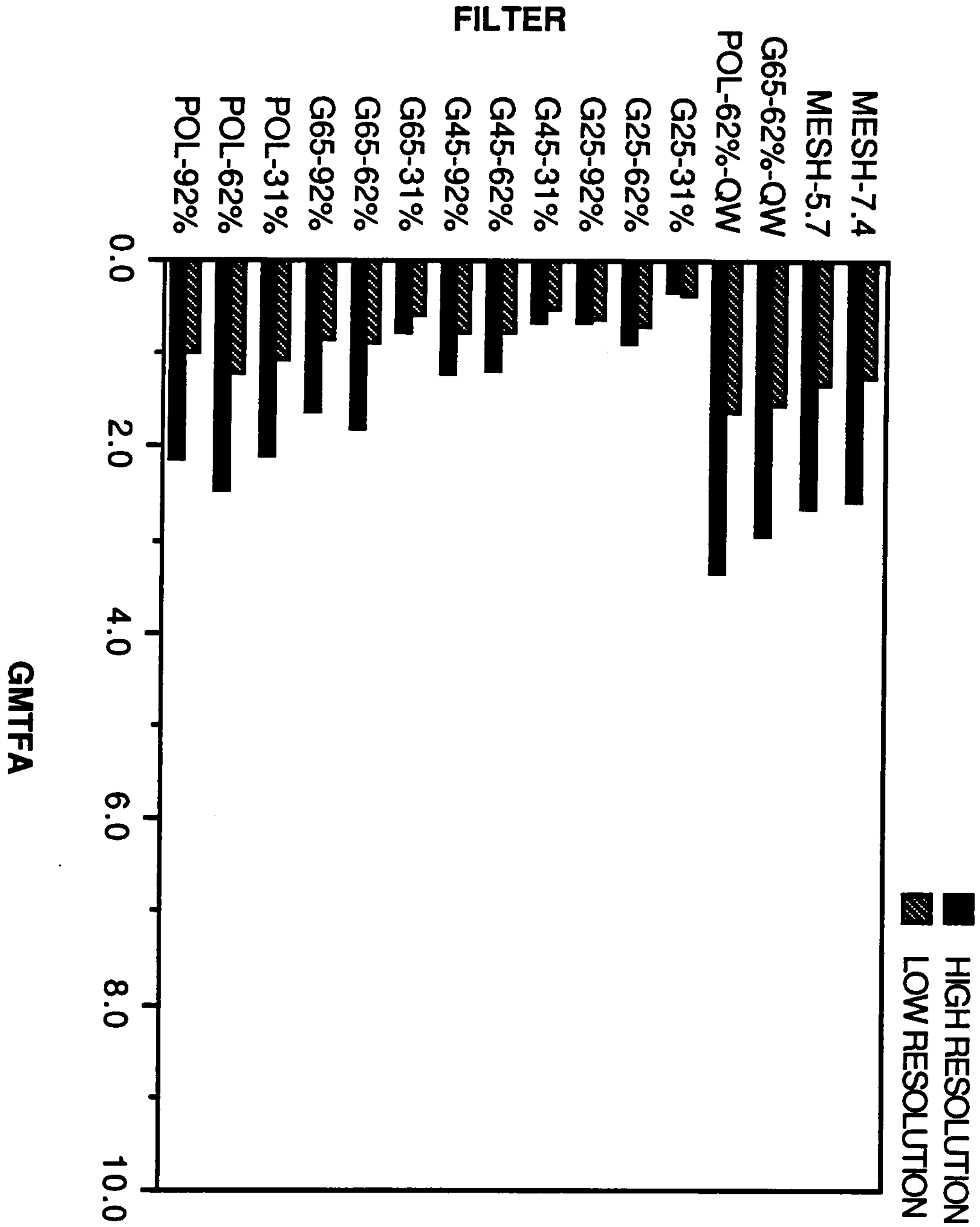


Figure 15. GMTFA as a function of filter and display resolution under specular lighting conditions.

Table 2

Results of the Image Quality Measurements - Phase I

	Low Res. - Dark		Low Res. - Diffuse		Low Res. - Specular	
Filter	MTF Area	MTFA	GMTF Area	GMTFA	GMTF Area	GMTFA
POL-92%	0.638	4.823	0.228	1.686	0.138	0.984
POL-62%	0.638	4.784	0.294	2.193	0.170	1.214
POL-31%	0.633	4.764	0.397	2.968	0.154	1.080
G65-92%	0.598	4.552	0.219	1.634	0.118	0.843
G65-62%	0.616	4.665	0.284	2.110	0.125	0.894
G65-31%	0.598	4.499	0.373	2.823	0.083	0.573
G45-92%	0.573	4.370	0.205	1.512	0.108	0.772
G45-62%	0.560	4.267	0.256	1.925	0.107	0.767
G45-31%	0.545	4.163	0.338	2.555	0.076	0.523
G25-92%	0.504	3.870	0.164	1.232	0.087	0.633
G25-62%	0.539	4.157	0.242	1.825	0.095	0.680
G25-31%	0.463	3.539	0.267	2.018	0.053	0.364
POL-62%-QW	0.635	4.811	0.288	2.150	0.223	1.629
G65-62%-QW	0.610	4.644	0.280	2.092	0.212	1.559
MESH-5.7	0.598	4.542	0.292	2.205	0.178	1.306
MESH-7.4	0.583	4.428	0.287	2.147	0.169	1.237

Note: MTF Area and MTFA measurements were only made under dark lighting conditions and that, under dark conditions GMTF Area and GMTFA were equivalent to MTF Area and MTFA, respectively.

(continued)

Table 2 (continued)

	High Res. - Dark		High Res. - Diffuse		High Res. - Specular	
Filter	MTF Area	MTFA	GMTF Area	GMTFA	GMTF Area	GMTFA
POL-92%	1.220	9.039	0.450	3.178	0.324	2.144
POL-62%	1.232	9.104	0.571	4.087	0.361	2.465
POL-31%	1.222	9.073	0.750	5.489	0.309	2.085
G65-92%	1.151	8.541	0.437	3.046	0.240	1.601
G65-62%	1.215	8.999	0.564	4.018	0.269	1.798
G65-31%	1.009	7.518	0.607	4.472	0.120	0.776
G45-92%	1.027	7.592	0.333	2.370	0.179	1.198
G45-62%	1.049	7.810	0.454	3.221	0.177	1.193
G45-31%	0.894	6.741	0.515	3.774	0.105	0.678
G25-92%	0.747	5.581	0.199	1.419	0.100	0.688
G25-62%	0.878	6.526	0.340	2.426	0.133	0.899
G25-31%	0.632	4.716	0.306	2.267	0.054	0.348
POL-62%-QW	1.250	9.293	0.592	4.267	0.477	3.330
G65-62%-QW	1.197	8.882	0.541	3.860	0.417	2.947
MESH-5.7	1.198	8.902	0.562	4.081	0.379	2.652
MESH-7.4	1.149	8.523	0.529	3.827	0.364	2.570

Note: MTF Area and MTFA measurements were only made under dark lighting conditions and that, under dark conditions GMTF Area and GMTFA were equivalent to MTF Area and MTFA, respectively.

Directly evident from Figure 13 is that, in the dark, low resolution yielded smaller GMTFA values than did the high resolution for all filters evaluated. This could be expected, as the high resolution spot was about half the width of the low resolution and thus contained higher spatial frequency information. It is also apparent that as the harshness of etch was increased, that GMTFA decreased. That is, the polished surfaces produced the highest GMTFA, followed by the 65 gloss, the 45 gloss, and then the 25 gloss for both low and high resolutions. Again, this result reflects the loss of high frequency information as the image becomes increasingly blurred. This effect, however, was not so extreme at low resolution, and indicates the low-pass filtering characteristics of the etch. Therefore, because the high resolution contained more high frequency information, the low-pass filtering had a greater effect on GMTFA. The coated and mesh filters failed to show any degradation in measured image quality when compared with their polished and etched, non-coated, baselines.

An unexpected result was the strong performance of the 62% transmission filter at the 65, 45, and 25 gloss conditions. Analytically, these filters should have yielded lower GMTFAs than the 92% filters and somewhat higher GMTFAs than the 31% filters at each of the respective etch levels. A rather simple explanation for these results exists. Each of these panels was picked from a pool of similar panels in an attempt to include only filters with nearly identical gloss values at each of the three levels of etch. However, the gloss values reported by the company performing the etching were significantly different than those obtained during our own gloss measurements of these panels. It appeared that the producers of the etched panels had used inappropriate procedures during their gloss measurements and had thus delivered panels that were misleadingly labeled. In particular, the harshness of the etch on the 62% filters was not as extreme as their gloss numbers indicated. While for simplicity, these filters have been referred to in this report as gloss 65, 45, and 25, the 62% transmission filters have actual gloss values of 74, 49, and 35, respectively. Thus, they do not exhibit as much image degradation as would normally be expected. This anomaly has implications for the human performance statistical analyses and will be discussed again.

Figure 14 shows the GMTFA results obtained under diffuse lighting. Note the addition of diffuse glare and subsequent loss of contrast produced an overall reduction in GMTFA when compared with the GMTFA in the dark. Again, low resolution yielded consistently smaller GMTFA values than high resolution and there were large reductions in GMTFA with increasing etch. Also of interest is the finding that for both low and high resolution, decreasing the transmission of the filter served to increase GMTFA. That is, as filter transmission decreases under diffuse lighting, the displayed information increases in contrast, the higher contrast being reflected by corresponding increases in GMTFA. Again, the coated filters did not differ significantly from their non-coated, 62% baseline filters and the two mesh filters performed well due to the contrast enhancement they offer.

Figure 15 shows the effect of specular lighting on GMTFA. Immediately evident is that GMTFA is reduced over both the dark and diffuse lighting conditions, but that the mesh and coated filters have been affected to a lesser extent than the other filters. Here, the anti-reflection characteristics of the mesh and quarterwave coatings have reduced the first-surface reflected luminance and therefore increased the contrast. Interesting, however, is that increased gloss did not yield poorer GMTFAs under these specular lighting conditions. Etches have traditionally been used to eliminate reflection, but it appears that they do so ineffectively. While increasing the level of etch passes less high frequency information through the filter from the display under any lighting condition, under specular lighting the etch also serves to redistribute the reflected image across the face of the filter. This redistribution creates a veiling glare and essentially lowers the contrast of the image. Here again, the superior performance of the 62% transmission filters is indicative of the reduced etch harshness of these filters.

RTF Area results. Figure 16 shows the results of the RTF Area measurements. The numerical values are listed in Table 3. Note that in Figure 16, log RTF Area is plotted for each filter. This transform was necessary as the range of actual RTF Area values is about four orders of magnitude and could not be easily represented on the same figure. Here, larger (more positive) values represent a decreased ability of the filter to eliminate reflection and are thus considered a measure of filter noise under specular conditions.

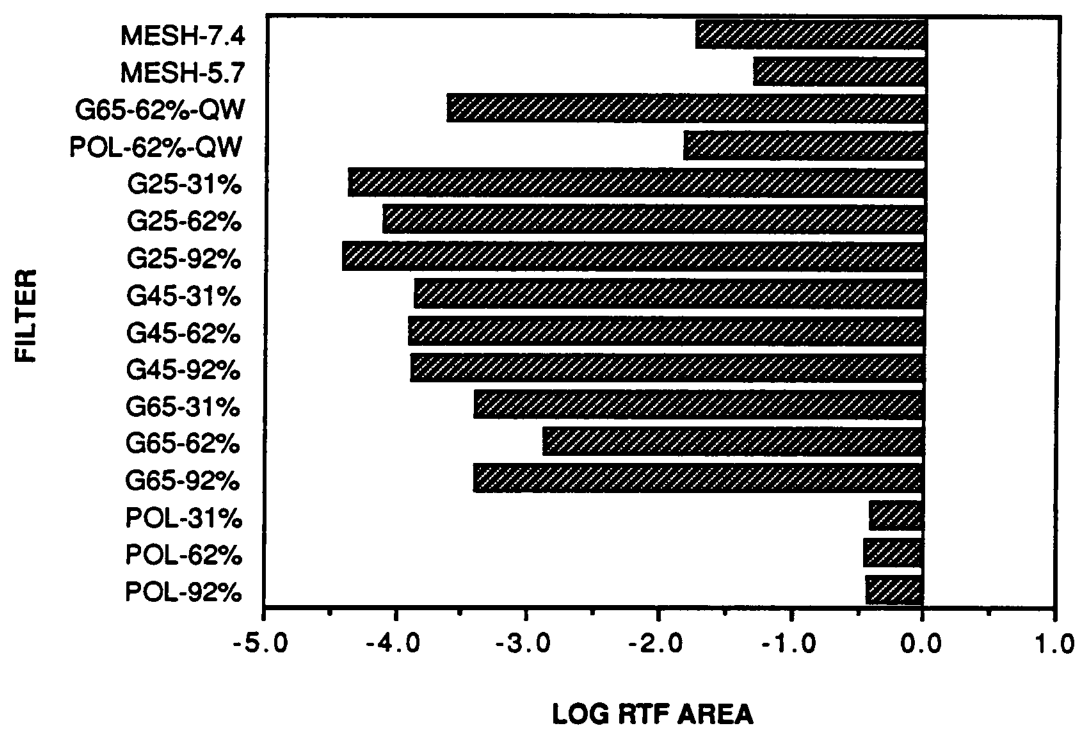


Figure 16. Log RTF Area as a function of filter. More negative values indicate reduced first surface reflectivity.

Table 3

Results of Noise Measurements for Phase I

Filter	RTF Area	Gloss	WS Area	VWSA
POL-92%	.383400	85.8	.009805	.017560
POL-62%	.367400	87.7	.005390	.006274
POL-31%	.406000	85.2	.000024	.000025
G65-92%	.000402	65.9	.010330	.015050
G65-62%	.001386	74.0	.005144	.006768
G65-31%	.000396	65.0	.003360	.003540
G45-92%	.000130	45.6	.008826	.012070
G45-62%	.000129	49.1	.006242	.008029
G45-31%	.000143	46.1	.003627	.001999
G25-92%	.000040	24.6	.010110	.015950
G25-62%	.000079	35.1	.005735	.006806
G25-31%	.000042	26.0	.003666	.004136
POL-62%-QW	.015120	37.9	.006326	.009196
G65-62%-QW	.000239	30.9	.006989	.008978
MESH-5.7	.051080	18.0	1.724300	.222500
MESH-7.4	.018500	7.3	.874800	.313100

Figure 16 reveals some interesting findings. First, as expected, the uncoated polished filters yielded the highest RTF Areas, followed by the uncoated 65 gloss, 45 gloss, and 25 gloss. Again, the reduction in etch harshness for the 62% transmission filters is evident by the higher RTF Areas found for these panels. The mesh filters (placed over a polished panel) produced RTF Area values about one order of magnitude less than the polished, but not as low as any of the etched panels. As would be expected due to the increased absorption and scattering characteristics of the higher spatial frequency filter, the 5.7 lines/mm mesh was not as effective as the 7.4 lines/mm filter. The coated panels both reduced RTF Area significantly over their baseline, comparison 62% transmission filters. However, while the mesh and coated filters appear effective in reducing the RTF Area, the greatest gains were recognized with etching.

Gloss measurement results. The results of the gloss measurements are shown in Figure 17 and listed in Table 3 as well. Similar to the RTF Area results, there were clear reductions in gloss as the harshness of the etch was increased. The coated filters also showed significant reductions in gloss when compared to their uncoated, baseline filters. The mesh filters appear to be the most effective in gloss reduction, yielding the smallest gloss values of all filters evaluated. Again, the 62% transmissivity panels produce gloss values indicating that they are more reflective than the 31% and 92% panels at each level of etch harshness.

Wiener spectrum results. The results of the Wiener spectrum measurements are shown in Figure 18 and are listed in Table 3. As a measure of noise inherent in the filter, larger values of WS Area or VWS Area indicate poorer image quality. Figure 18a shows that the log WS Area results were very similar for nearly all filters evaluated. The polished, 31% filter stood out in that it apparently contained only very small luminance fluctuations, while the mesh filters produced WS Area results about two orders of magnitude greater than the rest of the filters. The mesh results are not surprising, as the measurement is sensitive to luminance modulation created by the carbon fibers making up the mesh itself. Therefore, the Wiener spectrum contains substantial power at the fundamental spatial frequency of the mesh (i.e., 5.7 and 7.4 cycles/mm).

The visually weighted WS Area (VWSA) results are shown in Figure 18b. For the most part, these results are virtually identical to the WS Area results shown above. Notable, however, are the

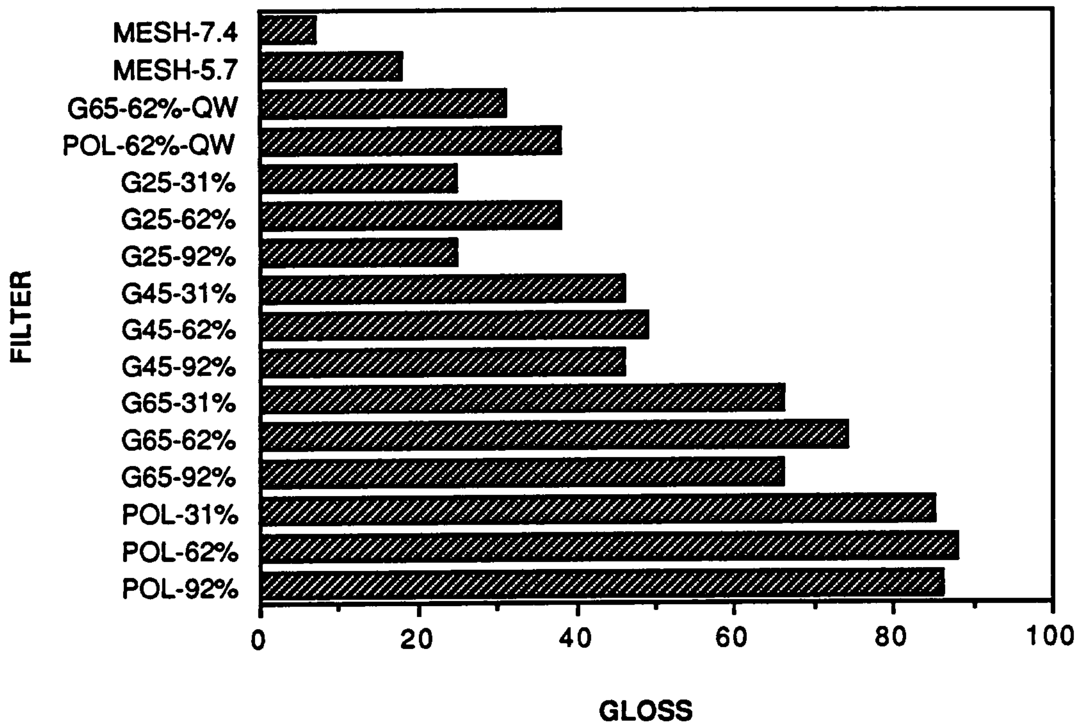
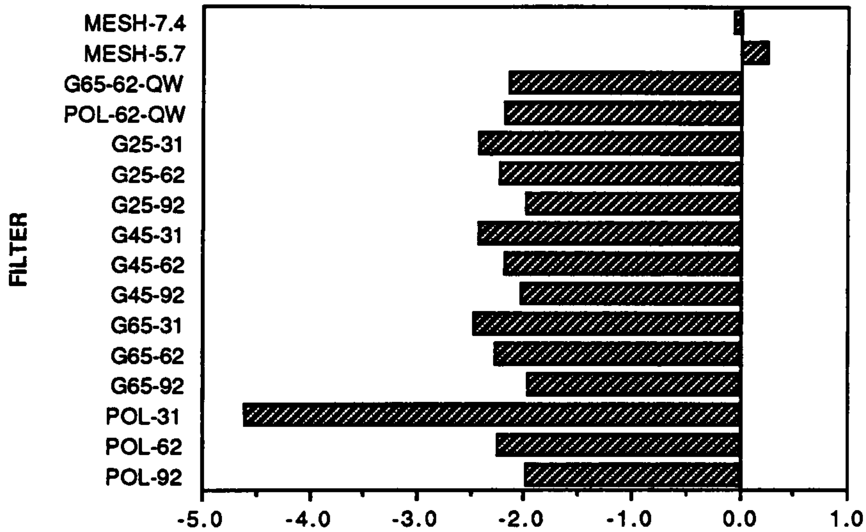
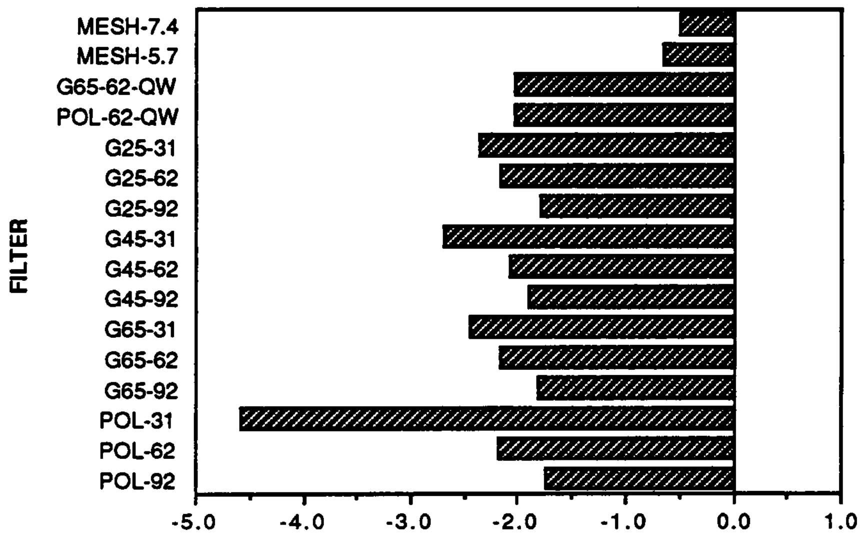


Figure 17. GLOSS as a function of filter. Smaller values indicate reduced first surface reflectivity.



a)

LOG WS AREA



LOG VWS AREA

b)

Figure 18. a) Log WS Area and b) Log VWSA as a function of filter. More negative values indicate reduced noise.

improvements in VWSA found for the two mesh filters. Explanation for this results lies in the fact that when the WS is weighted by the CSF of the human observer, the importance of high spatial frequency information in the WS is drastically reduced. For the mesh filters, much of the power of the WS located in "spikes" at 5.7 and 7.4 cycles/mm (about 46.0 and 59.0 cycles/deg at 46.0 cm viewing distance) is given a near-zero weighting. In theory, these high frequencies should be only minimally detectable at levels of modulation approaching unity.

Signal-to-noise ratio results. Due to the large number of signal-to-noise ratios (SNRs) formed, it is not useful to present figures of each or present the values from the SNR calculations in tabular form. Therefore, the SNRs are shown only in terms of their calculational formulas in Table 4. These formulas, along with the signal and noise values presented in Tables 2 and 3, respectively, should enable the interested reader to re-create these SNRs. Note that in most instances it is not meaningful to indicate the units making up the SNR and that comparisons among the various ratios is also inappropriate as they are contain no common scale. To discuss the results of each SNR calculation here would be inefficient and tedious, as the value of the SNR lies in its ability to predict human performance. Therefore, SNRs found to predict reading time, search time, or image quality ratings will be identified in a later section concerned with performance modeling.

Correlations among the metrics. The Pearson product-moment correlation coefficients defining the relationships among the image quality metrics at both low and high resolution are shown in Table 5. These are important in that they can reveal collinearities among the metrics and can, therefore be used to help determine which metrics should be included during model-building and human performance prediction.

A number of interesting relationships emerge from these correlations. First, the relatively high, positive correlations (from .42 to .58) between the measures of signal strength (MTFA and MTF Area) and measures regarded as indicators of noise (RTF Area and GLOSS) are important. It is difficult to argue that when such collinearities exist, that the MTF-based signal measures and the noise metrics are not, in fact, describing common filter characteristics. This relationship can be explained fairly simply in that etched filters were found to degrade the MTF of the display and also to reduce the values recorded for RTF Area and GLOSS. Therefore, as signal is reduced by the

Table 4

Signal-to-noise ratio calculational formulas

MTF Area / WS Area
MTF Area / VWSA
MTF Area / RTF Area
MTF Area / GLOSS
MTF Area / WS Area + GLOSS
MTF Area / VWSA + GLOSS
MTF Area / WS Area + RTF Area
MTF Area / VWSA + RTF Area
GMTF Area / WS Area
GMTF Area / VWSA
GMTF Area / RTF Area
GMTF Area / GLOSS
GMTF Area / WS Area + GLOSS
GMTF Area / VWSA + GLOSS
GMTF Area / WS Area + RTF Area
GMTF Area / VWSA + RTF Area
MTFA / WS Area
MTFA / VWSA
MTFA / RTF Area
MTFA / GLOSS
MTFA / WS Area + GLOSS
MTFA / VWSA + GLOSS
MTFA / WS Area + RTF Area
MTFA / VWSA + RTF Area
GMTFA / WS Area
GMTFA / VWSA
GMTFA / RTF Area
GMTFA / GLOSS
GMTFA / WS Area + GLOSS
GMTFA / VWSA + GLOSS
GMTFA / WS Area + RTF Area
GMTFA / VWSA + RTF Area

Table 5

Pearson Product-Moment Correlations among Image Quality Measurements at Low and High Resolution

LOW RESOLUTION

	GMTF Area	MTFA	GMTFA	LOG MTFA	LOG GMTFA	RTF Area	GLOSS	WS Area	VWSA
MTF Area	.18	.99	.17	.99	.22	.54	.58	.04	.07
GMTF Area	--	.18	.99	.18	.96	.10	.07	.04	.04
MTFA	--	--	.16	.99	.22	.52	.55	.04	.07
GMTFA	--	--	--	.16	.96	.08	.06	.04	.04
LOG MTFA	--	--	--	--	.22	.49	.54	.05	.08
LOG GMTFA	--	--	--	--	--	.10	.05	.07	.08
RTF Area	--	--	--	--	--	--	.69	-.12	-.10
GLOSS	--	--	--	--	--	--	--	-.55	-.53
WS Area	--	--	--	--	--	--	--	--	.86

HIGH RESOLUTION

	GMTF Area	MTFA	GMTFA	LOG MTFA	LOG GMTFA	RTF Area	GLOSS	WS Area	VWSA
MTF Area	.35	.99	.33	.99	.42	.44	.43	.19	.23
GMTF Area	--	.35	.99	.34	.93	.16	.13	.09	.10
MTFA	--	--	.33	.99	.42	.44	.42	.19	.23
GMTFA	--	--	--	.32	.92	.15	.12	.08	.09
LOG MTFA	--	--	--	--	.42	.41	.42	.19	.23
LOG GMTFA	--	--	--	--	--	.18	.13	.12	.14
RTF Area	--	--	--	--	--	--	.69	-.12	-.10
GlOSS	--	--	--	--	--	--	--	-.55	-.53
WS Area	--	--	--	--	--	--	--	--	.86

etch, so is the amount of noise. Conversely, as signal strength is increased, the amount of noise also increases. In theory, the findings are contrary to one another and it may be inappropriate to only consider the RTF Area metric solely as a measure of display noise.

As expected, the correlations between the display dependent MTF- based image metrics (MTF Area and GMTF Area) and their counterpart metrics that consider the display and human observer (MTFA-LOG MTFA and GMTFA- LOG GMTFA, respectively) were extremely high in all cases. Certainly, the strength of these relationships indicates that simultaneous inclusion of more than one of these highly correlated metrics into a regression model is probably inappropriate. Along these same lines, the correlations between WS Area and VWSA also define the metrics to be highly nonorthogonal to one another.

Human Performance Results

The results of the human performance measurements are presented this section. As discussed previously, the 62% transmission filters at the 25, 45, and 65 gloss levels were found to be inappropriately labeled in terms of the harshness of the etch that had been applied to them. This inconsistency led to the exclusion of these filters from the subsequent analysis of variance procedures, in that the results of these analyses would have been confounded and may have led to inaccurate and misleading interpretations. However, the data from these filters were included in the model-building, as the image quality measurements and human performance data generated for these filters are still valid and meaningful.

Search results. Table 6 shows the results of a 2 x 3 x 13 (Resolution x Lighting x Filter) repeated-measures, within-subjects analysis of variance (ANOVA) performed on search times. Two second-order effects (R x F and R x L) reached statistical significance and will be discussed in some detail below. The Resolution, Lighting, and Filter main effects are best understood in the context of these interactions.

Table 6

Analysis of Variance Summary for Search Time - Phase I

Source of Variance	df	MS	F	p
<u>Between Subjects</u>				
Subjects (S)	7	132.997		
<u>Within Subject</u>				
Resolution (R)	1	3355.532	269.00	0.0001
R x S	7	12.474		
Lighting (L)	2	426.922	26.71	0.0001
L x S	14	15.983		
Filter (F)	12	63.438	4.07	0.0001
F x S	84	15.587		
R x L	2	140.588	5.19	0.0206
R x L x S	14	27.076		
R x F	12	41.814	3.42	0.0004
R x F x S	84	12.222		
L x F	24	13.496	0.93	0.5600
L x F x S	168	14.494		
R x L x F	24	14.433	1.37	0.1258
R x L x F x S	168	10.497		
Total	623			

Figure 19 shows the R x F effect on search time. Note that the figure represents delta search time values that have been scaled to the high resolution, polished-92% filter search times. Thus, the baseline condition is zero, and all other conditions are positive or negative difference times from that baseline. It is important to recognize in this figure and all subsequent figures presenting delta search time or delta reading time data, that more positive (to the right) values denote poorer (longer) performance.

The results of simple-effect F-tests on Filter at low and high resolution levels are shown in Table 7. It can be seen that Filter was only significant at high resolution. Therefore, a post hoc Newman-Keuls paired comparisons test was conducted and the results are presented in Table 8. Note that at high resolution, a number of groupings among filters exist. First, it was found that while the gloss25-31% filter was not different than the gloss65-31%, gloss25-92%, and gloss45-31% filters, it did produce significantly longer search times (ST) than all other filters. Also, at high resolution, no filter yielded shortened STs over the polished-92% baseline filter (4.61 s). In general, from Figure 19 it is clear that legibility (ST) is improved for all filters when the larger, low resolution characters are displayed. Further, these larger characters appear to be more resistant to image degradations due to the etched filters. That is, at high resolution, the low-pass filtering characteristics of the etched filters markedly increase the length of time required to locate random alphabetic characters and thereby decrease the legibility of the display. Finally, no filter improved performance over the baseline, although a number of filters yielded STs that, in a practical sense are equivalent to this baseline.

The significant R x L interaction is shown in Figure 20 and the results of the corresponding simple-effect F-test on Resolution at dark, diffuse, and specular lighting are presented in Table 9. Under all three lighting conditions, the low resolution produced significantly faster STs than did the high resolution, the difference being more extreme under dark lighting than either of the glare conditions. Also shown in Table 9 are the results of the simple-effect F-test performed on Lighting at low and high resolution. Significant differences among the levels of lighting were shown to exist at high resolution and the results of the Newman-Keuls analysis on Lighting at high resolution are

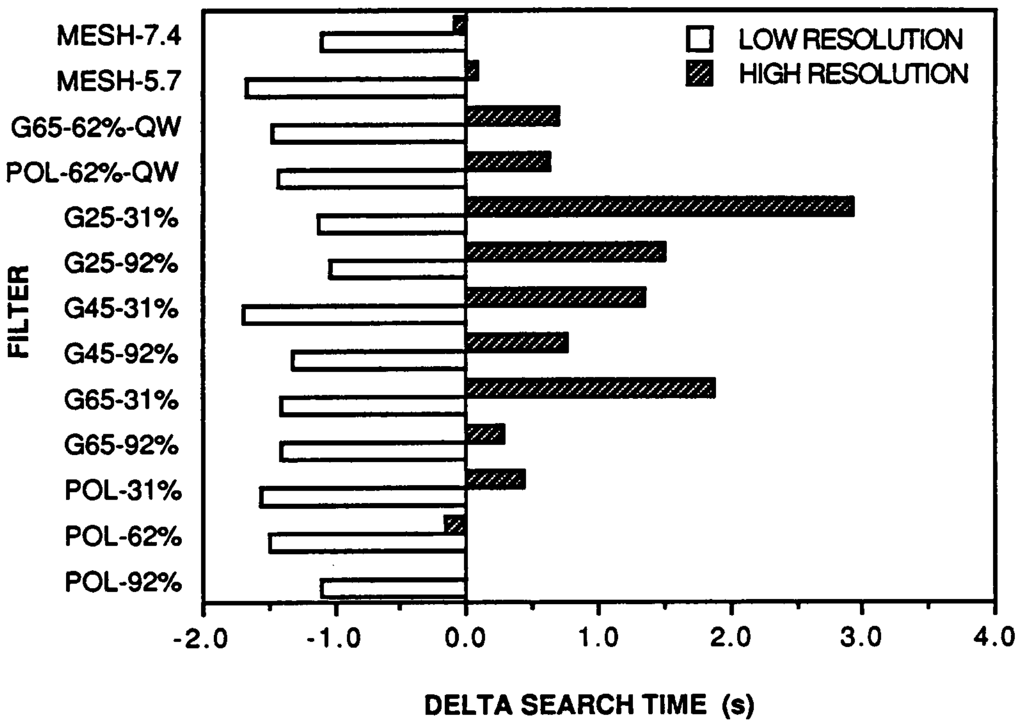


Figure 19. The R x F interaction for delta search time. All values are scaled to the high resolution, POL-92% condition (4.61 s).

Table 7

Simple-effect F-test for Filter at Low and High Resolution - Search Time

Source of Variance	df	MS	F
F at Low Resolution	12	8.561	0.55
F at High Resolution	12	96.691	6.20 **
F x S	84	15.587	

** $p < .01$

Table 8

Newman-Keuls Comparisons for Filter at High Resolution - Search Time

Filter	Mean ST (s)				
POL-62%	4.46	A			
MESH-7.4	4.52	A	B		
POL-92%	4.61	A	B		
MESH-5.7	4.69	A	B		
G65-92%	4.89	A	B	C	
POL-31%	5.04	A	B	C	
POL-62%-QW	5.24	A	B	C	
G65-62%-QW	5.32	A	B	C	
G45-92%	5.37	A	B	C	
G45-31%	5.96		B	C	D
G25-92%	6.12		B	C	D
G65-31%	6.49			C	D
G25-31%	7.54				D

Note: Means accompanied by the same letter are not significantly different, $p > .05$.

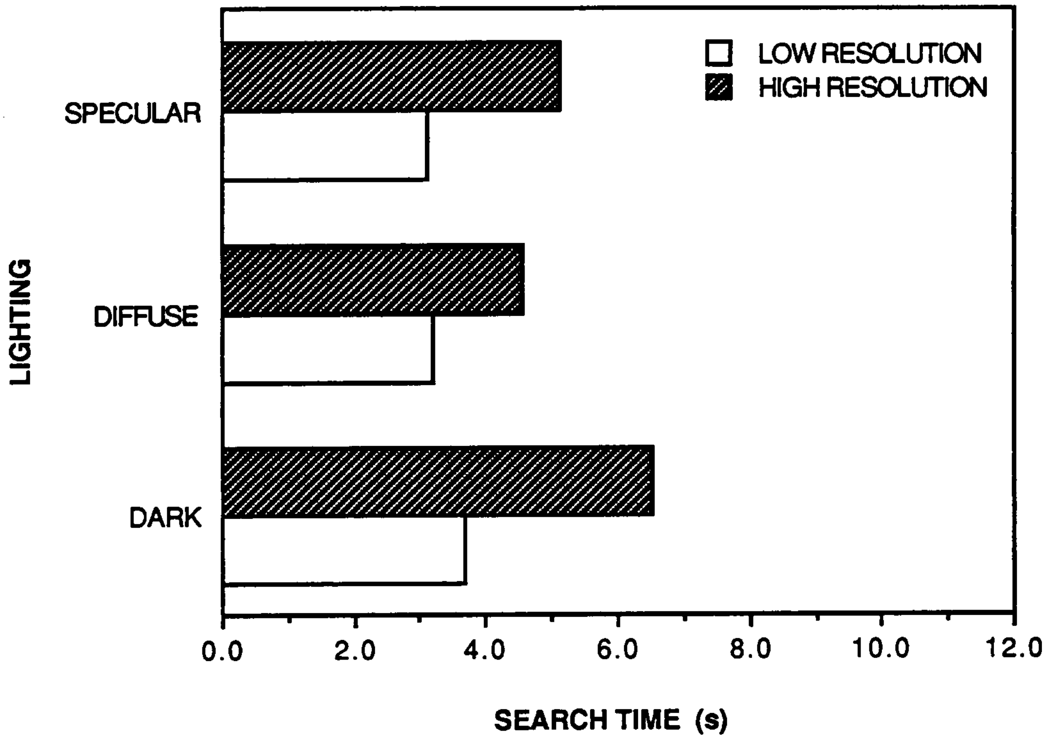


Figure 20. The R x L interaction for search time.

