# Visual Display Technology

**D. L. Post and H. L. Task**

Air Force Research Laboratory, Wright-Patterson Air Force Base, OH45433-7022, USA

## 1. INTRODUCTION

Many of the design problems encountered by human factors professionals involve electronic visual displays. Although modern human factors training often covers the psychophysical, perceptual and cognitive aspects of information display, education about display hardware — which is often needed to understand and solve display design problems and work productively with display engineers — is rare. This problem is exacerbated by the fact that most of the literature on display hardware is written for engineers and many aspects of the field are evolving quickly. This entry summarizes the major human factors parameters that are relevant to electronic visual displays, explains the principles of operation for the display technologies of greatest importance today and describes their characteristics insofar as they affect visual performance, and discusses the most common types of display systems that incorporate these technologies.

## 2. DISPLAY HUMAN FACTORS PARAMETERS

### 2.1. Photometric Characteristics

One of the most fundamental display characteristics is brightness. Brightness is assessed usually by measuring its psychophysical correlate, luminance, with a photometer. Another basic characteristic is contrast, which is a ratio of on and off pixel luminances; Table 1 lists common measures. Ambient illumination often reduces display contrast by adding reflected light to that coming from the display. Display screens are usually treated to reduce this problem, although most treatments reduce display luminance.

Grayscale is the number of luminances the display can produce at a fixed setting of the brightness and contrast controls, expressed usually as the base-2 logarithm of that number with units called "bits." For example, "8 bits" of grayscale denote 256 levels. The visual system's response to differing luminances is compressive, so technologies having exponentially increasing video-input-to-luminance response functions come closest to providing equally discriminable grayscale levels (although single steps may not actually be discriminable).

For most display technologies, grayscale limits are not inherent and result instead from built-in digital drivers. The exceptions are technologies having pixels that can only be fully on or off. A few technologies — notably CRT — have analog grayscale, which means their luminance can vary continuously and grayscale is limited, if at all, only by their video input signal. Grayscale can be produced in technologies having only on and off pixels by temporal multiplexing, which involves turning pixels on for varying fractions of a frame period, or by dithering, which is a type of spatial multiplexing that involves turning groups of adjacent pixels on and off so the group average has the desired luminance. Usually, dithering is used only as a supplement to temporal multiplexing. Emissive on–off technologies can be dimmed only by using up some of their grayscale (non-emissives can be dimmed by adjusting the illuminant), so they may exhibit contrast artifacts if they are dimmed.

The angular distribution of light emission for some display technologies is approximately Lambertian, which means their luminance is essentially independent of viewing angle; others depart from this ideal in ways that usually make the display harder to read as one moves off axis. LCD often exhibit additional anomalies off axis, namely reduced or even reversed contrast and color shifts, although manufacturers are using increasingly sophisticated methods to reduce these problems.

**Table 1. Common measures of contrast.**

| Name | Defining equation[1] |
|---|---|
| Contrast ratio[2,3] | $L_{max} / L_{min}$ |
| Relative contrast[2,3] | $(L_{max} - L_{min}) / L_{min}$ |
| Luminance contrast[3] | $(L_{max} - L_{min}) / L_{max}$ |
| Luminance modulation or Michaelson contrast | $(L_{max} - L_{min}) / (L_{max} + L_{min})$ |

[1] $L_{max}$ is the greater of the two luminances; $L_{min}$ is the lesser.

[2] The contrast ratio and relative contrast are undefined when $L_{min} = 0$.

[3] For symbology, $L_{max}$ and $L_{min}$ are replaced sometimes by the symbol and background luminances, respectively, in the first three equations, which creates the possibility of contrast ratios < 1 and negative values in the other two cases.

### 2.2. Spatial Characteristics

"Resolution" is used widely, but inconsistently, to describe the amount of spatial detail a display can produce. Display designers often specify resolution in cycles or pixels per unit distance on the display (e.g. cycles/cm), where a cycle is one complete sine wave. Sometimes, the unit distance is implicitly the display's width or height, in which case the specification gives the total number of cycles or pixels that can be produced. Specifications for television sets often state only the number of raster lines that can be displayed. Although conventional television broadcasting provides only 525 lines in the USA (only 483 are visible, though) and 625 lines in the UK and Europe (575 visible), the capacity to display more implies higher resolution within each line.

