USAARL Report No. 98-31

Are New Image Quality Figures of Merit Needed for Flat Panel Displays?

By

Victor Klymenko
and
Thomas H. Harding

UES, Inc.
Dayton, Ohio

and

Clarence E. Rash

Aircrew Health and Performance Division

June 1998

U.S. Army Aeromedical Research Laboratory
Fort Rucker, Alabama 36362-0577

Notice

Qualified requesters

Qualified requesters may obtain copies from the Defense Technical Information Center (DTIC), Cameron Station, Alexandria, Virginia 22314. Orders will be expedited if placed through the librarian or other person designated to request documents from DTIC.

Change of address

Organizations receiving reports from the U.S. Army Aeromedical Research Laboratory on automatic mailing lists should confirm correct address when corresponding about laboratory reports.

Disposition

Destroy this document when it is no longer needed. Do not return it to the originator.

Disclaimer

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation. Citation of trade names in this report does not constitute an official Department of the Army endorsement or approval of the use of such commercial items.

Reviewed:

MORRIS R. LATTIMORE, JR.
Colonel, MS
Director, Aircrew Health &
Performance Division

JOHN A. CALDWELL, Ph.D.
Chairman, Scientific Review
Committee

Released for publication:

for CHERRY L. GAFFNEY
Colonel, MC, SFS
Commanding

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release, distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) USAARL Report No. 98-31	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION U.S. Army Aeromedical Research Laboratory 6b. OFFICE SYMBOL (If applicable) MCMR-UAD	7a. NAME OF MONITORING ORGANIZATION U.S. Army Medical Research and Materiel Command
6c. ADDRESS (City, State, and ZIP Code) P.O. Box 620577 Fort Rucker, AL 36362-0577	7b. ADDRESS (City, State, and ZIP Code) Fort Detrick Frederick, MD 21702-5012
8a. NAME OF FUNDING / SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
62787A	30162787A879	PB	178

11. TITLE (Include Security Classification)
(U)Are new image quality figures of merit needed for flat panel displays?

12. PERSONAL AUTHOR(S)
Klymenko, V., Harding, T.H., and Rash, C.E.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM TO	1998 June	33

16. SUPPLEMENTAL NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
figures of merit (FOM), flat panel display (FPD), cathode ray tube (CRT), image quality, visual perception

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
Image quality figures of merit (FOMs)are explained. The example of spatial resolution FOMs are used to illustrate the theoretical issues in the development and use of FOMs. FOMs for cathode ray tube (CRT)and flat panel display(FPD)technologies are compared. It was concluded that new FOMs need to be developed for FPDs, particularly for characterizing aspects of dynamic imagery.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Chief, Science Support Center	22b. TELEPHONE (Include Area Code) (334) 255-6907 22c. OFFICE SYMBOL MCMR-UAX-SI

DD Form 1473, JUN 86 — *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

Table of Contents

List of figures

Table of contents (continued)
List of figures (continued)

Page

List of tables

Figures of merit (FOMs) quantify various aspects of the image quality of visual displays

Visual displays are an increasingly important method of efficiently conveying information in military as well as civilian environments. It is important that the display present the information in an accurate and easily perceived manner. Ensuring operational efficiency and safety in informationally intensive environments, such as military cockpits, requires measures of the capacity of displays to transfer visual information to the human observer. Also, physical criteria evaluating the merits of old and new display technologies are needed in cost effectiveness analyses, system development, and procurement decision making. For the purpose of allowing consistent comparisons between displays, FOMs have been developed as criteria in order to objectively quantify various aspects of the image quality of displays.

A number of image quality metrics, or FOMs, have been developed to quantify the "goodness" of an image presented on a display. These FOMs have been used to quantify the quality of images displayed by cathode ray tubes (CRTs), a mature and ubiquitous technology, which, until recently, was the primary display technology. Newer technologies, known collectively as flat panel displays (FPDs), have been, and are being, developed to overcome some of the limitations of CRTs, such as the CRT's physical depth, weight, power consumption, and electromagnetic interference. FOMs are being both borrowed and newly developed to meet the need to assess these newer technologies. Here, we are concerned with the visually relevant FOMs, and not with other physical aspects of the display, such as power consumption, weight, etc.

An example of an image quality metric is one which concerns the range of luminances that can be presented simultaneously in an image. This range, described by the concept of *contrast*, defines the relationship between the minimum and maximum possible luminance values. There are various ways to formulate and define contrast for different purposes, such as contrast ratio, modulation contrast, etc. (Klymenko et al., 1997). Additional FOMs concern other aspects of displays such as its color, spatial, or temporal properties. Table 1 gives a list of common FOMs for CRTs classified into the spatial, temporal, luminance and spectral domains. As alternative display technologies emerge, different image quality metrics may need to be developed to capture new variables affecting the image.

FOMs are technology related

The displays based on the new flat panel technologies use different physical principles to present the image, and often introduce new variables that can affect the perception of the image. These new technologies include liquid crystal displays (LCDs), electroluminescent (EL) displays, plasma display panels (PDPs), light emitting diode (LED) displays and field emission displays to name a few (Harding et al., 1996; Castellano, 1992; and Tannas, 1995). If new variables are introduced, they need to be assessed. For example, the luminance and contrast of images on many LCDs vary as a function of viewing angle, where there is a reduction in off-axis luminance. Therefore, when evaluating these LCDs, the reduction in luminance as a function of viewing

Table 1.
Common CRT FOMs classified into four categories.

Spatial	Spectral	Luminance	Temporal
Viewing distance Resolution Spot size and shape Modulation transfer function Luminance Uniformity Signal/noise ratio Display size Aspect ratio Number of scan lines Interlace ratio Scan line spacing Linearity Focus	Color gamut Color purity	Luminance Gray shades Contrast ratio Halation Ambient illuminance Gamma Dynamic range	Frame rate Field rate Bandwidth

angle needs to be specified. The FOM quantifying luminance as a function of viewing angle is somewhat unique to this technology (Harding et al., 1996; Dragon, 1993) and was not a consideration in the traditional more familiar CRT displays (which are Lambertian sources). A totally new FOM was designed involving, among other things, measuring contrast as a function of spherical coordinates. Table 2 gives a list of common FOMs for FPDs.

Table 2.
Common FPD FOMs classified into four categories.

Spatial	Spectral	Luminance	Temporal
Pixel resolution (H x V) Pixel size Pixel shape Pixel pitch Subpixel configuration Number of defective (sub)pixels	Spectral distribution Color gamut Chromaticity	Peak luminance Luminance range Gray levels Contrast (ratio) Uniformity Viewing angle Reflectance ratio Halation	Refresh rate Update rate Pixel rise/fall times

The number of defective pixels is also a new FOM in FPDs because of the difficulties in manufacturing totally defect-free displays. This FOM will be standardized by the International Standards Organization (ISO) sometime in 1998 and will likely include a number of allowable defective pixels, and how they are clustered on the display. The proposed North Atlantic Treaty Organization (NATO) Standardization Agreement (STANAG) (Study# 7095), for example, requires no more than 0.01 percent pixel defects, and the ratio of display pixels in square centimeters to number of cluster defects (two or more adjacent defective pixels) to be not less than 16 to 1. Also, defective rows or columns are unacceptable, and cluster defects where critical information is presented are unacceptable. FOMs are being developed based on what is visually important, and potential problems with a new technology, which can be anything from its method of image generation to the quality control of its manufacturing process.

The development of new image quality FOMs is a complex theoretical issue involving a number of measurement questions: What is the nature of the image quality being measured, and what are the most appropriate metrical units to quantify it? Do the units capture relevant variations in image quality? And, finally, is the FOM comprehendible in terms of the human user?

What FOMs measure

A number of different image quality aspects need to be measured. For each, one needs to consider the nature of the image quality being measured, and the most appropriate metric to quantify it. As an example, we would use a luminance measure (e.g., foot-lamberts), not a radiance measure, to quantify the "brightness" of a display, since it is intended for human observers (not sensors). Consider how one would quantify the number of distinct luminances, or gray levels, between the maximum and minimum luminances of the display. This is an important factor in characterizing the quality of images a display can generate. This quantity describes how realistic and how informative the images can be. By tradition, this quantity has been approximated as the number of square root of two increments in luminance between minimum and maximum luminance. This quantity, known as the shades of gray (SOG) FOM, reflects a compromise between engineering requirements (for simplicity) and psychophysical data. The SOG FOM is derived from the luminance range, or contrast, FOM of the display, by formula. The formula assumes the luminance range can be continuously sampled. This is reasonable and appropriate for analog CRTs, which can present any incremental gray level between the luminance extremes, (i.e., the luminance can take on any value over a continuous range). However, the use of the SOG FOM to characterize the new digital FPDs is misinformative because the continuous sampling assumption does not apply to FPDs, where gray scales are generated differently. FPD gray scales are discrete in that the luminance values are limited to specific values in the luminance range, (i.e., the range is digitally sampled). As a result, any computed SOG value, would be totally misleading as a descriptor of the display's grey scale capacity. What the display's capacity actually is, as we shall discuss later, depends on the placement of the discrete samples within the luminance range. [See Klymenko et al. (1997) for an extensive discussion.]

Another example of a descriptor of image quality, which when first used misled the general public and professionals alike, is the advertisement tactic of many computer graphic workstation manufacturers describing their displays as capable of producing millions of colors, all at once. Although these numbers are derived simply by multiplying the numbers of red, green and blue driving levels controllable by software, it implied, in the minds of many buyers, enormous new vistas of color sensation. In fact, in terms of human vision, this number is of no consequence whatsoever. The display might generate all these colors, but how large is the range (or gamut) of these colors in terms of human color space, and how many different colors can we actually see. A million levels of indistinguishable cyans is just that. Failure to consider the observer can be very misleading.