Table 9

Simple-effect F-test for Resolution at Dark, Diffuse, and Specular Lighting and for Lighting at Low and High Resolution - Search Time

Source of Variance	df	MS	F
R at Dark Lighting	1	1051.528	167.73 **
R at Diffuse Lighting	1	492.794	39.50 **
R at Specular Lighting	1	2092.388	84.29 **
R x S	7	12.474	

** $p < .01$

Source of Variance	df	MS	F
L at Low Resolution	2	48.505	3.03
L at High Resolution	2	519.006	32.47 **
L x S	14	15.983	

** $p < .01$

shown in Table 10. Search times were much longer in the dark than under specular (about 1.4 s) or diffuse (about 2.0 s) lighting and all were significantly different. These results again indicate that larger characters yielded better legibility regardless of the lighting under which the display was viewed. Further, they point out that small characters should not be used in environments where no illumination is incident on the display surface. Possible explanations for this final point will be offered in later sections.

While other main effects were statistically significant, their interpretation must be made only when considering the higher-order interactions that have been described above. Therefore, these main effects will not be considered at this point, but can be reconstructed from the means of the second-order effects.

Reading results. The results of the repeated-measures, within-subjects ANOVA on the reading time (RT) data are shown in Table 11. The R x L x F interaction was significant, as were all second-order interactions and main effects. While these lower-order effects are interesting and important, they are most meaningfully explained in terms of the R x L x F interaction. Therefore, to avoid redundancy and to simplify the presentation of the results, the lower-order effects will not be specifically discussed here.

Due to the complexity of the third-order interaction, it was broken into more manageable two-way interactions for representation and analysis. Figures 21, 22, and 23 show the R x F interactions under specular, dark, and diffuse lighting, respectively. Simple main effects F-tests were performed on these R x F interactions and the results of these tests are listed in Table 12. These results indicate that only under specular lighting did the interaction reach statistical significance. All RTs in Figures 21, 22, and 23 have been scaled to the high resolution, polished-92% (5.46 s) mean RT. Therefore, this baseline condition was set equal to zero and all other conditions are delta RTs about this baseline.

To interpret the third-order effect, it is necessary to examine Figure 21, the R x F interaction under specular lighting, more closely. Table 13 shows the results of the simple-effect F-test performed on Filter at low and high resolution. Significant differences among filters were detected at

Table 10

Newman-Keuls Comparisons for Lighting at High Resolution - Search Time

Lighting	Mean ST (s)			
Diffuse	4.57	A		
Specular	5.12		B	
Dark	6.52			C

Note: Means accompanied by the same letter are not significantly different, $p > .05$.

Table 11

Analysis of Variance Summary for Reading Time - Phase I

Source of Variance	df	MS	F	p
<u>Between Subjects</u>				
Subjects (S)	7	630.400		
<u>Within Subject</u>				
Resolution (R)	1	288.379	14.27	0.0069
R x S	7	20.214		
Lighting (L)	2	97.471	19.00	0.0001
L x S	14	5.128		
Filter (F)	12	81.355	6.62	0.0001
F x S	84	12.291		
R x L	2	46.019	22.39	0.0001
R x L x S	14	2.055		
R x F	12	24.297	5.84	0.0001
R x F x S	84	4.157		
L x F	24	13.606	5.05	0.0001
L x F x S	168	2.692		
R x L x F	24	11.551	4.34	0.0001
R x L x F x S	168	2.663		
Total	623			

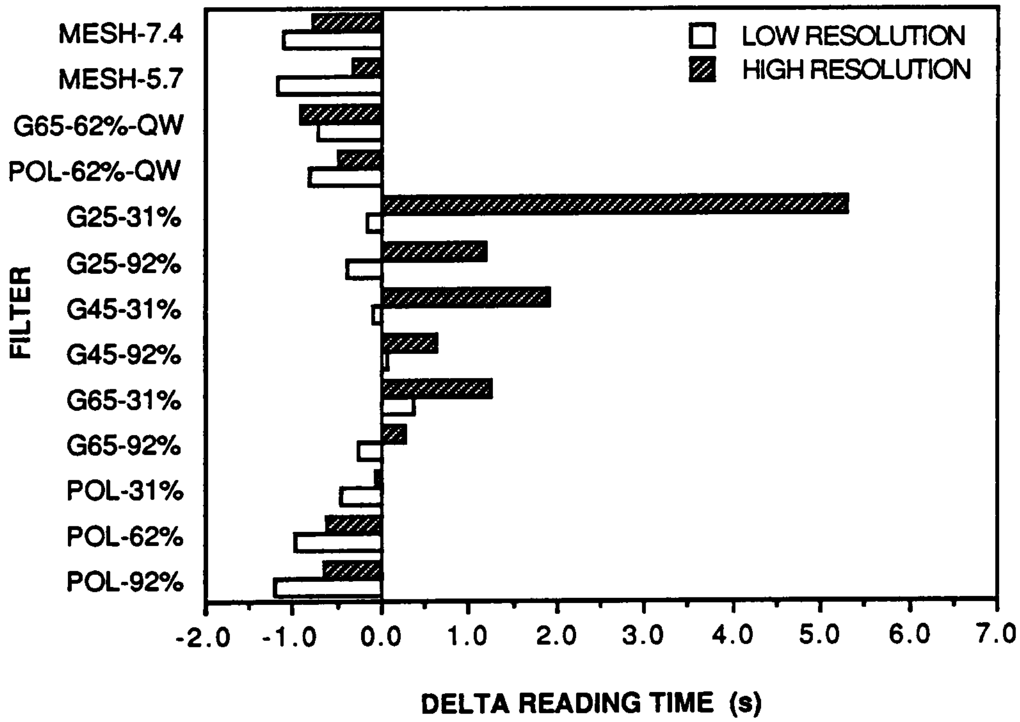


Figure 21. The R x F interaction for delta reading time under specular lighting. All values are scaled to the dark, high resolution, POL-92% condition (5.46 s).

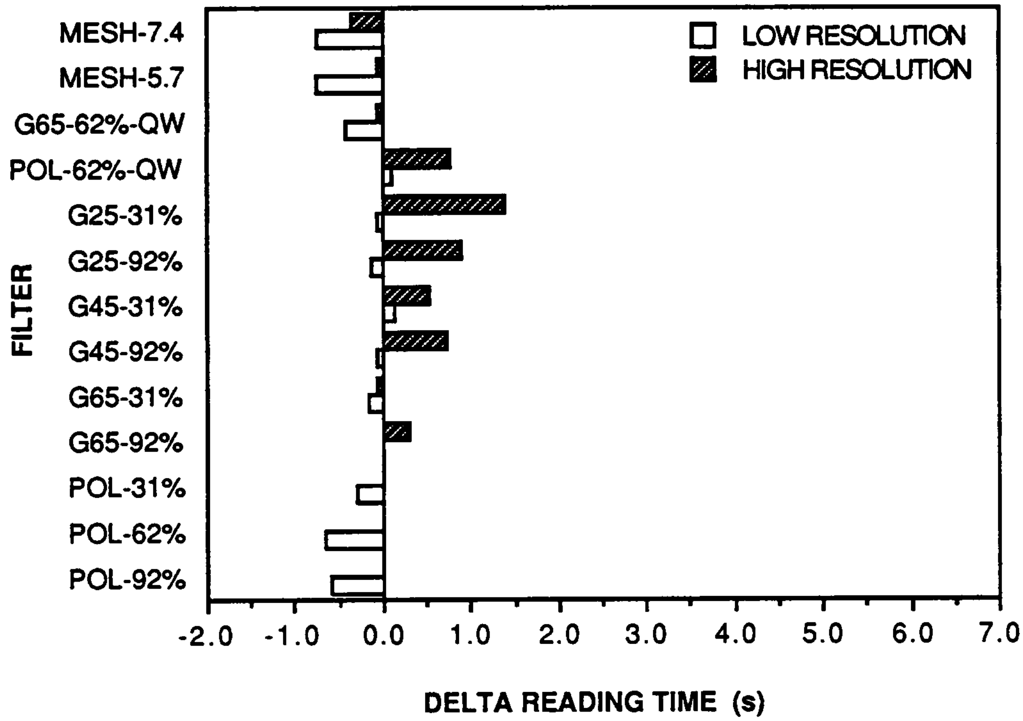


Figure 22. The R x F interaction for delta reading time under dark lighting. All values are scaled to the dark, high resolution, POL-92% condition (5.46 s).

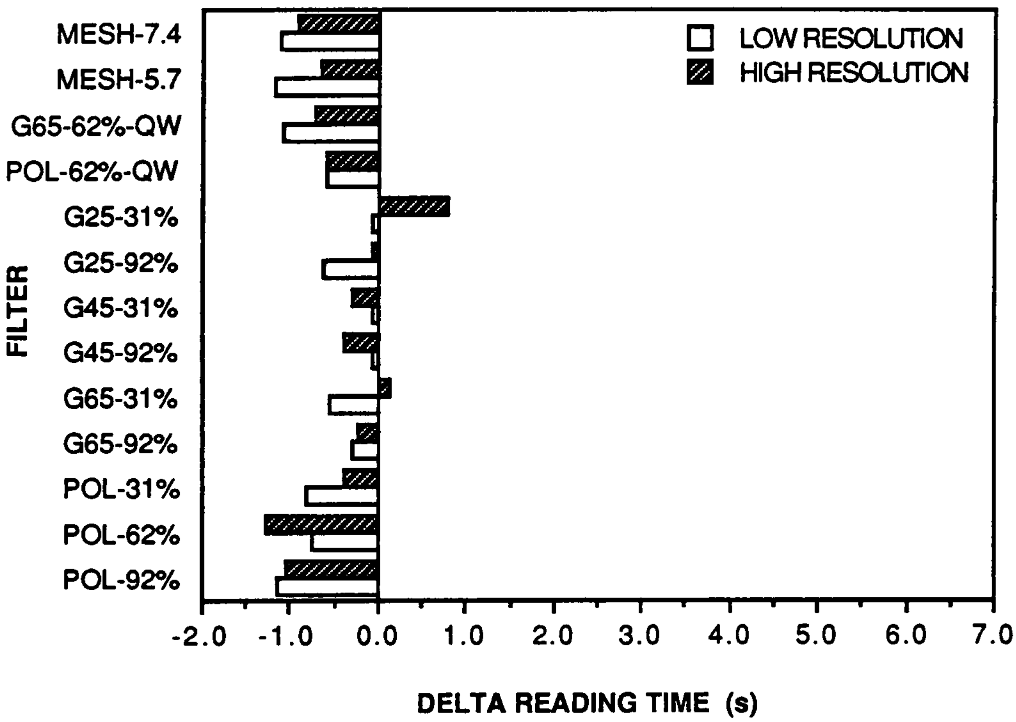


Figure 23. The R x F interaction for delta reading time under diffuse lighting. All values are scaled to the dark, high resolution, POL-92% condition (5.46 s).

Table 12

Simple-effect F-test for R x F at Dark, Diffuse, and Specular Lighting - Reading Time

Source of Variance	df	MS	F
R x F at Dark Lighting	12	2.491	0.59
R x F at Diffuse Lighting	12	3.334	0.80
R x F at Specular Lighting	12	41.157	10.00 **
R x F x S	84	4.157	

** $p < .01$

Table 13

Simple-effect F-test for Filter at Low and High Resolution under Specular Lighting - Reading Time

Source of Variance	df	MS	F
F at Low Resolution	12	10.342	0.84
F at High Resolution	12	112.906	9.18 **
F x S	84	12.292	

** $p < .01$

high resolution, but not at low resolution under specular lighting. Therefore, a Newman-Keuls analysis was conducted and the differences among filters at high resolution while under specular lighting are presented in Table 14. Here, the gloss25-31% filter is shown to have yielded significantly longer RTs than all other filters, but no other differences among the filters were found. Generally, though, when high resolution characters are presented under specular glare, the etched filters produce degradations in readability that are not evident for either the polished, mesh, or coated filters. This degradation is especially true when the etched filter is also combined with low transmission levels. This finding can probably be attributed to the softening of the image that occurs with etching and the loss of character luminance that accompanies the increased filter density. Note that such degradations were not found for the larger, low resolution characters and that none of the filters evaluated were shown to improve readability over the polished-92% baseline.

Figure 22 shows the R x F effect under dark lighting. This interaction was not statistically significant when analyzed through simple-effect F-tests, nor were either of the main effects of Filter and Resolution for this interaction. However, in a practical sense, it can be seen in Figure 22 that the etched filters generally produced poorer readability than did the polished, coated, and mesh filters. None of these filters proved to yield shorter RTs than the polished-92% conditions.

Figure 23 presents the R x F effect under diffuse lighting. Again, this interaction was not significant when analyzed with a simple-effects F-test, nor were the main effects of Filter and Resolution. While no statistical differences existed, it is interesting to note that the same general trend showing reduced readability with the introduction of the etched filters was evident under diffuse lighting as well as under the dark and diffuse conditions. It is also meaningful to note (by comparing Figures 21, 22, and 23) that generally longer RTs were found for the dark condition when compared to those obtained under specular and diffuse lighting.

In summary, reading times were found to be longer under dark lighting than either diffuse or specular for a majority of the filters examined. Also, the larger, low resolution characters served to improve readability under specular lighting when the image was very degraded (i.e., with the gloss25-31% filter). Under moderate, diffuse lighting and the dark condition, no differences among

Table 14

Newman-Keuls Comparisons for Filter at High Resolution under Specular Lighting - Reading Time

Filter	Mean RT (s)		
G65-62%-QW	5.62	A	
MESH-7.4	5.75	A	
POL-92%	5.90	A	
POL-62%	5.91	A	
POL-62%-QW	6.05	A	
MESH-5.7	6.23	A	
POL-31%	6.50	A	
G65-92%	6.80	A	
G45-92%	7.18	A	
G25-92%	7.74	A	
G65-31%	7.79	A	
G45-31%	8.47	A	
G25-31%	11.85		B

Note: Means accompanied by the same letter are not significantly different, $p > .05$.

filters or between resolution levels were evident. Finally, while it was determined that no filters improved readability over the polished-92% baseline filter, RTs for the etched filters were consistently slower than for the polished, mesh, or coated filters. One filter, the gloss25-31% was found to severely degrade performance under nearly all viewing conditions (although only significantly so under specular lighting and at high resolution).

Rating results. The results of the repeated-measures, within-subjects ANOVA for the image quality ratings is shown in Table 15. No third-order interaction was present, and therefore the interpretation of these results will focus on the three significant second-order effects.

Figure 24 presents the L x F interaction and Table 16 lists the results of the corresponding simple-effect F-test for Filter under dark, diffuse, and specular lighting. The effect of Filter was found to be significant at each level of lighting. Thus, it was necessary to perform Newman-Keuls analyses for each, the results of these analyses being presented in Table 17. Under dark lighting, it can be seen that the gloss25-31% filter was rated poorer than all other filters, but was followed closely by the gloss25-92% filter. Both of these gloss 25 filters were found to degrade perceived image quality over the rest of the filters evaluated, but no other differences among these remaining filters were identified. Also, as found in the ST and RT analyses, the general trend in the data indicated degradations due to the etched filters. Of interest, the highest-rated filter proved to be the polished-92% baseline (7.81 of a possible 9.00) by which all other filters were compared.

Under diffuse lighting, Table 17 reveals that the gloss25-31% filter again yielded the lowest ratings, followed by the gloss25-92% filter. As under dark lighting, the gloss25-31% was rated worse than all other filters and, with the exception of the gloss25-92% filter, no other differences were evident. Note that the polished-92% filter was no longer rated higher than all others (as it was in the dark), but was instead rated lower than six of the other filters. While not statistically significant, this relationship is important because each of the six higher-rated filters is of reduced transmission, thereby pointing out the subjects' general preference for enhanced contrast when diffuse glare was present.

Table 15

Analysis of Variance Summary for Image Quality Ratings - Phase I

Source of Variance	df	MS	F	p
<u>Between Subjects</u>				
Subjects (S)	7	8.493		
<u>Within Subject</u>				
Resolution (R)	1	45.231	2.93	0.1306
R x S	7	15.425		
Lighting (L)	2	260.376	28.51	0.0001
L x S	14	9.133		
Filter (F)	12	44.712	23.68	0.0001
F x S	84	1.888		
R x L	2	7.101	14.90	0.0003
R x L x S	14	0.476		
R x F	12	3.783	4.69	0.0001
R x F x S	84	0.806		
L x F	24	7.422	8.12	0.0001
L x F x S	168	0.914		
R x L x F	24	0.361	0.82	0.7091
R x L x F x S	168	0.441		
Total	623			

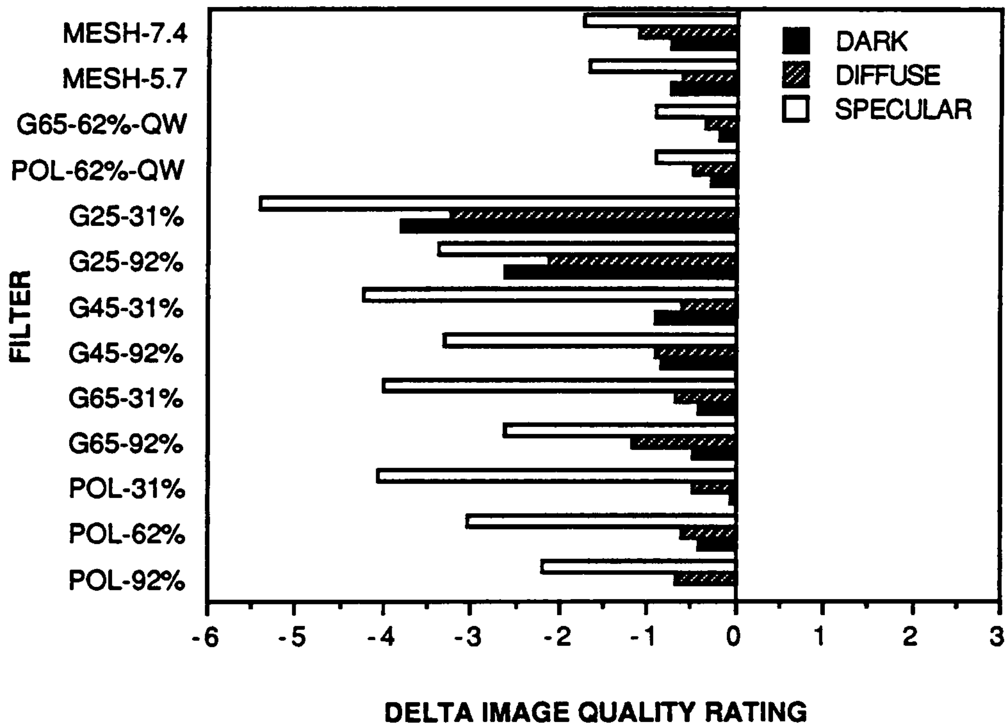


Figure 24. The L x F interaction for delta image quality ratings. All values are scaled to the dark, POL-92% condition (7.81 rating).

Table 16

Simple-effect F-test for Filter at Dark, Diffuse, and Specular Lighting - Image Quality Ratings

Source of Variance	df	MS	F
F at Dark Lighting	12	19.276	10.21 **
F at Diffuse Lighting	12	10.514	5.57 **
F at Specular Lighting	12	29.764	15.76 **
F x S	84	1.888	

** $p < .01$

Table 17

Newman-Keuls Comparisons for Filter under Dark, Diffuse, and Specular Lighting - Image Quality Ratings

Dark Lighting					Diffuse Lighting				
Filter	Mean Rating				Filter	Mean Rating			
POL-92%	7.81	A			G65-62%-QW	7.44	A		
POL-31%	7.75	A			POL-31%	7.31	A		
G65-62%-QW	7.62	A			POL-62%-QW	7.31	A		
POL-62%-QW	7.50	A			POL-62%	7.19	A	B	
POL-62%	7.38	A			G45-31%	7.19	A	B	
G65-31%	7.38	A			MESH-5.7	7.19	A	B	
G65-92%	7.31	A			POL-92%	7.12	A	B	
MESH-5.7	7.06	A			G65-31%	7.12	A	B	
MESH-7.4	7.06	A			G45-92%	6.88	A	B	
G45-92%	6.94	A			MESH-7.4	6.69	A	B	
G45-31%	6.88	A			G65-92%	6.62	A	B	
G25-92%	5.19		B		G25-92%	5.69		B	
G25-31%	4.00			C	G25-31%	4.56			C

Note: Means accompanied by the same letter are not significantly different, $p > .05$.

(continued)

Table 17 (continued)

Specular Lighting

Filter	Mean Rating					
POL-62%-QW	6.88	A				
G65-62%-QW	6.88	A				
MESH-5.7	6.12	A				
MESH-7.4	6.06	A	B			
POL-92%	5.62	A	B	C		
G65-92%	5.19		B	C		
POL-62%	4.75			C	D	
G45-92%	4.50			C	D	
G25-92%	4.44			C	D	
G65-31%	3.81				D	
POL-31%	3.75				D	
G45-31%	3.56				D	
G25-31%	2.38					E

Note: Means accompanied by the same letter are not significantly different, $p > .05$.

It can be seen in Table 17, that under specular lighting the gloss25-31% filter was again rated as yielding the poorest image quality of all filters evaluated. Also, the introduction of specular lighting was found to magnify the differences among the filters to a greater extent than either the dark or diffuse lighting. The four poorest performing filters have a single commonality -- they are all of 31% transmission. The next highest-rated group of filters consists of those with either 92% or 62% transmittance. However, the four filters rated as yielding the best image quality all have either a quarterwave coating or mesh front surface serving as an antireflection treatment. While none of these coated or mesh filters produced ratings that were significantly higher than the polished-92% baseline, the fact that the gloss65-62%-QW and polished-62%-QW filters were rated better than the polished-62% filter reveals that the their high ratings were due to the antireflection treatment and not to reduced transmission. In fact, under specular conditions just the opposite is true. As mentioned during discussion of the ST results, if no antireflection treatment is applied, the 4% reflectance of the front surface produces a constant reflected luminance, but the lower transmission of the filter reduces character luminance (and thus lowers contrast).

Figure 25 shows the R x F interaction for image quality ratings, and Table 18 lists the results of the simple-effect F-test performed on Filter at low and high resolution. These tests revealed that significant differences existed among the filters at both resolutions, and therefore Newman-Keuls analyses were used to isolate these differences. The results of these Newman-Keuls analyses are shown in Table 19. At low resolution, it can be seen that the gloss25-31% filter was rated lower than all other filters and that the gloss25-92% filter was perceived to be of poorer quality than the gloss65-62%-QW filter. No filters, however, were shown to improve perceived image quality over the polished-92% baseline filter. It is notable, though, that the five filters receiving the highest ratings consisted of the coated, mesh, and the polished-92% baseline, and thus were rated higher than all etched filters.

At high resolution, the Newman-Keuls analysis revealed a number of differences among the filters. Again, the gloss25-31% and gloss25-92% filters were rated a good deal lower than nearly all the other filters. Also, while none of the filters was shown to yield better perceived image quality

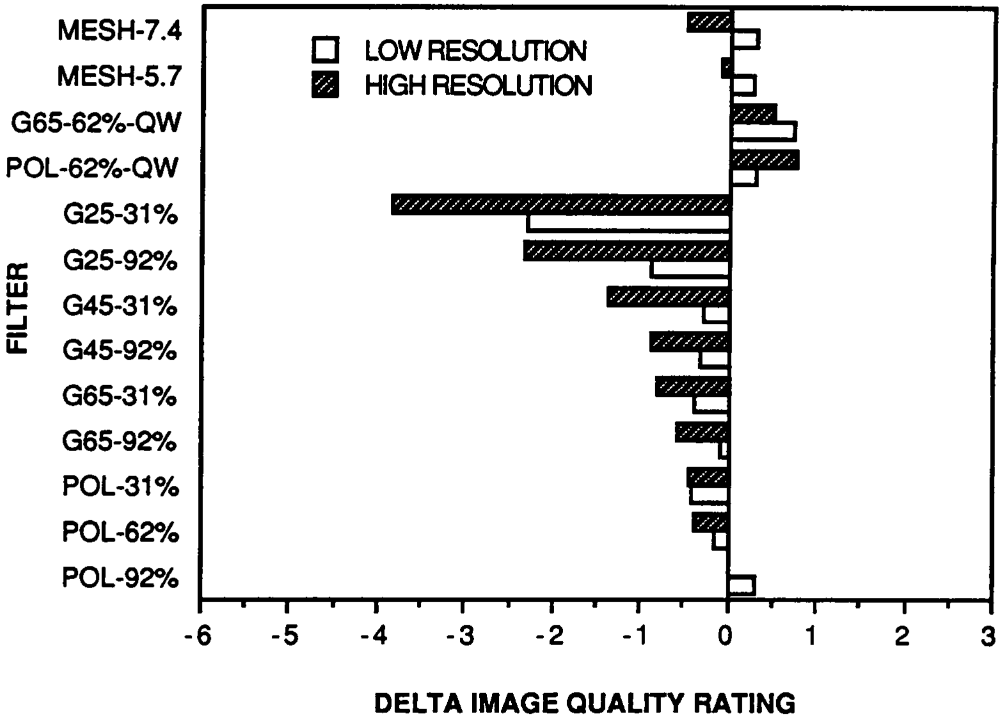


Figure 25. The R x F interaction for delta image quality ratings. All values are scaled to the high resolution, POL-92% condition (6.71 rating).

Table 18

Simple-effect F-test for Filter at Low and High Resolution - Image Quality Ratings

Source of Variance	df	MS	F
F at Low Resolution	12	13.379	7.90 **
F at High Resolution	12	35.115	18.60 **
F x S	84	1.888	

** $p < .01$

Table 19

Newman-Keuls Comparisons for Filter at Low and High Resolution - Image Quality Ratings

Low Resolution Filter	Mean Rating		High Resolution Filter	Mean Rating	
G65-62%-QW	7.42	A	POL-62%-QW	7.46	A
POL-92%	7.00	A B	G65-62%-QW	7.21	A B
POL-62%-QW	7.00	A B	POL-92%	6.71	A B C
MESH-7.4	7.00	A B	MESH-5.7	6.62	A B C
MESH-5.7	6.96	A B	POL-62%	6.33	B C D
G65-92%	6.62	A B	POL-31%	6.25	B C D
POL-62%	6.54	A B	MESH-7.4	6.21	B C D
G45-31%	6.42	A B	G65-92%	6.12	B C D
G45-92%	6.38	A B	G65-31%	5.88	C D
G65-31%	6.33	A B	G45-92%	5.83	C D
POL-31%	6.29	A B	G45-31%	5.33	D E
G25-92%	5.83	B	G25-92%	4.38	E
G25-31%	4.42	C	G25-31%	2.88	F

Note: Means accompanied by the same letter are not significantly different, $p > .05$.

than the baseline, polished-92% filter, it is meaningful to recognize that a clear dichotomy exists between the etched and non-etched filters. That is, the top seven filters are either polished, coated, or mesh and the bottom six filters are all etched in varying degrees. This relationship is immediately evident in Figure 25. In general, it seems that subjects were extremely sensitive to the the differing capabilities of the filters to pass high spatial frequency information, this sensitivity being accentuated as resolution was increased.

The R x L interaction for image quality ratings is represented in Figure 26 and the results of the simple-effect F-test performed on Lighting at low and high resolution are shown in Table 20. Lighting was found to be significant at both resolutions. The results of the Newman-Keuls analyses performed on Lighting at low and high resolution are shown in Table 21. At both resolutions, the image quality under specular glare was rated significantly lower than under either the dark or diffuse lighting. Simple-effect F-tests performed on Resolution at each of the three lighting levels revealed no significant differences. These results indicate that image quality was perceived to be degraded when a specular glare source was present, but that only small differences existed between the dark and diffuse lighting conditions. Evidently, the losses in contrast due to the moderate, diffuse lighting were not extreme enough to appreciably degrade the displayed information. Finally, minimal differences in ratings existed between low and high resolution for the majority of filters evaluated, the exception being those that were harshly etched.

Modeling results

Approach. The goal of the performance modeling was to develop quantitative descriptions of the human performance results using the image quality metrics and SNRs described earlier as the predictors. If successful, such expressions could allow display and filter designers as well as human factors engineers to select appropriate anti-glare filters based on relatively simple, objective display measurements. Clearly, the advantage in such an approach is the ability to make sound design and selection decisions without conducting costly and repetitive human performance studies.

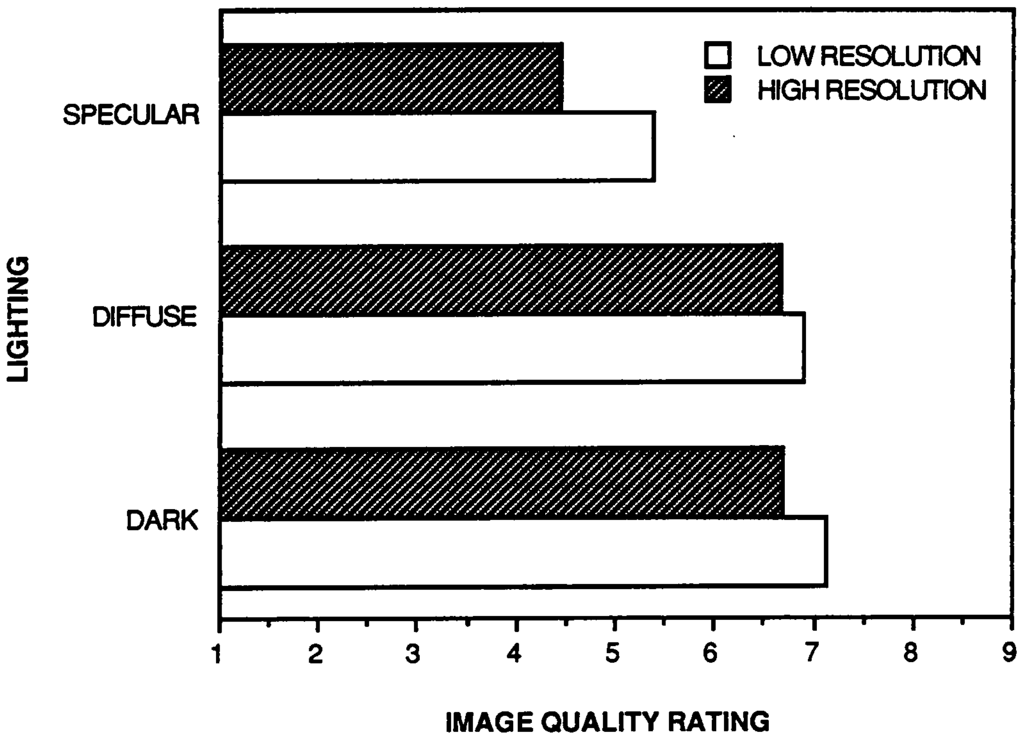


Figure 26. The R x L interaction for image quality ratings.

Table 20

Simple-effect F-test for Lighting at Low and High Resolution - Image Quality Ratings

Source of Variance	df	MS	F
L at Low Resolution	2	92.705	10.15 **
L at High Resolution	2	174.772	19.14 **
L x S	14	9.133	

** $p < .01$

Table 21

Newman-Keuls Comparisons for Lighting at Low and High Resolution - Image Quality Ratings

Low Resolution			High Resolution		
Lighting	Mean Rating		Lighting	Mean Rating	
Diffuse	7.12	A	Dark	6.70	A
Specular	6.91	A	Diffuse	6.67	A
Dark	5.39	B	Specular	4.44	B

Note: Means accompanied by the same letter are not significantly different, $p > .05$.

The approach taken toward model selection combined traditional model-selection techniques with newer, less widely used methods. The details of this approach are discussed in in Myers' (1986) text and the reader is directed to this source for further development. Briefly, the approach consisted of using a number of more contemporary tests during model selection, rather than examining only the coefficient of determination (R^2). In general, a systematic procedure was used to examine all possible regression models given various subsets of the regressor variables, under each of the lighting and resolutions conditions of the study. Mallow's Cp-statistic (a bias estimate), the PRESS-statistic (prediction error estimate), and the mean square error (MSE) or residual estimate of variance of the model were all jointly considered. In addition, the predicted R^2 of the model and an indicator of multicollinearity, the variance inflation factor (VIF), of each parameter were also examined in the selection process. Traditional model selection techniques have focused on the use of R^2; however, it has been shown (Myers, 1986) that while R^2 can be an important statistic to consider, it can be unduly influenced by the number of regressor variables and the range of the regressors. As mentioned above, for each combination of lighting and resolution, models of all possible linear combinations of a subset of regressor variables (image metrics) were examined for each of the three dependent variables. By subset, it is meant that only theoretically meaningful combinations of metrics were included. For example, only one signal metric was included in any of the models, and noise metrics that were believed to measure the same filter characteristics were never used simultaneously (such as RTF Area-Gloss or WS Area-VWS Area). Inclusion of these redundant terms would certainly violate assumptions concerning multicollinearity, and would also render the model less useful in a theoretical sense.

Subsequent sections concerned with discussion of the regression results will focus on the values obtained for PRESS, Cp, MSE, and R^2 and will not report the individual statistical significance of the equations and parameter estimates. Suffice it to say that all equations were significant with probabilities less than or equal to 0.05 except for model ST = f(MTFA, RTF Area) and all parameter estimates were significant with probabilities less than or equal to 0.25. Also, all VIFs were less than 1.95 and thus need not be considered further. Predicted R^2 values behaved as expected relative to the computed R^2 values. For the reading time data, minimum deflation relative to R^2

was about 12% for the Low Resolution-All Lighting model and maximum deflation was about 39% for the Low Resolution-Diffuse Lighting model. For the search time data, minimum deflation of about 10% was found for the High Resolution-Diffuse Lighting model, while maximum deflation occurred for the Low Resolution-Diffuse Lighting model where predicted R^2 values approached zero. For the perceived image quality data, a minimum deflation of about 4% was found for the High Resolution-Dark Lighting model and a maximum deflation of about 48% occurred for the Low Resolution-Diffuse Lighting model. However, while the deflation range for the rating models was comparable to those of the reading and search time models, it is important to note that six of the eight rating models yielded deflations in R^2 of less than 15% (far superior to the reading and search time data).