**Table 2. Contemporary digital resolution standards.**

| Label | Aspect ratio[1] | Resolution (H × V) in pixels |
|---|---|---|
| QVGA | 4:3 | 320 × 240 |
| VGA | 4:3 | 640 × 480 |
| SVGA | 4:3 | 800 × 600 |
| XGA | 4:3 | 1024 × 768 |
| SXGA | 5:4 | 1280 × 1024 |
| UXGA | 4:3 | 1600 × 1200 |
| SDTV[2] | 4:3 or 16:9 | 704 × 480 |
| HDTV[3] | 16:9 | 1920 × 1080 |
| | 16:9 | 1280 × 720 |

[1] Aspect ratio is the ratio of image width to height.

[2] SDTV approximates the resolution of the 525-line analog system used currently in the USA. It requires rectangular pixels in one of two possible sizes to achieve the stated aspect ratios. All other entries in Table 2 use square pixels.

[3] HDTV offers two possible resolutions.

Psychophysicists usually prefer cycles or pixels per unit angle subtended at the viewer's eye (e.g. cycles/degree) because these measures account for the effects of viewing distance and display size on image appearance. Thus, two displays having equal resolution from a display designer's perspective may have different resolutions from a psychophysicist's perspective and vice versa.

Additional confusion is introduced sometimes by failing to distinguish between the resolution of the display and that of the video system driving it or the image it is displaying. For example, using an SXGA (Table 2) graphics card does not give a display SXGA resolution, nor does showing a VGA image yield VGA resolution. Another poor practice is confusing resolution with the amount of visual information in an image. For example, a photograph may be described as having higher "resolution" than a drawing of the same scene.

CRT produce images by drawing on a screen with a continuously moving electron beam. Their resolution is mainly a function of their video bandwidth, refresh rate and electron-beam spot size. These parameters' effects are described most completely by the spatial modulation transfer function, which graphs the response to sinusoidal inputs as a function of their frequency, in a manner analogous to the frequency–response plots one often sees for audio equipment.

Other display technologies consist of a matrix of discrete pixels that are set to the desired luminances by addressing them one or possibly several rows at a time, with all columns loaded simultaneously — methods called "line by line" and "multi-line" addressing, respectively or, more generally, "matrix addressing." For these technologies, resolution is simply a function of the number of horizontal and vertical pixels. Another addressing method is to connect each pixel to its own driver. This "direct" addressing technique allows each pixel to be driven throughout each frame, but is practical only for simple applications, requiring only a few dozen pixels or segments (e.g. watches, calculators, and home appliances) because otherwise the number of connections and drivers becomes prohibitive.

Matrix displays come in two types: passive, which have only a pair of electrodes at each pixel, and active, which add a switch (typically, a thin-film transistor, or "TFT," but sometimes a diode) at each pixel so voltage can be maintained throughout the frame period. Passive-matrix displays are cheaper but can apply voltage to the pixels only while their row is selected. Therefore, their luminance and contrast tend to be inferior to otherwise identical active-matrix displays and diminish as the number of rows increases, thus limiting the number of rows that can be provided.

## 2.3. Temporal Characteristics

It is obvious that changing a display image requires redrawing it, but it is less obvious that, for most display technologies, the image fades unless it is refreshed. (The exceptions are bi-stable technologies, which have pixels that remain on or off until they are switched.) In the USA, television refreshes at 30 Hz by alternately drawing all the odd- and then even-numbered lines at 60 Hz. Each set of lines is a "field," both fields constitute a "frame," and this alternating system is called "2:1 interlacing." (In the UK and Europe, the field rate is 50 Hz, yielding a 25-Hz frame rate.) Interlacing reduces bandwidth requirements while avoiding flicker for typical television images, but tends to produce jitter in text. Therefore, computer images are usually non-interlaced (this method is known also as "progressive scanning") and refresh at 50 Hz or more. The rate needed to prevent flicker increases as display luminance, contrast, and angular subtense increase. Sometimes, the rate must also be adjusted to prevent beats with fluorescent lighting from becoming visible.

Display luminance is greatest if the pixels emit throughout each frame, but this can cause image smear if the observer tracks a moving object. Therefore, motion rendition is best for display technologies that produce adequate time-averaged luminance while emitting only for small fractions of a frame period. A display having pixels that cannot switch completely during one frame will exhibit contrast losses for moving images, so it is also important that the pixels respond quickly. Most LCD technologies are troublesome in this regard. Speed is even more crucial for technologies that rely on temporal multiplexing to achieve grayscale or use temporal integration to produce color, because in these cases the pixels must switch more than once during each frame.