Many of the different FOMs quantify display variables in one or more of three ways: Range, resolution, and (form of) sampling. Range is the maximum extent of a variable. For example, the maximum contrast of a display, or luminance range, represents the full range over which the luminance can vary. Resolution is how finely the range of the variable is divided. This includes spatial and temporal resolution, and for luminance, the number of gray levels available in the luminance range, e.g., 16 or 256. Sampling is the placement of the levels within a range. For example, where in the luminance range are those 16 or 256 gray levels? In analog CRTs, this can appropriately be specified by the gamma curve. A gamma curve can be misleading in digital displays with few grey levels. Which of these three ways---range, resolution, or sampling---is the basis of a particular FOM parameter should be kept clearly in mind.

For example, the previously mentioned SOG FOM has been inappropriately used to characterize digital FPDs. This should have been avoided by a consideration of the units of the FOM variables. Converting from physical contrast, a physical range FOM, to SOG, a (visual) brightness resolution FOM, was natural for analog CRTs, given the assumption of continuous (physical) sampling, the virtually unlimited resolution of the luminance levels of the CRT's video signal. But, because digital CRTs and FPDs use discrete numbers of driving levels (dls) to code luminance, the SOG FOM is not directly applicable. The luminance range, without the continuous sampling assumption, does not imply the SOG formulated number of (visually) available brightness levels. Another example is the millions of colors advertisement tactic noted above, which is based on quantifying physical color resolution (number of dls), but misleads the public into assuming it is somehow related to the perceptual color range seen by the observer.

As we have seen, we need to explicitly consider what a metric means in terms of the human observer. The SOG FOM incorporates the fact that the human brightness scale is not linearly related to the physical luminance scale. [The square root of two rule reflects the fact that greater luminance steps are required to noticeably increase the brightness of brighter objects than dimmer objects.] When assessing the image quality of different technology displays, metrics of image quality should be either formulated, or, at least comprehendible, in terms of their value to the human user. This means its value should be relevant in terms of human vision per se (the ability to see it) and/or in terms of human performance (the human's capacity to respond to it appropriately and efficiently). The ultimate goal of displays is the useful conveyance of visual information to the human observer. For our purposes, image quality metrics should be considered primarily in terms of visual information, with special attention to the difference between the potential informational capacity of the displayed image and the actual visual information that the image can convey to the human observer. This is based on how well the properties of the display match the visual requirements of the observer.

FOMs sometimes reflect visual information conveyable to human observers

How much information does a displayed image present, and, how much of that information is conveyed to the observer? First, let us consider what constitutes visual information. Basically, visual information, is patterned spatial information conveyed (primarily) by differences in luminance, known as contrast. The potential information content within a single displayed image can be (loosely) defined as the product of the following factors: Number of gray (luminance) levels, number of different colors for polychrome displays, and the number of picture elements, or pixels. This result multiplied by the update rate of the display (how fast the image can change) gives the informational flow rate of the display (Biberman and Tsou, 1991). This defines abstractly how many different patterns the device can present. However, this abstract description of potential visual information does not take into account the magnitude or type of information that human observers can usefully receive. The human observer's visual/cognitive system has specific visual (and cognitive) capacities and limitations. Human characteristics include, for example, the human's capacity for global pattern perception and insensitivity to minute local differences, the human's differential sensitivity across the spatial frequency spectrum, and the human's threshold limitations in distinguishing luminance levels (as compared to a photometer).

For a display to be useful in transferring information, there must be a match between the capacity for visual information presentation by the display and the capacity for visual information reception by the observer. FOMs are most useful in quantifying and distinguishing displays when they reflect this match.

As an example of the need to incorporate the match between display and human, consider the case of a monochrome matrix display, where each pixel can be either black or white. In terms of information, we would give the factor, the number of gray levels, a value of two, because this image, known as a binary image, has two luminances, or gray levels. It is, therefore, limited in the quality of the real imagery it can present because every pixel in the picture is either black or white. Many things that normally are distinguishable with unaided vision, such as the difference between a black circle and a shaded sphere, are, with only two gray levels, two indistinguishable black circles. Many items would become completely unidentifiable, that is many of the subtle pattern differences used by human observers would be lost. This is especially crucial in difficult tasks such as distinguishing and identifying small targets. Now consider, displaying the same image with 256, rather than two, gray levels. The image, with near photographic realism, becomes almost as salient as the real world, where targets are easier to see.

But, now, consider the case where a poor quality display device presents the same image, also with 256 gray levels, but the human user can not discern the difference between the lowest 251 levels, which all look equally black. Physically, as measured with a photometer, the image presents 256 distinct gray levels, but, visually, as seen by the user, the image is only conveying six discernibly different gray levels to the human visual system, what we define as perceivable gray levels (PGLs). Therefore, to the human visual system, the image is indistinguishable from an image containing much less information because the sampling of the gray scale is a suboptimal match to human requirements for receiving visual information. In the conveyance of information, there needs to be a match between the presentation of image information and the reception of visual information. [Here, we focus on the perceptual aspects of the reception of visual information, although we note there are cognitive aspects concerning the requirements of human attention and memory.]

In the different FOMs developed for characterizing the image quality of CRTs, this match between the display and the observer is sometimes explicit, as in the SOG FOM, and sometimes implicit and/or undefined, as in the display descriptions claiming millions of colors. Often additional data need to be incorporated in order to understand what the FOM means in terms of the human observer. This is perhaps the most difficult and controversial part in the development of image quality FOMs. One needs to understand how to interpret often esoteric and difficult psychophysical research when incorporating the requirements of the human user. Below, as an illustration, we discuss the development of spatial resolution FOMs. As we shall see, spatial resolution FOMs used to characterize CRT displays (the standard display technology throughout most of this century) are still, even now, controversial in the ways in which they incorporate the requirements of the human observer. In many cases, what the value of a particular FOM means in terms of the observer is not known, or only known over a limited range or a limited set of experiments. As an example of this, refer to the Design Handbook for Imagery Interpretation (Farrell and Booth, 1984), the best compendium of display requirements when it was published, where innumerable relevant psychophysical experiments are described, but where caveats limit the findings to the specific experimental conditions.

As we have seen, the different types of image display FOMs, the spatial, spectral, geometric, temporal, and luminance FOMs for CRTs and FPDs, have implications for visual performance based on the match between the characteristics of the image and the requirements of the user. The requirements of the user encompasses more than just the accommodation of the display to the visual capacity of the user. It also means taking into account the visual task required of the observer. In other words, taking into account the purpose for which the display is used by the observer.

FOMs and the purpose of the visual display

The purposes for which a display is used will determine the most crucial FOMs for characterizing the image quality of a particular technology. Consider some examples. For the purpose of medical diagnosis in radiology, where image quality is tied to the user's ability to detect subtle targets, Roehrig et al. (1990) consider the most essential parameters, FOMs, for describing CRT display quality to be spatial resolution, noise, characteristic curve, and absolute luminance, with the additional parameters of brightness uniformity, veiling glare and distortion also being important. For the purpose of computer graphics, where subjective aesthetic quality and realism are of paramount importance, Oakley (1984) states that the image quality of CRTs is most usefully described by the parameters of size, resolution, contrast ratio, geometric distortion, flicker, video bandwidth, and shades of gray (and chromaticity and convergence for color displays). Oakley (1984), also, states that additional important variables are the spot size and shape, which are determined by how well the beam is controlled and focused, and the physical characteristics of the phosphors including their luminance efficiency, spectral responses, and temporal response properties, such as luminance rise time and decay rate. For the purposes of U.S. Navy display systems, Meister (1984) has compiled a data base on operator performance in using electronic displays, particularly CRTs, to serve as a guide for their design. There he states that the eight most critically important factors affecting operator performance with CRTs are frame rate, contrast ratio, ambient illumination, target/symbol size, resolution, bandwidth, registration, and phosphor type. For use in Army aviation, Rash and Becher (1982) consider the most important factors bearing on CRT image quality as luminance, contrast, frame and field rate, resolution, spot size and shape, and modulation transfer function (MTF).

Images convey pattern information to users. A basic measure of the capacity of displays to generate patterned images is spatial resolution. To illustrate the issues we have introduced, we discuss next the example of the development of spatial resolution FOMs.

The example of spatial resolution FOMs

First, let us consider an aspect of the spatial domain, the size of a display's pixels in terms of visual angle. It takes about 3000 pixels per square degree of visual angle for a displayed image to be indistinguishable from reality, because of the limited resolving power of the human visual system (Silverstein et al., 1990). Complex images will look more realistic as they approach this; however, realism is a subjective quality and is not of paramount importance for most visual tasks in terms of the usable information generated. More important in terms of reliably conveying visual information, and probably the most important parameter, is the spatial resolution of the display, which is the ability of the display to generate finely detailed patterns.

Displays based on CRT technology operate on the principle of sweeping an electron beam across a phosphor screen, where the image on the display is the result of light being emitted from the phosphors when excited by the electrons. Spatial patterns, made up of luminance differences within the image, are accomplished by the modulation of the electron beam which results in a modulation of the luminance. Because of the nature of CRT technology, smaller details become increasingly difficult to display at a given contrast, because of the physical limitations of modulating the beam amplitude. The smaller the detail, the faster the beam needs to be modulated as it sweeps across the monitor at a constant rate. Increasing power is required to make equivalent amplitude changes more rapidly as the beam sweeps at its constant rate. Physically, the maximum power output is set, that is, it will be independent of the frequency of modulation in a frame (i.e., independent of the spatial frequency). Therefore, more rapid modulations (higher spatial frequency) will have lower maximum amplitudes, which means that smaller details will have lower maximum contrast. This is problematic and a major factor in CRT quality control because those small details are where the human visual system requires greater contrast. A number of measures have been developed to quantify this spatial resolution.

Historically, the most commonly used FOMs for CRT spatial resolution have been the shrinking raster resolution, the television resolution (TV limiting response), and the MTF (Biberman, 1973).

To determine the shrinking raster resolution of a monitor, a raster of equally spaced lines are written on the display. The line spacing is reduced or shrunk until the lines almost blend together to form an indistinguishable blur. This flat field condition occurs, for an experienced observer, when the line spacing is approximately two standard deviations, where one standard deviation is the spot radius at the 60 percent amplitude of the spot intensity distribution, which has a Guassian profile.