Reading time models. Table 22 shows the most optimal fits to the reading time data using MTFA, GMTFA, LOG MTFA, and LOG GMTFA as signal metrics and RTF Area, Gloss, WS Area, and VWSA as noise metrics. Large values of PRESS, Cp, and MSE indicate poorer prediction and fits, while larger values of R^2 indicate greater proportions of variance accounted for by the model. It is clear from Table 22 that the modeling was only marginally successful for the majority of lighting and resolution conditions. In theory, all signal measures should carry with them coefficients with negative signs and noise measure coefficients should be positively signed when applied in models of time data. Deviations from this convention could indicate multicollinearity and instability among the regressors (Myers, 1986) or that the theoretical assumptions underlying the regressors are inaccurate under certain sets of conditions. Figure 27 presents the RT data under the dark, diffuse, and specular lighting conditions as a function of GMTFA for both low and high resolution and is included to give the reader an idea of the trends and variability evident in the data. Figures representing the relationships between GMTFA and the dependent variables of ST and image quality rating will be presented in later sections. Any of the signal metrics could have been used as a regressor, but GMTFA was chosen because it yielded a distinct value for each RT, ST, or rating data point. MTF Area or MTFA (because they were only measured in the dark and do not consider lighting in their calculation) would not have shown the reaction of the dependent variables to lighting changes as clearly as GMTFA (or GMTF Area). Finally, the logarithmic

Table 22

Best-fit Regression Models for Reading Time - Phase I

Low Resolution - All Lighting

RT = 6.0930 + .0845 GMTFA - 1.3302 RTF Area - 2.665 VWSA

PRESS = 4.09 $R^2 = 0.56$ $C_p = 4.03$ MSE = 0.0808

High Resolution - All Lighting

RT = 13.3280 - 3.1877 LOG MTFA - 0.3361 WS Area

PRESS = 32.36 $R^2 = 0.45$ $C_p = 1.62$ MSE = 0.5632

Low Resolution - Dark Lighting

RT = 6.5060 - 1.1448 RTF Area - 2.4316 VWSA

PRESS = 0.64 $R^2 = 0.75$ $C_p = 1.24$ MSE = 0.0271

High Resolution - Dark Lighting

RT = 9.8609 - 1.4376 LOG MTFA or LOG GMTFA - 0.3297 WS Area

PRESS = 2.21 $R^2 = 0.54$ $C_p = 0.92$ MSE = 0.1189

Low Resolution - Diffuse Lighting

RT = 7.5062 - 0.3195 MTFA - 0.7936 RTF Area - 2.2122 VWSA

PRESS = 1.38 $R^2 = 0.57$ $C_p = 6.47$ MSE = 0.0764

High Resolution - Diffuse Lighting

RT = 9.5716 - 1.5638 LOG MTFA - 0.3377 WS Area

PRESS = 2.10 $R^2 = 0.58$ $C_p = 4.19$ MSE = 0.1182

Low Resolution - Specular Lighting

RT = 6.7139 - 0.5665 GMTFA - 1.0376 RTF Area - 2.0867 VWSA

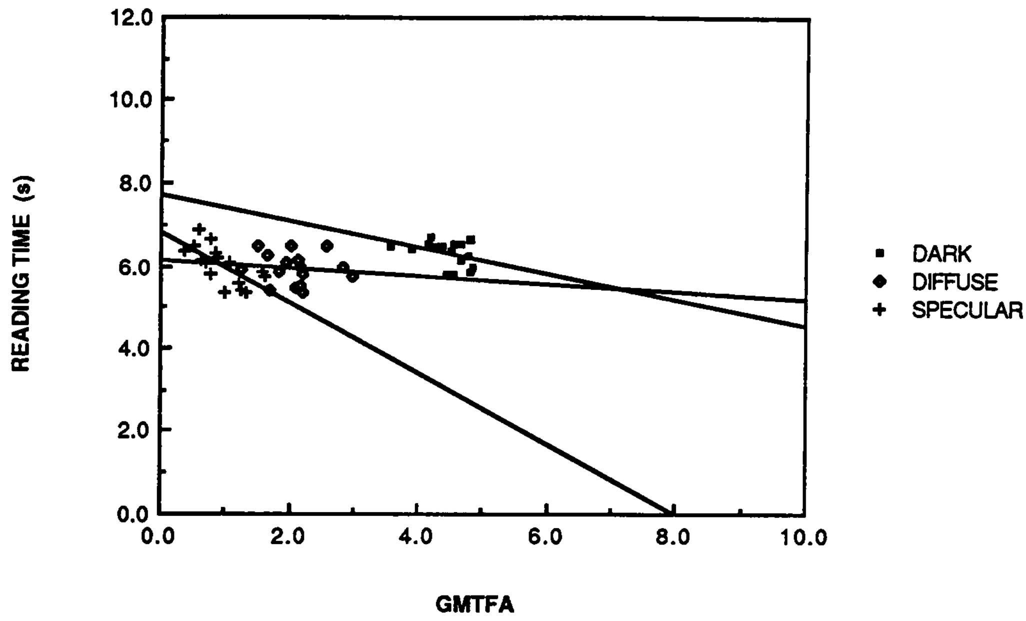
PRESS = 1.81 $R^2 = 0.67$ $C_p = 3.22$ MSE = 0.0885

High Resolution - Specular Lighting

RT = 7.7765 - 2.0800 LOG GMTFA

PRESS = 13.48 $R^2 = 0.80$ $C_p = 1.31$ MSE = 0.4928

a)



b)

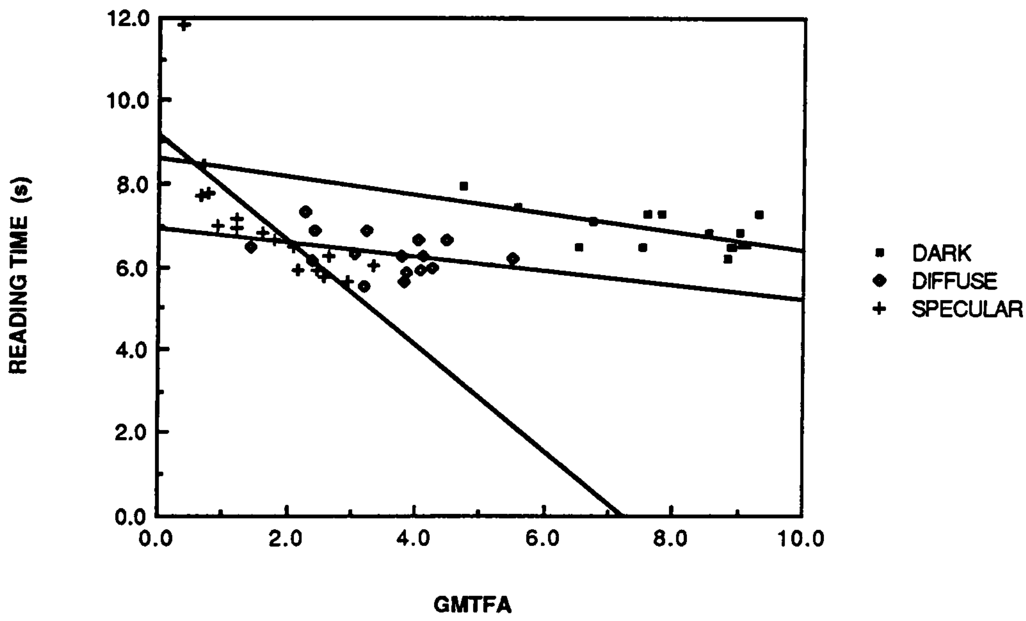


Figure 27. Reading time as a function of GMTFA for dark, diffuse, and specular lighting for a) low resolution and b) high resolution. Regression lines are fit to 16 data points and some points may be hidden.

transforms of the MTF-based metrics were avoided due to their non-linear representation of the relationship.

In general, the RT data seem to be best described by models including a signal and at least one of the noise metrics. In most instances, RT is shown to vary inversely with the MTF-based signal metrics, but the same is often found for the noise terms as well. No truly consistent trends exist among the eight models presented, and even the model yielding the highest R^2 (high resolution - specular lighting) proved to have the one of the poorest predictive values (PRESS = 13.48) and high residual variance (MSE = .4928). Explanation for these poor fits can best be seen in Figure 27. Here, it is clear that for both low and high resolution, even with large changes in image quality (GMTFA), RT was found to remain quite constant. This was especially true when the larger, low resolution characters were used. Additionally, within each of the separate lighting conditions, the variation about the simple regression lines is quite large. Referred to earlier in the ANOVA results, but presented more clearly in Figure 27 is the poorer readability as a result of dark lighting; this occurring even though the highest measured image quality was found in the dark.

Search time models. The best-fit models under the various resolution and lighting conditions for the search time data are shown in Table 23. Again, legibility seems to best be modeled by one of the MTF-based signal terms combined with one or more of the noise metrics. However, the signs of the noise coefficients were not consistent from one lighting condition to the next, this pointing out the instability in the models. Further, most models yielded high PRESS-statistics as well as MSE values, indicating poor predictive capability and high residual variance, respectively.

As in the RT modeling results, inspection of Figure 28 lends insight into why the attempts at modeling the ST data met with only limited success. At high resolution, the data at each lighting level did contain some linear component; however, the variation about the regression line was large when compared to any systematic change in ST. This effect is magnified when the low resolution data are considered. Here, the data were tightly packed and search time was extremely consistent for all three lighting conditions. These results indicate, in a most general sense, that the search time data vary inversely with the MTF-based signal metrics. Again, it appears that legibility was main-

Table 23

Best-fit Regression Models for Search Time - Phase I

Low Resolution - All Lighting

ST = 3.2400 + 0.1635 GMTFA - 0.0057 GLOSS - 1.5316 VWSA

PRESS = 8.15 $R^2 = 0.34$ $C_p = 3.42$ MSE = 0.1524

High Resolution - All Lighting

ST = 6.2356 - 0.2176 GMTFA - 0.0290 GLOSS - 8.5903 VWSA

PRESS = 59.75 $R^2 = 0.39$ $C_p = 3.65$ MSE = 1.1754

Low Resolution - Dark Lighting

ST = 5.1102 - 0.2599 MTFA or GMTFA - 0.7835 RTF Area

PRESS = 2.45 $R^2 = 0.22$ $C_p = 15.88$ MSE = 0.1553

High Resolution - Dark Lighting

ST = 8.2640 - 0.0284 GLOSS - 1.7568 WS Area

PRESS = 12.32 $R^2 = 0.51$ $C_p = 0.65$ MSE = 0.6159

Low Resolution - Diffuse Lighting

ST = 6.8407 - 2.4713 LOG MTFA

PRESS = 3.02 $R^2 = 0.24$ $C_p = 0.61$ MSE = 0.1487

High Resolution - Diffuse Lighting

ST = 12.8185 - 4.0283 LOG MTFA

PRESS = 5.03 $R^2 = 0.70$ $C_p = 0.83$ MSE = 0.2887

Low Resolution - Specular Lighting

ST = 3.0034 - 0.2552 LOG GMTFA + 0.7834 RTF Area - 1.1672 VWSA

PRESS = 0.82 $R^2 = 0.52$ $C_p = 2.04$ MSE = 0.0460

High Resolution - Specular Lighting

ST = 15.9658 - 5.5072 LOG MTFA + 0.0092 GLOSS

PRESS = 7.11 $R^2 = 0.78$ $C_p = 0.96$ MSE = 0.3454

---

a)

b)

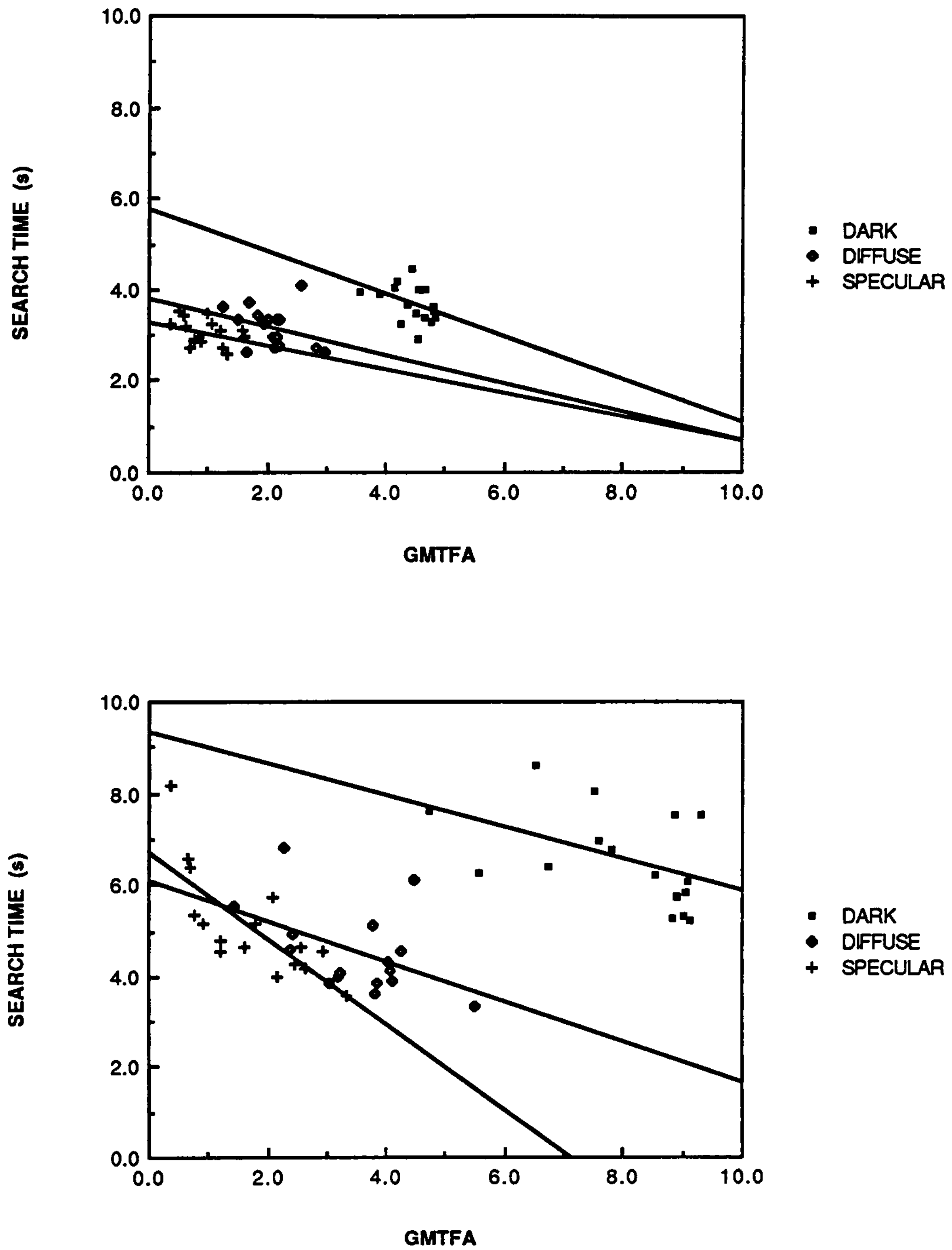


Figure 28. Search time as a function of GMTFA for dark, diffuse, and specular lighting for a) low resolution and b) high resolution. Regression lines are fit to 16 data points and some points may be hidden.

tained at a fairly constant level, regardless of the changes in image quality at the display. This finding was especially true at low resolution. Also interesting to note is the reduced legibility (i.e., longer search times) for the data collected under dark conditions. Certainly, it is not feasible to model the entire data set (all lighting conditions at once) with a monotonic function when degradations in performance occur at the highest measured image quality. Nonmonotonic functions applied to the data would not be meaningful in either a theoretic or applied sense.

Rating models. The models yielding the best fits to the image quality rating data are presented in Table 24 for all resolution and lighting conditions. Figure 29 shows these data plotted as a function of GMTFA. Examination of both the table and the figure indicates that some rather extreme differences existed between modeling of objective measures of human performance (RT and ST) and modeling of perceived image quality. Notable are the consistent and systematically linear trends found in Figure 29 for both the low and high resolution data, as well as the generally excellent fit of the models listed.

It is evident from Table 24 that attempts to model the data under all lighting conditions combined were minimally successful and the high values obtained for the PRESS statistic point out that use of these models in a predictive fashion is probably inappropriate. However, under the individual lighting conditions, combinations of a single signal metric (the log transform in all cases) and one or more noise measures produced high-quality fits to the perceived image quality data regardless of resolution level. In contrast to the time data, the signal metric should now carry with it a positive sign and the noise metric should be inversely related to image quality ratings. This relationship was maintained for six of the eight models, and in all cases the ratings varied positively with the signal metric. In addition to the high R^2 and low MSE values, in most instances, the PRESS-statistic indicated the models to be of good predictive quality. The goodness of fit to the data is further emphasized by values of R^2 exceeding 0.90 in two instances, indicating the proportion of variance in perceived image quality that was accounted for by the model approached unity.

Signal-to-noise ratio models. Best-fit, single-term SNR regression models were found for the same lighting and resolution conditions described in the preceding sections. It was hoped that by forming an SNR with a single signal metric and one or more of the noise metrics, that better fits

Table 24

Best-fit Regression Models for Image Quality Ratings - Phase I

Low Resolution - All Lighting

RATE = 5.7255+ 1.1410 LOG GMTFA

PRESS = 27.18 $R^2 = 0.56$ $C_p = 2.52$ MSE = 0.5412

High Resolution - All Lighting

RATE = 3.9661 + 1.6450 LOG GMTFA

PRESS = 45.33 $R^2 = 0.67$ $C_p = 0.34$ MSE = 0.9274

Low Resolution - Dark Lighting

RATE = -4.7084 + 7.9600 LOG MTFA

PRESS = 2.89 $R^2 = 0.79$ $C_p = 0.16$ MSE = 0.1304

High Resolution - Dark Lighting

RATE = -7.0134 + 6.8087 LOG MTFA - 0.7686 RTF Area - 3.4759 VWSA

PRESS = 2.35 $R^2 = 0.95$ $C_p = 3.34$ MSE = 0.0951

Low Resolution - Diffuse Lighting

RATE = -1.9847 + 6.0521 LOG MTFA - 0.8516 RTF Area

PRESS = 3.36 $R^2 = 0.66$ $C_p = 0.04$ MSE = 0.1297

High Resolution - Diffuse Lighting

RATE = -2.1687+ 4.3546 LOG MTFA - 0.4165 WS Area

PRESS = 3.76 $R^2 = 0.83$ $C_p = 0.77$ MSE = 0.1725

Low Resolution - Specular Lighting

RATE = 6.0482 + 2.3487 LOG MTFA - 3.1911 RTF Area - 0.0184 WS Area

PRESS = 9.07 $R^2 = 0.75$ $C_p = 2.00$ MSE = 0.3624

High Resolution - Specular Lighting

RATE = 4.9195 + 2.3132 LOG GMTFA - 0.0254 GLOSS - 0.7926 WS Area

PRESS = 4.27 $R^2 = 0.92$ $C_p = 4.11$ MSE = 0.2067

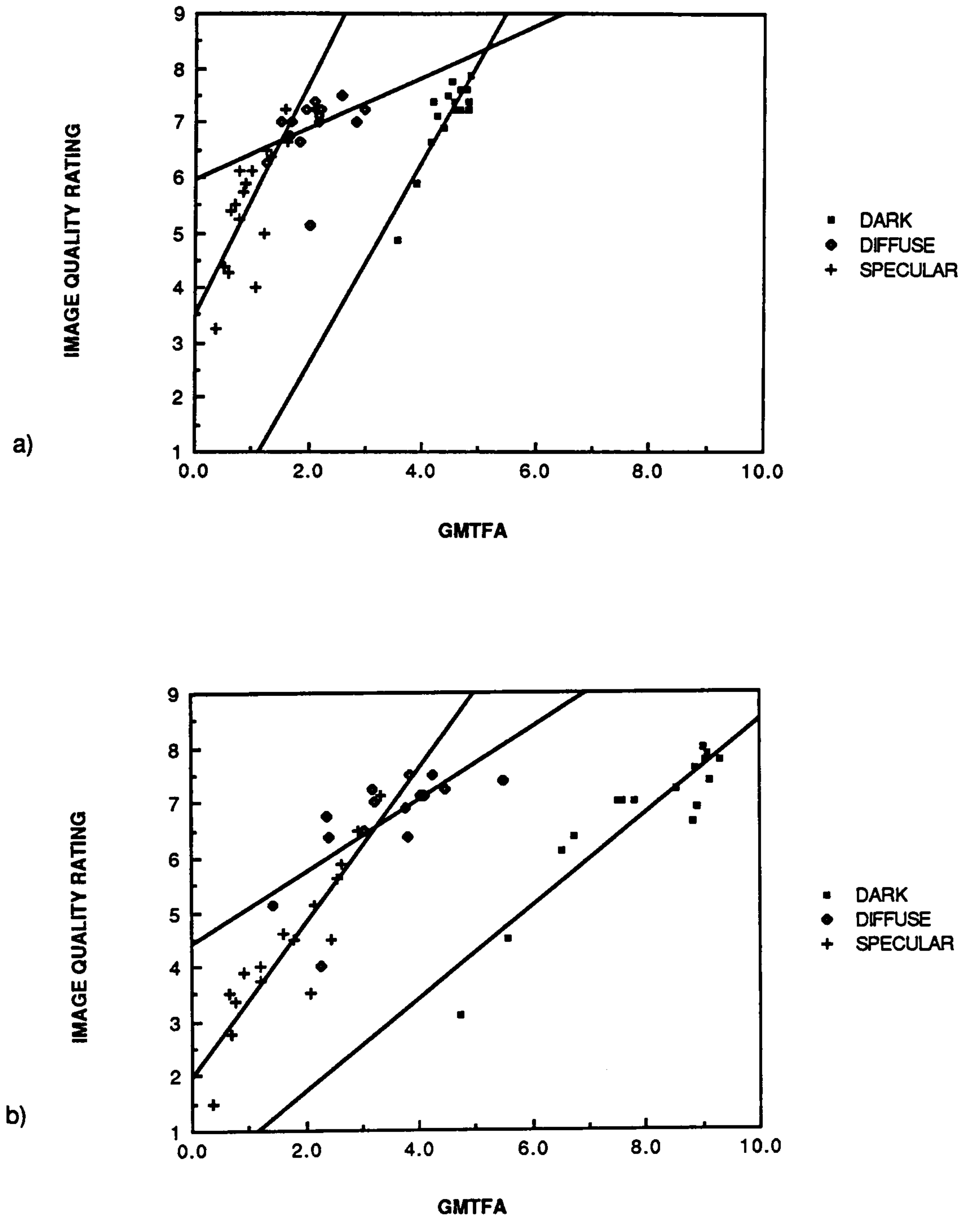


Figure 29. Image quality ratings as a function of GMTFA for dark, diffuse, and specular lighting for a) low resolution and b) high resolution. Regression lines are fit to 16 data points and some points may be hidden.

would be found to the RT, ST, and rating data sets. While such an approach is theoretically appealing, it did not yield higher quality description or prediction than those models made up of additive combinations of the signal and noise ratios. In fact, none of the best-fit SNR models provided adequate fits to the RT, ST, or rating data. Further, the SNR term always carried with a sign opposite to that which was theoretically or logically expected (i.e., negative for the rating data and positive for the timed measures). Clearly, it would not be meaningful to propose a predictive expression indicating that as the ratio of signal-to-noise in the image increases, that both subjective and objective performance is degraded.

On a final note, the reader should recognize that it was not meaningful to attempt to model the combined low and high resolution data sets due to the confound that existed with varying character sizes. That is, while the high resolution condition produced higher values for the MTF-based metrics and would therefore be considered of correspondingly higher image quality, the small character size (about 16.1 arcminutes) yielded reduced readability and legibility. Conversely, the low resolution condition (32.2 arcminute characters) yielded lower measured image quality, but enhanced readability and legibility. Regressions on the combined low and high resolution data produced poor quality models with virtually no theoretical or practical significance.

CONCLUSIONS - PHASE I

A number of important conclusions can be drawn from the previous discussion of the image quality measurement, human performance, and modeling results. In addition, the finding that dark illumination yielded slower search and reading times than did either the diffuse or specular glare conditions deserves further attention. Therefore, this section summarizes the results obtained in Phase I of this research and, where necessary, offers possible explanations for findings that were unexpected and not included in the original focus of this work.

Anti-glare Treatments

One of the primary goals of this research was to identify the effects, if any, that various anti-glare treatments have on human performance. While the image quality measurements collected in the present investigation supported previous work that found rather large differences among the filters, there had been little conclusive research on whether these measurements related directly to readability, legibility, or subjective differences. Certainly, the present study has pointed out that degradation in timed performance and perceived image quality can be attributed to changes in the anti-glare treatment applied to the CRT. While no treatments were found to improve performance over a clear, polished glass filter that served as a baseline for comparison, a number of treatments were identified that severely degraded performance.

Generally speaking, etches with gloss values of 45 or lower should be not be used on monochrome CRTs, regardless of the lighting conditions the display will be operated in or the resolution of the unit. It is especially critical, however, that etched filters not be used on displays considered to be of high resolution. Under high resolution and specular lighting, large decrements in RT, ST, and perceived image quality were found when filters with etched front surfaces were used. It would be safe to assume that as the resolution of the display is increased, the less tolerant it will become to the loss of high spatial frequency information that accompanies the etching process.

Equally notable to the above-mentioned degradations in performance was the lack of improvement in readability and legibility when the anti-glare treatments were presented under the diffuse and specular glare conditions. Here, none of the filters was found to produce shorter reading or search times than a clear, polished glass surface that served as a control condition. When the same filters were evaluated for their effects on perceived image quality, the differences among the filters were more well-behaved and predictable, but only under specular lighting did any of the filters (the coated and mesh) meaningfully improve ratings over the baseline. Even then, the differences were not as extreme as might have been expected. The rating results also indicated (though not statistically significant in most cases) clear preferences for the polished, coated, and mesh filters over those with etched surfaces. It could be concluded, then, that there is little benefit to anti-glare treatments when the display is to be operated in an environment similar to that of the typical office (i.e., moderate diffuse and specular glare levels). Evidently, the value of these treatments appears limited primarily to providing enhanced perceived image quality under specular glare conditions. Certainly, the importance of users' perceptions of the device cannot be ignored from a user-satisfaction viewpoint (Stammerjohn et al., 1981) or in a related sense, in terms of marketing considerations. Also, it should be recognized that this research investigated the filters' effects on performance over extremely limited lengths of time. For each of the tasks in the present research, participants were exposed to each filter only for short periods, the reading, search, and rating tasks normally requiring under 10.0 s to complete. The effect on performance of these anti-glare treatments over longer periods of time is not known and future, similarly well-controlled research should investigate this issue.

Prediction Models

Another of the goals of this research was to determine if linear regression equations could be developed that would offer good fits to the RT, ST, and rating data, using only the image quality metrics as predictive terms. Quantifying such relationships would make available to the display and human factors engineers a powerful tool that could be used to predict subsequent human performance based on relatively simple display measurements.

The regression analyses investigated all theoretically meaningful combinations of the various signal and noise metric regressor variables. Using modern model selection techniques, models yielding the best-fit and predictive capabilities were identified for each of the lighting and resolution conditions studied. However, the results of these analyses indicated that for the reading and search time data, the poor fits found for most of the models greatly limits their use as descriptive equations. The use of these models to predict performance in other environments is certainly not an option.

It should be noted, that when classical model selection techniques such as forward selection, backward elimination, and stepwise regression (Myers, 1986) were used to develop models for the RT and ST data, R^2s with values ranging from .53 to .74 were produced. However, the models arrived at through these methods contained a greater number of highly nonorthogonal regressor terms than did those models selected using more modern techniques. In terms of the number of predictors and the R^2 values produced, these models were similar to those reported in previous research investigating the relationship between image quality and human performance (see Snyder, 1985). It appears that the primary differences between the modeling results of this study and those reported in the past can be explained by the more stringent selection constraints used in the present work.

A number of more specific explanations can be offered for the comparatively poor fits and lack of predictive capability found with the models for the RT and ST data sets. First, with the exception of those data recorded at high resolution and for harshly etched filters, the dependent measures did not exhibit a large range of variation. Without some significant variation in the dependent variable, the slopes of the regression lines are small and little of the total variation is due to the regression itself. Further magnifying this lack of regression effect, both the RT and ST data exhibited large residual variation about the regression lines. The combination of small regression effects and relatively high variability rendered the majority of the models of little practical value. Also, this research attempted to model performance differences that were produced under more widely divergent display and environmental conditions than those found in previous research (Snyder, 1985). It is likely, given the poor modeling performance, that numerous display-dependent variables exist that have not yet been identified, but need to be included into models of human performance under these conditions.

While the timed dependent measures were notable for their inability to be modeled accurately and consistently, the regression fits to the perceived image quality data were of exceptionally high quality in both a descriptive and predictive sense. The most accurate fits to the data were found when the lighting and resolution conditions (i.e., specular lighting - high resolution) were evaluated individually, but acceptable models were also produced when the data were examined under all three lighting levels simultaneously. Generally, it was determined that the data varied positively with one of the MTF-based image quality signal metrics (or log transform thereof) and inversely with the various noise metrics included in the models.

The success in modeling the perceived image quality data occurred due to a number of reasons. First, as opposed to the RT and ST data, variations in measured image quality at the display produced equally large variations in participants' ratings of these changes. That is, the subjects maintained fairly constant reading and search times for the majority of conditions presented to them, but were found to be quite sensitive to perceived image quality over the same display and environmental conditions. Second, while the variability in the rating data explained by inclusion of the image quality metrics as regressor variables was large, the residual error variation about the regression line was small for most of the models. It can be concluded, then, that while the modeling of the timed performance data was marginally successful at best, the models developed to describe the rating data were of sufficient quality to be used by other investigators as predictive devices.

A final comment should be made about the performance modeling results. It is possible that had the ranges of specular and diffuse illuminance been increased beyond those normally found in the office environment, that a more easily modeled situation would have resulted. Here, larger decrements in reading and search times would have undoubtedly been observed, and the fit to the data would have been made easier. However, such was not the goal of this research and the conclusions, therefore, must be restricted to office-type environments due to the limits placed on the lighting extremes.

The Low Illumination Issue

Finally, the finding that search and reading times were generally slower in the dark than under either the diffuse or specular glare conditions run so contrary to the expected performance that it must be discussed in greater detail. The recently accepted ANSI specification (ANSI, 1988) concerning the human factors issues involved in VDT workplaces indicates that no scientific reasons exist for imposing restrictions as to the maximum allowable display contrast or minimum background luminance levels. However, it has been the contention of numerous European researchers (e.g., Cakir, et al., 1980) that acuity is degraded when background luminance levels fall below 2 to 6 cd/m^2. These same researchers indicate that optimal relative acuity is found when reflected luminance from the phosphor of the CRT reaches 15 to 23 cd/m^2. While these results appear counterintuitive, enough is known about basic visual processes to offer a few plausible explanations for these findings.

First, a well-known phenomenon called the autokinetic effect can occur when luminous point sources are imposed on a very dark background (Kaufman, 1974). The resultant visual illusion is that the sources appear to drift across the visual field of the observer although their physical position remains entirely constant. The cause of this illusion is under debate, but some research has indicated the gradual, involuntary saccadic movements of the eye are an important contributor (Kaufman, 1974). While the autokinetic effect may not apply in a strict and classical sense to the present research (in that the stimulus is not a point source), a similar process may be at work. One explanation, then, for the longer timed performance in the dark is that the subjects must continually adjust to the perceptually drifting image on the display.

Another possibility is that a proper accommodative stimulus is not available to the observer when the luminance of the background approaches zero. The accommodative system, then, is in a state of continual flux in search of the optimal focus (Campbell and Durden, 1983; Moses, 1987). The issue has been raised that the Gaussian luminance distribution of the CRT spot already offers a poor stimulus for accommodation, and any further blurring of this spot due to filter etch would only serve to accentuate this effect. When the background is illuminated by some ambient source, the phosphor surface and faceplate of the CRT both offer some detailed cues for the visual system as to the distance of the displayed information from the eye. In the dark, however, the only cues available to the observer are those produced by the already Gaussian-shaped "fuzzy" spots making up the displayed information.

A final explanation for the degraded readability and legibility in the dark is the simple pupillary response of the eye to changes in retinal illumination. With reductions in the incident illuminance at the eye, the diameter of the pupil can change from a minimum of about 2.0 mm up to a maximum of about 7.0 mm. With this dilation of the pupil come reductions in visual acuity due to the resultant losses in visual depth of focus (Moses, 1987). It is possible, then, that the performance losses are due more to the reduction in total illuminance at the eye than due to the near zero luminance of the background. The issue could be easily investigated by through the use of varying diameters of artificial pupils that would serve to adjust the depth of focus, without altering the zero background luminance and contrast of the display.

METHODS - PHASE II

Phase II of this study investigated the interactive and singular effects that filter transmission and diffuse illuminance have on readability, legibility, and perceived image quality. Quantitative regression models were developed that incorporated measures of displayed image quality such as luminance modulation, contrast ratio, GMTFA, and GMTF Area to describe the human performance differences. Additionally, the levels of illuminance and filter transmission were chosen to complement and help understand the results obtained from Phase I regarding the effects of low illuminance and contrast enhancement on performance.

Measurement Apparatus and Procedure

Contrast enhancement filters. Five contrast enhancement filters were evaluated in terms of their effect on image quality during Phase II of this research. The filters ranged in neutral-density transmission from 11% to 92% (11, 31, 58, 75, and 92). All were polished and no anti-reflection treatments had been applied. To cover the wide range of transmission, and yet do so at roughly equal intervals, the 11% and 58% filters were composed of two filters permanently bonded together with optical cement (31%-31% and 92%-62%, respectively). This was necessary because the glass used for CRT filters is only available at 31%, 62%, 75%, and 92% transmissions.

Imaging system. The imaging system used for Phase II was identical to that of Phase I and consisted of a 50-cm diagonal, monochrome P4 phosphor CRT, a video signal generator, and an IBM-PC. The monitor was again run in a 60-Hz non-interlaced mode and at an addressability of 1024 x 1024 pixels. Unlike the Phase I research, however, Phase II evaluated the CRT only at the high resolution setting (i.e., optimal focus, minimum spot size, and single pixel lines). Again, all filters were attached to the CRT via the mounting apparatus and the resultant cavity between the filter and CRT faceplate was filled with glycerin.

Measurement system. The microphotometric measurement system used in Phase I also served for measurements in Phase II. However, the system was only used for validation purposes, as all image quality metrics were determined analytically rather than empirically. The reasoning behind this analytic approach is simple. Because only filter transmission and ambient illuminance were varied, only the maximum luminance and background luminance of the image differed among conditions (i.e., spot width and shape were constants). These luminance values are easily determined through calculation if trace luminance, phosphor reflectance, filter transmission, and ambient illuminance are known. Had the filters been etched or treated with some other anti-reflection technique, the procedure would not have been straightforward and may not have been feasible at all.

Measures of filter noise were not applicable to the Phase II research. That is, because no specular glare source was presented, the RTF Area and Gloss measurements were inappropriate. Further, since all filters were polished and had no anti-reflection treatment applied, luminance non-uniformities were minimal and measurement of the Wiener spectrum would add little information. Therefore, only the MTF-based metrics were calculated.

Diffuse glare source. A single 122-cm, four-lamp, fluorescent light fixture served as the diffuse glare source for both the image quality and human performance measurements. The lamps were General Electric MaxiMiser, 40-watt with a color temperature of about 4100 deg K and maximum single lamp outputs of 3375 lumens. The standard fluorescent ballasts were replaced with General Electric dimming ballasts (# 8G5007-W) and the output of the lamps was continuously adjustable with a Leviton Manufacturing Company dimmer switch (# 6676-1). At the geometric position relative to the CRT shown in Figure 30, the source was capable of illuminance levels ranging from about 300 to 2800 lx.