## 2.4. Colorimetric Characteristics

The range of colors a display can produce is called its "color gamut" and is determined by its red, green and blue (RGB) primaries. The closer the primaries come to being monochromatic, the more saturated they become and the larger the gamut becomes; however, no finite set of primaries can reproduce the entire range of colors humans can perceive. Furthermore, increasing a primary's saturation requires reducing its spectral bandwidth, which typically reduces its luminous efficacy; therefore, primary selection often involves a tradeoff between gamut and luminance.

Television broadcasting uses a standard set of primaries that have been chosen for compatibility with available CRT phosphors. Developers of other color display technologies generally try to match or exceed television's gamut. For emissive technologies, one approach uses a white emitter with a patterned RGB color filter in front; another uses RGB emitters, possibly enhanced by a patterned color filter. For non-emissive technologies, the common approach uses a white illuminant and color filters. Luminous efficacy can often be increased by selecting an illuminant having its radiance concentrated in three RGB wavebands and matching the filters' transmittances to those wavebands.

Every color display has a "white point," which is the chromaticity it produces when its RGB channels are driven equally and may or may not be user-adjustable. Television broadcasting and a few other applications use standard white points, defined in terms of Commission Internationale de L'Éclairage (CIE) chromaticity coordinates. Differences in white points are often at least partly responsible for problems in obtaining accurate color transfer when images are viewed on different displays, printed, or scanned.

Most displays produce color by spatial integration: The screen is divided into RGB subpixels, which subtend a visual angle so small that the diffraction patterns they produce on the retina overlap and mix. Varying the subpixel luminances produces what appear to be full-color pixels having any color an RGB mixture can produce.

A similar method, used by most projection displays, is addition: Separate RGB images are produced by three monochrome displays and superimposed internally or at the screen. Addition can yield higher resolution than spatial integration because every pixel can take on any color, but it requires three displays and is therefore bulkier and more expensive.

Color can also be produced by temporal integration — a technique often called "field sequential": the display generates spatially superimposed RGB fields in rapid sequence, within a single frame period, and the visual system integrates them temporally to form a full-color image. Temporal integration has the same resolution advantage as addition, but requires fast displays, sacrifices luminance because each color field is displayed for no more than one-third of each frame, and the images can break up into their RGB components if there is relative motion between the image and the viewer's eye.

Finally, a subtractive process can produce color: white light passes through a stack of three transmissive displays, and each display modulates the R, G, or B component of the light so a full-color image emerges at the front. Subtraction also has a resolution advantage, but requires three displays, careful design to minimize transmission and contrast losses as the light proceeds through three apertures, thin displays to avoid parallax in direct-view applications, and may require special optics to focus the layers simultaneously in head-mounted and projection applications.

### 2.5. Size and Weight

CRT tend to be the heaviest and bulkiest of the display technologies because they are funnel-shaped glass vacuum tubes. Their shape results from the need to provide space for the electron beam to scan the screen. Their weight increases exponentially with screen size because the glass must be thickened to withstand the atmospheric pressure resulting from the internal vacuum. Most other technologies consist only of thin, lightweight layers sandwiched between two flat pieces of glass, so they are more compact and their weight is proportional with screen size. LCD usually require a backlight, though, which can increase their weight and depth substantially.

## 3. EMISSIVE DISPLAY TECHNOLOGIES

### 3.1. Cathode-ray Tube (CRT)

A CRT produces images by modulating an electron beam as it scans across a phosphor screen. The electrons are produced by heating a cathode, accelerated toward the screen, and focused and scanned by dynamic electromagnetic or electrostatic fields. Dynamic focusing compensates for the changing distance to the screen as the beam scans. Most CRT scan the beam in a repeating "raster" pattern, consisting of closely spaced horizontal lines; however, oscilloscope CRT — and some that are used exclusively to show symbols and line graphics — use a "stroke" pattern that moves the beam as if it were a pen. Stroke writing provides increased luminance (because the phosphor can be excited more times per s) and resolution.

Most color CRT scan three electron beams across a patterned screen containing RGB phosphor dots or stripes, thus producing color by spatial integration. A metal structure ("shadowmask") behind the screen, having one opening for each RGB triplet, helps ensure that each beam strikes only the appropriate phosphor; adjustable magnets also assist. Older, "delta-gun" CRT arrange the beams in a triangular pattern and include user-adjustable circuitry to ensure a condition called "convergence" that, if violated, produces color fringes that are noticeable especially around white objects on the screen and reduce resolution. Modern, "in-line gun" CRT arrange the beams in a row and use circuitry and magnets, which are set at the factory and rarely adjusted afterward, to achieve convergence. Color can also be produced by temporal integration, using a white-phosphor CRT in conjunction with a switchable or rotating color filter — this approach is used sometimes in head-mounted systems.