To determine the TV limiting response, a television wedge pattern is displayed and the spot size is measured by observing the point at which the lines of the wedge are just detectable. This is equivalent to the square wave modulation function and is often referred to as the limiting square wave response. The point at which the wedge is just detectable, given in terms of the number of TV lines per unit distance, is the measure of the limiting resolution.

These first two techniques, and other similar techniques, described in Biberman and Tsou (1991) and Verona (1992) have the disadvantage of being subjective and therefore subject to observer error as described in Verona (1992). And they do not allow one to predict observer performance (Biberman and Tsou, 1991). The MTF, therefore, came into use.

The display's MTF to quantify display spatial resolution

The MTF does not suffer from the mentioned disadvantages in that it is objective and it has a natural interpretation in terms of human vision. The MTF is an FOM which characterizes the efficiency of a CRT display device in converting voltage (scene contrast data) into displayed image contrast over of range of spatial frequencies, where spatial frequency refers to the number of modulations per unit length. Also known as the sine wave response (SWR) curve, when each spatial frequency is measured individually, this FOM uses maximum modulation contrast sine wave gratings of different sizes as input data (Figure 1). Loosely speaking, the different spatial

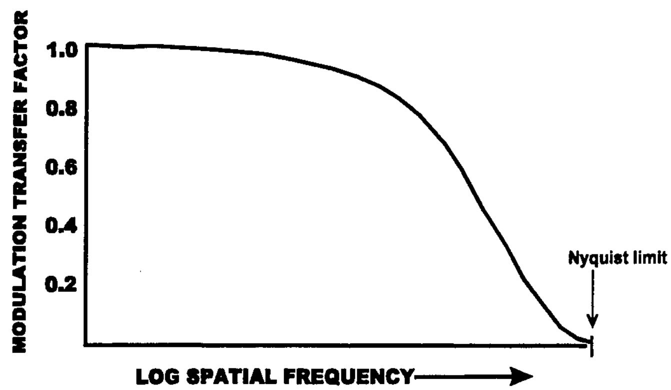

Figure 1. Typical modulation transfer function curve.

frequencies of the sine wave grating test the device's efficiency at different scales, or sizes. The measure of the input/output efficiency for each frequency is given by the magnitude of the reduction in contrast of each of the different sized sine waves in the displayed image. And the overall plot of the reduction in contrast (that is amplitude) as a function of frequency is the MTF.

A CRT display's MTF curve typically is a monotonic function, maximum at the lowest spatial frequency available (determined by the display width) and decreasing to the limiting highest spatial frequency of the display (known as the Nyquist limit). This means that smaller scales, representing smaller details and objects, can not be displayed with the same high contrast as larger ones. Figure 2 shows a representative measured MTF of a CRT.

The observer's CSF to model the observer

Similar to the physical MTF for display devices, there are psychophysical functions for humans, which characterize the visual system's efficiency in transmitting contrast, in this case a physical stimulus contrast into a perception of that contrast. For humans, this curve is (typically) obtained by measuring the contrast detection thresholds of sine wave gratings over a range of spatial frequencies, i.e., the minimum contrast required to see the grating at each of the different spatial frequencies. The result is the contrast sensitivity function (CSF), sometimes referred to as the contrast transfer function, which can be considered as a (rough) psychophysical analogue of the MTF. [Whereas the MTF depicts the contrast output for different spatial frequencies when full contrast is input, the CSF depicts the minimum contrast required to see different spatial

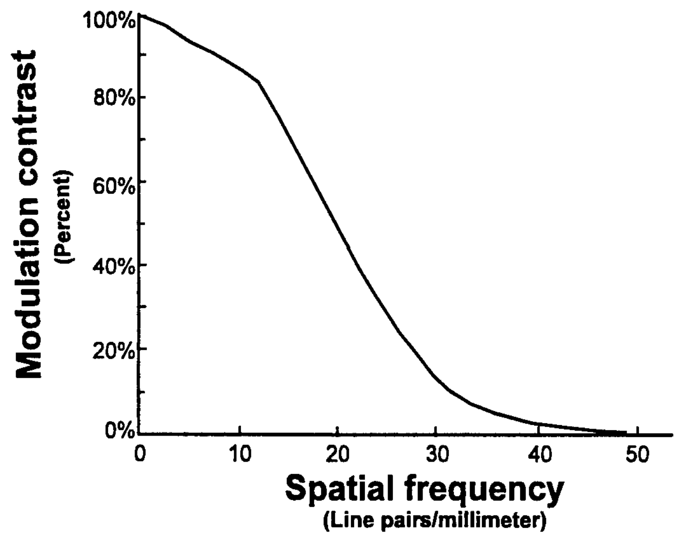

Figure 2. A representative measured modulation transfer function for a CRT.

frequency sine waves.] Unlike the typical monotonic MTF for CRTs, which show maximal efficiency at the lowest spatial frequencies, the CSF shows that the human visual system is maximally efficient at intermediate spatial frequencies, of around 4 cycles per degree of visual angle, and drops off in efficiency at higher and lower frequencies, where greater contrast is required to see the grating. Figure 3 shows a typical CSF.

Combining the MTF and CSF into an FOM to quantify conveyable information

Sine wave gratings are a convenient stimulus in generating both human and CRT efficiency functions because the mathematical tools available (Fourier analysis and linear systems theory) allow one to generalize the results to a wide range of stimuli and imaging conditions, (for example, the optical transfer function is an analogue of the MTF for characterizing the imaging quality of lenses). The mathematics of linear systems allow one, also, to cascade the transfer functions of various components of a complex optical system, such as a helmet mounted display, and to integrate the final component, the human perceiver, into a description of the total imaging context.

This last point---integrating the human perceiver--- is a complex issue: How does one combine the CSF of the human visual system with the MTF (or sine wave response curve) of a display device to derive an image quality metric based on units directly interpretable in terms of

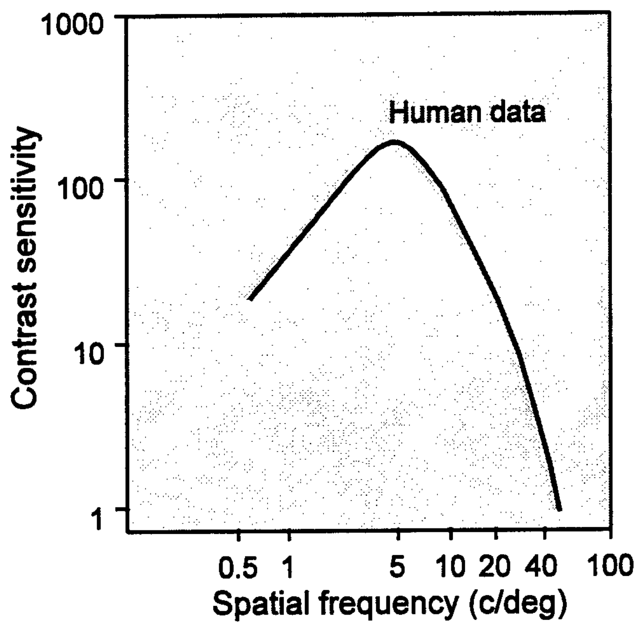

Figure 3. The contrast sensitivity function.

observer perception. That is, how do we obtain an overall FOM for the efficiency of a display-human system which starts by converting voltage, or scene contrast data, into a displayed image, which conveys a perceived image. Both functions are based on the contrast response of the respective systems, human and display, to sine wave gratings of a range of spatial frequencies. This question of how to combine them, not as straightforward as it may appear at first glance, has generated additional issues and some theoretical debate and visual experiments.

The most straightforward means of combining the human CSF and the display device MTFs is the image quality metric known as the modulation transfer function area (MTFA), shown in Figure 4, advocated by Synder (1973, 1980). Here the two curves are simply plotted and superimposed on the same spatial frequency scale. The point of intersection of the two curves is the upper spatial frequency limit, the point at which the device can no longer generate the contrast required for the human visual system to see the modulation. The human visual system can see the contrast generated by the device for the spatial frequencies to the left of this point. The MTFA metric, measuring the overall image quality of the device for spatial resolution, is quantified as the area between these two curves. This is the area that the device can display contrast which the human eye can see. This area, bounded on the bottom by the lowest contrasts the eye can see, and on the top by the highest contrasts the display can present, is taken to represent the integrated sum (over spatial frequency) of the amount of spatially resolvable information potentially available to the observer. The original assumption being that the greater area represented greater distance above threshold vision and hence better image quality. However, it is not obvious how the two axes of the MTFA graph should be scaled.

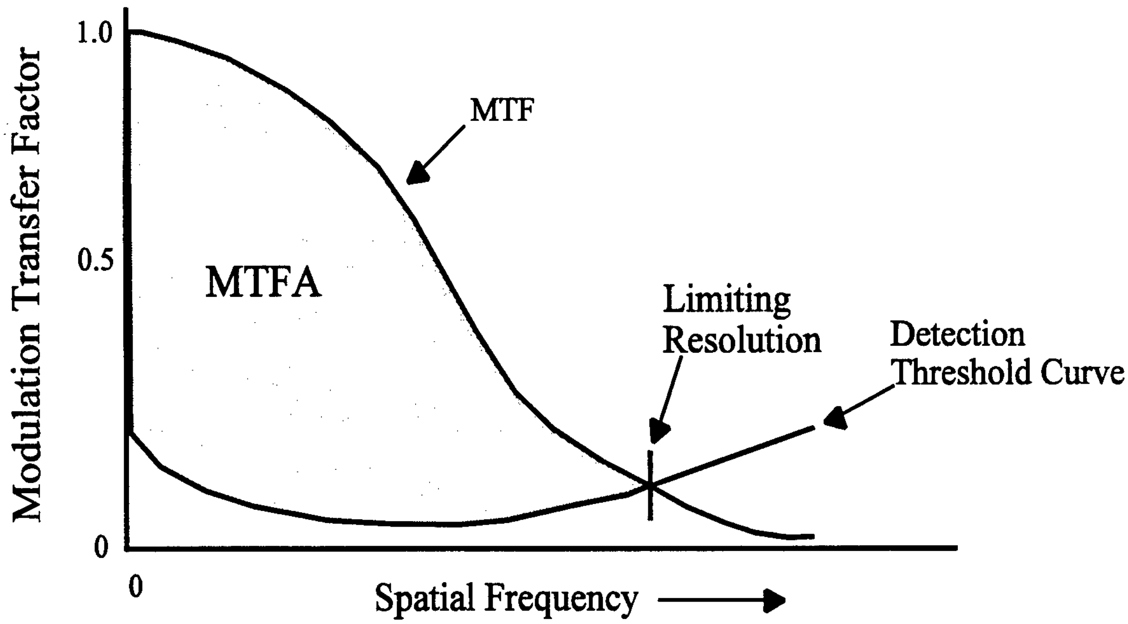

Figure 4. The modulation transfer function area.