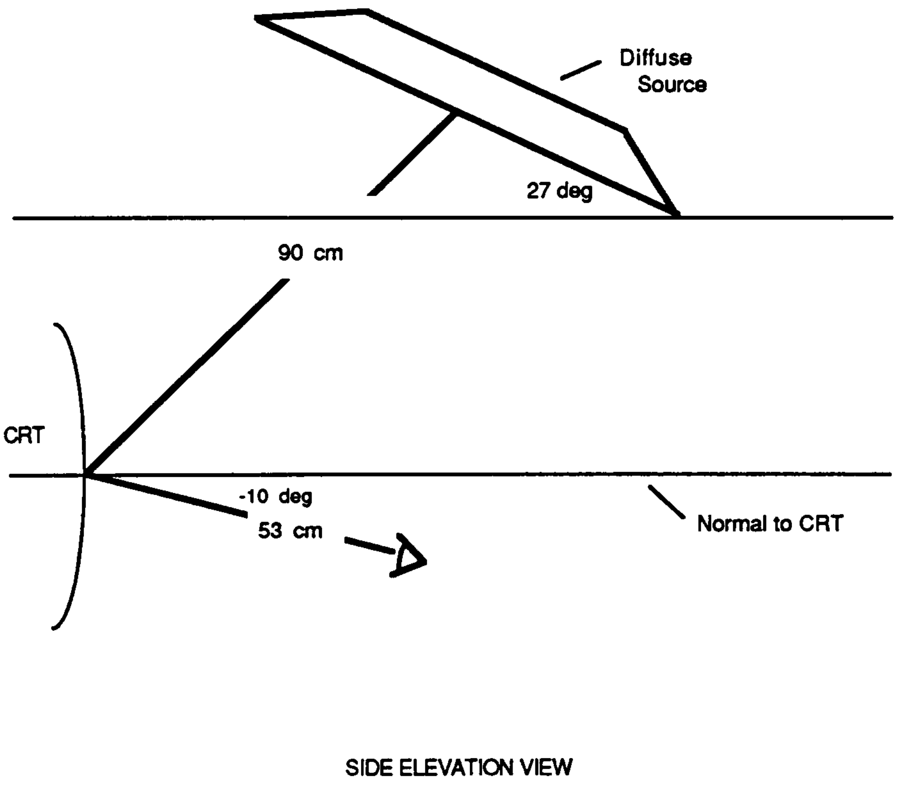


Figure 30. Relationship among the observer, the CRT, and the diffuse glare source for Phase II.

Human Performance Apparatus

The human performance measurements of Phase II used the same adjustable, fluorescent glare fixture, filters, and imaging and display system as described in the preceding discussion of image quality measurements. Subjects were seated and positioned in a location identical to that in Phase I. The glare source was controlled remotely by the experimenter via the dimmer switch, and produced five discrete levels of diffuse illuminance (0, 320, 1200, 2000, and 2800 lx). The levels were chosen to allow evaluation of performance across the range of intensities that might be encountered in the office environment. As mentioned previously, the display was always operated in the high resolution mode and the 7 x 9 Huddleston characters subtended 16.1 arcminutes at the 53-cm viewing distance.

Human Performance Procedure

Subjects. Eight different subjects (four female) participated in the Phase II research and were paid for their time. Screening procedures were identical to those employed in Phase I in terms of vision and language requirements. Subjects ranged in age from 18 to 25 yr with a mean of about 21 yr. The group was composed entirely of persons from the university community and averaged about 4.6 hr per week using computers with CRT displays.

Experimental design. The experimental design used for the human performance study in Phase II is shown in Figure 31. Each of the eight subjects was exposed to all 25 experimental conditions over a three-day period. Order effects for both filter transmission and illuminance level were controlled by random selection of presentation orders from two 5 x 5 Latin squares. Subjects alternated as to whether they performed the reading task or search task first but, as in Phase I, image quality ratings were always completed prior to either of the time measures. Subjects were presented a single filter on their first day of participation and two filters for their second and third days with rest periods given between filters. Also, the procedures for the reading, search, and rating tasks were identical to those described in Phase I of this research and will not be reiterated.

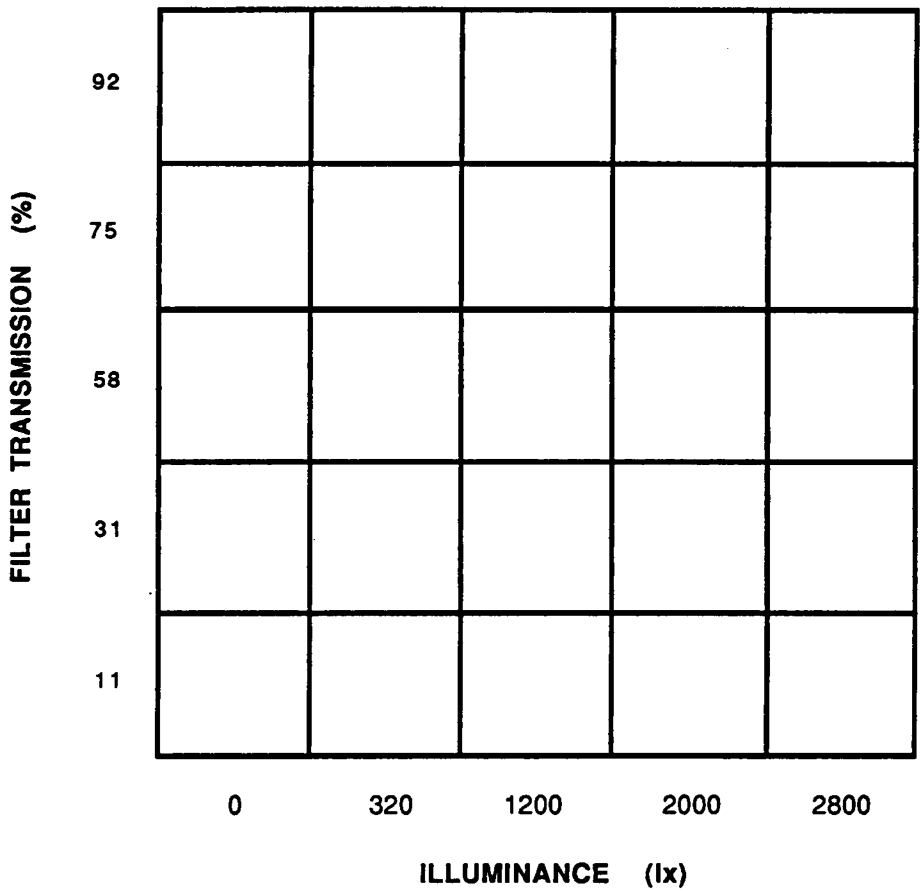


Figure 31. Experimental design for the Phase II human performance evaluation.

RESULTS AND DISCUSSION - PHASE II

Image Quality Measurements

As no etching, mesh filters, or anti-reflection coatings were used in Phase II, and the monitor was operated only in the high resolution mode, differences in the image quality metrics were due solely to the changes in filter transmission and diffuse illumination. As in Phase I, the area of the MTF of the display system that falls below the CTF of the "average" human observer was essentially a constant for all MTFA and GMTFA calculations. Therefore, the difference between a metric that considers only the display (GMTFA Area) and a metric that accounts for the human observer (GMTFA) is primarily due to scaling from cycles/mm to cycles/deg, respectively. As the metrics are nearly proportional to one another, to avoid redundancy, only the GMTFA results will be presented in this discussion.

Figure 32 shows the results of the GMTFA measurements for each of the 25 (5 Transmission x 5 Illuminance) experimental conditions. The values of GMTF Area, GMTFA, luminance modulation, and contrast ratio are listed in Table 25. It is clear from Figure 32 that illuminance had a strong inverse effect on GMTFA. That is, as the illuminance increased from 0 to 2800 lx, GMTFA decreased for each of the five transmission levels. This decrease, of course, indicates a degradation in measured image quality and was due to the loss of luminance modulation across all spatial frequencies of the MTF. Similar losses in GMTFA were shown for all but the 0 lx condition when filter transmission was increased. Again, this reduction in image quality was due to the lower modulations found as the filter allowed more ambient light to reach the phosphor surface of the

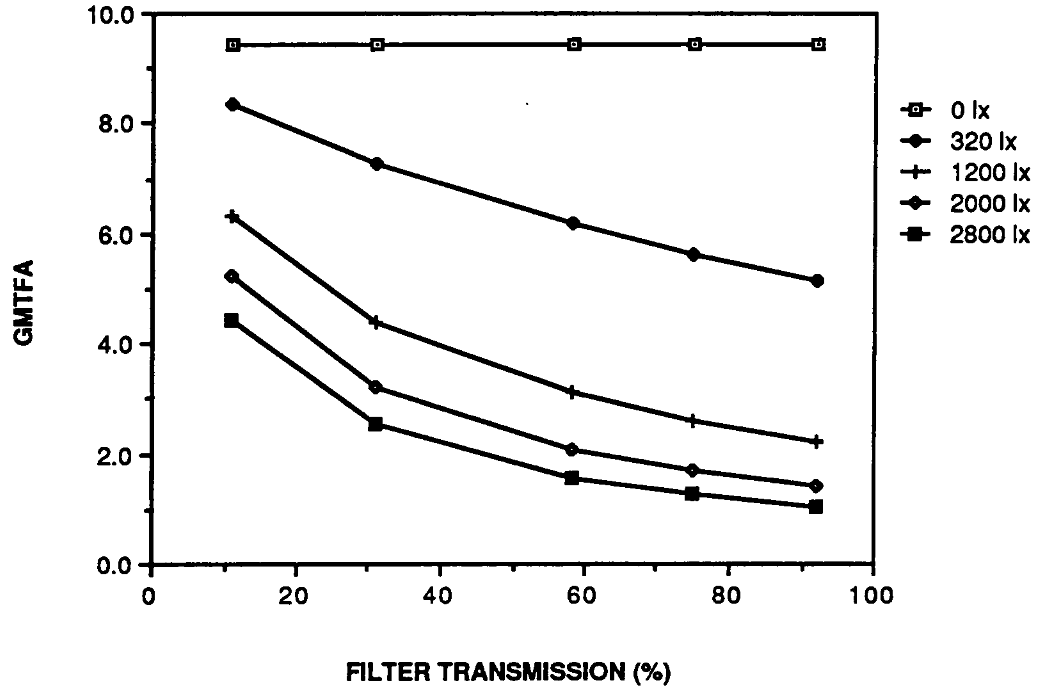


Figure 32. GMTFA as a function of filter transmission and illuminance.

Table 25

Results of Image Quality Measurements - Phase II

Illum.	Trans.	MOD	CR	GMTF Area	GMTFA	LOG GMTFA
<u>0 lx</u>						
	11%	0.97	63.0:1	1.2609	9.4174	2.2426
	31%	0.97	74.0:1	1.2615	9.4222	2.2431
	58%	0.98	97.3:1	1.2624	9.4288	2.2438
	75%	0.98	83.5:1	1.2629	9.4329	2.2442
	92%	0.98	81.2:1	1.2634	9.4370	2.2446
<u>320 lx</u>						
	11%	0.86	13.4:1	1.1211	8.3394	2.1210
	31%	0.76	7.3:1	0.9851	7.2761	1.9846
	58%	0.65	4.8:1	0.8427	6.1808	1.8214
	75%	0.60	3.9:1	0.7703	5.6303	1.7282
	92%	0.55	3.4:1	0.7099	5.1607	1.6411
<u>1200 lx</u>						
	11%	0.66	4.9:1	0.8591	6.3082	1.8418
	31%	0.47	2.8:1	0.6097	4.3945	1.4804
	58%	0.34	2.0:1	0.4385	3.0926	1.1290
	75%	0.29	1.8:1	0.3717	2.5875	0.9507
	92%	0.25	1.7:1	0.3228	2.2218	0.7983
<u>2000 lx</u>						
	11%	0.55	3.5:1	0.7212	5.2480	1.6578
	31%	0.35	2.1:1	0.4564	3.2301	1.1725
	58%	0.24	1.6:1	0.3050	2.0929	0.7386
	75%	0.20	1.5:1	0.2527	1.7022	0.5319
	92%	0.17	1.4:1	0.2157	1.4281	0.3563
<u>2800 lx</u>						
	11%	0.47	2.8:1	0.6120	4.4126	1.4844
	31%	0.28	1.8:1	0.3633	2.5310	0.9286
	58%	0.18	1.4:1	0.2342	1.5617	0.4458
	75%	0.15	1.3:1	0.1915	1.2514	0.2243
	92%	0.12	1.3:1	0.1619	1.0357	0.0351

display. Note, however, that under 0 lx illuminance, a similar trend was not found. Here, because the background luminance of the display was near zero for all transmission levels, the area under the MTF of the display remained constant across transmission.

In terms of prediction, the above findings indicate that performance should be degraded as ambient illuminance is increased, and that (with the exception of the dark condition) similar degradation should occur as filter transmission is increased. While the first point concerned with loss of contrast due to ambient illuminance is straightforward, the second implies that the trace luminance of the display is not critical, and that contrast alone is responsible for performance differences.

Correlations among measurement results. Table 26 shows the Pearson product-moment correlation coefficients among the image quality measurement of GMTFA, GMTF Area, LOG GMTFA, modulation, and contrast ratio and the factors of Illuminance and filter Transmission. As would be expected, the image quality metrics are all highly intercorrelated, this due to the fact that each was primarily driven by the luminance difference between the character and the background. The high degree of collinearity among these metrics also precludes their simultaneous use in the regression models to be described in subsequent sections.

Human Performance Results

Search results. A 5 x 5 (Transmission x Illuminance) repeated-measures, within-subjects ANOVA was performed on the search time data and the summary of this analysis is shown in Table 27. Figure 33 presents the T x I interaction. Immediately evident are the longer STs under 0 lx illuminance across nearly all transmission levels. As the Transmission x Illuminance (T x I) interaction was statistically significant, simple-effect F-tests were performed on Transmission at each illuminance level and, correspondingly on Illuminance at each transmission level. The results of these analyses are shown in Table 28. It can be seen that Illuminance was significant at 58% and 92% transmission and that Transmission failed to reach statistical significance at any lighting level. The results of the Newman-Keuls analyses performed on Illuminance at both 58% and 92%

Table 26

Pearson Product-Moment Correlations among Image Quality Measurements - Phase II

	MOD	CR	GMTF Area	GMTFA	LOG GMTFA
MOD	--	.80	.99	.99	.96
CR		--	.80	.80	.66
GMTF Area			--	.99	.96
GMTFA				--	.96
LOG GMTFA					--

Table 27

Analysis of Variance Summary for Search Time - Phase II

Source of Variance	df	MS	F	p
<u>Between Subjects</u>				
Subjects (S)	7	44.456		
<u>Within Subject</u>				
Transmission (T)	4	16.778	0.94	0.4541
T x S	28	17.804		
Illuminance (I)	4	114.226	6.18	0.0011
I x S	28	18.491		
T x I	16	26.064	1.77	0.0436
T x I x S	112	14.700		
Total	199			

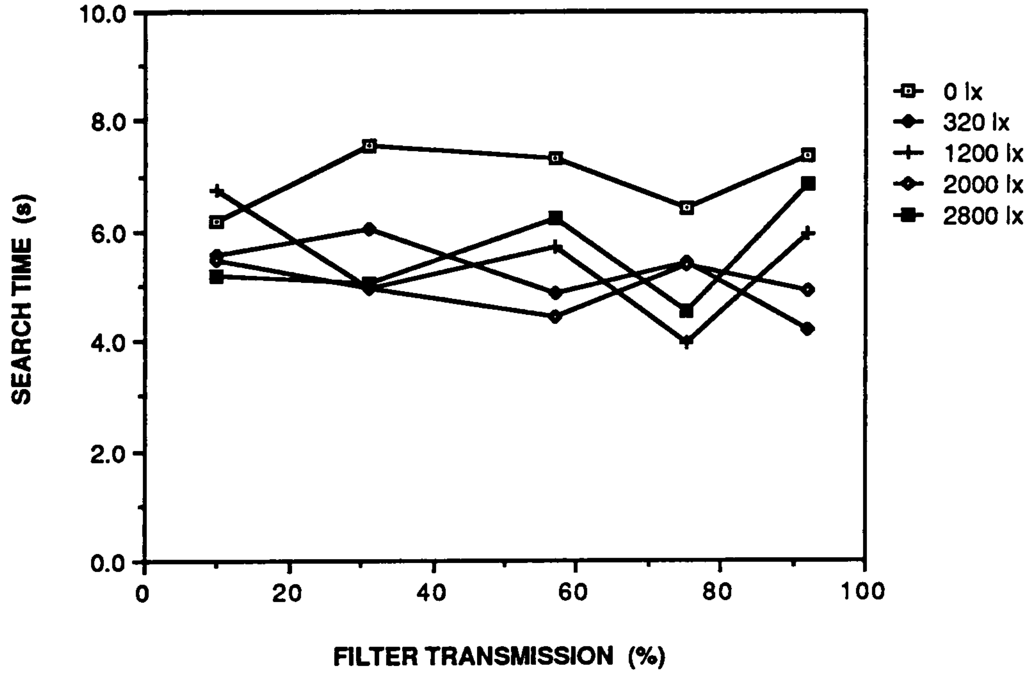


Figure 33. The T x I interaction for search time.

Table 28

Simple-effect F-test for Transmission at All Levels of Illuminance and for Illuminance at All Levels of Transmission - Search Time

Source of Variance	df	MS	F
T at 0 lx	4	13.861	0.78
T at 320 lx	4	20.017	1.12
T at 1200 lx	4	44.230	2.48
T at 2000 lx	4	6.480	0.36
T at 2800 lx	4	35.706	2.01
T x S	28	17.804	

Source of Variance	df	MS	F
I at 11%	4	16.615	0.89
I at 31%	4	49.811	2.69
I at 58%	4	51.016	2.75 *
I at 75%	4	39.753	2.14
I at 92%	4	70.465	3.81 *
I x S	28	18.491	

* $p < .05$

transmission are shown in Table 29. At 58% transmission, STs under 0 lx illuminance (7.33 s) were longer than those under 2000 lx illuminance (4.47 s), but no other comparisons were significant. At 92% transmission, both the 0 lx and 2800 lx yielded longer STs (7.41 s and 6.84 s, respectively) than the 320 lx (4.20 s) condition. These findings indicate that legibility was generally degraded at either extremely low (0 lx) or moderately high (2800 lx) levels of diffuse illuminance. While the loss in legibility under high ambient conditions was expected, the reduced legibility under low illuminance supports the somewhat surprising results of Phase I. Figure 34 clearly represent the effect of illuminance on ST.

Reading results. A summary of the repeated-measures, within-subjects ANOVA performed on the reading time data is given in Table 30. The main effects of Transmission and Illuminance were significant and are represented in Figures 35 and 36, respectively. The results of the corresponding Newman-Keuls analyses are listed in Tables 31 and 32.

Referring to Figure 35, it is evident that the 11% transmission level produced the slowest mean RT, and was found to be significantly slower than that of the 75% condition. The difference between these conditions, though under 1.0 s, is nevertheless important when it is considered that only four short lines of text made up each of the reading passages and that the difference in time is about 12%. Should this effect hold true when full displays of text are presented, the increased time spent in reading would be non-trivial. It would appear, then, that while the 11% filter yielded the highest contrast, the loss in display luminance that accompanied the introduction this dense filter produced the slower RTs.

Figure 36 reveals that performance was again degraded under dark viewing conditions. Here, at 0 lx, RTs were about 0.5 s (or 6%) slower than for the 300 lx condition, and this difference proved to be statistically significant. This result supports the general findings of Phase I, and is also in agreement with the ST results of Phase II that were discussed earlier. It does not appear, however, that readability was degraded by increased levels of diffuse illuminance up to and including 2800 lx, as RTs remained quite stable across the the entire range of ambient lighting. It appears then, that the contextual nature of the reading task made it more resistant to losses in contrast than

Table 29

Newman-Keuls Comparisons for Illuminance at 58% and 92% Transmission - Search Time

58% Transmission

Illuminance (lx)	Mean ST (s)		
2000	4.47	A	
320	4.88	A	B
1200	5.72	A	B
2800	6.21	A	B
0	7.33		B

92% Transmission

Illuminance (lx)	Mean ST (s)		
320	4.20	A	
2000	4.91	A	B
1200	5.98	A	B
2800	6.84		B
0	7.41		B

Note: Means accompanied by the same letter are not significantly different, $p > .05$.

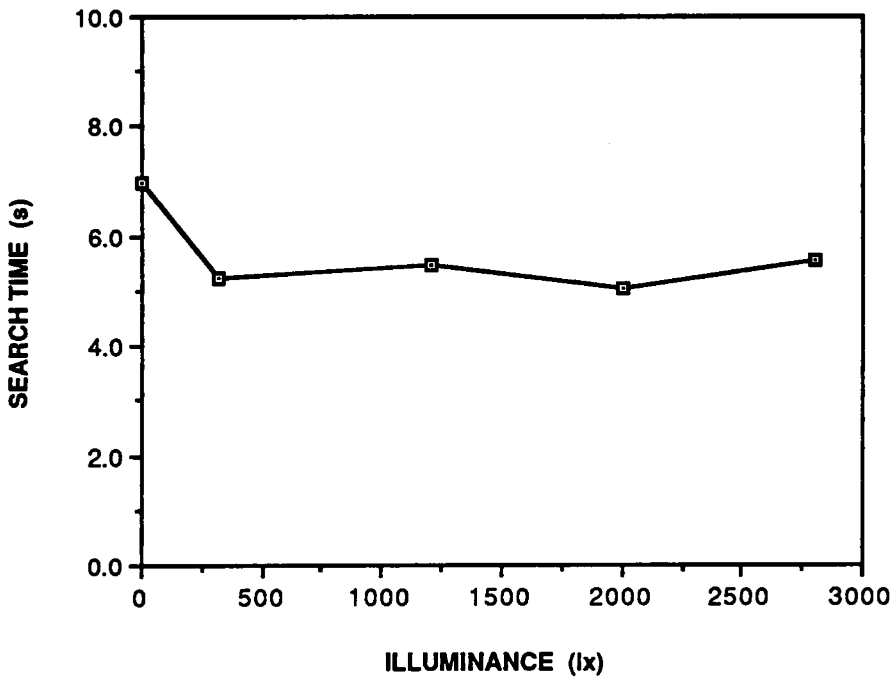


Figure 34. The effect of illuminance on search time.

Table 30

Analysis of Variance Summary for Reading Time - Phase II

Source of Variance	df	MS	F	p
Between Subjects				
Subjects (S)	7	307.849		
Within Subject				
Transmission (T)	4	22.361	2.94	0.0377
T x S	28	7.594		
Illuminance (I)	4	4.975	2.86	0.0419
I x S	28	1.740		
T x I	16	3.590	1.37	0.1687
T x I x S	112	2.617		
Total	199			

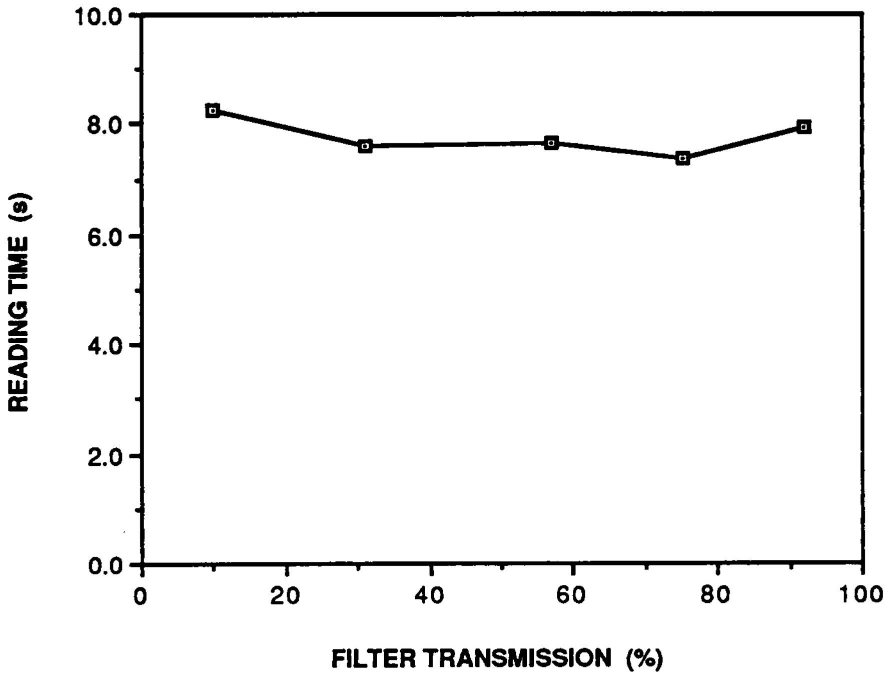


Figure 35. The effect of transmission on reading time.

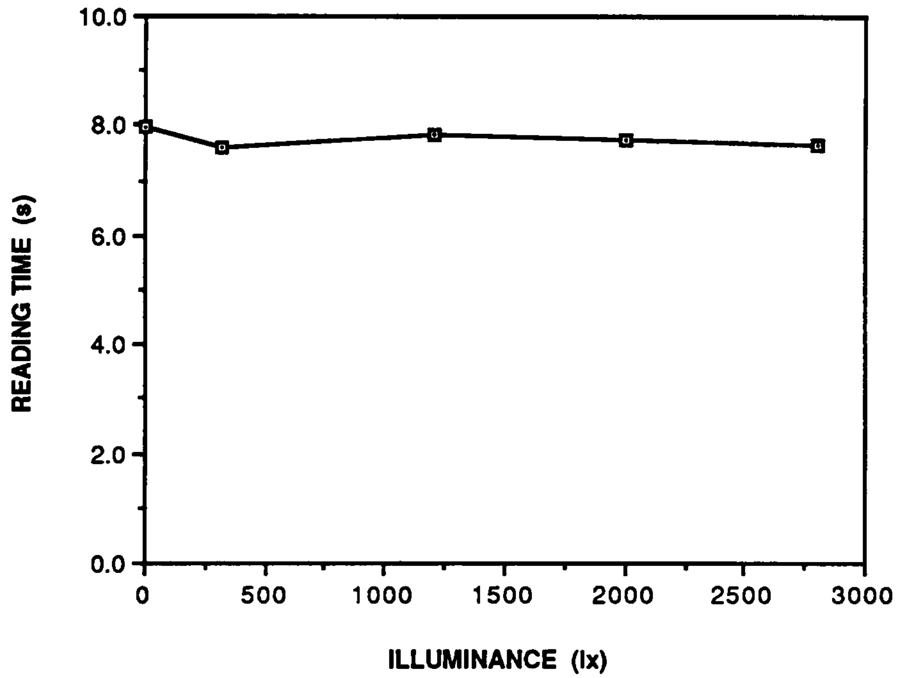


Figure 36. The effect of illuminance on reading time.

Table 31

Newman-Keuls Comparisons for Transmission - Reading Time

Transmission (%)	Mean RT (s)		
75	7.37	A	
31	7.59	A	B
58	7.64	A	B
92	7.92	A	B
11	8.24		B

Note: Means accompanied by the same letter are not significantly different, $p > .05$.

Table 32

Newman-Keuls Comparisons for Illuminance - Reading Time

Illuminance (lx)	Mean RT (s)		
320	7.59		B
2800	7.64	A	B
2000	7.71	A	B
1200	7.81	A	B
0	7.99	A	

Note: Means accompanied by the same letter are not significantly different, $p > .05$.

the noncontextual, alphabetic search task. These finding are consistent with those reported by previous researchers (Snyder and Maddox, 1978).

Rating results. The results of the repeated-measures, within-subjects ANOVA performed on the image quality rating data are shown in Table 33. Figure 37 presents the significant T x I interaction that clearly makes evident the orderly nature of the various illuminance curves across filter transmission. Simple-effect effect F-tests were performed on Illuminance at the levels of transmission and on Transmission at the levels of illuminance. The results of these analyses are presented in Table 34 and show that Illuminance was significant at 92% transmission, but that no other differences were found. The results of the Newman-Keuls results for illuminance at 92% transmission are shown in Table 35 and revealed the 320 lx condition to have yielded better image quality ratings (mean = 7.25) than the 2800 lx condition (mean = 3.25). Generally, it can be seen in Figure 37 that image quality ratings were inversely related to illuminance and the corresponding losses in luminance modulation that accompany higher illuminance (with the exception of the 0 lx condition). Further, the roll-off in each of the illuminance curves at 11% transmission seems to imply that the contrast enhancement offered by very dense filters may be offset by the resultant reductions in character luminance. Figures 38 and 39 present the main effects of Transmission and Illuminance, respectively, and are included as simplifications to Figure 37. Analysis and discussion of of these effects would be redundant as they are most meaningfully interpreted in terms of the T x I interaction described above.

Model Selection

Approach. The approach followed for model selection in Phase II was identical to that of Phase I. That is, Mallow's Cp-statistic, the PRESS-statistic, the MSE, the predicted R^2, and the R^2 were all considered in the process designed to determine the model with the best fit to the data as well as good predictive capability. However, in Phase II, the process was simplified due to the reduced number of regressor variables available for inclusion in the model. The signal metrics included in the analysis were GMTFA, MTFA, luminance modulation (MOD), and contrast ratio (CR). Again, only one of these terms could be included in any single model. Since no noise metrics

Table 33

Analysis of Variance Summary for Image Quality Ratings - Phase II

Source of Variance	df	MS	F	p
Between Subjects				
Subjects (S)	7	39.266		
Within Subject				
Transmission (T)	4	4.082	3.18	0.0285
T x S	28	1.385		
Illuminance (I)	4	45.408	9.44	0.0001
I x S	28	4.810		
T x I	16	1.011	1.93	0.0247
T x I x S	112	0.524		
Total	199			

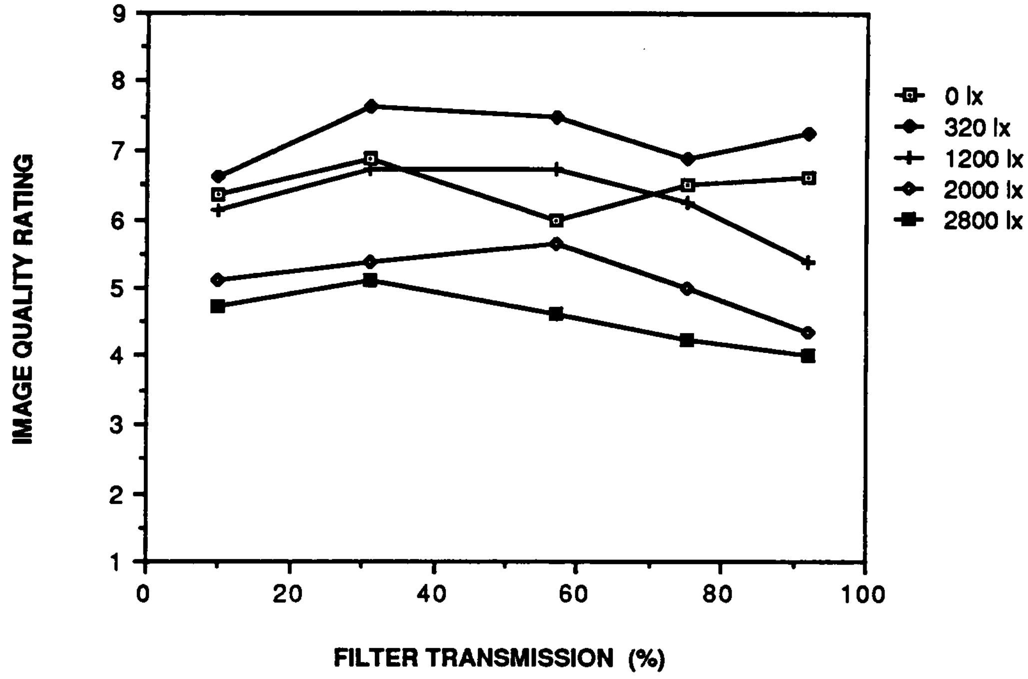


Figure 37. The T x I interaction for image quality ratings.

Table 34

Simple-effect F-test for Transmission at All Levels of Illuminance and for Illuminance at All Levels of Transmission - Image Quality Ratings

Source of Variance	df	MS	F
T at 0 lx	4	0.838	0.65
T at 320 lx	4	1.412	1.10
T at 1200 lx	4	2.562	1.99
T at 2000 lx	4	1.775	1.38
T at 2800 lx	4	1.538	1.20
T x S	28	1.285	

Source of Variance	df	MS	F
I at 11%	4	5.350	0.89
I at 31%	4	9.025	1.88
I at 58%	4	9.587	1.99
I at 75%	4	9.775	2.03
I at 92%	4	15.712	3.27 *
I x S	28	18.491	

* $p < .05$

Table 35

Newman-Keuls Comparisons for Illuminance at 92% Transmission - Image Quality Ratings

Illuminance (lx)	Mean Rating		
320	7.25	A	
0	6.62	A	B
1200	5.38	A	B
2000	4.38	A	B
2800	4.00		B

Note: Means accompanied by the same letter are not significantly different, $p > .05$.

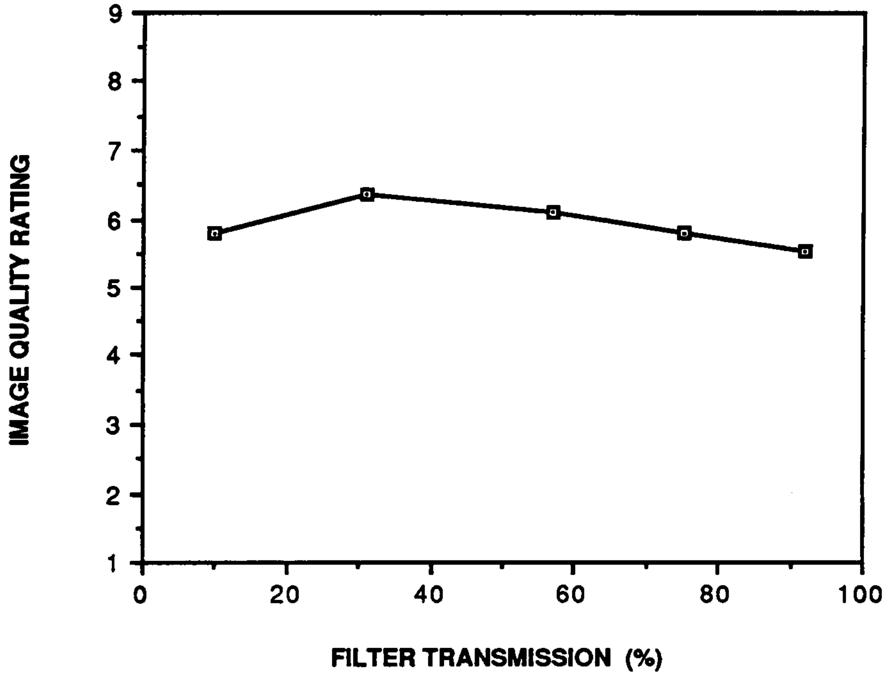


Figure 38. The effect of transmission on image quality ratings.

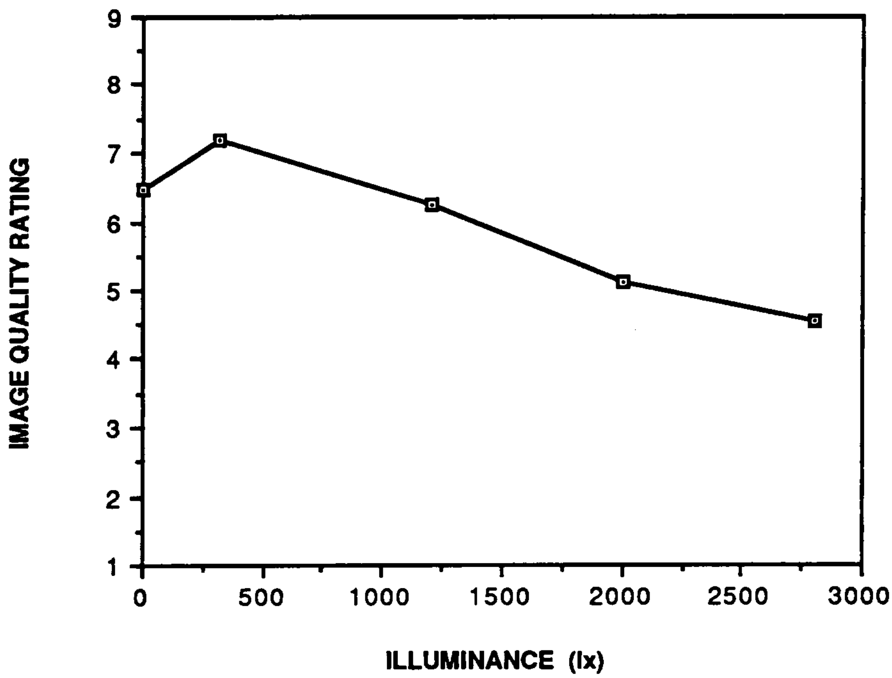


Figure 39. The effect of illuminance on image quality ratings.

were used in Phase II, all resultant models consisted of single-term, simple linear regressions. In further contrast to Phase I, the data were only modeled intact (i.e., they were not modeled separately for each illuminance level).