### 3.2. Vacuum Fluorescent Display (VFD)

The VFD is a type of flat CRT that excites a low-voltage phosphor using electrons produced by heating a cathode grill consisting of fine, heated wires. The flow of electrons is controlled by a wire grid that lies between the cathode and a phosphor-coated opaque anode array. The photons then pass through the grid, cathode, and a cover glass. Another arrangement, which deposits the phosphor on the cover glass, is also used sometimes. The anode array can be matrix addressed or, more commonly, arranged in segments. Usually, a blue-green emitting phosphor is used because it has the highest luminous efficacy, but a good range of hues up to red are available.

VFD are used widely as simple numeric, alphanumeric, and bargraph displays measuring a few square centimeters in products like VCRs, but it has not been practical to > ~160 cm$^2$ for more general-purpose use because this would require adding internal support to control the gap between the cathode and anode, which would interfere with the electron flow. Grayscale and dimming can be produced by varying the anode voltage. Motion rendition is good due to low response time and the emission is Lambertian, so viewing angle is not a problem. Color can be produced by spatial or temporal integration, but multicolor units consisting of differently colored monochrome areas are far more common.

### 3.3. Field-emission Display (FED)

The FED is another type of flat CRT, in which a phosphor-coated glass screen is excited by electrons produced by an addressable matrix of tiny cathodes. There are basically two types: low-voltage (40–100 V), which require special phosphors and have not yet matched the luminances available from CRT, and high-voltage (5–10 kV), which use conventional phosphors and can match CRT luminances but are vulnerable to damage and failure caused by arcing. Grayscale and dimming can be produced by temporal multiplexing, cathode modulation, or dithering, although multiplexing supplemented by cathode modulation usually yields the best result. The response time is low and the emission is Lambertian. Color is produced by spatial integration, using RGB phosphors, although temporal integration is also feasible.

### 3.4. Image Intensifier Tube (I$^2$T)

The I$^2$T is a unique combination of sensor and display in one package. Contemporary third-generation I$^2$T consist of three primary elements: (1) a photo-cathode screen, which converts photons to electrons; (2) a micro-channel plate (MCP) that accelerates and multiplies the electrons; and (3) a phosphor screen — coated on a fiber-optic or plain glass faceplate — that converts the multiplied electrons to photons. Typically, a power supply is included, to convert battery power to the relatively high voltage needed by the MCP. In operation, an image is focused onto the photo-cathode, which then emits electrons in the general direction of the adjacent MCP. The MCP contains millions of tiny tunnels (micro-channels), which amplify the incident electrons in proportion to their number. After exiting the MCP, they are accelerated toward the nearby phosphor screen by an electric field. The phosphor nearest each micro-channel glows with a luminance that is proportional to the number of incident electrons, thus producing a monochrome image that replicates the image focused on the photo-cathode, except that its luminance is greatly amplified. See the section on NVG for information on typical uses of $I^2T$.

### 3.5. Light-emitting Diode (LED)

An LED is basically a solid-state material — typically gallium phosphide or gallium arsenide — that has been treated so it consists of two layers that form a P-N junction diode, and passing a current across the layers causes them to emit light. A reflector is placed behind the diode to direct rear-emitted light toward the front, and a lens is often incorporated at the front to further shape the output. The diodes can be arranged in segments to form numeric, alphanumeric, or bargraph displays or laid out as an addressable matrix. A full range of colors, including RGB, is available, so color displays using spatial integration can be created. Grayscale and dimming can be produced by varying the current or temporal multiplexing. The reflectors and lenses are usually designed to produce non-Lambertian emission, to increase on-axis luminance. Large matrices of LED for use as general-purpose desktop color displays have not proven to be practical, mainly because of the difficulty of assembling discrete devices at the required densities (producing different colors on one substrate has proven impossible, thus far), but are used widely as message signs. Numeric units — which were once common in wristwatches and calculators until being replaced largely by LCD — are used in clocks and, along with alphanumeric and bargraph units, as instrument readouts.