Task and Verona (1976) have argued that since human vision is not linearly related to the modulation axis in the MTFA, it should be transformed into more visually relevant units. They transformed the contrast axis into square root of two incremented gray shades. The resulting Gray Shade Response (GSR), indicates how many gray shades are visible as a function of spatial frequency. This scale is more in keeping with the human's visual limitations in distinguishing gray levels. The new metric, defined as the area between the visual threshold curve and the GSR (instead of the MTF) is referred to as the Gray Shade Frequency Product (GFP).

To determine which FOM was more appropriate for describing image quality, Task and Verona (1976) performed a visual experiment to compare the predictions of the two metrics. The visual task was to increase the size of a target on the monitor until the subject could identify it. They used two versions of both the MTFA and the GFP. The two versions consisted of using linear and log scales for the spatial frequency axis in computing the values of the FOMs resulting in a total of four FOMs. In total, there were two scale conversions of each axis of the original MTFA graph. The experiment ran as follows: Physically manipulating image quality by changing the electron beam to generate three spot sizes on the monitor, then computing the resulting indices for the four FOMs, and then visually testing the monitor with the three spot sizes to see which FOM best predicted the visual performance results. They conservatively reported that the visual performance results indicated that the (log scaled) GFP was at least as good a measure as the (log scaled) MTFA (Task and Verona, 1976). As opposed to analytically defining an FOM in advance, this is the empirical way to determine the value of an FOM; one derives the correlation between the value of the FOM for a number of image display conditions and the results of human visual performance.

In addition to the GSR, a number of other modifications of the MTFA have been proposed. Based on a series of visual experiments, Barten (1990), for example, advocates his Square Root Integral (SQRI) Method as an improvement over the MTFA. His modification is essentially a different computational formula for deriving the FOM value based on certain assumptions about what is important to vision. For example, the MTFA weights changes in all the spatial frequencies equally when computing the metric--a change in one spatial frequency will be the same as in another--, while the SQRI method uses a nonlinear scaling across spatial frequencies to weigh the scale according to visual importance (Barten, 1990). The Integrated Contrast Sensitivity (ICS) metric, another computational variation of the MTFA, has been suggested (Van Meeteren, 1973) when one needs to take into account changes in human sensitivity, which vary with viewing condition. It is claimed that changes in the human contrast transfer function, such as occurs for adaptation level, are not accounted for in the MTFA, but are in the ICS. Beaton and Farley (1991) discuss the relative advantages and disadvantages of the MTFA, the ICS and the SQRI metrics. The research controversy with spatial resolution FOMs chiefly concerns how well behaved they are in tracking the effect of changes in the display conditions on visual perception. Barten (1990) claims the SQRI FOM tracks changes in a large number of viewing and image parameters.

Olson and Balram (1996), in their discussion of image quality metrics for display resolution, suggest that the limiting resolution and the MTFA, because they both emphasize high spatial frequency performance, are more appropriate for evaluating military displays, while the SQRI, which was designed to heavily emphasize low spatial frequencies that play a major role in subjective quality, is primarily suited for evaluation of consumer displays. The SQRI de-emphasizes the higher spatial frequency information, which is of great importance in military tasks such as target detection and recognition.

Task (1979) tested a number of these metrics to determine how well they predicted visual performance. He correlated the metric values of images with observer performance in three detection/recognition studies in which image quality was varied by changing the system MTF. Of the FOMs he tested, he found that the MTFA appeared to be the best predictor of visual performance. Many of these spatial resolution measures differ in the underlying assumptions as to the most important factors, and many of these methods differ simply in the computation of the FOM. The American National Standards Institute and the Human Factors Society, who have developed standards for monochrome displays published as the American National Standard for Human Factors Engineering of Visual Display Terminal Workstations in 1988 have adopted the MTFA as the standard measure for spatial resolution.

The effect of visual tasks on spatial resolution requirements

Whatever the particular spatial resolution metric, observer performance dependent on spatial resolution will increase with increases in display resolution to some upper limit, where observer performance will reach an asymptote, caused by either a ceiling effect of the performance variable or by the display resolution exceeding the spatial sensitivity of the observer (Snyder, 1980). There is a task dependency, in that, in general, the simpler the task, the lower the resolution at which the asymptote is reached. For example, if the task is an alphanumeric character recognition task, a minimum of 12 television lines per symbol height is necessary for high recognition. Character recognition accuracy increases as TV lines increase from 8 to 16 lines per symbol. Also, for silhouette recognition, as raster lines per symbol height increase, so does identification accuracy (Baker and Nicholson, 1967).

Because of the complexity of the human visual system (including the visual cognitive system), which is not completely understood, it is not possible to make many unqualified general statements about the display's physical image quality requirements. Human factors psychologists have developed a body of sophisticated psychophysical techniques ranging from simple detection tasks, requiring no expertise, to more sophisticated techniques, which measure performance by experts in a particular visual task. An example is a spatial resolution study by Seeley et al. (1987) to determine the display requirements for making accurate medical diagnoses of a particular class of medical images on CRT monitors. This is a specific visual/cognitive task performed by highly trained experts. The accuracy of expert radiologists making patient diagnoses by viewing digital CRT images of chest X-rays was measured by receiver operating characteristic (ROC) analysis. In this task, where the spatial resolution of the displayed image was varied, they found that radiologists' performance peaked at a spatial resolution of 1.25 line pairs per millimeter (lp/mm). Therefore, 1.25 lp/mm was the minimum resolution standard they recommended for radiologists viewing chest films on a CRT.

As we have seen, the image quality requirements for even simple parameters such as spatial resolution can be very task specific. And, which FOM more accurately reflects the quality of a display will depend on how the display is to be used. For example, some consider the SQRI FOM as a better metric for commercial purposes, and the MTFA FOM a better metric for military purposes.

The effect of technology on spatial resolution FOMs

Originally, the MTF, and FOMs based on it, were developed for analog displays. Feltz (1990) and Moon (1986) have claimed that the MTF, originally developed for analog CRTs, can be applied to discrete matrix displays, including FPDs, by mathematical modifications in the way the MTF is calculated (see Beaton and Farley, 1991). They assume that despite the many differences between the analog quality of CRTs (such as the Guassian luminance profile of its image spots) and the digital pixel structure of FPDs (including the variations in pixel geometry), the resulting MTF adequately can serve to predict visual utility of all display technologies. This is a controversial position. For example, Balram and Olson (1996) disagree and advocate the use of a new metric, the multivalued MTF or MMTF, which incorporates spatial phase as well as frequency, which is claimed to be necessary for a full characterization of matrix displays.

If we know a display's image quality in terms of its MTF, can we assume that in any visual performance task relying on spatial resolution (ranging from simple detection to more complex identification), the metric will be equally predictive of the visual performance results for the analog as well as for the digital display? The hidden assumption here is that all the different subjective quality aspects and the physical differences between displays that are not captured by the MTF are not as relevant or important in spatial resolution-limited tasks. This is currently a hotly debated assumption. Given the difficulties in arriving at new standards, an important question is whether or not FOMs previously developed for CRT displays can be applied to FPDs.

FOMs for CRTs and FPDs

We have seen the list of common FOMs used for CRTs and FPDs listed in Tables 1 and 2, and, using the example of spatial resolution, the issues involved in the development of FOMs including the issue of new digital technologies. Here, we look at which CRT FOMs can be directly applied to the newer FPD technologies, and we point out data voids, where new FOMs need to be developed for FPDs.

Differences of FPD technology

CRTs are a technology based around the modulation of a sweeping electron beam as described above. FPDs are electro-optical displays which derive their name from the flatness of the viewing surface and the reduced depth behind the surface. Unlike electron beam CRT technology, FPDs are based on a number of different technologies using varying physical mechanisms to produce the displayed image. This has consequences for visual perception and for what FOMs should be used to characterize them.

The most prevalent current FPD technologies are LCDs, EL displays, plasma display panels, LED displays and field emission displays (FEDs). Each of these display types generally consists of a rectangular spatial array of picture elements (pixels) known as a matrix (Figure 5). Each pixel is independently controlled by electronic drivers. FPDs differ from CRTs, and each other, essentially in the physics of how each pixel generates or modulates light energy (Tannas, 1985). Independently of the physics associated with these technologies, FPDs may differ in the way the pixels are electronically addressed. Different addressing schemes will primarily affect the temporal parameters of the image and the type of noise likely to be encountered in the image. [In the context of military display updates, a recent survey of flat panel technologies can be found in Harding et al. (1996).]