All equations were significant with probabilities less than or equal to 0.10 except for model RT = f(CR) and model RATE = f(CR). As in Phase I, the predicted R^2 values behaved as expected relative to the computed R^2 values. That is, poor quality models yielded the greatest deflation with predicted R^2 values that approached zero, while better models (i.e., for rating data) produced deflations ranging from 10% to 20%.

Reading time models. The four best-fit models to the RT data for each of the four signal metrics are presented in Table 36 and Figure 40 shows these RT data plotted as a function of GMTFA (any of the other metrics would have served equally well). Table 36 indicates the poor fit of all four models to the data set, and this lack of fit is clearly presented in the corresponding figure. Little variation in RT occurred across the extremes of GMTFA and therefore, the slope or effect of the regression was minimal. The range of values obtained for GMTFA do approach those of Phase I, but the subsequent changes in readability were not evident. This points out that when the MTF of the display is measured under glare conditions, similar MTF values can be obtained through low-pass filtering (Phase I) caused by etching or through contrast reduction (Phase II) due to the addition of high illuminance levels. Although the photometric measures would indicate similar degrees of image quality, comparison of the human performance results from Phases I and II reveals that the readability of the two situations is quite different. Apparently, the degradation in image quality due to the blurred presentation of the information has a greater effect on readability than does similar degrees of degradation due to loss of contrast only.

Search time models. Table 37 contains the best-fit models for the ST data and Figure 41 presents these data plotted as a function of GMTFA. Again, none of the four models produced adequate fits to the ST data set and therefore cannot be considered useful in terms of description or prediction. The positive signs of the regressor coefficients indicate that increases in displayed image quality yielded poorer legibility. The variability about the regressor line, however, makes any interpretation of any of these effects highly questionable. Interesting to note is the increase in ST

Table 36

Best-fit Regression Models to the Reading Time Data - Phase II

	MODEL	R^2	PRESS	MSE	C_p
1.	RT = 7.5122 + 0.0484 GMTFA	0.12	4.38	0.1614	7.77
2.	RT = 7.5030 + 0.4700 MOD	0.12	4.39	0.1620	7.86
3.	RT = 7.6819 + 0.0038 CR	0.08	4.49	0.1698	9.26
4.	RT = 7.5152 + 0.1712 LOG GMTFA	0.08	4.64	0.1680	8.93

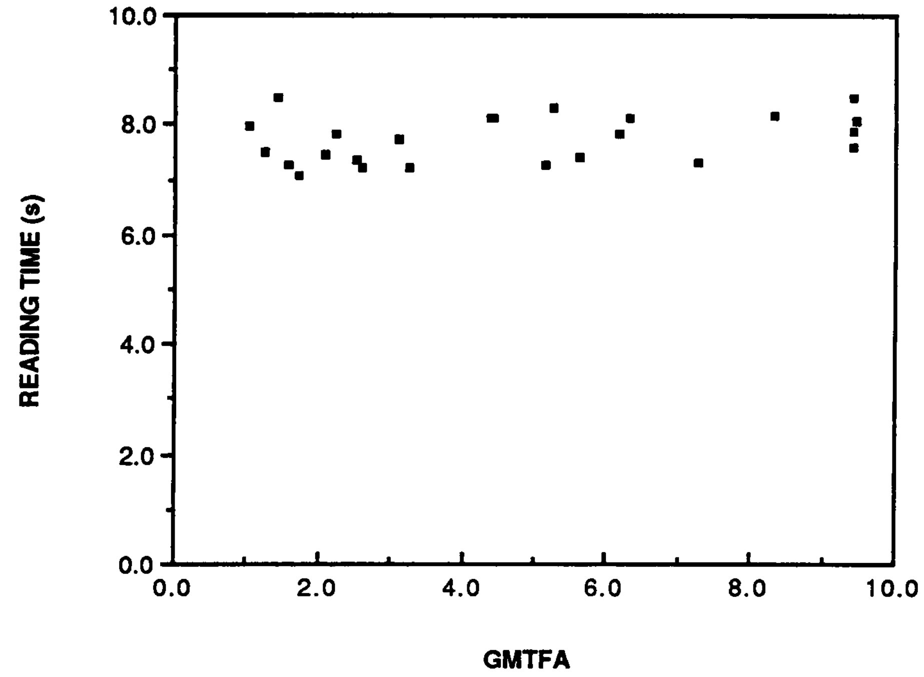


Figure 40. Reading time as a function of GMTFA (n = 25). Some points may be hidden.

Table 37

Best-fit Regression Models to the Search Time Data - Phase II

	MODEL	R^2	PRESS	MSE	C_p
1.	ST = 5.2474 + 0.0222 CR	0.47	14.42	0.5436	1.96
2.	ST = 4.7332 + 0.1867 GMTFA	0.32	19.27	0.7004	8.58
3.	ST = 4.6913 + 1.8400 MOD	0.32	19.36	0.7034	8.71
4.	ST = 4.8453 + 0.5869 LOG GMTFA	0.18	24.05	0.8463	14.74

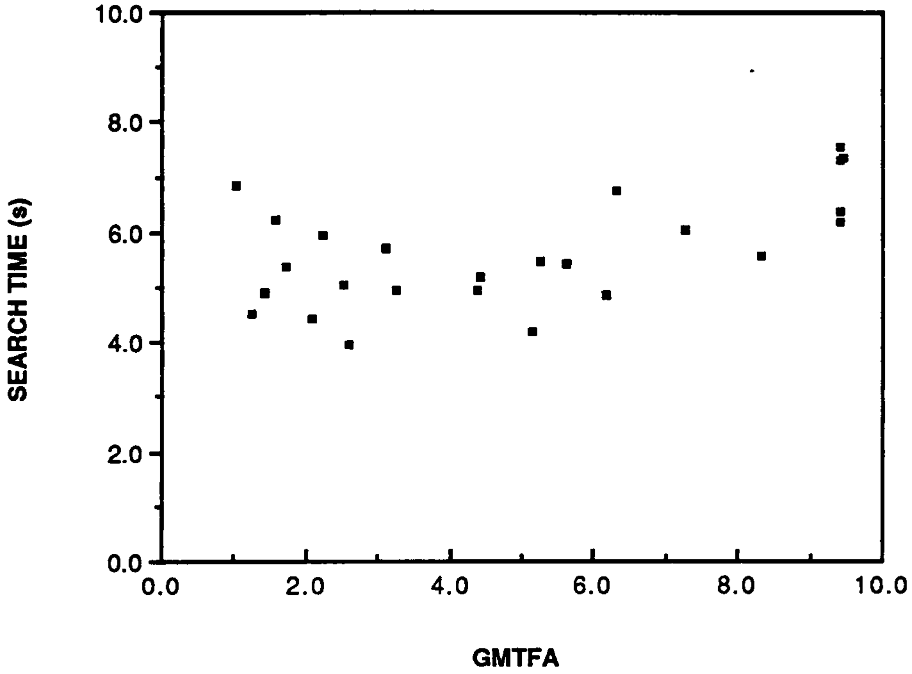


Figure 41. Search time as a function of GMTFA (n = 25). Some points may be hidden.

as GMTFA approaches its maximum value of near 10.0. The five tightly bunched data points that correspond to the highest levels of GMTFA were, of course, collected under 0 lx ambient lighting. These data add even more support to the case that performance seems to be degraded when displayed contrast levels are very high.

Rating model. The best-fit regression models for the image quality rating data are presented in Table 38 and these data are plotted in Figure 42 as a function of GMTFA. Although LOG MTFA, MOD, and GMTFA produced models with reasonably high R^2 values, the high values found for PRESS, MSE, and Cp make application of these models highly questionable. While the variability shown in Figure 43 make interpretation difficult, it appears that very high levels of luminance modulation (GMTFA is driven only by modulation in Phase II) produce lower perceived image quality than more moderate levels. This relationship has not been tested in a statistical sense. No appreciable improvements in the fits of the models occurred when these data obtained under dark conditions were deleted from the analysis. Therefore, it was concluded that none of the models examined were of sufficient quality to be used in application.

Table 38

Best-fit Regression Models to the Image Quality Rating Data - Phase II

	MODEL	R^2	PRESS	MSE	C_p
1.	RATE = 4.3714 + 1.1215 LOG GMTFA	0.58	12.74	0.4874	19.67
2.	RATE = 4.6920 + 2.3400 MOD	0.45	16.86	0.6347	31.97
3.	RATE = 4.7580 + 0.2340 GMTFA	0.45	17.03	0.6415	32.54
4.	RATE = 5.7181 + 0.0106 CR	0.09	27.17	1.0534	66.92

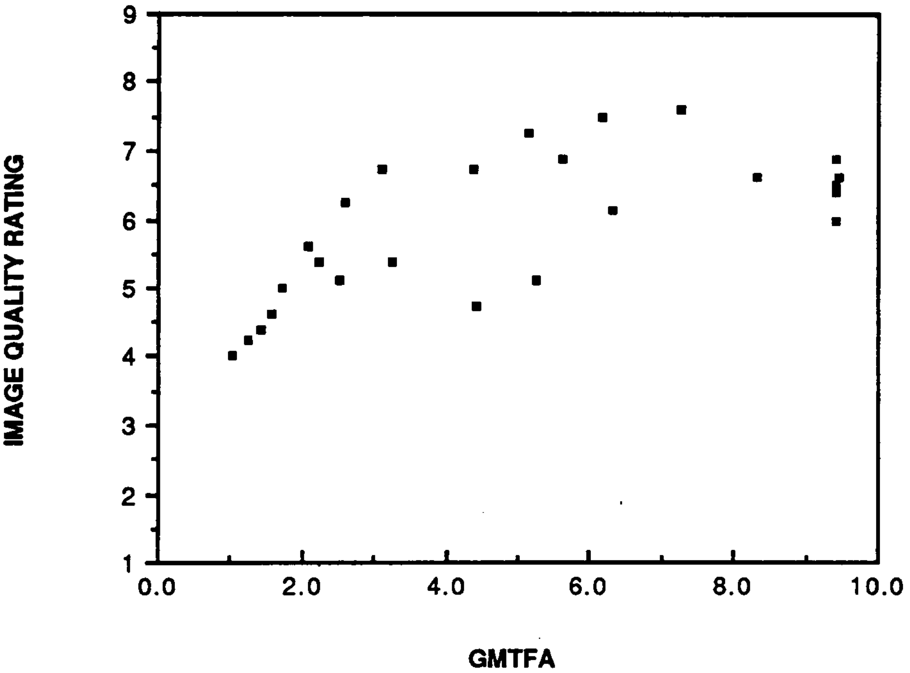


Figure 42. Image quality ratings as a function of GMTFA (n = 25). Some points may be hidden.

CONCLUSIONS - PHASE II

A number of interesting conclusions can be drawn from the results of the RT, ST, and perceived image quality analyses. However, attempts to model the data for each of these dependent measures using single-term regression models met with only minimal success. The results of the Phase II analyses are especially interesting when considered in the context of the Phase I results.

Human Adaptability

The first major conclusion to be drawn from the Phase II findings concerns the general lack of variation in the ST and RT values as filter transmission and ambient illuminance were changed. For all but the 0 lx illuminance, STs and RTs remained quite stable as illuminance varied from 320 lx to 2800 lx (though some degradation did occur at 2800 lx). As indicated in Phase I, these findings point out the ability of the human observer to adapt to a variety of degraded viewing conditions and yet yield essentially the same response to these conditions. Under this diffuse glare, luminance modulation ranged from 0.12 to 0.86 (contrast ratios of 1.3:1 to 13.4:1), and yet the performance differences between these extremes were minimal. These findings are not in support of those reported by previous researchers (Howell and Kraft, 1959; Snyder and Maddox, 1978) who found legibility to be reduced when modulation fell below 0.90 in non-contextual situations and below 0.75 for contextual.

In further support of the findings of Phase I, the rating task proved to be the most sensitive to changes in image quality and yielded a larger range of response than either RT or ST. Here, perceived image quality was generally degraded (with the exception of 0 lx) as illuminance was increased. In a practical sense, ratings of image quality were not significantly reduced until illuminance levels of 2000 lx were reached (see Figure 37). However, it is difficult to determine a specific contrast or modulation value at which meaningful degradations in perceived image quality begin to occur. Further, as the differences in ratings were subtle for each illuminance as transmission was varied, no optimal transmission can be recommended. Figure 37 does show, however, a trend of reduced ratings when illuminance levels of 1200 lx and above are combined with high transmission filters.

Evident in Table 25 and Figure 37, modulations can vary quite widely across transmission for each illuminance level, but with little change in ratings. That is, perceived image quality is dependent upon both the ambient illuminance and filter transmission, these variables both interacting upon the modulation of the information as well. For example, the 320 lx/ 92% condition produced a luminance modulation of about 0.55 and yielded an image quality rating of 7.25. Correspondingly, the 2000 lx/11% condition also produced a modulation of 0.55, but yielded a rating of only 5.12. Therefore, specification of minimum modulation levels must be made with caution if the display is to be operated under glare conditions.

Unsuccessful Modeling

Modeling the performance differences using luminance modulation, contrast ratio, GMTFA, and LOG GMTFA as single-term predictors proved to be unsuccessful for the RT and ST data, and only moderately successful when fits to the rating data were attempted. The signs of these regressor variables were not consistent with theory, the variance about the regression was large, and the effect of the regression itself was quite small for the timed data. Although the rating data covered a larger relative range than did either the RT or ST data, the fits to the data were only slightly better, and predictive quality was not improved. Certainly, as mentioned in discussion of the Phase I results, had the range of illuminance been increased beyond the 2800 lx maximum found in Phase II, it is likely that larger differences would have been evident and the situation more easily modeled.

However, the 0 lx to 2800 lx range incorporated in this research is greater than the extremes found in the most office-type environments to which this research was designed to generalize.

Low Illumination Performance

Finally, the results of Phase II lend further credibility to those of Phase I that showed both readability and legibility to be reduced in dark (0 lx) viewing environments. While it could be argued that the differences were small between the dark and the other lighting conditions, they proved to be consistent and repeatable, and as such cannot be ignored. Modulation values under this dark condition were about 0.97 regardless of the filter transmission used, and contrast ratios were found to range from 81:1 to 63:1, for the 92% and 11% transmission filters, respectively. It seems, although contrary to current thought, that contrast ratios can be made too high on CRTs and that background luminance levels approaching zero should be avoided.

SUMMARY

Filter Selection

Reduced transmission. The results of this research indicate that across a fairly wide range of diffuse illuminance (0 lx to 2800 lx), readability, legibility, and perceived image quality are only minimally affected by the contrast enhancement qualities of the filter. However, when specular glare is present and reduced transmission filters are to be incorporated, it is important to include an effective (AR coating or mesh) antireflection treatment to the front surface of the filter as the lower character luminance and darker background can magnify the degradation caused by the reflected image. Finally, the finding that "fuzzy" images produced by etched filters yielded poor overall performance also has implications for the use of reduced transmission filters. That is, if the user must adjust the character brightness to compensate for the loss in luminance caused by the filter, the effective resolution of the display is lowered -- a situation analogous to adding a low-pass filter. Therefore, it is recommended that filters with high transmittance (i.e., 75% to 95%) be used in environments where only moderate diffuse or specular glare is present.

Antireflection treatments. Of the three antireflection treatments (chemical etching, quarterwave coating, and micromesh filters) evaluated in this research, none were found to improve readability or legibility over a polished, 92% transmission baseline regardless of lighting condition. Notable, however, was that both reading and search times were longer when the etched treatments were applied. This effect was particularly dramatic when low transmission, harshly etched surfaces were presented at the high resolution display format. Correspondingly, the same etched filters were perceived to yield poorer image quality than the polished, coated, or mesh filters. Finally, while the ratings of the coated and mesh filters did not statistically differ from those of the baseline filter under any lighting condition, both treatments yielded meaningful improvements in quality when a specular glare source was present (the coated were rated slightly higher than the mesh). Therefore, in environments with moderate glare, the use of high quality mesh filters and quarterwave antireflection coatings will probably not provide noticeable improvements in readability or legibility, but can yield higher perceived image quality when specular sources are present. The use of etched filters (65 to 25 gloss) under any lighting conditions is not recommended, especially when high resolution information is to be displayed.

With Respect To Previous Findings

The results of the image quality measurements are consistent with those reported by previous researchers (Beaton and Snyder, 1984; Hunter et al., 1987; Morse, 1985; Rancourt et al., 1986). Each study indicated large physical differences among the anti-glare techniques evaluated. As noted in the introductory section of this dissertation, these measured differences have not consistently been shown to yield well-behaved changes in human performance. Numerous studies reported only minimal objective performance differences when timed measures such as reading and keystroke rates were measured (Habinek, et al., 1982; McVey, et al., 1984). The results of this study support these researchers' contentions that, in general anti-glare filters do not significantly enhance nor degrade performance when used in office-like environments (the exception being the degradation found for harshly etched filters). However, in terms of perceptual differences among the treatments, the present findings of only minimal improvement due to anti-glare treatments do not correspond well with those of Habinek et al. (1982), Hunter et al. (1987), and Morse (1985). Beaton et al. (1985) did report similar findings of only slightly higher ratings for coated and mesh filters, but found larger improvements with filters designed to enhance contrast. With the exception of Beaton et al. (1985), each of the above studies contained either methodological problems (Habinek et al., 1982; Morse, 1985) or such limited scope (Hunter et al., 1987) to make direct comparisons to the present findings inappropriate.

The finding that the search time and reading time data could not be accurately and meaningfully modeled by the image quality measurements does not support the results of previous research (Snyder and Maddox, 1978; Snyder, 1985). However, it is the author's contention that the classical model selection techniques (e.g., stepwise regression) popular at the time of this research and thus employed in this earlier work, led to the selection of models that contained both unacceptable collinearities among the predictor variables and produced inflated R^2 values due to overfit. The models reported in this dissertation were selected via more modern model selection methods that optimized both fit and prediction. Additional constraints were imposed on the combinations of predictor variables that were allowed to enter the models, requiring the final models to be based within a strict theoretical framework. While the models of timed human performance were not deemed acceptable in this study, excellent fit and prediction were obtained when the rating data were modeled. Evidently, over the range of display and environmental conditions studied by this investigation, the image metrics employed are more applicable to the description and prediction of perceived image quality than objective performance.

Future Research

The results of this dissertation point out a number of areas in which research is needed. Well-controlled research should be conducted to determine the cause of poorer readability and legibility under low illuminance conditions and to identify whether existing human factors recommendations concerning illuminance or display contrast should be duly revised. With the increased use of color CRTs in the office environment, a series of studies should be performed to investigate whether the results of this research can be generalized to these devices. The effects of glare and anti-glare filters on the image quality and human performance issues related to negative contrast display formats should also be pursued. Finally, studies employing a similar methodology to this research should be conducted to determine whether the same results (i.e., minimal readability or legibility improvements with glare filters) are found when performance is measured over longer periods of time.

REFERENCES

Abramson, S. R. and Snyder, H. L. (1984). *Operator performance on flat-panel displays with line and cell failures* (Tech. Report HFL-83-3). Blacksburg, VA: Virginia Polytechnic Institute and State University, Human Factors Engineering Laboratory.

American National Standards Institute (1973). *American national standard practice for office lighting.* New York: American National Standards Institute, ANSI A132.1. In *The Journal of IES*, October, 1973.

American National Standard Institute (1988). *American National Standard for human factors engineering of visual display terminal workstations.* New York: American National Standards Institute, ANSI/HFS 100-1988.

Baggen, E. A. (1987). *A human factors evaluation of current touch entry devices.* Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Baggen, E. A., Snyder, H. L., and Miller, M. R. (1988). A human factors evaluation of current touch entry technologies. In *Society for Information Display International Symposium Digest of Technical Papers.*

Bauer, D., Bonacker, M., and Cavonius, C. R. (1981). Influence of VDU screen brightness on the visibility of reflected images. *Displays*, 242-244.

Beaton, R. A. (1984). *Human performance based evaluation of quality metrics for hard-copy and soft-copy digital imaging systems.* Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Beaton, R. A., Murch, G. M., and Knox, S. T. (1985). *On the selection of glare filters for CRT-based workstations: Part I. Human performance studies* (Tech. Report HFL-604-02). Beaverton, OR: Human Factors Research Laboratory, Tektronix, Inc.

Beaton, R. J. and Snyder, H. L. (1984). The display quality of glare filters for CRT terminals. In *Society for Information Display 1984 Digest of Technical Papers* (pp. 298-301).

Blackwell, R. H. (1946). Contrast thresholds of the human eye. *Journal of the Optical Society of America, 36* (11), 624-643.

Brauninger, U. and Grandjean, E. (1983). Lighting characteristics of visual display terminals from an ergonomic point of view. In *CHI '83 Proceedings* (pp. 274-276).

Brown, B. S., Dismukes, K., and Rinalducci, E. J. (1982). Video display terminals and vision of workers. *Behaviour and Information Technology, 1*(2), 121-140.

Cakir, A., Hart, D. J., and Stewart, T. F. M. (1980). *Visual display terminals.* New York: Wiley.

Campbell, F. W. and Durden, K. (1983). The visual display terminal issue: a consideration of its physiological, psychological and clinical background. *Ophthalmology and Physiological Optics, 3* (2), 175-192.

Campbell, F. W. and Maffei, L. (1974). Contrast and spatial frequency. *Scientific American, 281,* 106-114.

Campbell, F. W. and Robson, J. B. (1968). Applications of Fourier analysis to the visibility of gratings. *Journal of Physiology* , *197*, 551-566.

Dainty, J. C. and Shaw, R. (1974). *Image science.* New York: Academic Press.

DeBoer, J. B. (1977). Performance and comfort in the presence of veiling reflections. *Lighting Research and Technology*, *9*(4), 169-176.

Garcia, K. D. and Wierwille, W. W. (1985). The effect of glare on performance of a VDT reading-comprehension task. *Human Factors* , *27*(2), 163-173.

Gould, J. D. (1968). Visual factors in the design of computer controlled CRT displays. *Human Factors*, *10*(4), 359-376.

Habinek, J. K., Jacobsen, P. M., Miller, W., and Suther, T. W. (1982). A comparison of VDT antireflection treatments. In *Proceedings of the Human Factors Society 26th Annual Meeting* (pp. 285-289). Santa Monica, CA: Human Factors Society.

Howell, W. C. and Kraft, C. L. (1959). *Size, blur and contrast as variables affecting the legibility of alphanumeric symbols on radar-type displays* (Tech. Report 59-536). Wright-Patterson AFB, OH: Wright Aeronautical Development Center.

Hunter, M. W., Pigion, R. G., Bowers, V. A., and Snyder, H. L. (1987). MTF and perceived image quality for three glare-reduction techniques. In *Society for Information Display International Symposium Digest of Technical Papers* (pp. 30-33).

Hunter, M. W., Reger, J. J., Farley, W. W., and Snyder, H. L. (1986). *Photometric evaluations of the glare suppression and image quality characteristics of various CRT surface/filter combi-*

nations (Tech. Report). Blacksburg, VA: Virginia Polytechnic Institute and State University, Human Factors Engineering Laboratory.

Isensee, S. H. and Bennett, C. A. (1983). The perception of flicker and glare on computer CRT displays. *Human Factors, 25* (2), 177-184.

Kaufman, L. (1974). *Sight and mind. An introduction to visual perception.* New York: Oxford University Press.

Kelly, D. H. (1977). Visual contrast. *Optica Acta, 24*(2), 107-129.

McVey, B. W., Clauer, C. K., and Taylor, S. E. (1984). A comparison of anti-glare contrast-enhancement filters for positive and negative image displays under adverse lighting conditions. In E. Grandjean (Ed.), *Ergonomics and health in modern offices* (pp. 405-409).

Morse, R. S. (1985). Glare filter preference: influence of subjective and objective indices of glare, sharpness, brightness, contrast and color. In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 782-786). Santa Monica, CA: Human Factors Society.

Moses, R. A. (1987). Accommodation. In R. A. Moses and W. M. Hart (Eds.), *Adler's physiology of the eye* (pp. 291-310). St. Louis: R. V. Mosby.

Myers, R. H. (1986). *Classical and modern regression with applications.* Boston, MA: Duxbury Press.

National Research Council. (1983). *Video displays, work, and vision.* Washington D. C.: National Academy Press.

Rancourt, J., Grenawalt, W., Hunter, M. W., and Snyder, H. L. (1986). Quantitative evaluation of the effect of an antireflection filter. In *Society for Information Display International Symposium Digest of Technical Papers* (pp. 420-422).

Reitmaier, J. (1979). Some effects of veiling reflections in papers. *Lighting Research and Technology, 11*(4), 204-209.

Sach, G. M. (1970). The effect of filters on contrast and readability of CRT displays. In *Proceedings of the Society for Information Display* (pp. 177-186).

Snyder, H. L. (1973). Image quality and observer performance. In L. M. Biberman (Ed.), *Perception of displayed information*. New York: Plenum Press.

Snyder, H. L. (1974). Image quality and face recognition on a television display. *Human Factors, 16*, 300-307.

Snyder, H. L. (1980). *Human visual performance and flat panel display image quality* (Tech. Report HFL-80-1/ONR-80-1). Blacksburg, VA: Virginia Polytechnic Institute and State University, Human Factors Engineering Laboratory.

Snyder, H. L. (1984). Lighting, glare measurement, and legibility of VDTs. In E. Grandjean (Ed.), *Ergonomics and health in modern offices*. London: Taylor and Francis.

Snyder, H. L. (1985). Image quality: Measurement and visual performance. In L. E. Tannas (Ed.), *Flat-panel displays and CRTs*. New York: Van Nostrand Reinhold Company.

Snyder, H. L. and Maddox, M. E. (1978). *Information transfer from computer-generated dot-matrix displays* (Tech. Report HFL-78-3). Blacksburg, VA: Virginia Polytechnic Institute and State University, Human Factors Engineering Laboratory.

Stammerjohn, L. W., Smith, M. J., and Cohen, B. G. F. (1981). Evaluation of workstation design factors in VDT operations. *Human Factors, 23*(4), 401-412.

Stevenson, J. R., Fendley, J. R., and Wallner, J. (1986). Optical characterization of anti-glare surfaces on cathode ray tubes. In *Society for Information Display International Symposium Digest of Technical Papers* (pp. 424-427).

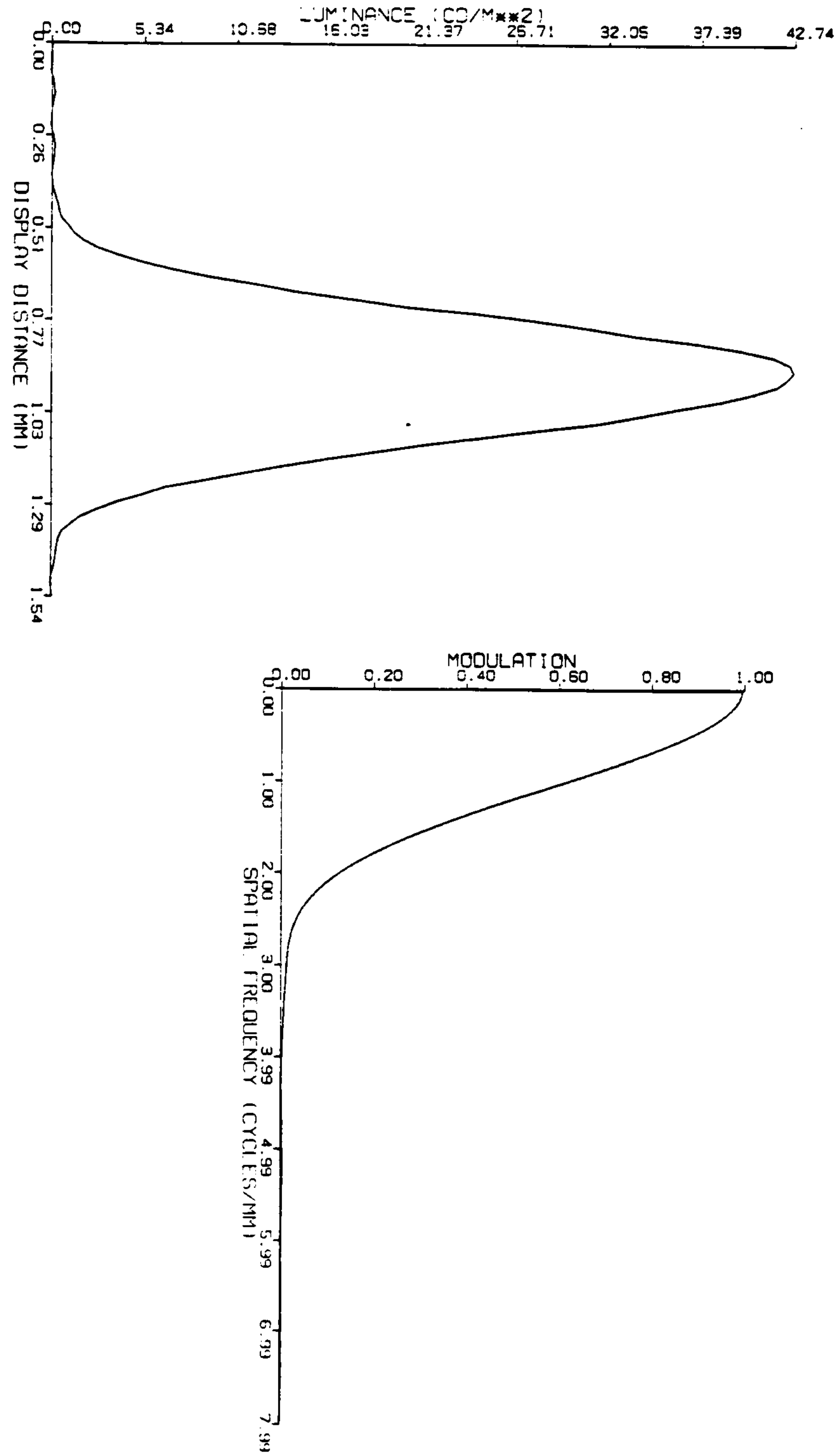
Stocker, A. C. (1964). Displays, papers, and lighting. The visual system in command centers. *Information Displays, 1*, 16-26.

Task, H. L. (1979). *An evaluation and comparison of several measures of image quality for television displays* (Tech. Report AMRL-TR-79-7). Wright-Patterson AFB, Ohio: Aerospace Medical Research Laboratory.

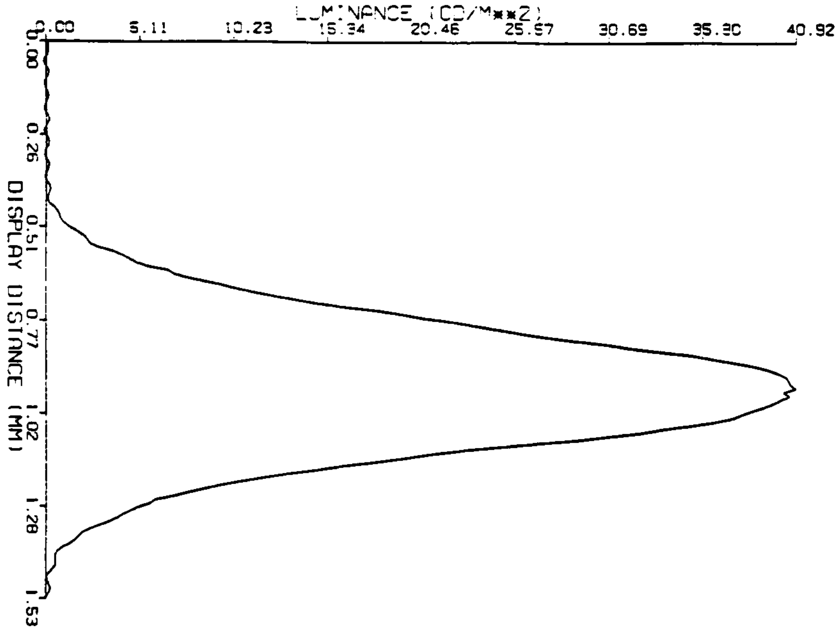
Tinker, M. A. (1955). *Examiner's manual for tinker speed of reading test.* Minneapolis, MN: University of Minnesota Press.

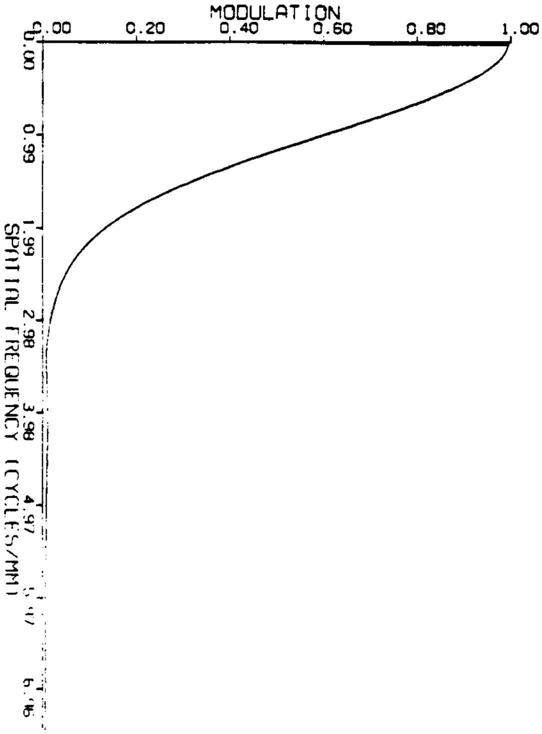
Tinker, M. A. (1958). Length of work periods in visual research. *Journal of Applied Psychology, 42*, 343-345.

Appendix A. LSF AND MTF MEASUREMENT PLOTS - PHASE I

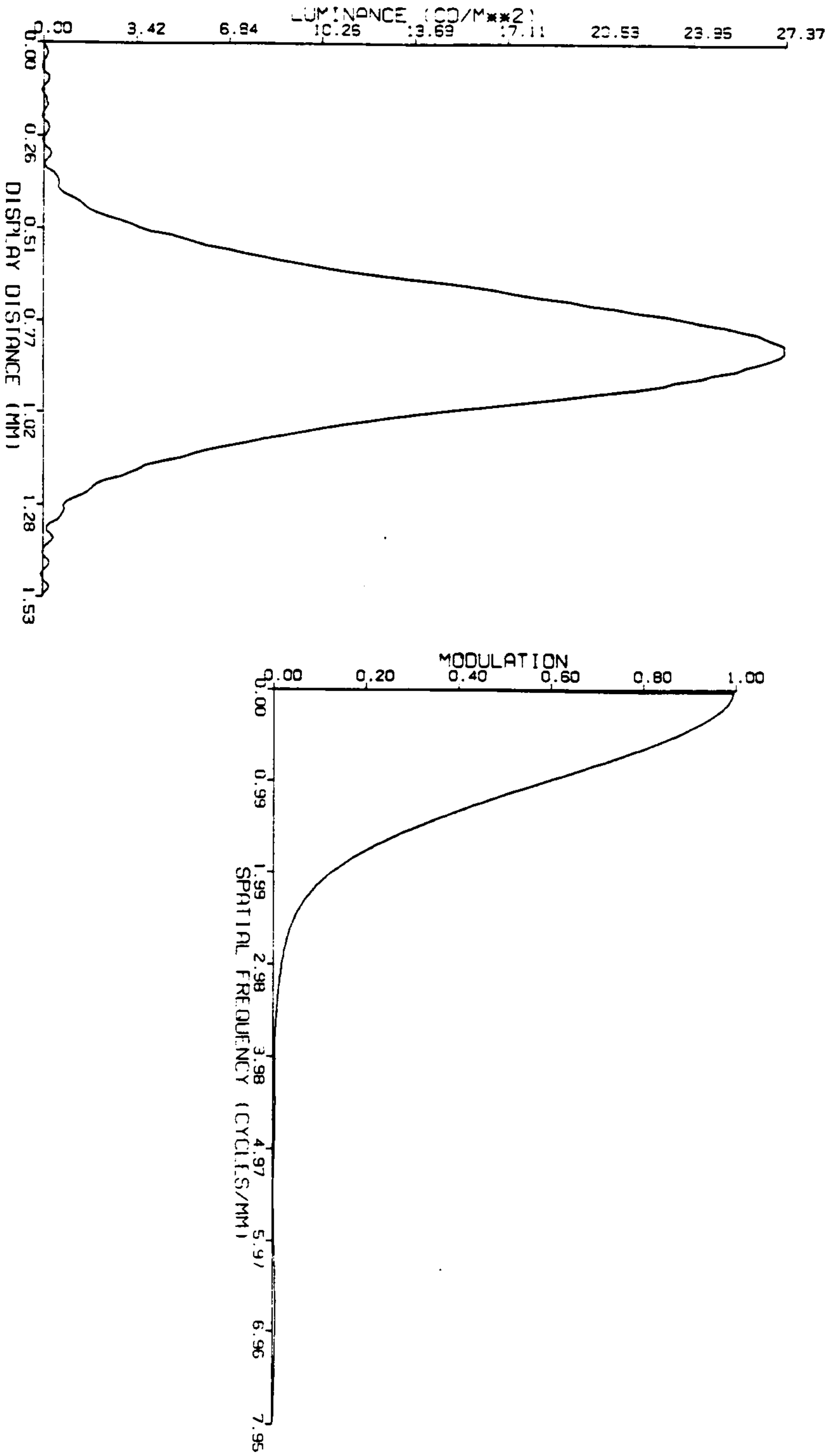


LSF and MTF for dark, non-filtered CRT at high resolution.