Flat panels differ from CRTs and from each other in a number of significant ways that have been shown to have consequences for vision. As discussed, some have argued that the spatial resolution FOMs developed for CRTs can be used for FPDs, when modified for discrete matrix displays. The MTF of FPDs can be predicted based upon pixel density, spacing, and pixel geometry (Barten, 1991, 1993; Infante, 1993). However, in addition to the resolution of the pixels, their geometric arrangement, as well as pixel shape and pitch (distance between them),

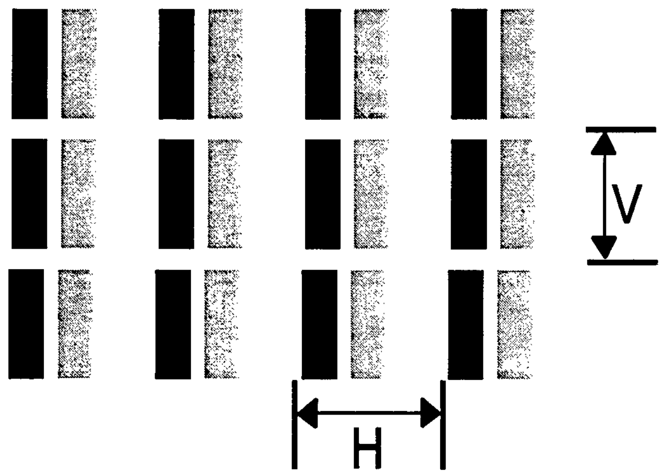

Figure 5. A rectangular spatial matrix of triad color (RGB) picture elements (pixels).

need to be specified in new FOMs, as these all have visual consequences. For example, Silverstein et al. (1990) have shown how these factors can introduce spatial artifacts, and affect display resolution and visual performance. Harding et al. (1997), controlling for visual angle, luminance, and contrast, have shown that the various pixel characteristics of FPDs can effect visual performance. They found that visual legibility was affected by pixel geometry, pixel noise, and screen filter characteristics. Another pixel characteristic, pixel fill factor, defined as the area of the active pixel divided by the total pixel area, also is important. Infante (1993) has shown that pixel fill factor affects screen resolution, where displays with smaller fill factors have a higher resolution as measured by the MTF. Also, the number of defective pixels, as mentioned, is a new variable needing particular attention because of the current quality control difficulties in manufacturing totally defect-free FPDs.

Which CRT FOMs apply to FPDs?

A number of variables effect CRT image quality as discussed above. They have been classified according to one tripartite scheme (Task, 1979) as follows: (1) Geometric variables such as display size and aspect ratio, number of scan lines and interlace ratio, illuminance, scan line spacing, and linearity; (2) electronic variables such as bandwidth, dynamic range and contrast ratio, signal/noise ratio, frame rate and field rate; and (3) photometric variables such as luminance and contrast ratio, gray shades, halation, MTF, color, resolution, spot size and shape, and luminance uniformity. Another four part scheme was shown in Table 1. Many of these variables are tied to the specifics of the CRT's electron beam technology.

According to Curtin and Infante (1996), the most important factors affecting CRT image quality are listed in Table 3. Most have a relatively straightforward application to FPDs. Since pixel positions are fixed in FPDs, without the CRT variable of beam control, the issues concerning factors 4 (linearity) and 7 (convergence) do not apply to FPDs. Geometric distortion does not apply to FPDs which have fixed pixel positions, nor does convergence of colors, although the pixel packing geometry of FPDs will affect perception due to color fringes and spatial artifacts. [Note: these effects can be of concern at the highest resolutions (Silverstein et al., 1990).]

Factors 1 and 2, luminance and contrast FOMs, are directly applicable to FPDs, with the caveat that the ambient luminance must be taken into account. In general, brighter is better, with the required luminance of the display dependent on the environment, e.g., typical office displays require a luminance of 50 to 100 candelas per meter squared to achieve sufficient brightness. Brighter displays are needed for viewing in direct sunlight. Reflective and transreflective LCDs, which will compensate for increased light levels, need backlighting for low ambient light levels. Whether the FPD display is emissive or passive will also determine how ambient light affects the contrast. For both luminance (display brightness) and contrast FOMs, the measures are the same for CRTs and FPDs, although differences in the effect of ambient lighting should be noted. In some FPDs, particularly LCDs, these measures need to be given as a function of viewing angle. Luminance uniformity across the display is an image quality issue with CRTs, but is less of a concern in most FPDs because of the nature of the addressing, by rows and columns, and the uniformity of the pixel activation.

Table 3.

Most important variables affecting CRT image quality (adapted from Curtin and Infante, 1996).

Factor	Perceptual Parameter	Physical Parameter	Relevant for FPDs
1	Brightness	Luminance	Yes
2	Contrast	Contrast	Yes
3	Sharpness, crispness	Resolution, addressability	Yes
4	Geometry, linearity	Linearity, straightness of lines	No
5	Vividness of colors	Color saturation, color contrast	Yes
6	Fidelity of colors	Color accuracy	Yes
7	Convergence	Convergence	No
8	Freedom from flicker	Refresh rate	Yes

Factor 3, resolution, as discussed above, is typically characterized by MTF-based FOMs. These translate in a relatively straightforward way from CRTs to FPDs. In the case of matrix displays, attention must be given to the pixel geometry which can cause visual artifacts and may effect perception at the highest resolutions. Resolution metrics, such as the MTFA, will be similar for CRTs and FPDs, with modifications in the calculations due to the pixel matrix structure. Although as noted, the details here are controversial.

Metrics describing factors 5 and 6, concerning colors, should be the same for CRTs and FPDs. Color quality measurements, characterizing color saturation (purity of color) and chromaticity (color coordinates), and color gamut (range of colors) present no new conceptual issues for FPDs. Although, color fringing may be introduced at the highest resolutions with some subpixel patterns (Silverstein et al., 1990). Color anisotropy stands for the change in color as a function of viewing angle. In FPDs, if luminance and contrast FOMs need to incorporate viewing angle, so will color FOMs. Color uniformity across the display may need to be a concern to the degree that luminance uniformity is. Color tracking, the chromaticity changes as a function of luminance, may need to be specified to ensure color accuracy throughout the luminance range.

For factor 8, the conventional standard is that the display should be flicker free for direct viewing for 90 percent of the population. This should be true of CRTs and FPDs. This factor is a function of refresh rate, update rate and other (technology) specific temporal variables such as phosphor persistence. Of more concern, which we discuss below, is how these temporal variables contribute to the quality of dynamic imagery.

Another issue we have not touched on is the visual noise in the display (Biberman, 1973). Where the noise is in the signal source, the issue does not concern us here; where it is display technology dependent (e.g., "snow" or white noise produced by analogue circuits in CRTs, cross talk in some flat panel displays, or defective pixels), separate technology-specific FOMs may need to be used to quantify the visual effects of the noise.

Recent texts which cover many technology-based measurement issues include Holst (1995), Karim (1992), Keller (1997), and MacDonald and Lowe (1997). However, adequate FOMs to quantify and predict the visual consequences resulting from the temporal aspects of different displays are lacking.

The importance of measuring certain temporal aspects of displays such as phosphor persistence, pixel rise time, and so on, are recognized (Keller, 1997). Caveats on the need to be aware of dynamic image quality differences due to technology have been expressed (Parker, 1997; MacDonald and Lowe, 1997; Rabin and Wiley, 1995; Bitzakidis, 1995). And, dynamic imagery perception has been theoretically analyzed in the abstract (e.g., Lindholm, 1992). However, what is lacking are measures of dynamic image quality which can rationally characterize the visual consequences with dynamic imagery that result from the specific variations in the temporal parameters of different technology displays. FOMs need to be developed which can clearly delineate the visual quality of the dynamic imagery produced by different technologies. To do this, we need to consider what aspects of dynamic imagery are most important? Spatial resolution and gray scale, and how effectively these are matched to the observer, are the aspects of most importance in terms of visual information. How are these affected by dynamic imagery?

The differences in the quality of dynamic imagery is the most important new factor of concern with FPDs. When image motion is displayed on FPDs, the spatial resolution and the gray scale are affected, sometimes severely. [Whatever affects the gray scale also generally affects aspects of the color.] Exactly how they are affected needs to be quantified in a way that is comprehensible in terms of visual perception. This currently is the major data void in FPD FOMs.

Even with the long dominant CRT technology, the quality of dynamic imagery in certain applications has been a concern. Rash and Verona (1987) discuss the need to consider the display system's time constants because of its critical effect in pilotage and target acquisition. When modulation contrast degrades below a certain threshold level, targets begin to blend with the background and the user loses the ability to discriminate targets from their background. In dynamic images, pilots may fail to see tree branches and gunners mistake tanks for trucks. They suggest indexing image quality by dynamic MTFs, with modulation contrast reduction as a function of spatial frequency and drift rate. In order to predict the utility of CRTs for moving target detection and recognition, Rash and Becher (1982) have analyzed how various parameters of CRT technology contribute to the quality or degradation of dynamic images. The temporal parameters of the CRT, such as the phosphor characteristics (luminance rise and decay time, and phosphor persistence) and certain electronic variables (bandwidth, scan rate, frame rate, and field rate), contribute to image degradation of moving targets by blurring and smearing. Rash and Becher (1982) have developed equations modeling how target velocity, scan rate and phosphor persistence affect the MTF of the display.

How the phosphor persistence factor affects the MTF of CRTs can be seen by comparing Figures 6 and 7. In general, for relatively fast phosphors, such as P43, the MTF for static images on a CRT display describes the spatial resolution of static images as well as the spatial resolution of dynamic images (Figure 6). In CRTs with slow phosphors, such as P1, the dynamic spatial resolution diverges from the static spatial resolution as shown in Figure 7.

The many novel factors involved in the generation of FPD images, each with their own novel time constants, will affect the dynamic MTFs. For example, Figure 8 shows how even a small ambient temperature change can induce large changes in the temporal response of an LCD display.

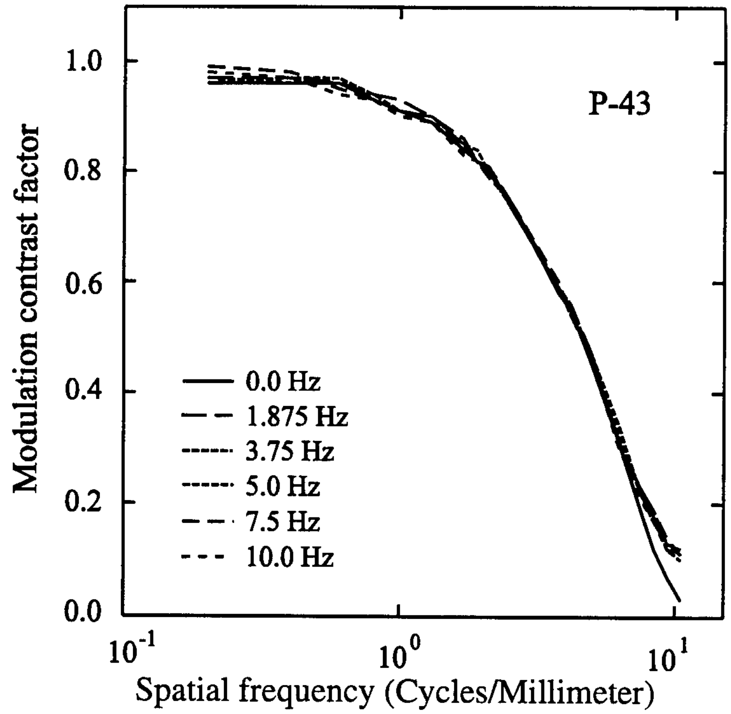

Figure 6. MTF curves for P43 phosphor
(Beasley et al., 1995).

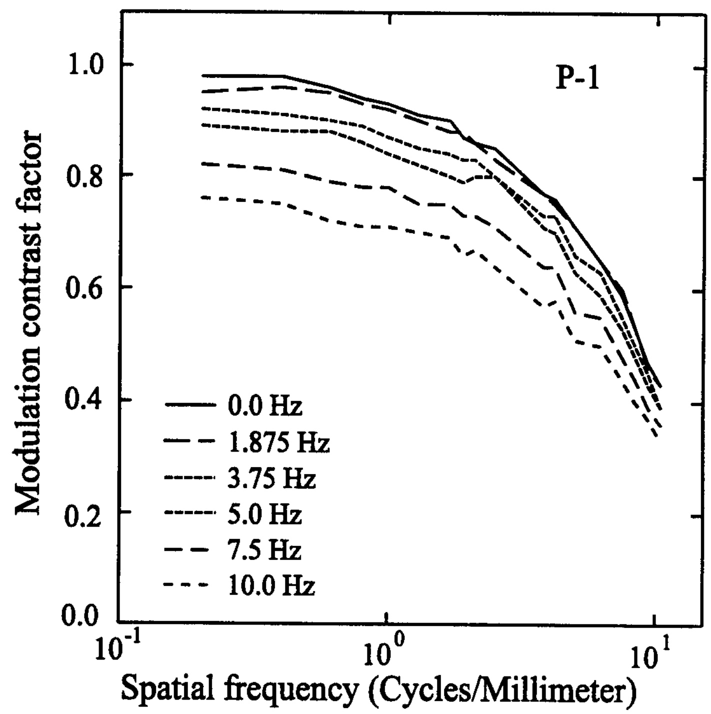

Figure 7. MTF curves for P1 phosphor
(Beasley et al., 1995).

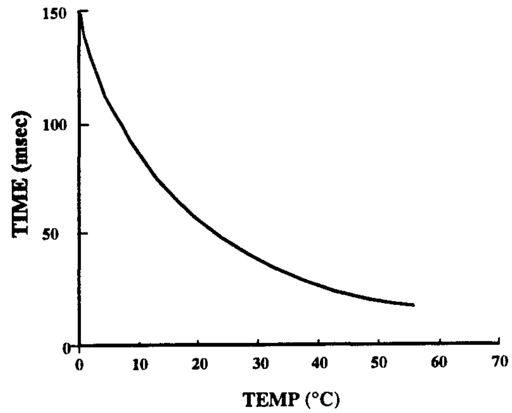

Figure 8. Temperature response of a liquid crystal display (Rash and Verona, 1987).

In order to clarify the issues of dynamic imagery on FPDs, we next delve into the essential factors at the root of a display's temporal parameters, which (along with the pixel factors discussed above) constitute the fundamental technology differences between CRTs and FPDs. These include methods of addressing, updating and gray scale generation.

Differences in image addressing and updating in CRTs and FPDs

A basic difference between the CRT and FPDs is the way displayed images are addressed and updated. Addressing refers to how an image location is coded and accessed in order to display a gray level (and color). Since the CRT uses periodic beam sweeps over a phosphor screen to generate the image, addressing is transparently accomplished by the simple association of a spatial location with the beam at a certain time in its cycle.

Updating the image in the CRT is equally transparent as the electron beam has a fixed rate of cycling through a complete image. The rate at which a complete CRT image is drawn, the time for the electron beam to complete a sweep to draw one frame, is known as the frame rate. The frame rate determines how smooth motion on the display will appear. CRTs are typically interlaced, where two fields are drawn for each frame, first the odd lines, and then the even lines. This allows the motion to appear smoother at a lower frame rate than if the images were non-interlaced, and this avoids flicker of the monitor. Thus, important parameters for CRTs are the frame rate, typically 30 Hz, as well as, the field rate, typically 60 Hz.

How fast the image is updated to change the image (the CRT's frame rate), and refreshed to avoid flicker (the CRT's field rate), is important to observer performance. Refresh rates should be over the critical flicker fusion (CFF) rate, when succeeding frames fuse together so that the image appears flicker free, but often it can not be much higher, because of the necessary increase in bandwidth. Additional contributing temporal characteristics of CRTs include phosphor persistence, horizontal scan rate, vertical refresh rate, and amplifiers' bandwidths. Rash and Becher (1982) have analyzed how these characteristics contribute to image smear with moving targets on CRTs.

In FPDs, addressing is done by a number of different electronic methods, usually involving some form of cycling through the column and row leads to the matrix array of individual pixels. While some older flat panels have inherent memory, where pixels remain on or off until addressed, most FPD technologies use periodic refreshing to avoid the perception of flicker. The required refresh rate to avoid flicker is based on the rise and decay rates of the pixels. If the pixel rise time and/or decay rate is too long, smearing may occur in dynamic images, as occurs in many LCDs. If too short, the refresh rate needs to be correspondingly faster to avoid flicker, particularly in those flat panel technologies which, like CRTs, use phosphors. These temporal characteristics of FPDs are pertinent when considering the conveyance of dynamic visual information, such as moving targets and changing imagery (Toms and Cone, 1995; Toms, Cone and Cavallaro, 1995; Rabin and Wiley, 1995). FOMs that are important here are the pixel rise and fall times for achieving a full luminance value, and the update and refresh rates.

Differences in image gray scales and contrast in CRTs and FPDs

As discussed, gray scales in CRTs are generated by modulating the amplitude of the electron beam. The resulting gray level corresponds to the amplitude of the transient voltage of the electron beam. Some of the older flat panel designs have used spatial aspects to generate gray scales by trading spatial resolution for gray scale (e.g., by partitioning pixels into subpixels, where luminance equals the subpixel area turned on, or by dithering, which involves turning on random pixels in small unit areas to control the overall luminance of each unit area). These methods are not used in the newer displays where spatial resolution is important. Instead, many FPDs generate gray scales by using temporal aspects of the display, therefore, potentially trading temporal, rather than spatial, resolution for gray scale.

Each of these FPD methods rely on electronic controllers to translate gray level data to the form used by the display's circuitry. In pulse-amplitude modulation, transient voltage levels directly control luminance at each pixel. Instead of voltage, a number of methods use timing. The luminance is controlled by the amount of time a pixel is turned off versus turned on. In pulse-width modulation, the variation in the width of the pulse to a pixel during an addressing cycle determines the pixel's luminance. Other methods use different variations of the timing method such as duty cycle modulation, multiple pulse widths per frame, or combinations of pulse-width and pulse-amplitude modulation. All these methods interact with the physics of luminance generation and, each trades temporal resolution for gray scale. This can become a problem if the duration of the timing intervals needed to generate the gray scale become too large. This occurs in some LCDs that use this technique over multiple frames to control the luminance, which results in severe image smearing during image motion.

CRT technology has a known and consistent reduction in the available contrast for the smallest targets. This can be seen in the shape of a typical CRT MTF, in which there is a drop off at the highest display spatial frequencies (Figure 2). But, because the rate of the beam sweep in CRTs is a constant, moving that target across frames in dynamic imagery will not greatly effect the contrast already available at a particular spatial frequency (unless the phosphor is inordinately slow as shown in Figure 7). This was demonstrated in a study by Verona et al. (1994) using temporally modulated sine waves as generic dynamic imagery (undergoing sinusoidal counterphase modulation). They tested two CRTs, with a slow phosphor (P-1) and a fast phosphor (P-44), and found the expected spatial MTF curve, which is the reduction in contrast for higher spatial frequencies. Only at the highest temporal frequencies (7.5 and 10 Hz) for the slow phosphor was there any effect of temporal frequency, where there was a slight reduction in contrast. Except for very slow phosphors, the MTF does not change shape appreciably as a function of temporal frequency.

However, the (relative) invariance of the shape of the MTF curve (quantifying spatial resolution) for image motion found in CRTs can not be taken for granted in FPDs due to the differences in technology. In the different types of flat panels, gray scales and contrast are generated in a number of different ways, often by pulse-width modulation alone, or in combination with amplitude modulation, or some variation of duty cycle modulation (Sobel, 1992; Klymenko et al., 1997). These methods may be completely adequate for static imagery. However, because of the timing requirements inherent in these methods, the gray scale and contrast available for dynamic imagery most likely will be affected. Thus, a problem in CRTs is that small targets have lower contrast and compressed grey scales. But, in many FPDs, this problem occurs for moving targets instead of small targets. This, combined with the smearing effect due to pixel persistence, may have disastrous consequences for perceiving dynamic visual information on some flat panels. That this is indeed the case has been shown by Rabin and Wiley (1995), who found that visual performance with dynamic imagery was degraded with LCDs compared to CRTs, more so for higher velocity targets. They concluded that the differences between CRTs and LCDs were due not to differences in luminance, color, or spatial resolution, but to the poorer capacity of the flat panels to generate sufficient contrast for the dynamic imagery.

To summarize, in CRTs the small size of targets (indexed by the high spatial frequency response of the display) reduces contrast, while the high velocity of targets (indexed by the high temporal frequency response) does not. Conversely, in some of the matrix addressed FPDs, there is a problem of reduced contrast for high velocity targets, but not for small sized targets. Velocity, not size, is the main concern. [This is perceptually unusual in that in the real world the physical contrast of an object does not change based on its velocity.] Therefore, there is a need for a more inclusive FOM, reflecting the effects of motion as well as size.