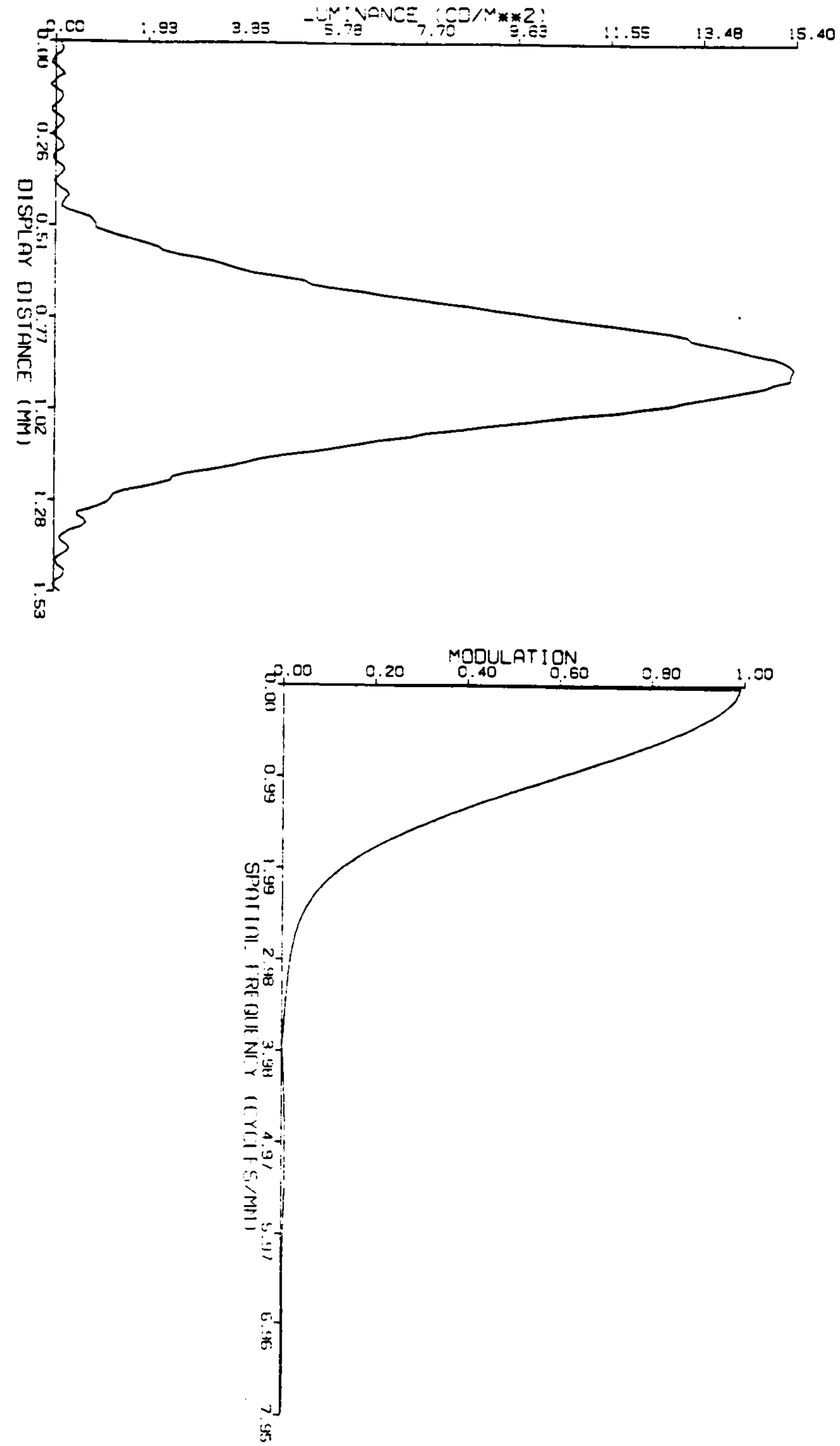




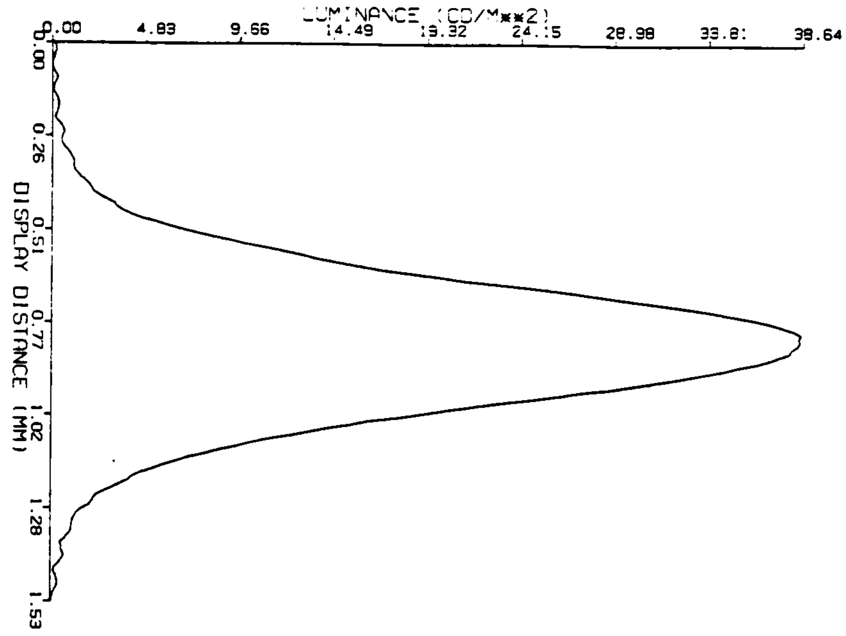
LSF and MTF for dark, POL-92% at high resolution.

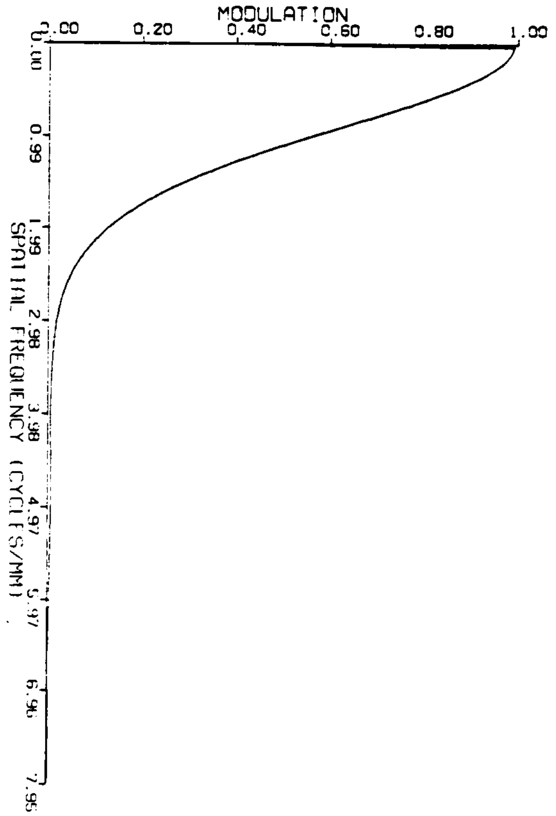


LSF and MTF for dark, POL-62% at high resolution.

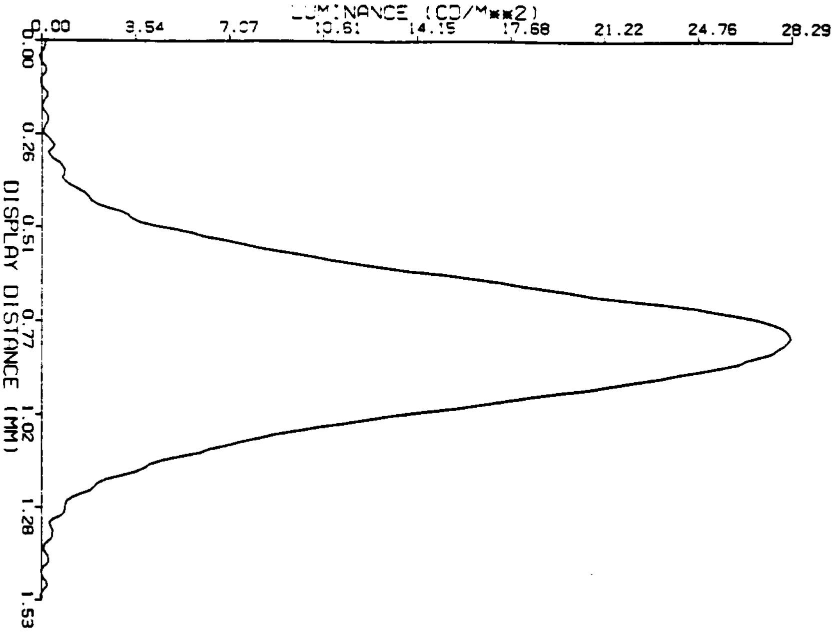


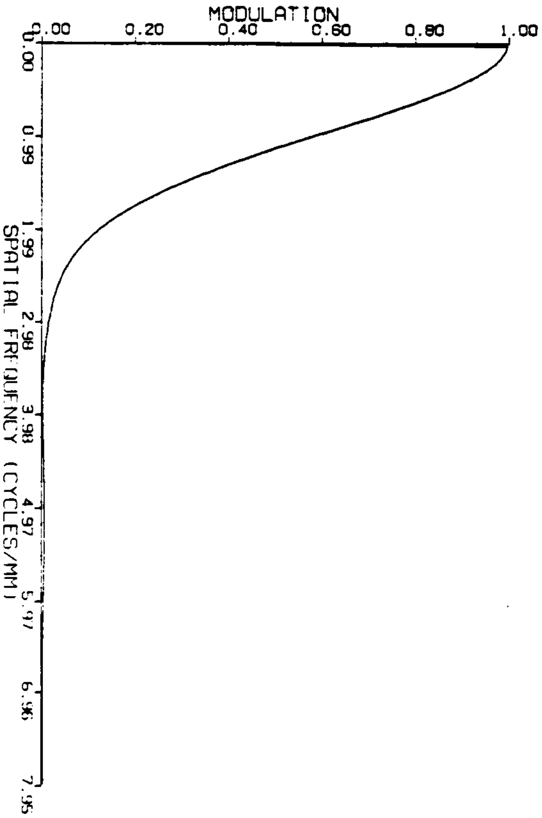
LSF and MTF for dark, POL-31% at high resolution.



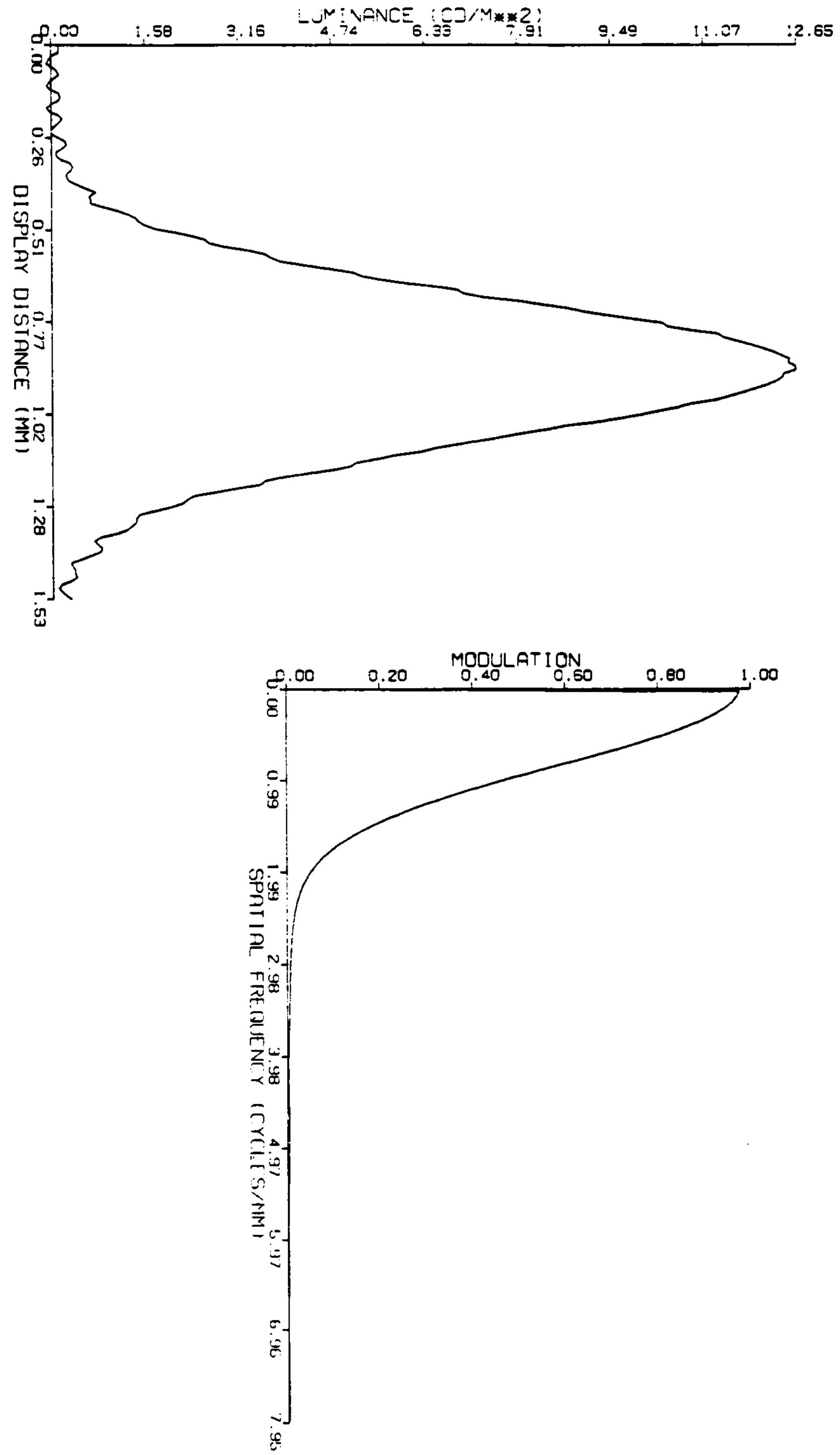


LSF and MTF for dark, G65-92% at high resolution.

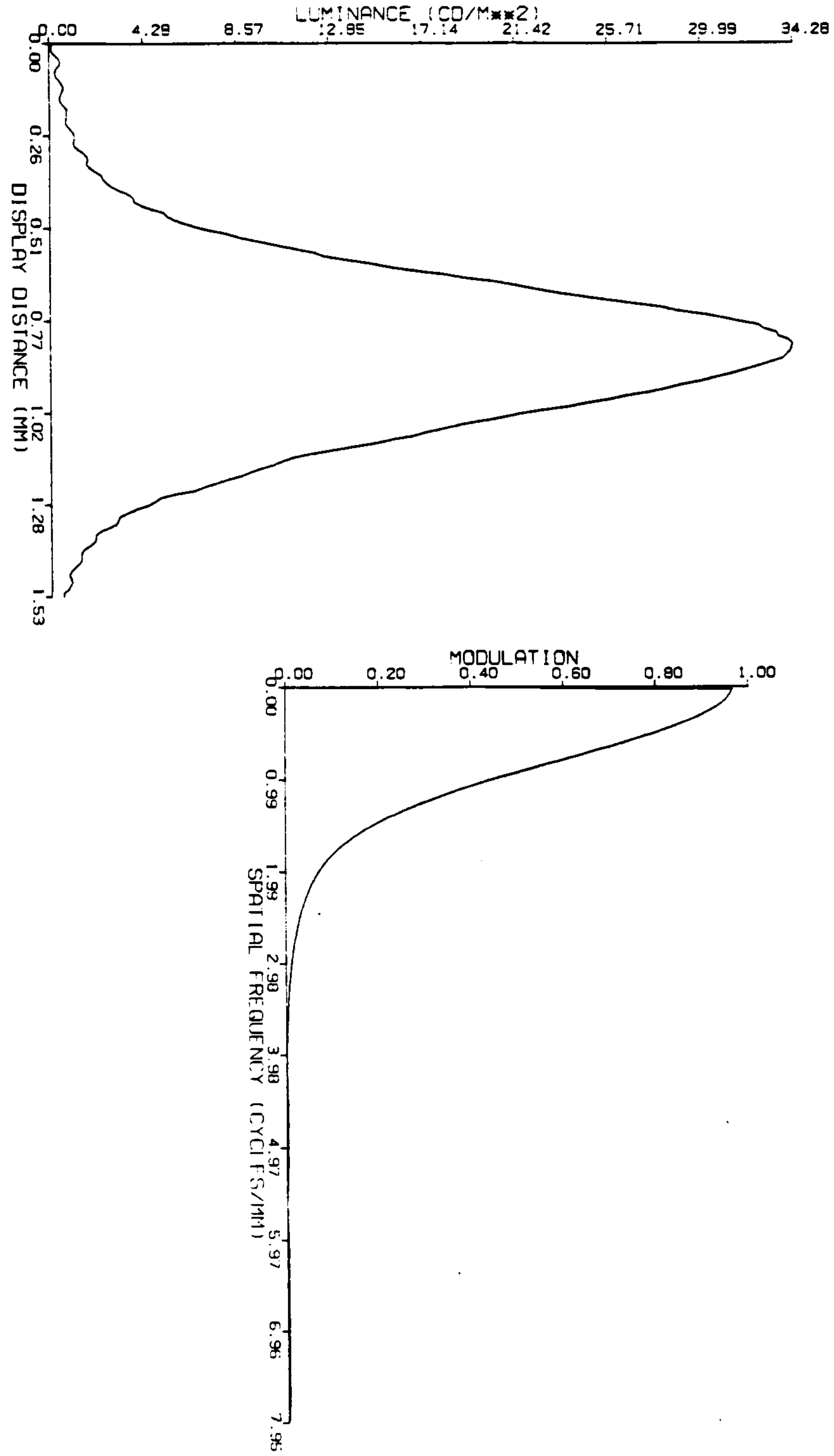




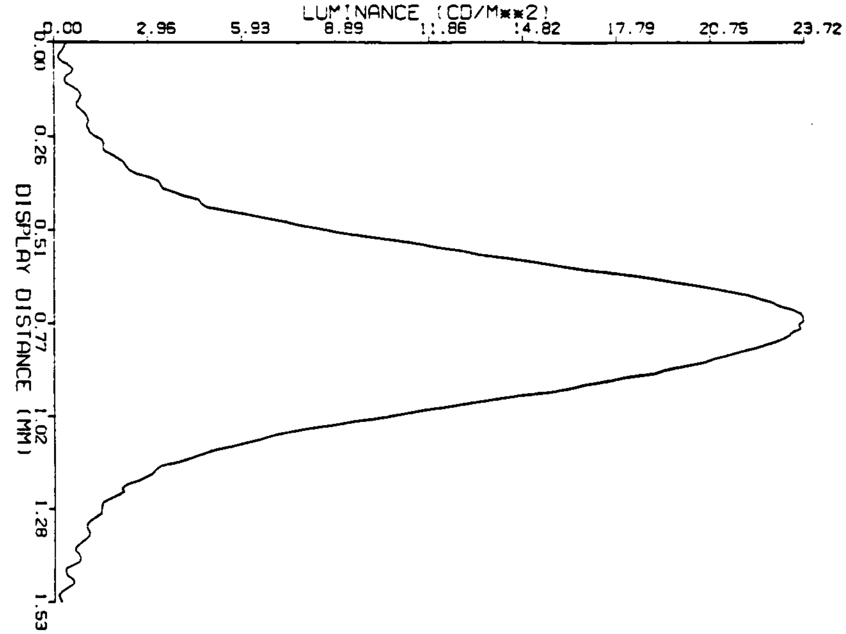
LSF and MTF for dark, G65-62% at high resolution.

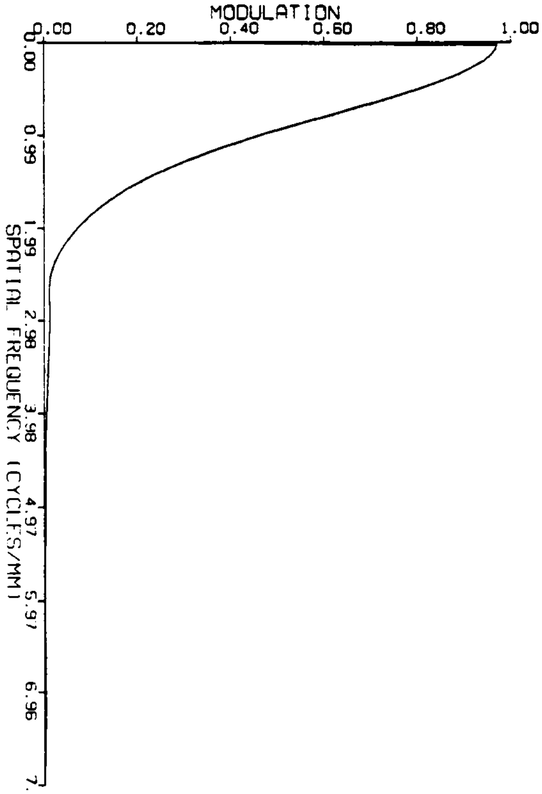


LSF and MTF for dark, G65-31% at high resolution.

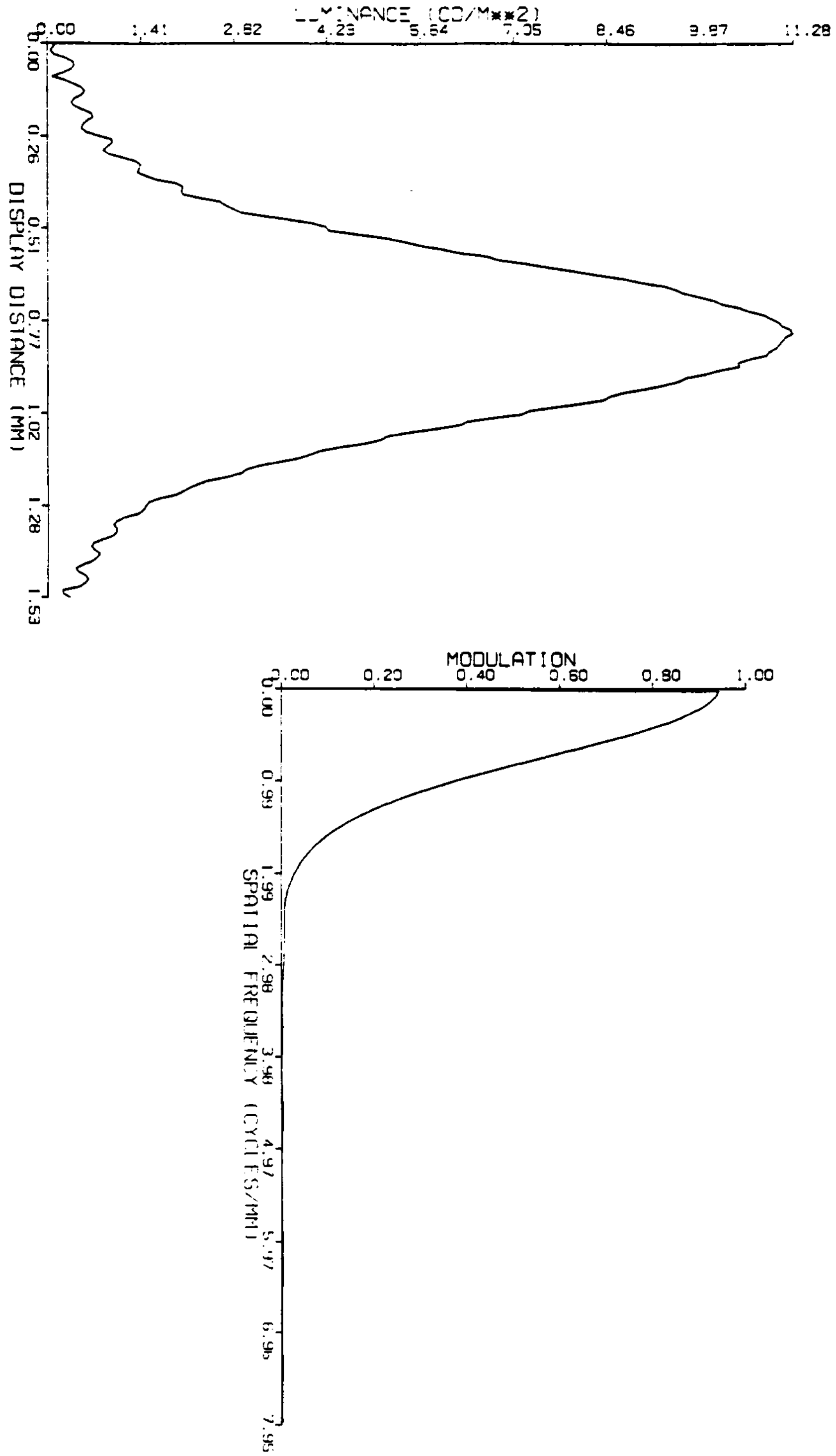


LSF and MTF for dark, G45-92% at high resolution.

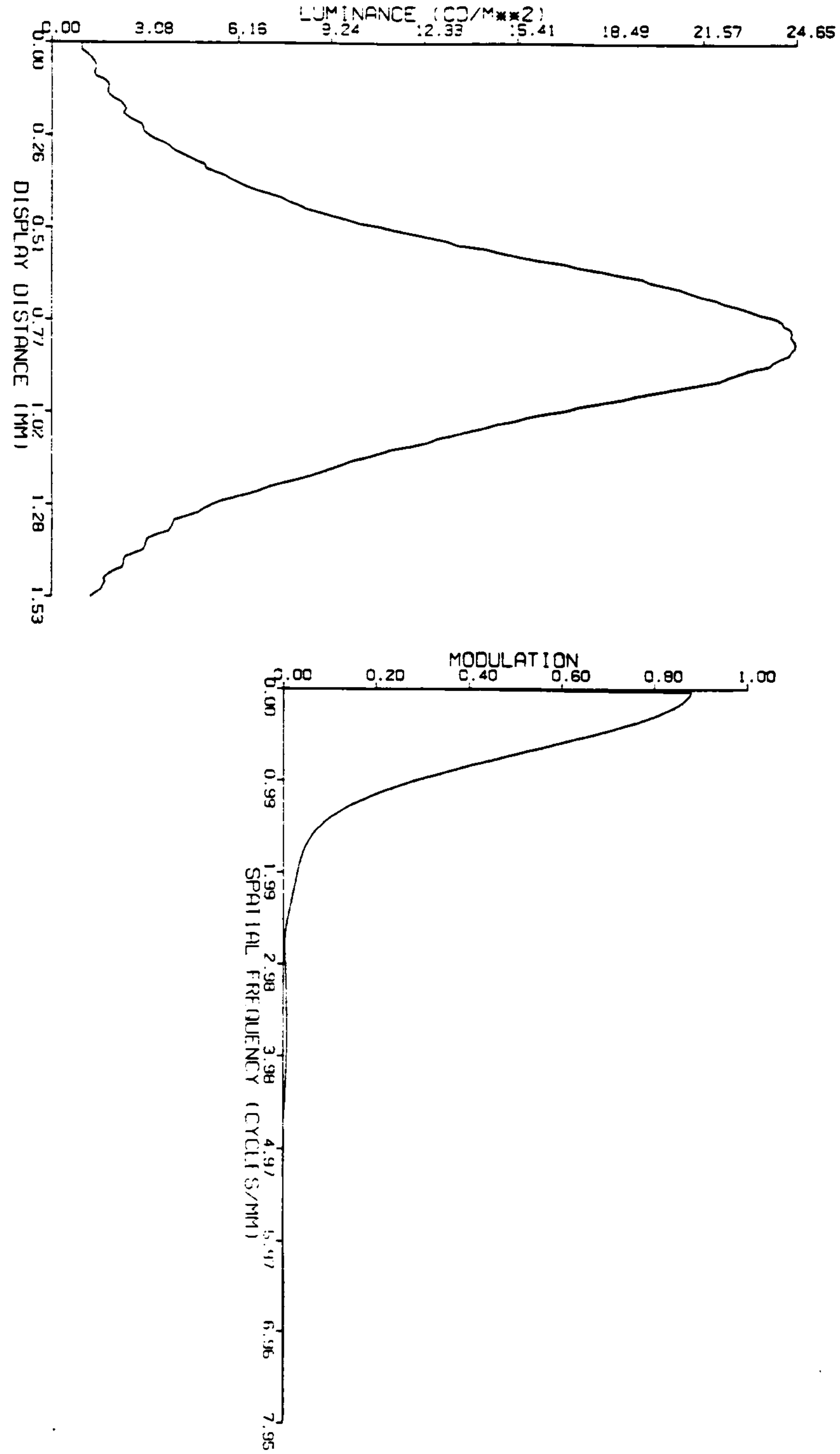




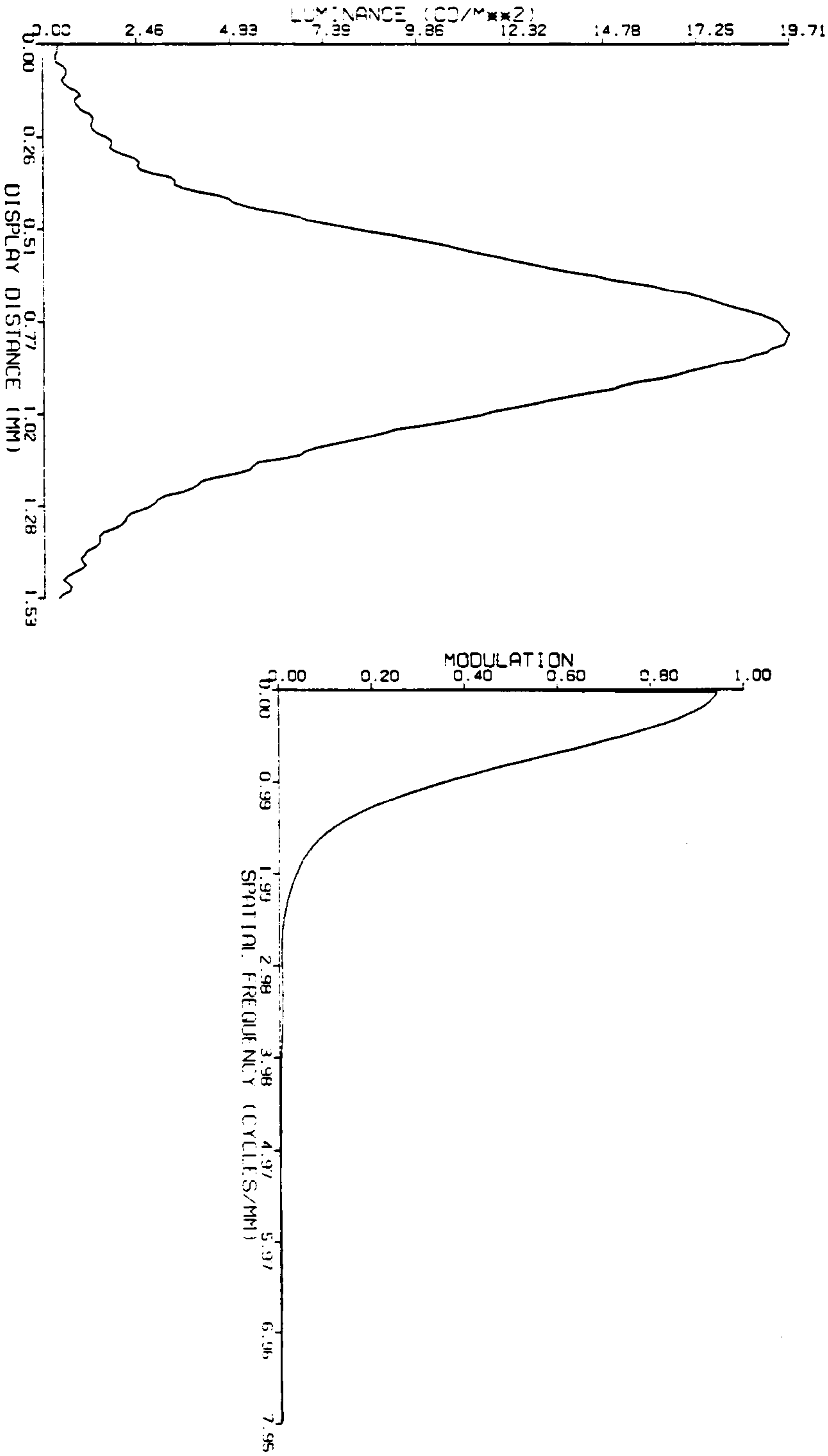
LSF and MTF for dark, G45-62% at high resolution.



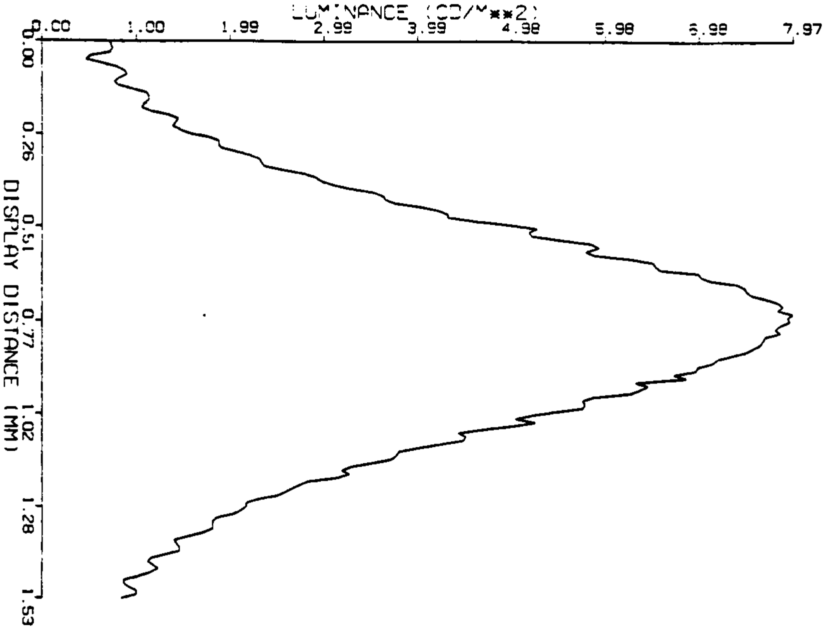
LSF and MTF for dark, G45-31% at high resolution.