New FOMs needed for dynamic imagery on FPDs

A number of commercial and government standards for CRTs, based more or less on psychophysical and human factors testing, have evolved to meet the requirements of the display community, including Underwriters Laboratories, Verband Deutscher Electrotechniken, the Federal Communications Commission, International Standards Organization (ISO), American National Standards Institute (ANSI), the Human Factors Engineering Society, National Information Display Laboratory, Society for Imaging Science and Technology, Society for Information Display, Society for Motion Picture and Television Engineers, Institute of Electrical and Electronic Engineers, and the National Institute of Standards and Technology's Flat Panel Display Laboratory. However, at the present time, there are no universal standards specifically for FPDs. Standards are being developed and proposed by the ISO (Greeson, 1996), but, because of the number of FPD technologies, and the continuing current development of the various technologies, it will still be some time before FPD standards are universally agreed upon in the same manner available for the mature CRT technology.

The standards of any new FOM need to mesh with the other display requirements. For example, it is mandatory for many military applications, particularly for displays that operate in day and night ambient lighting, that the display maintain its gray scale when dimmed over a 10 to 1 range, while maintaining an acceptable contrast ratio of 20 to 1 (Tannas, 1985). This is considered acceptable with the luminance either continuously variable, or controllable into at least 16 logarithmically spaced steps (64 for real world imagery to be aesthetically pleasing) Tannas (1985). Most flat panel displays do not use analog control for gray scale, but instead use temporal aspects of the image generation process such as pulse-width modulation, changing the element's duty factor on a frame time basis, etc. There are great demands on the contrast ratio and switching speeds already. Therefore, as we have indicated, there is a definite need to clearly assess the affect of temporal parameters on the spatial resolution and gray scale of dynamic imagery.

As we have seen, the influence of temporal factors would appear to be at the top of our list of concerns with the differences in FPD image quality that are not easily captured by any standardized metric. There are definite differences in the properties of dynamic images between CRTs and FPDs (and, of course, among FPDs). These differences are based on the ways the images are generated by the different technologies, and FOMs are needed to quantify this in order to evaluate the new technologies. Before the insertion of FPDs, to replace CRTs, these new FOMs need to be implemented for image quality control assessments. This is especially crucial in areas such as medical imaging involving motion (e.g., dynamic cardiac imaging), and military applications (e.g., detection, discrimination and identification of moving targets), and aviation, or piloting, (where, often, the entire scene may be moving). The way dynamic imagery affects spatial resolution and the gray scale of the display, are the key variables of concern we turn to next.

Gray scale and spatial resolution of dynamic imagery on FPDs

The capacity of new display technologies to present dynamic visual information, such as moving letters across a monitor, or moving targets, or a moving background landscape seen by a pilot wearing a helmet mounted display (HMD), can not be fully defined by the old FOMs. As seen in Tables 1 and 2, where the common FOMs are divided into standard categories, there are FOMs for measuring temporal and other properties independently, but no FOMs for assessing variables in other categories under dynamic conditions, for example, what happens to a display's luminance range with dynamic as opposed to static stimuli. In as recent a source on display measurement as Keller's (1997) text, temporal measurements are considered only independently, as a separate category from other measurements.

As we have seen, in some FPDs the shape of the MTF of the display changes based on the temporal characteristics of the image. Velocity in the image is likely to be the important variable, where higher rates produce more of a reduction in image quality. Also, as we have noted, in certain FPDs, especially some LCDs where the gray scale is generated by pulse-width modulation over multiple frames, the gray scale will be greatly impoverished for moving images, and the impoverishment will also likely be a function of velocity--a more degraded gray scale for faster moving objects. The quality of spatial resolution (measured by the MTF), and the quality of the gray scale (luminance range and/or resolution, possibly sampling) are affected by temporal factors. How should one quantify this?

First let us reiterate, the basic function of displays is to present patterned visual information as static and as dynamic images. Image quality FOMs for spatial resolution, such as the static MTF, inform us as to the fineness of the details in the patterns that the device can present. These, though not without controversy, appear to be, more or less, sufficient for their intended purposes for static images. FOMs quantifying contrast and grey scale inform us as to the range, and/or number, of luminance levels that make up the spatial patterns, in essence, the number of potentially different patterns that can be presented at each resolution. (These contrast FOMs are discussed in detail in Klymenko et al., 1997.) Generally these FOMs are adequate as far as they go, the quantification of the maximum static contrast of a display and the number of SOGs in static images of analog CRTs.

In dynamic images on FPDs, the spatial resolution and the gray scale of the display will often be degraded. This change in image quality with dynamic imagery needs to be specified precisely in a way that can be generalized over many imagery conditions, and be useful in assessing the relative merits of different displays for observers. What we need are FOMs to quantify the change in image quality, specifically the changes in the spatial resolution and gray scale of the device, as a function of the dynamic nature of the image. A further question concerns how to integrate the observer into these FOMs for dynamic imagery. In the case of gray scales, this may entail quantifying dynamic gray scale in terms of number of PGLs, and in the case of spatial resolution this may entail combining dynamic MTFs with dynamic CSFs.

Considerations in developing the new FOMs

One useful characteristic of many FOMs is that the FOM is a summary, or condensation, of information. For example, the MTFA FOM for spatial resolution, discussed above, gives a single number, which represents weighed averages of more information. How one assigns weights may be based on research, as in the example of Task and Verona's (1976) research which compared log and linear weighing of spatial frequencies; or, it may be based on what one considers important. The SQRI more heavily weighs low spatial frequencies, considered important for commercial viewing, while the MTFA is more biased toward high spatial frequencies, considered important for military applications. Reducing data this way, for the purpose of putting the various displays on a unidimensional scale so that they can be easily compared, has a drawback. While it reduces the need to consider the original data and many variables, it does lose information.

The new FOMs that need to be developed to quantify the way dynamic imagery affects FPDs will need to summarize the image quality effects of a large number of temporal conditions of a number of parameters. What happens to the MTF under different dynamic conditions? What happens to the gray scale of a standard target moving at various velocities? The maximum contrast, or luminance range of a gray scale, needs to be specified for dynamic conditions, in addition to the baseline static case. Is anything else going on when a gray scale is compressed; i.e., is there a change in the shape of the sampling distribution within the luminance range, in addition to the reduction in the luminance resolution?

For these data, one might parametize the dynamic aspect of the imagery in terms of temporal frequency, or in terms of velocity. In FOMs for dynamic image quality, temporal frequency, defined as cycles per second of a periodic stimulus modulated over time (moving or flickering), may be useful analytically in the way spatial frequency is useful. Here, the FOM might use moving or flickering sine wave gratings as a stimulus. Alternatively, velocity (defined as either distance per second, or degrees of visual angle per second) might be more useful in terms of directly understanding real images moving on the display.

How should the observer be modeled? For example, how is the number of perceptual units between the maximum and minimum luminance, PGLs, determined? Should previously published data, or a standard assumption similar to the square root of two SOG rule, be used to model this, or are new psychophysical data required? How does the number of PGLs in the luminance range depend on velocity? Also, how would these FOMs incorporate visual tasks? A dynamic FOM is likely to give very different results if it is based on a task such as identifying a target that is being tracked versus a task such as identifying a target that streaks by.

What is the relevance of perceptual units, whether spatial resolution or gray scale PGLs, under different dynamic and task conditions? Consider the gray scale. If less PGLs, are available for moving targets, then less visual information can be generated, and conveyed, about that target when it is moving. In terms of information, the number of PGLs determines the number of potential targets that can be discriminated by the observer. This is important to know, particularly for many military applications, where rapid identification of small, high velocity targets is crucial. A gray scale FOM for dynamic imagery should inform one as to how many gray levels, and how many PGLs, are available for targets moving at different rates.

Consider spatial resolution, a more complicated topic. Should one simply generalize previous FOMs such as the MTF for static displays, to now include data for different temporal frequencies and/or velocities of the standard stimuli, as Verona et al. (1994) and Beasley et al. (1995) have done for CRTs. If done for new technologies, this might produce a great deal of data, and some means of rationally reducing the data might be necessary. Issues on incorporating the observer, as has been done in the MTFA, would be far more complicated, as humans have their own quite complicated contrast sensitivity function, which changes based on temporal frequency and velocity. The human visual response depends on both the spatial and temporal properties of the stimulus, so the CSF is now a two dimensional spatio-temporal response surface. Incorporating the data modeling the observer into an FOM is, as we have seen, a non-trivial exercise. The question of how to combine physical MTFs (measured with moving stimuli and parametized in terms of spatial and temporal frequencies) with human CSF responses is likely to generate at least as many controversial ways as has the purely static case discussed earlier.

In summary, the overall question is how will these data be weighed and reduced to create simple, reliable and useful standard FOMs quantifying the quality of dynamic imagery on FPDs?

Conclusions

The functions and development of FOMs for visual displays have been discussed with spatial resolution FOMs serving as an illustrative example of the issues involved. These issues include what aspects of the image are measured and in what units, how FOMs are often technology-specific, and how FOMs may incorporate the user, and the purpose of the display, and the particular visual task. Visual information on displays was defined and the concept of conveyable information was discussed in terms of the need for some FOMs to indicate how well the properties of the display match the requirements of the observer.

Various new technologies under the general rubric of FPDs are competing with the old mainstay, the CRT, in display applications in the private, commercial, entertainment, aviation, military, medical and other arenas where displays are used. In some cases, FOMs used for CRTs can be applied to FPDs, and in other cases, the new FOM needed is rather straightforward, such as the defective pixel count FOM for FPDs, and the off-axis viewing contrast reduction FOM for LCDs.

FPDs have introduced a number of new parameters such as different methods of image addressing and updating, and image gray scale generation. Many of these affect temporal characteristics of displayed images in ways not encountered with CRTs. These temporal aspects play a large role in the new FPD technologies, and directly affect the quality of the imagery, particularly dynamic imagery.