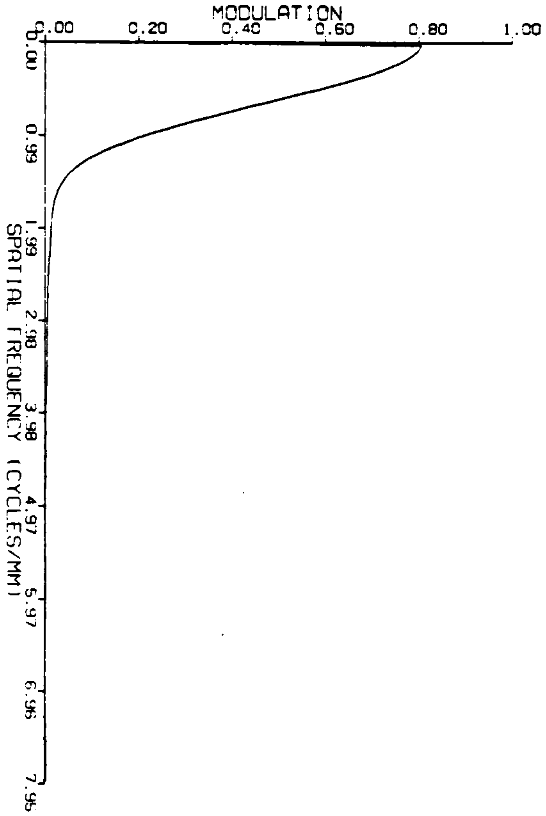


LSF and MTF for dark, G25-92% at high resolution.

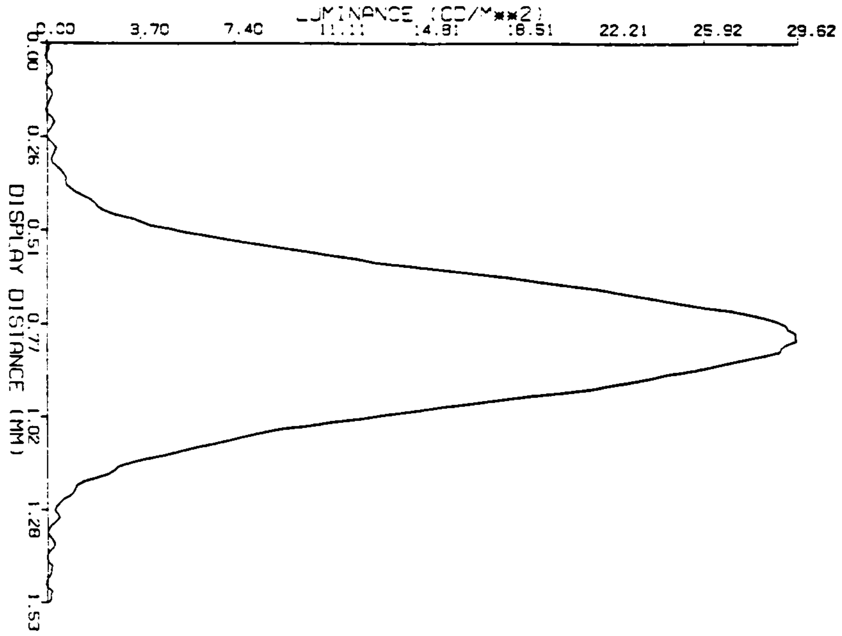


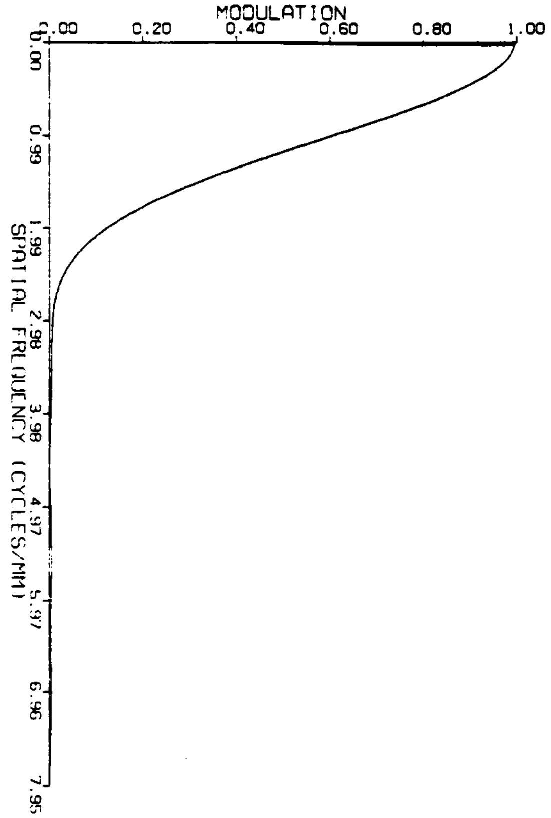
LSF and MTF for dark, G25-62% at high resolution.



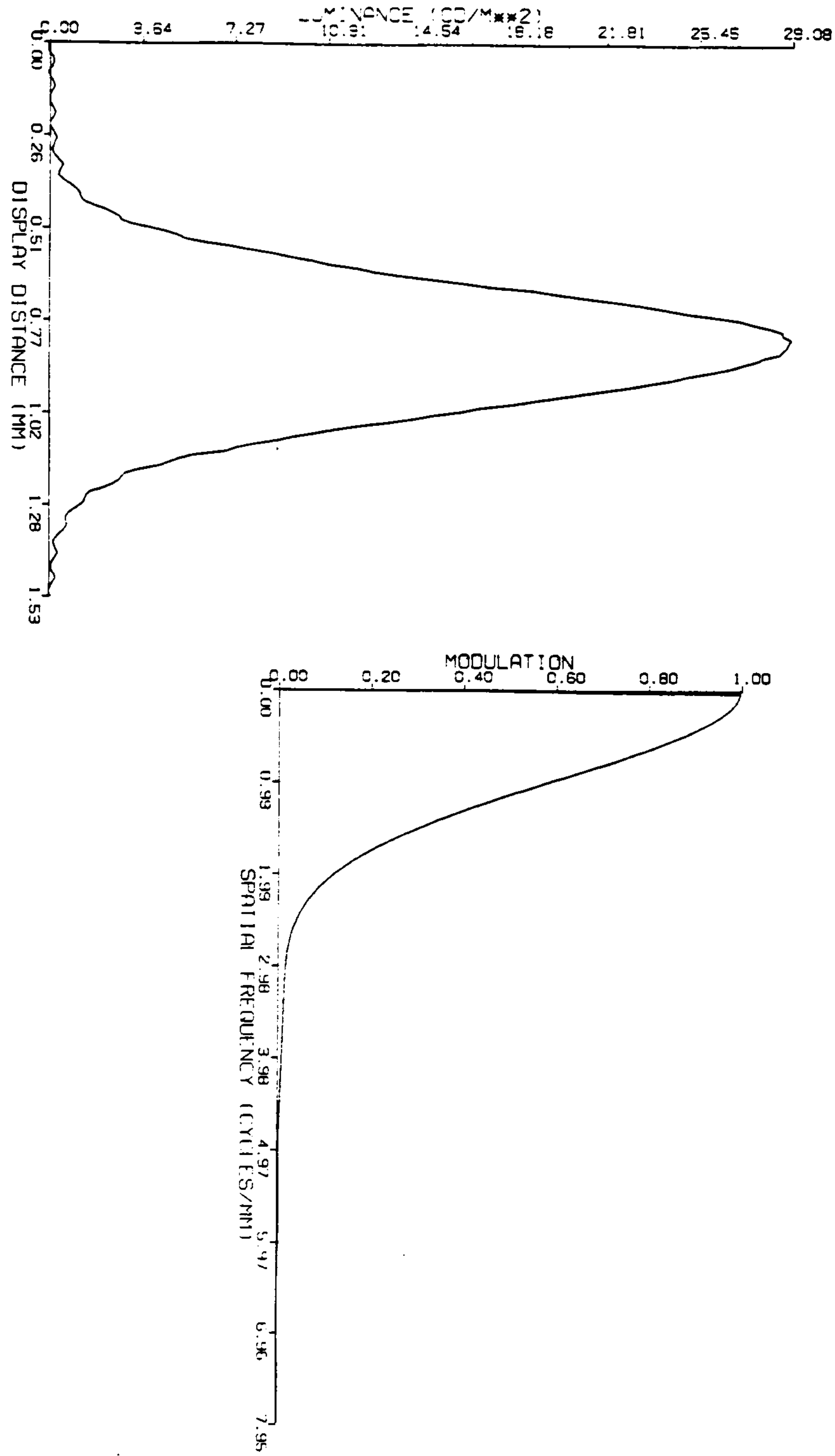


LSF and MTF for dark, G25-31% at high resolution.

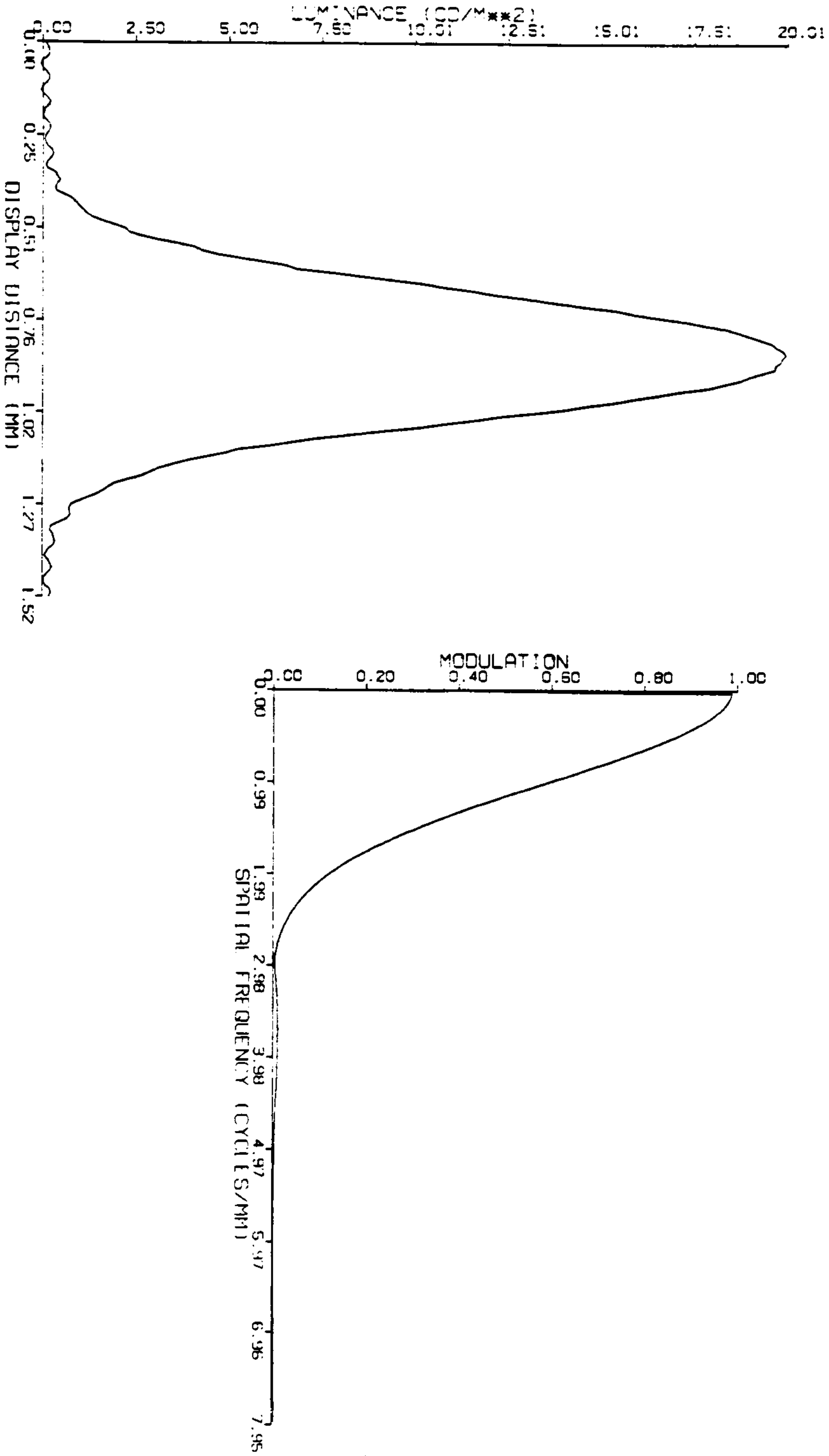




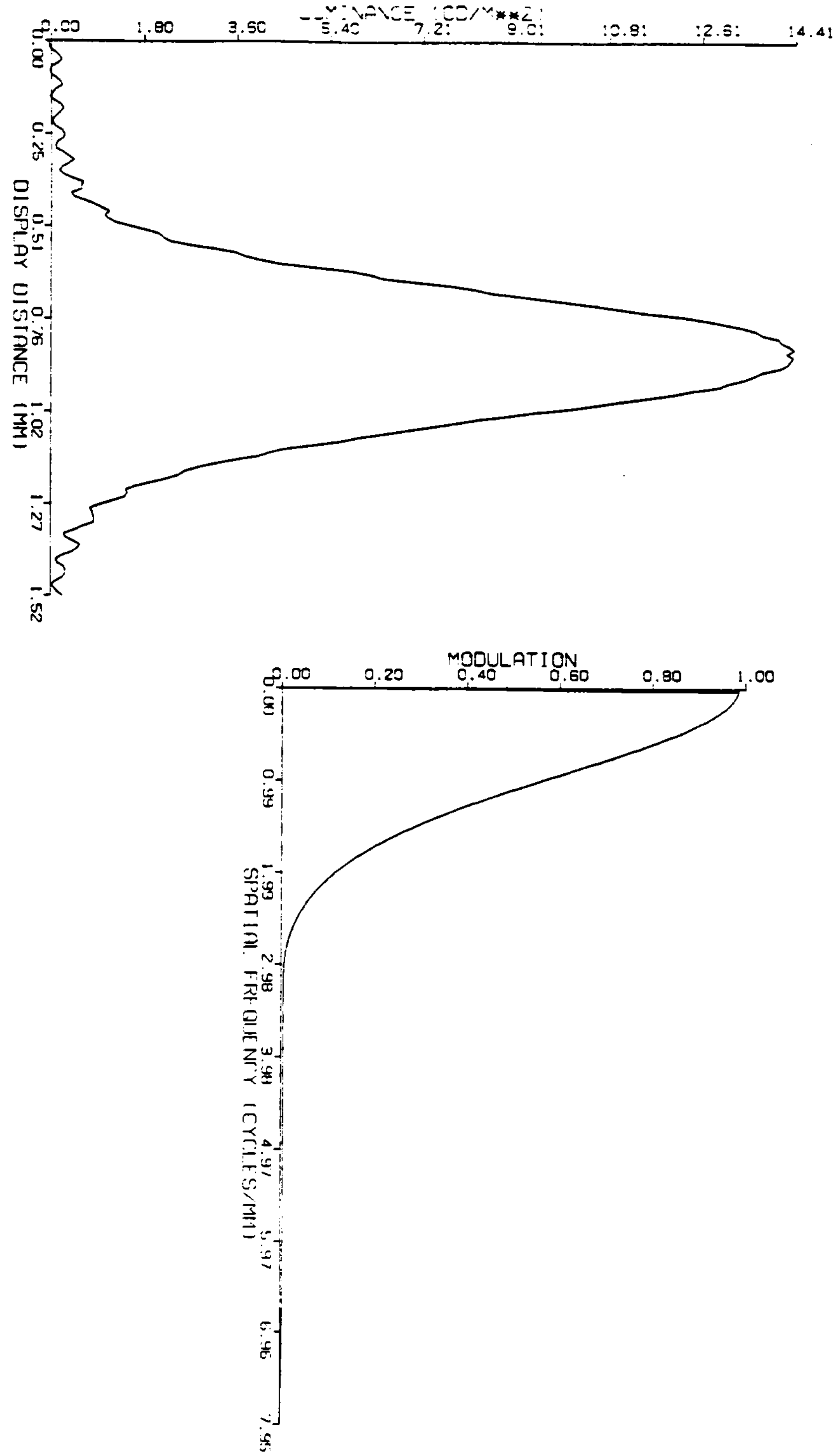
LSF and MTF for dark, POL-62%-QW at high resolution.



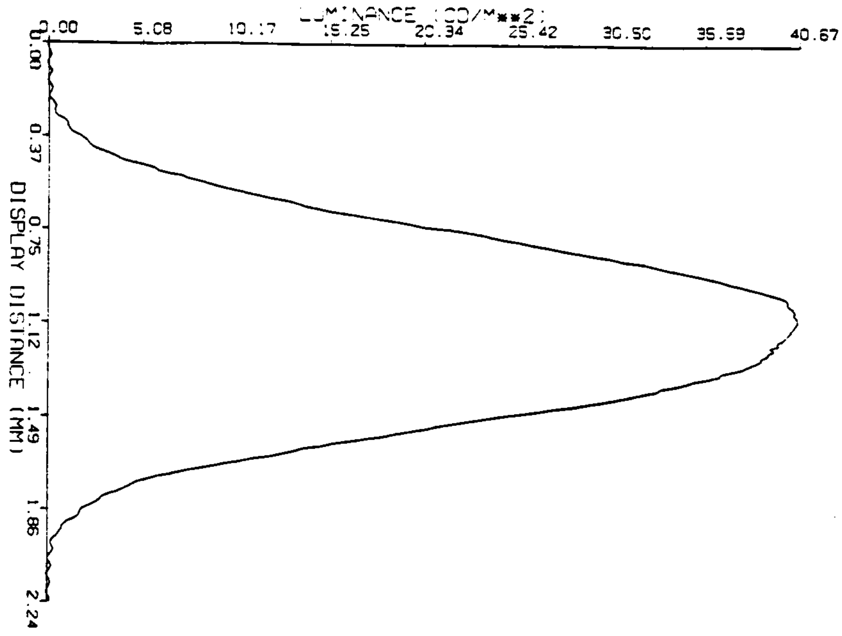
LSF and MTF for dark, G65-62%-QW at high resolution.

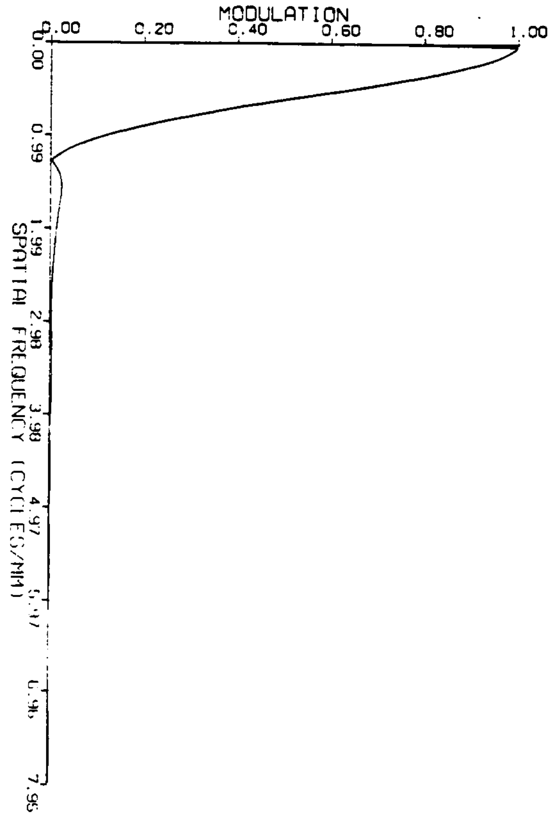


LSF and MTF for dark, MESH-5.7 at high resolution.

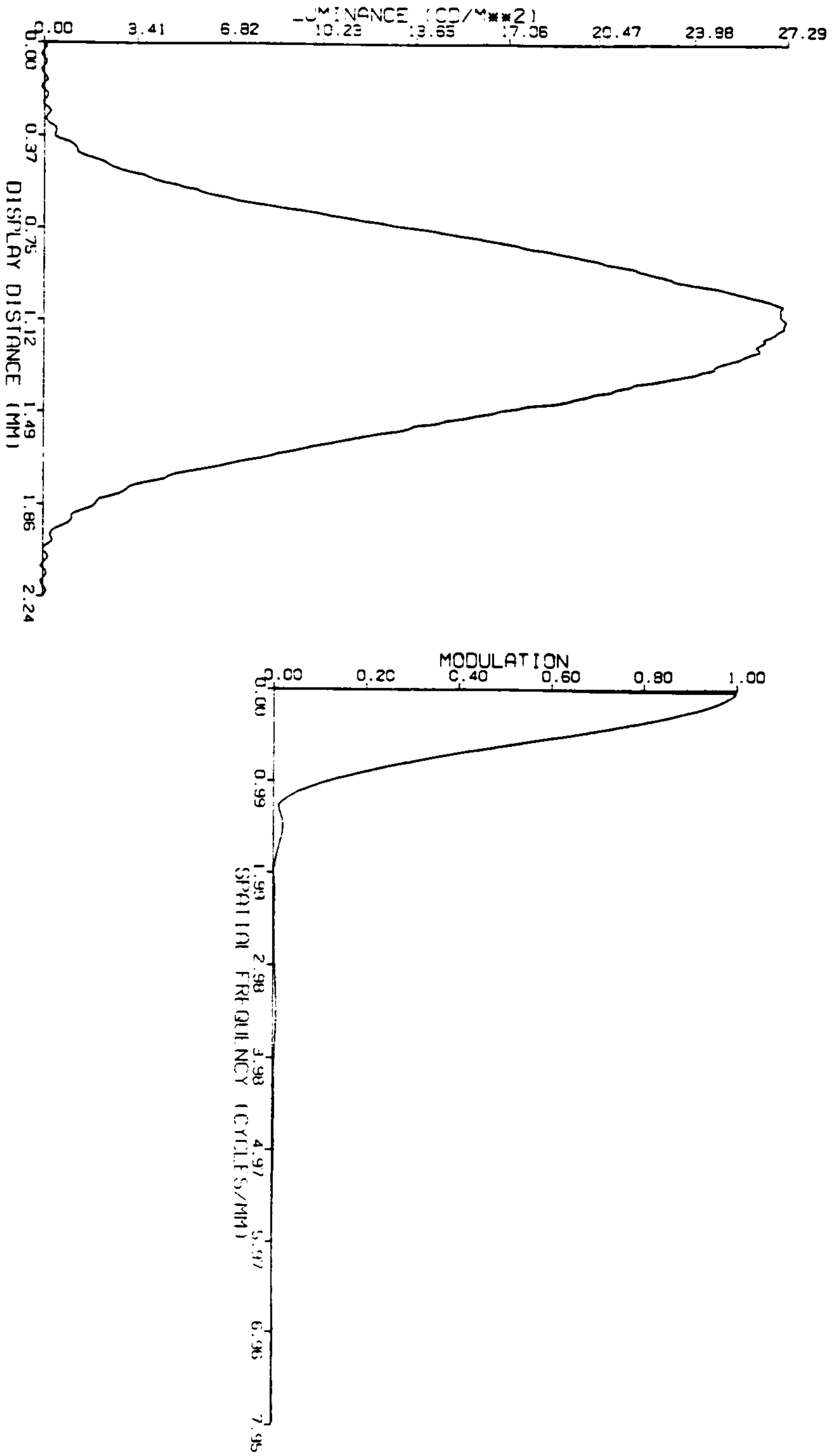


LSF and MTF for dark, MESH-7.4 at high resolution.

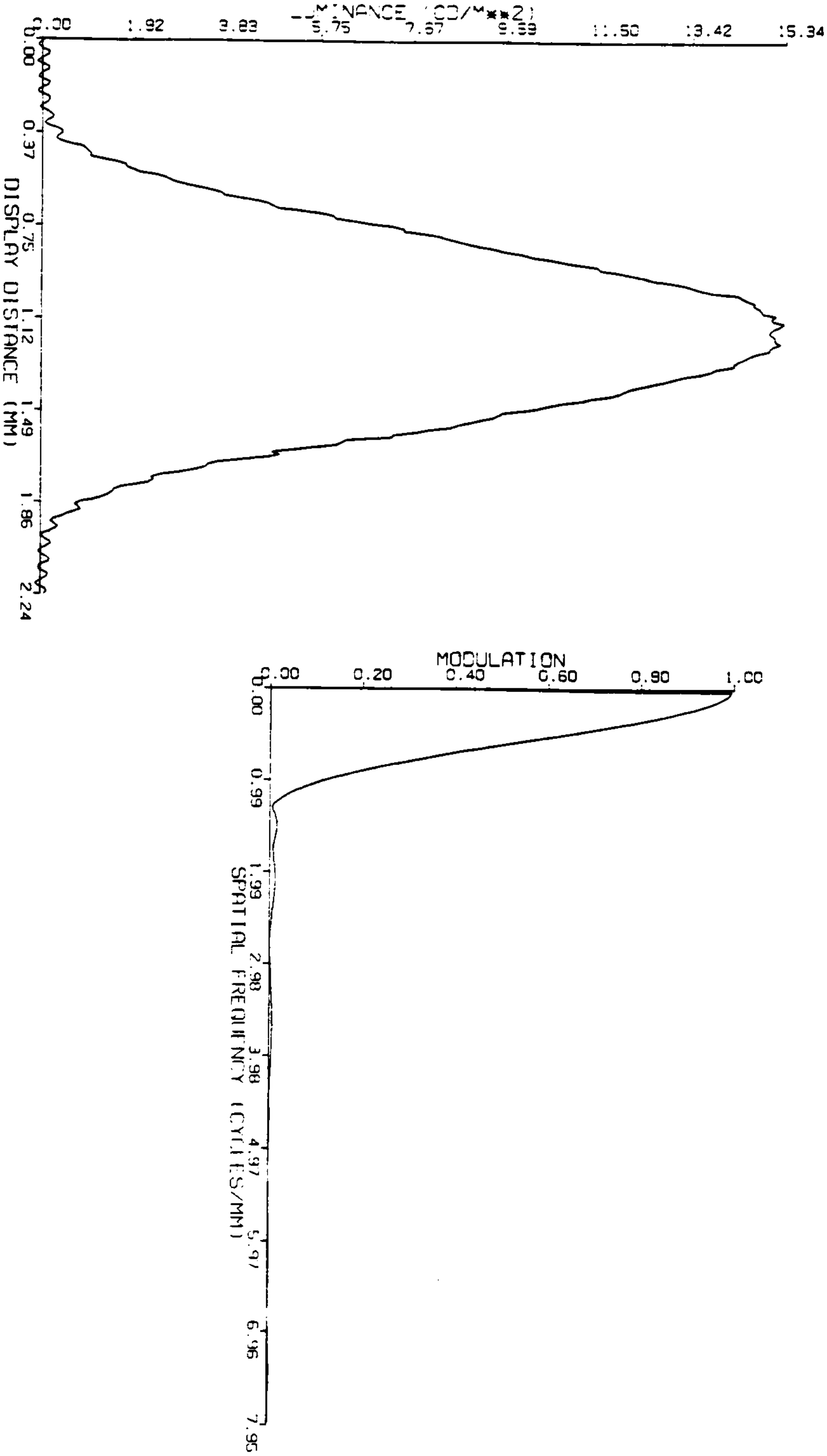




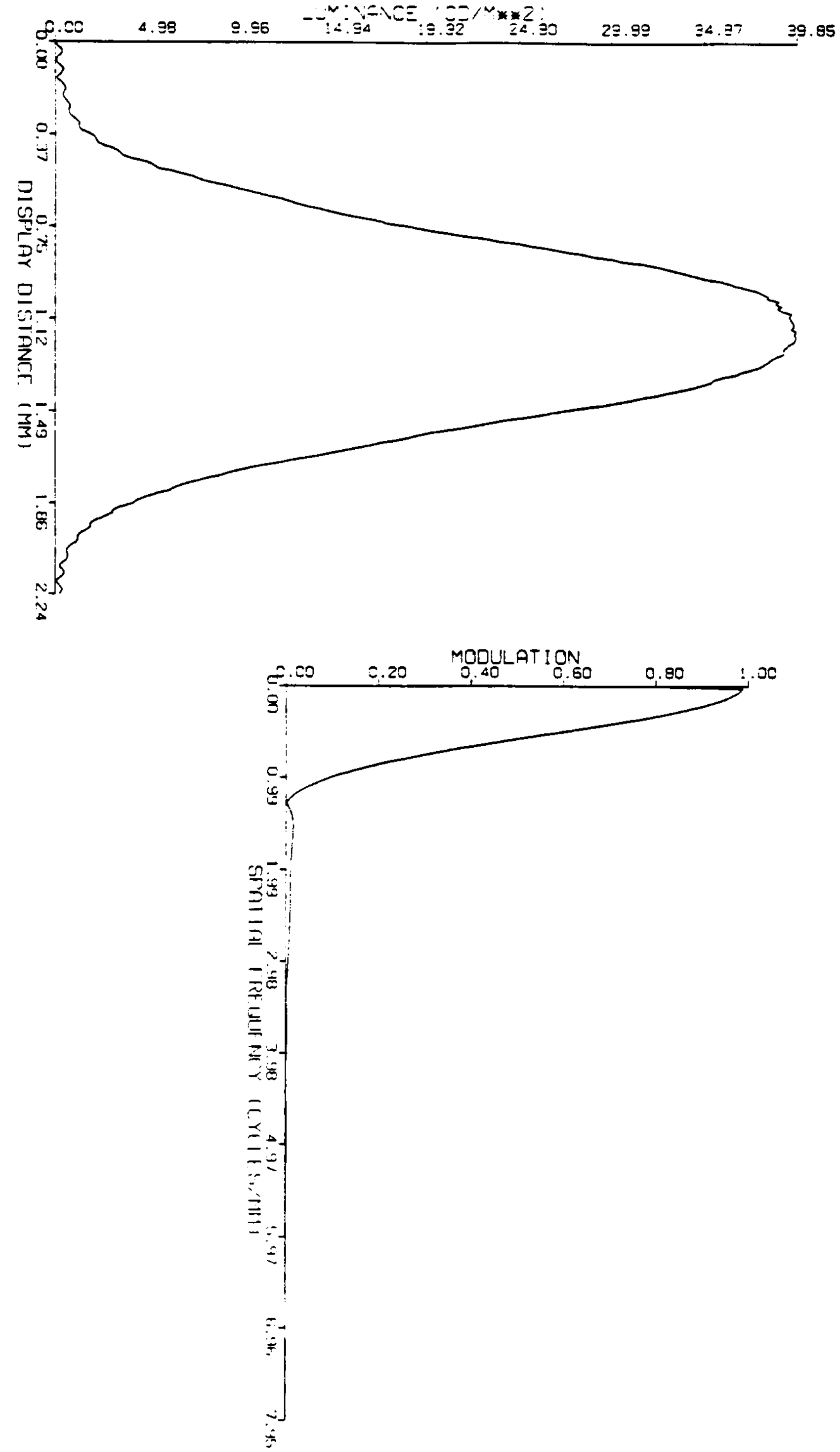
LSF and MTF for dark, POL-92% at low resolution.



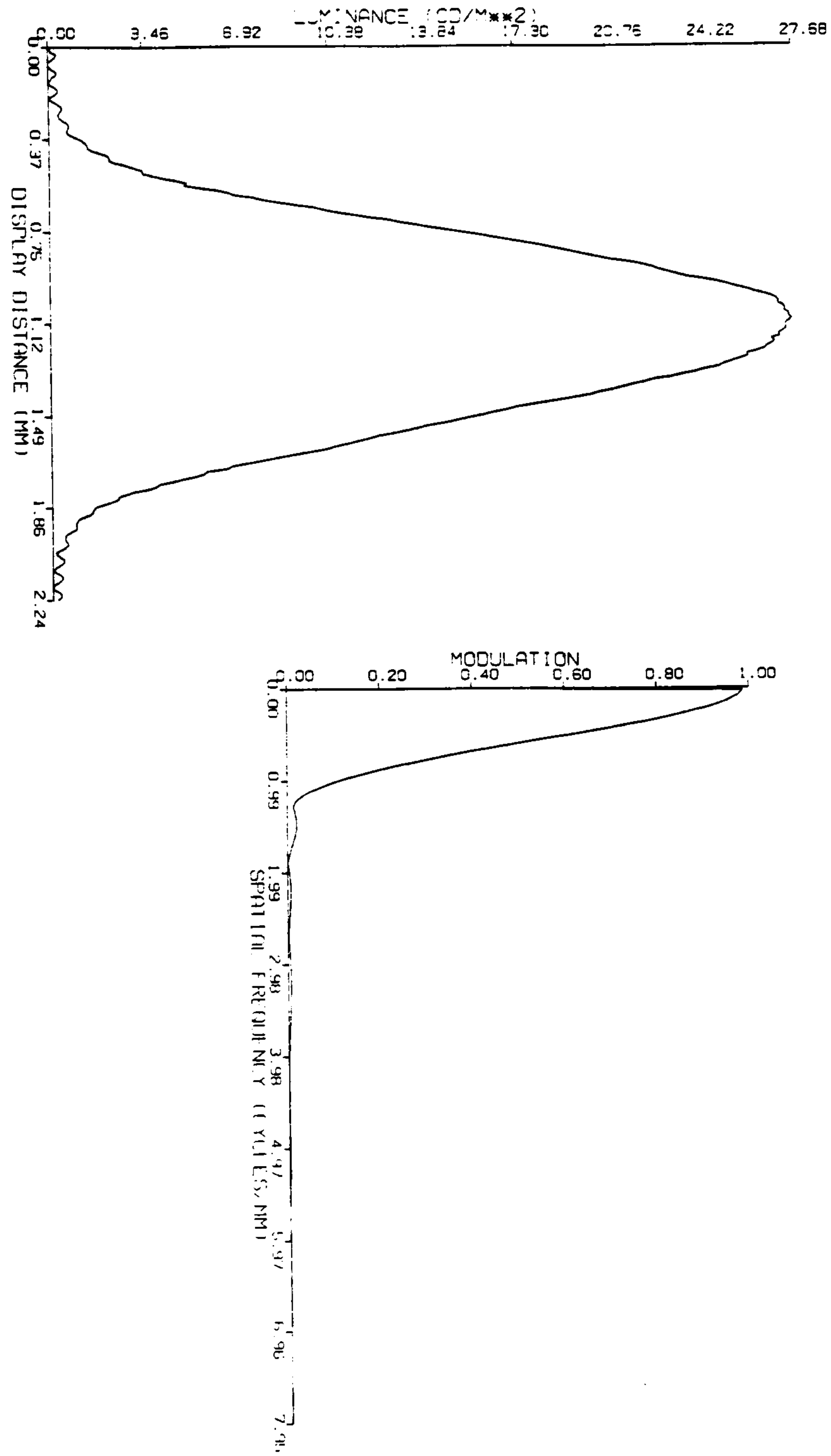
LSF and MTF for dark, POL-62% at low resolution.