There are no well developed FOMs to assess and quantify the resulting quality of dynamic imagery on FPDs. New FOMs need to be developed to fill this data void. We have discussed various issues which need to be considered when designing these new FOMs, including the specific needs for assessing gray scale and spatial resolution with dynamic imagery. We have suggested that these new FOMs should be developed to quantify gray scale and spatial resolution as a function of temporal frequency or velocity.

We have identified a crucial need for new FOMs to characterize the quality of dynamic imagery on FPDs. As we have seen, there are many complex considerations in the development of new FOMs. However, this data void needs to be addressed before the insertion of FPDs in critical areas such as medical and military applications.

References

American National Standard for Human Factors Engineering of Visual Display Terminal Workstations (ANSI/HFS 100-1988). 1988. Santa Monica, CA: Human Factors Society, Inc.

Baker, C. H. and Nicholson, R. 1967. Raster scan parameters and target identification. Proceedings of the 19th Annual National Aerospace Electronics Conference (NAECON) pp. 285-290. Dayton, OH: Institute of Electrical and Electronic Engineers.

Balram, N. and Olson, B. 1996. P-10: Multivalued modulation transfer function for flat-panel displays. Proceedings of Society for Information Display International Symposium Digest of Technical Papers, Vol. XXVII, pp. 429-432.

Barten, P. G. H. 1990. Evaluation of subjective image quality with the SQRI method. Journal of the Optical Society of America, (A), Vol. 7, pp. 2024-2031.

Barten, P. G. H. 1991. Resolution of liquid crystal displays. Proceedings of Society for Information Display International Symposium Digest of Technical Papers, Vol. XXII , pp. 772-775.

Barten, P. G. H. 1993. Effects of quantization and pixel structure on the image quality of color matrix displays. Journal of the Society for Information Display, Vol.1, No. 2, pp. 147-153.

Beaton, R. J. and Farley, W. W. 1991. Comparative study of the MTFA, ICS, and SQRI image quality metrics for visual display systems. Final Report, AL-TR-1992-0001.

Beasley, J. H., Martin, J. S., Klymenko, V., Harding, T. H., Verona, R.W., and Rash, C. E. 1995. A characterization of low luminance static and dynamic modulation transfer function curves for P-1, P-43, and P-53 phosphors. Fort Rucker, AL: U. S. Army Aeromedical Research Laboratory. USAARL Report No. 95-29.

Biberman, L. M. (Ed.) 1973. Perception of displayed information. New York: Plenum Press.

Biberman, L. M. and Tsou, B. 1991. Image display technology and problems, with emphasis on airborne systems. Institute for Defense Analysis, IDA paper P-2448.

Bitzakidis, S. 1995. Moving image enhancement on AMLCDs. Journal of the Society for Information Display. Vol. 3, No.4, 249-255.

Castellano, J. A. 1992. Handbook of display technology. Academic Press: San Diego, CA.

Curtin, C., and Infante, C. 1996. Fundamentals of Emissive Displays (short course s-3). Santa Ana, CA: Society for Information Display, p. 23.

Dragon, A. 1993. Sharp Application Note. Liquid Crystal Displays, Image quality: Measurements and definition.

Farrell, R. J. and Booth, J. M. 1984. Design Handbook for Imagery Interpretation Equipment. Seattle, WA: Boeing Aerospace Co.

Feltz, J. C. 1990. Development of the modulation transfer function and contrast transfer function for discrete systems, particularly charge-coupled devices. Optical Engineering, Vol. 29, No. 8, pp. 893-904.

Greeson, J. C. Jr. 1996. International standards organization ergonomic standards for displays (seminar F-6). Society for Information Display Seminar Lecture Notes, Vol. II. Santa Ana, CA.

Harding, T. H., Martin, J. S., Beasley, H. H., Rash, C. E., Garrard, J. A. 1996. A survey of flat panel display technologies. Fort Rucker, AL: U. S. Army Aeromedical Research Laboratory. USAARL Report No. 96-19.

Harding, T. H., Martin, J. S., Beasley, H. H., Rash, C. E. 1997. A visual evaluation near the threshold of acuity of five color liquid-crystal flat-panel displays. Fort Rucker, AL: U. S. Army Aeromedical Research Laboratory. USAARL Report No. 97-29

Holst, G. C. 1995. Electro-optical imaging system performance. Winter Park, FL: JCD Publishing and Bellingham, WA: SPIE Optical Engineering Press.

Infante, C. 1993. On the modulation transfer function of matrix displays. Journal of the Society for Information Display, Vol. 1 No. 4, pp. 449-450.

Karim, M. A. (Ed.). 1992. Electro-optical Displays. New York: Marcel Dekker, Inc.

Keller, P. A. 1997. Electronic display measurement. New York: John Wiley & Sons, Inc.

Klymenko, V., Harding, T. H., Martin, J. S., Beasley, H. H., Rash, C. E., and Rabin, J. C. 1997. Image quality figures of merit for contrast in CRT and flat panel displays. Fort Rucker, AL: U. S. Army Aeromedical Research Laboratory. USAARL Report No. 97-17.

Lindholm, J. M. 1992. Perceptual effects of Spatiotemporal sampling. In (M. A. Karim-Ed.) Electro-optical Displays. (Chapter 19, pp 787-808). NY: Marcel Dekker, Inc.

MacDonald, L. W., and Lowe, A. C. (Eds.) 1997. Display Systems: Design and Applications. New York: John Wiley & Sons, Inc.

Meister, D. 1984. Human engineering data base for design and selection of cathode ray tube and other display systems. San Diego, CA: Navy Personnel and Research Training Center. Report No. NPRDC TR 84-51.

Moon, D. 1986. Image quality for electro-optical displays: variables, metrics, and measurements. Ph.D. dissertation, University of Dayton, Ohio.

North Atlantic Treaty Organization. Design criteria for flat panel technology displays. Study 7095 for proposed NATO Standardization Agreement (STANAG7095).

Oakley, D. 1984. Raster graphics display technology: A review. Displays, Oct., pp. 229-234.

Olson, B. and Balram, N. 1996. Video quality on AMLCD versus Shadow-Mask CRT. Proceedings of SPIE: Cockpit Displays III, Aerosense'96 , Vol. 2734, pp.12-22.

Parker, D. 1997. The dynamic performance of CRT and LC displays. In (L. W. MacDonald and and A. C. Lowe-Eds.) Display Systems: Design and Applications. (Chapter 18, pp 353-364). New York: John Wiley & Sons, Inc.

Rabin, J. and Wiley, R. 1995. Dynamic visual performance: Comparison between helmet-mounted CRTs and LCDs. Journal of the Society of Information Display, Vol. 3 No. 3, pp. 97-100.

Rash, C. E. and Becher, J. 1982. Analysis of image smear in CT displays due to scan rate and phosphor persistence. Fort Rucker, AL: U. S. Army Aeromedical Research Laboratory. USAARL Report No. 83-5.

Rash, C. E. and Verona, R. W. 1987. Temporal aspects of electro-optical imaging systems. Imaging Sensors and Displays II, Proceedings of SPIE, Vol. 765, pp. 22-25.

Roehrig, H., Blume, H., Ji, T. and Browne, M. (1990). Performance tests and quality control of cathode ray tube displays. Journal of Digital Imaging, Vol. 3, No. 3, pp. 134-145.

Seeley, G. W., Fisher, H. D., Stempski, M. O., Borgstrom, M., Bjelland, J., and Capp, M. P. 1987. Total digital radiology department: spatial resolution requirements. American Journal of Radiology, Vol. 148 No. 2, pp.421-426.

Silverstein, L. D., Kranz, J. H., Gomer, F. E., Yeh, Y. and Monty, R. W. 1990. Effects of spatial sampling and luminance quantization on the image quality of color matrix displays. Journal of the Optical Society of America, A, Vol.7, No. 9, pp.1955-1968.

Snyder, H. L. 1973. Image quality and observer performance. In (L. Biberman) Perception of Displayed Information. New York: Plenum Press, pp 87-118.

Snyder, H. L. 1980. Human visual performance and flat panel display image quality. Virginia Polytechical Institute and State University Technical, Report HFL-80-1.

Sobel, A. 1992. Flat panel displays. In (M. A. Karim) Electro-optical Displays. NY: Marcel Dekker, Inc.

Tannas, L. E. Jr.(Ed.) 1985. Flat Panel Displays and CRT. New York: Van Nostrand Reinhold.

Tannas, L. E. Jr. 1995. Overview of Electronic Information Displays. SID Lecture Notes, Vol.1, pp.M1/1-M1/50.

Task, H. L. 1979. An evaluation and comparison of several measures of image quality for television displays. Wright Patterson AFB, OH: Aerospace Medical Research laboratory. Report No. AMRL-TR-79-7.

Task, H. L. and Verona, R.W. 1976. A new measure of television display quality relatable to observer performance. Wright Patterson AFB, OH: Aerospace Medical Research laboratory. AMRL-TR-76-73.

Toms, M. L. and Cone, S.M. 1995. Subjective LCD-CRT comparison of USAF transport aircraft cockpit electronic display formats. Proceedings of SPIE, Cockpit Displays II, Vol.2462, pp. 36-46.

Toms, M. L., Cone, S. M., and Cavallaro, J. J. 1995. Subjective image quality comparisons of AMLCD and CRT implementations of electronic flight formats. Final Report (WL-TR-95-3074). Veda, Inc.

Van Meeteren, A. 1973. Visual aspects of image intensification. Ph.D. Dissertation, University of Utrecht, The Netherlands.

Verona, R.W. 1992. A Comparison of CRT Display Measurement Techniques. Proceedings of SPIE, Helmet Mounted Displays III,, Vol. 1695, pp. 117-127.

Verona, R. W., Beasley, H. H., Martin, J. S., Klymenko, V. and Rash, C. E. 1994. Dynamic sine wave response measurements of CRT displays using sinusoidal counterphase modulation. Fort Rucker, AL: U. S. Army Aeromedical Research Laboratory. USAARL Report No. 94-22.