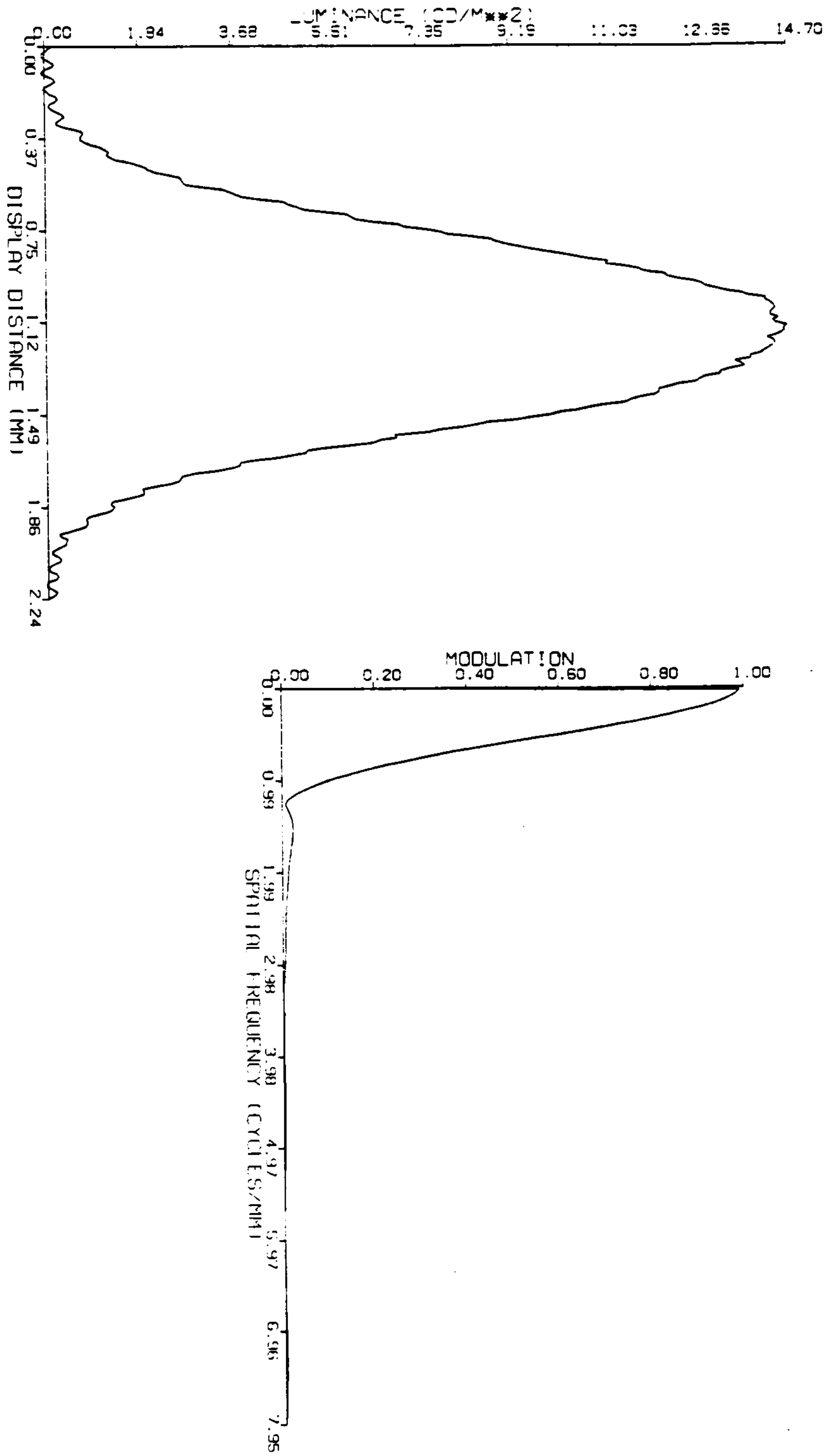
LSF and MTF for dark, POL-31% at low resolution.



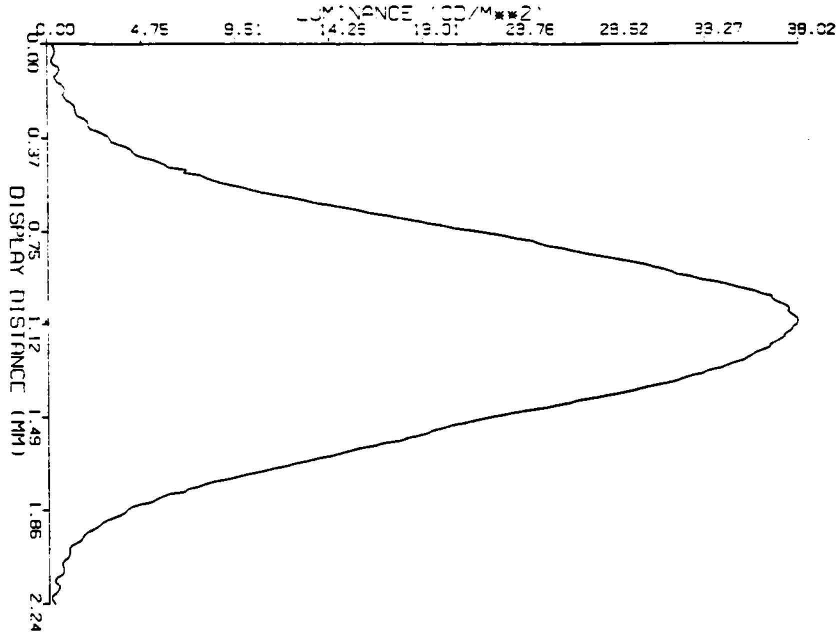
LSF and MTF for dark, G65-92% at low resolution.

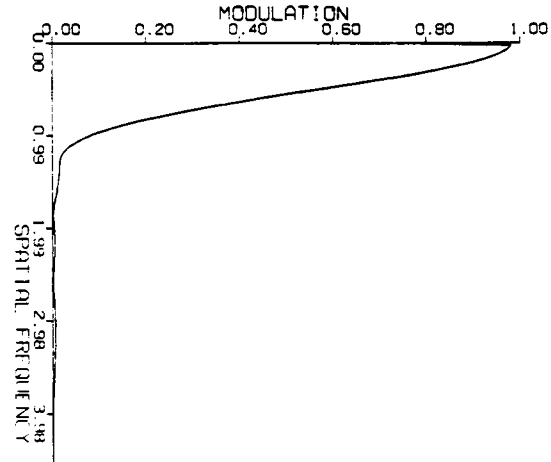


LSF and MTF for dark, G65-62% at low resolution.

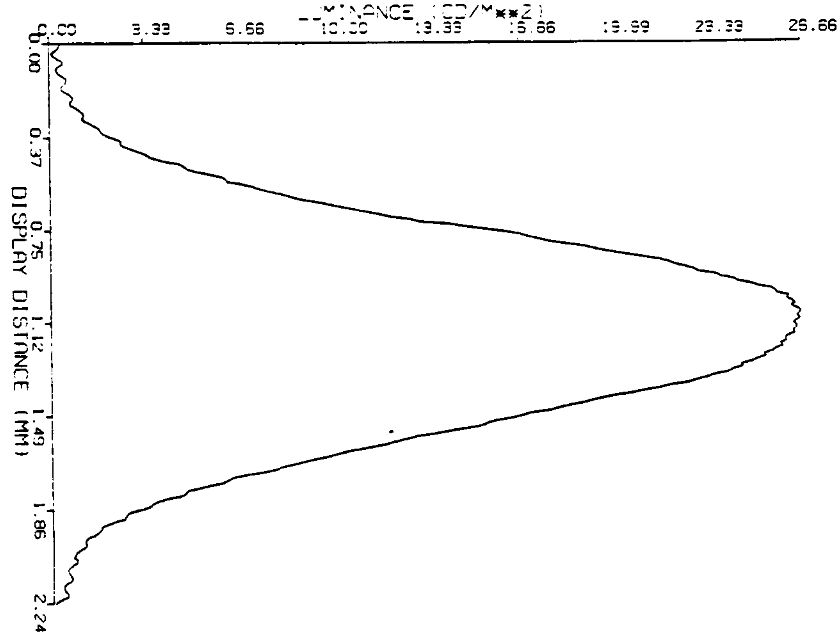


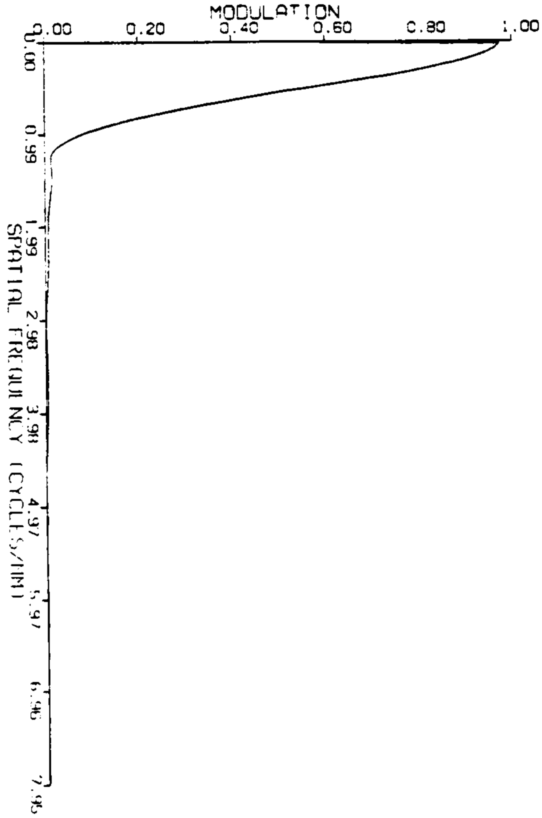
LSF and MTF for dark, G65-31% at low resolution.



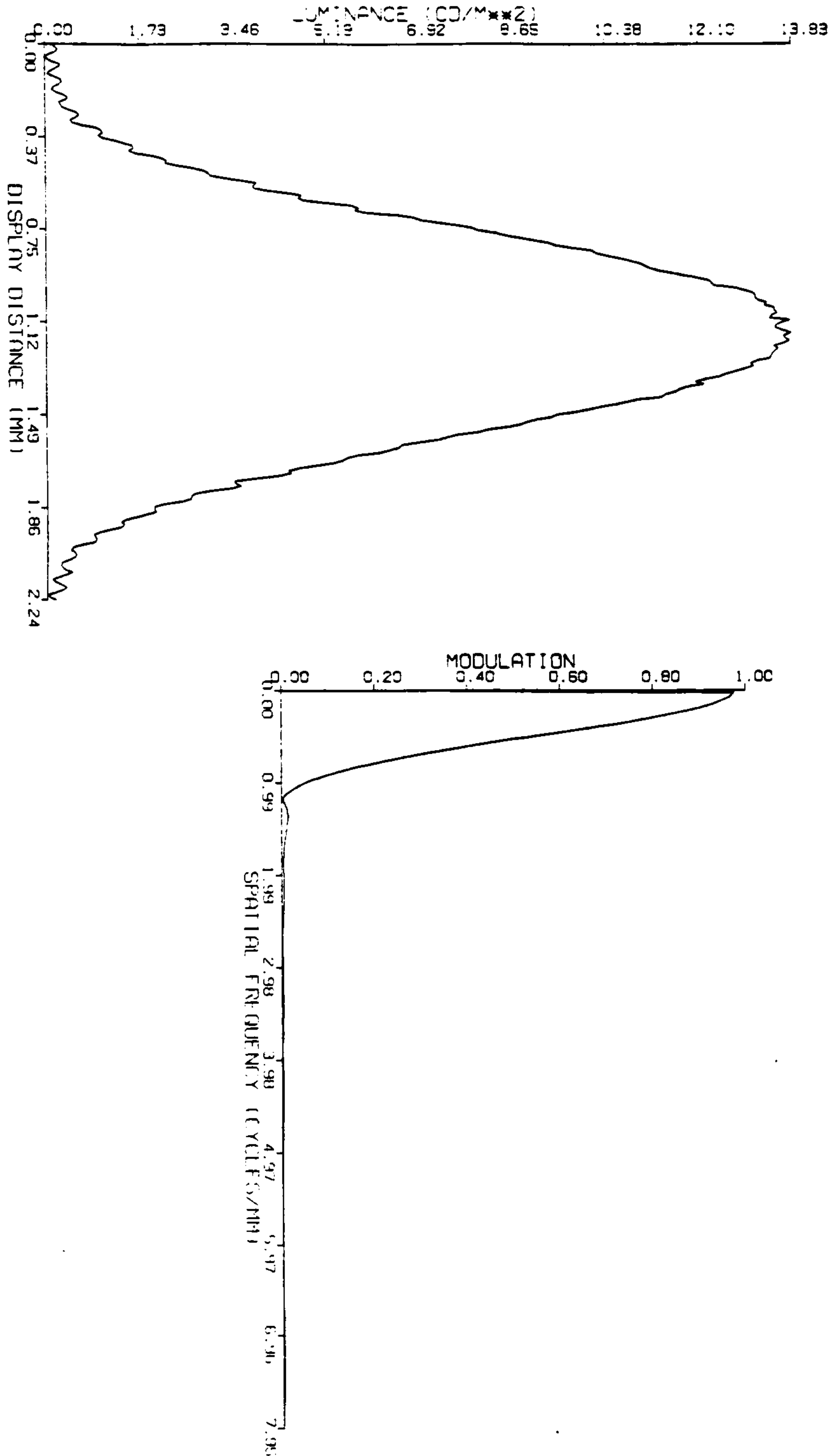


LSF and MTF for dark, G45-92% at low resolution.

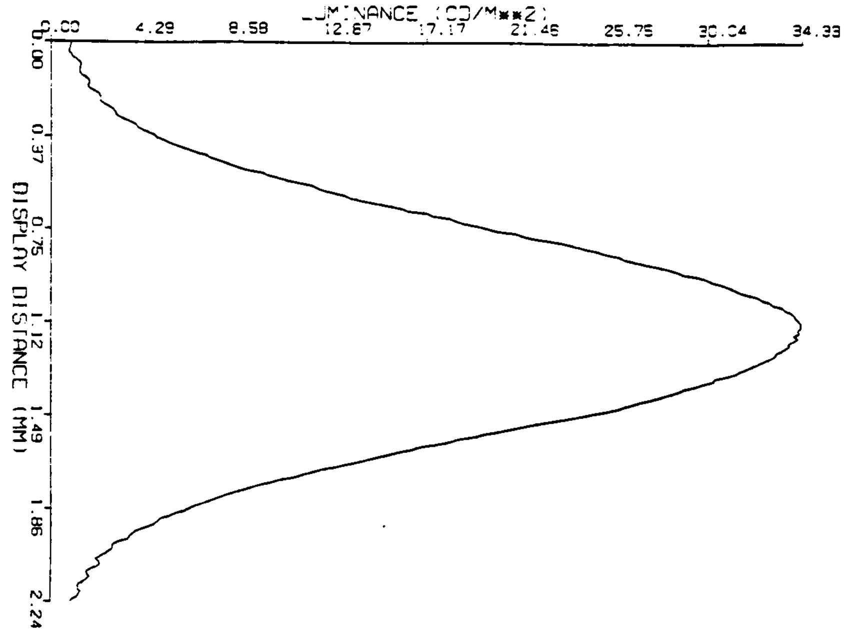


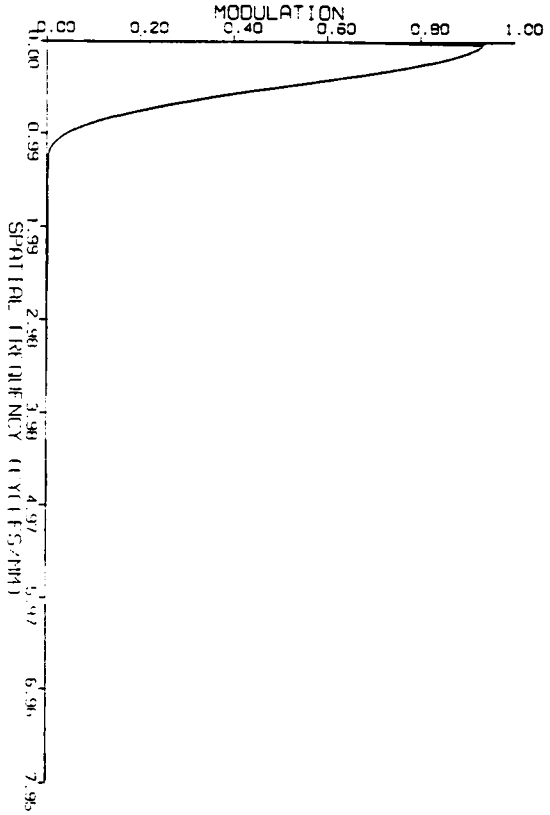


LSF and MTF for dark, G45-62% at low resolution.

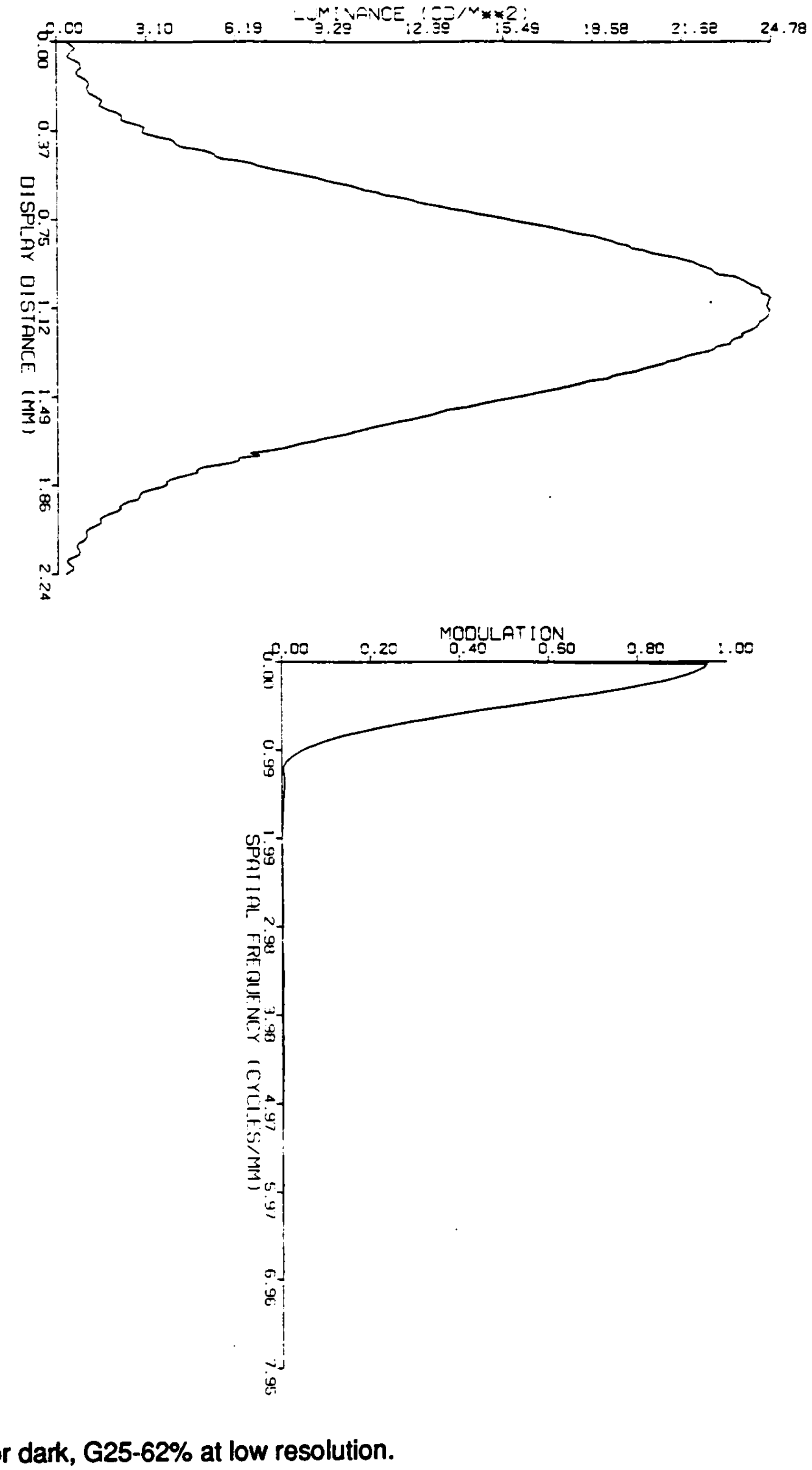


LSF and MTF for dark, G45-31% at low resolution.

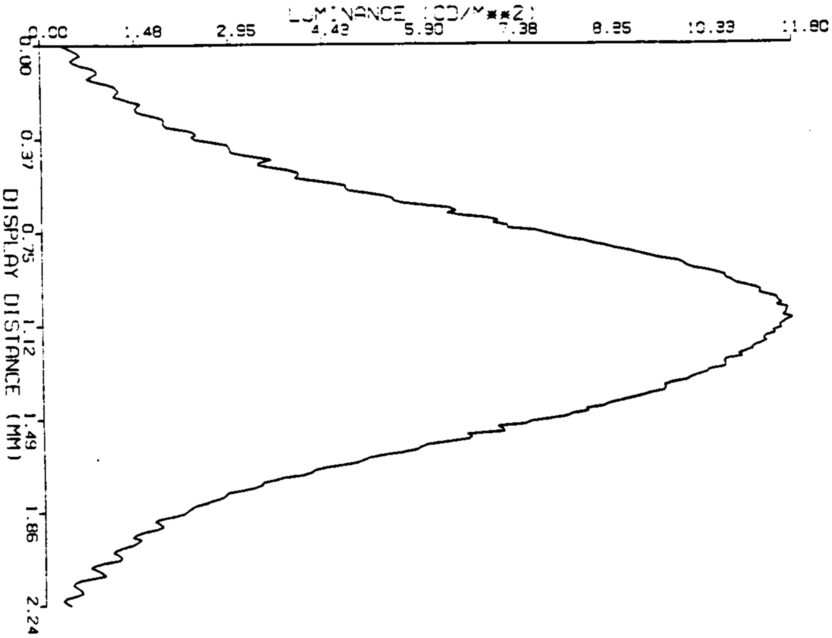


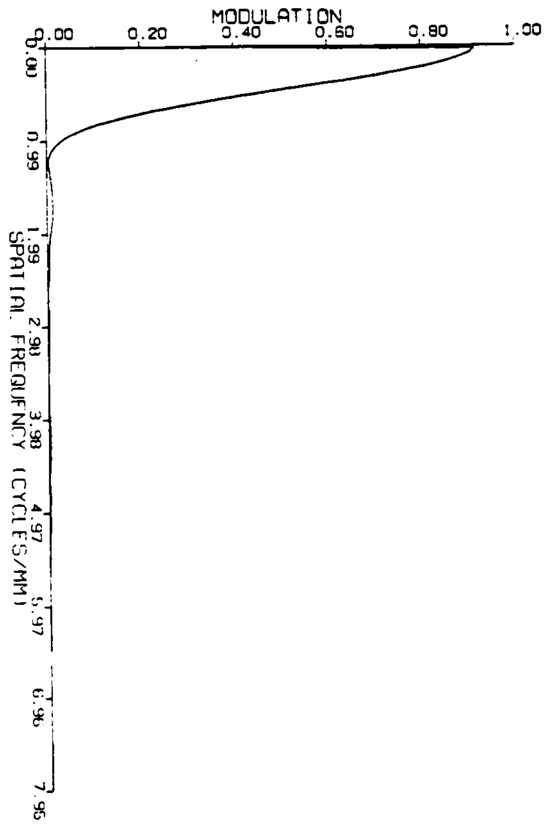


LSF and MTF for dark, G25-92% at low resolution.

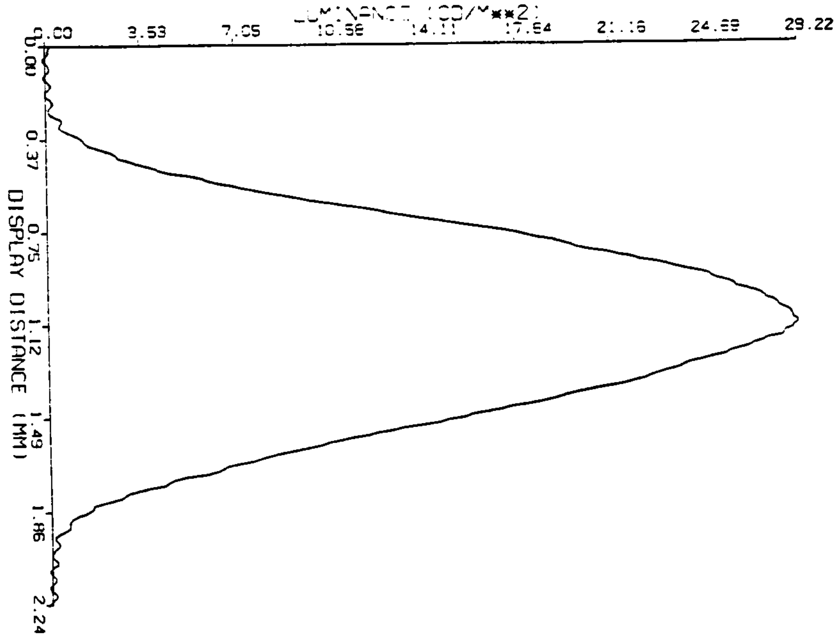


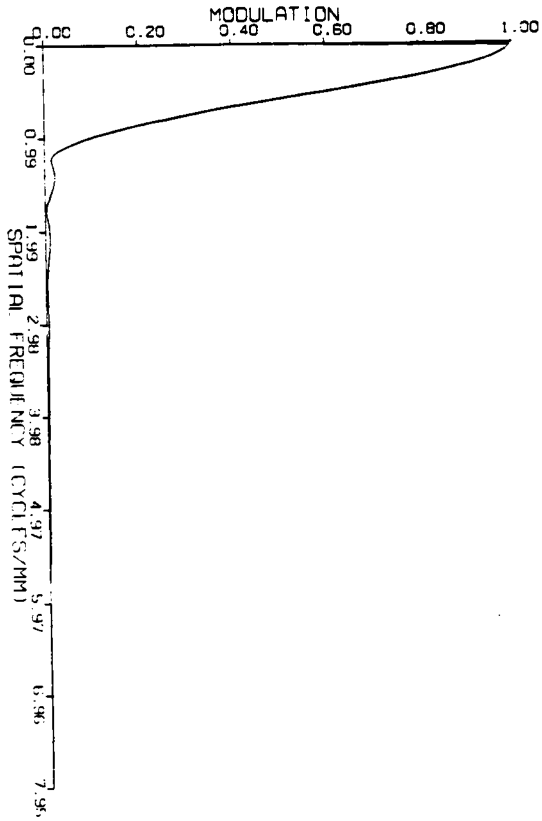
LSF and MTF for dark, G25-62% at low resolution.



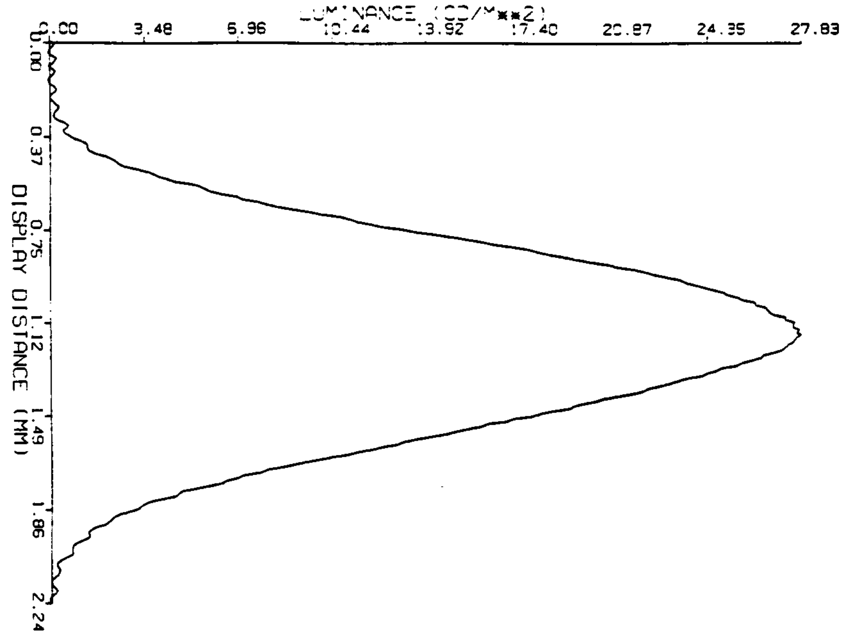


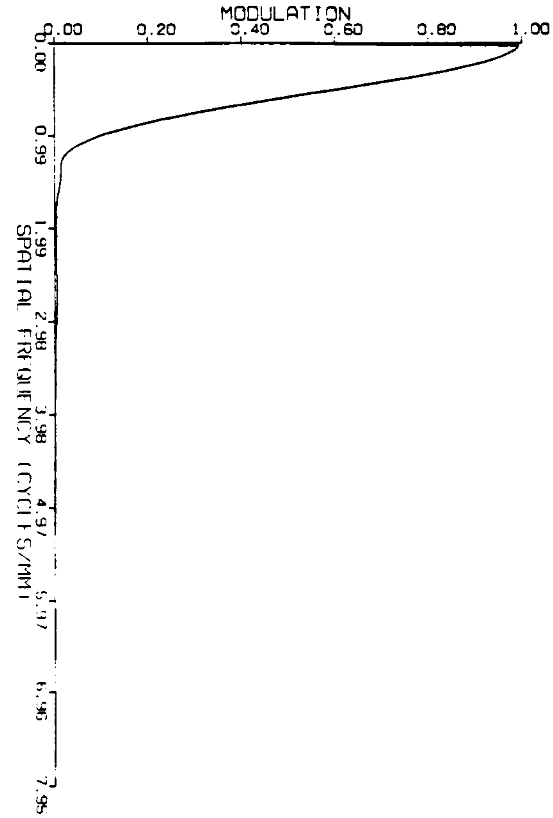
LSF and MTF for dark, G25-31% at low resolution.



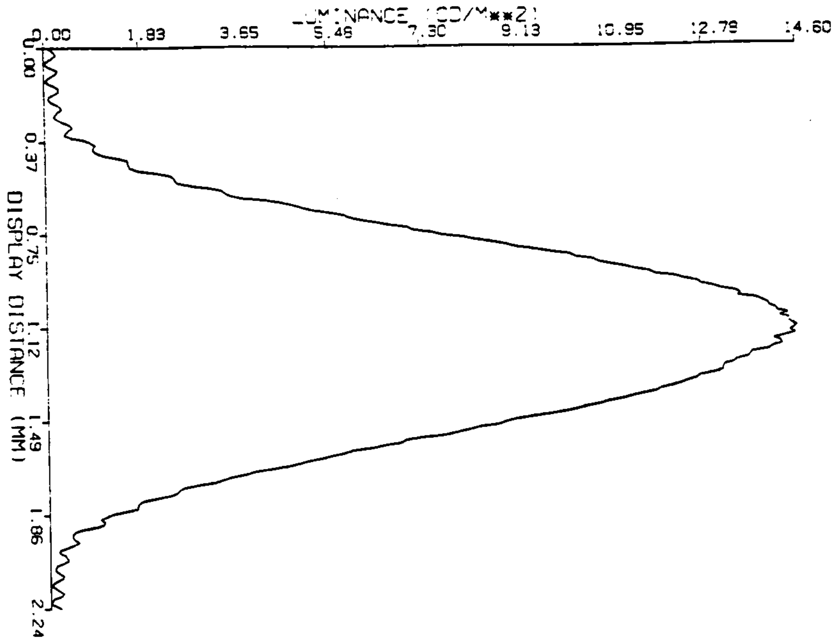


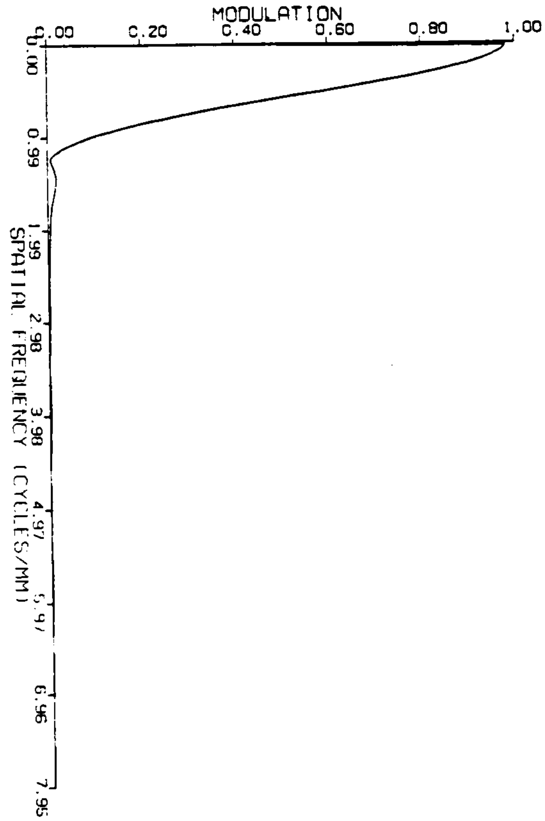
LSF and MTF for dark, POL-62%-QW at low resolution.



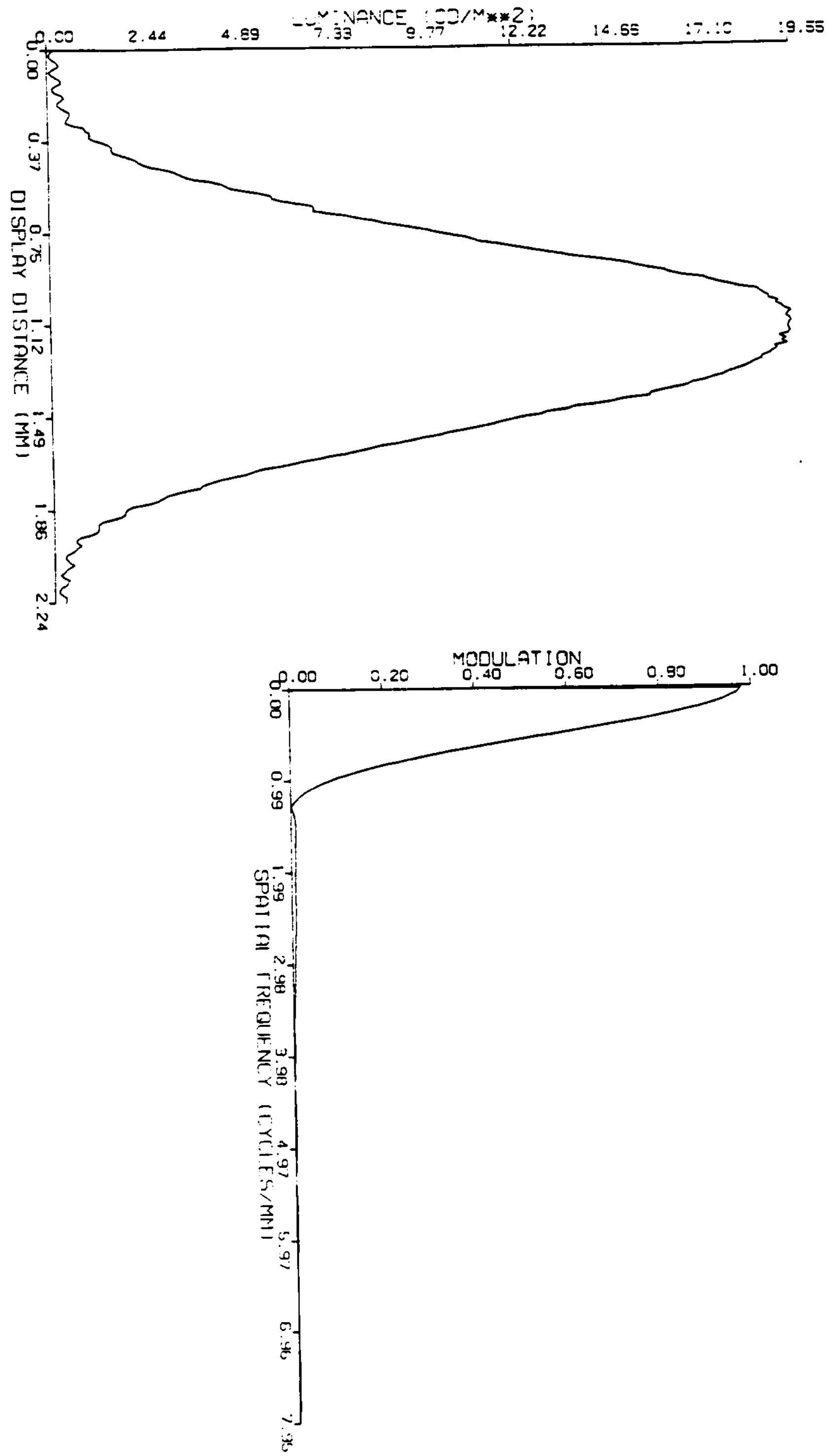


LSF and MTF for dark, G65-62%-QW at low resolution.





LSF and MTF for dark, MESH-5.7 at low resolution.



LSF and MTF for dark, MESH-7.4 at low resolution.

The vita has been removed from the scanned document