### 3.6. Organic Light-emitting Diode (OLED)

An OLED display consists of a thin transparent layer of organic semiconductors or polymers, sealed usually between two glass or plastic substrates bearing orthogonally oriented electrodes on their inner surfaces. The front electrode is transparent and the rear is reflective. Applying a current across the electrodes causes the organic layer to emit visible light. Plastic substrates may make it possible to produce flexible displays, which can be rolled up or folded. Silicon substrates have also been used in experimental, miniature active-matrix OLED designed for head-mounted applications. Currently, this technology is under development, and although it is regarded widely as very promising, there are few commercial products. Grayscale and dimming are produced by varying the electrode current. Viewing angle is good due to Lambertian emission. OLED materials capable of producing RGB light are available, so full color can be produced by spatial integration or, if the substrates are stacked, by addition; however, in the latter case, the stack must be very thin to avoid parallax.

### 3.7. Electroluminescent (EL) Display

The most common type of EL display today uses alternating current to excite a thin-film EL phosphor, yielding the "ACTFEL" structure. The phosphor is sandwiched between thin insulating layers and this multi-layer structure is deposited typically on a glass substrate bearing transparent electrodes. Reflective, orthogonally oriented electrodes are then applied and the phosphor is viewed through the glass. Thus, only one substrate is required and the display can be substantially thinner and lighter than other flat panel technologies. Ceramic and silicon (supported by ceramic) substrates are also used, although these designs require a cover glass. Miniature devices using active-matrix addressing have been produced for head-mounted use. The angular emission is Lambertian. Grayscale and dimming for EL use temporal multiplexing, enhanced sometimes by dithering. Color is achieved by spatial integration, using either patterned phosphors or a white phosphor plus color filter array.

### 3.8. Plasma Display Panel (PDP)

A PDP consists mainly of a gas that is sealed between two glass substrates bearing orthogonally oriented electrodes on their inner surfaces and, usually, a reflective coating at the rear. Passing a current across the electrodes ionizes the gas, producing a plasma that emits light. Monochrome PDP use neon, which emits reddish-orange light. In color displays, an ultraviolet-emitting gas is used and the ultraviolet excites phosphors that are coated on the inner surface of one of the substrates. Thus, a color PDP is similar to a fluorescent light bulb. Either alternating or direct current can be used, but the former is more common today. PDP are inherently bi-stable, so grayscale and dimming are produced by temporal multiplexing, enhanced sometimes by dithering. Response is very fast, so motion rendition is good. Color is produced by spatial integration. The emission angle is Lambertian for color panels but is isotropic for monochrome panels, causing luminance to increase with viewing angle.

## 4. NON-EMISSIVE DISPLAY TECHNOLOGIES

### 4.1. Illumination Sources

Non-emissive displays require a light source. For some applications, such as LCD wristwatches, ambient illumination reflected off the display may be sufficient, but usually a built-in source is needed. Most direct-view LCD use one or more tubular fluorescent lamps, mounted in a box containing a reflector or coupled to a light guide. Other approaches that provide more uniform illumination have been tried, though, including EL panels, flat and serpentine-shaped fluorescent lamps, and CRT. Projection systems require more light than can be provided compactly by a fluorescent source; they use xenon or metal-halide arc lamps. Xenon has a broad spectrum that must be filtered heavily to produce acceptable RGB primaries, thereby reducing overall luminous efficacy. Metal-halide lamps emit mainly in narrow wavebands dictated by the choice of halides, thereby reducing the need for spectral filtering; the fill gas is unavoidably excited too, though, adding at least some broad-spectrum light. Compact, solid-state lasers producing RGB wavelengths have become available recently. Their narrowband spectra are advantageous for many display purposes, and their prices, efficiencies, lifetimes, sizes, weights and cooling requirements are moving rapidly to a point that may make them attractive for projection uses.

### 4.2. Liquid Crystal Display (LCD)

Liquid crystals (LC) are elongated molecules that flow like liquids at typical ambient temperatures and tend to align with one another. There are many types of LC and ways of using them to make displays, but the most common today are nematics, sealed between two pieces of glass that have been treated so the LC lie parallel with the glass surfaces and undergo a rotation top to bottom, forming a helix — a configuration called "twisted nematic" (TN). In a transmissive TN-LCD, polarizers and transparent electrodes are added to the glass surfaces, and light passing through one polarizer rotates with the twist until it strikes the second polarizer, which either passes or absorbs it, depending on whether the polarizers are aligned with each other or crossed. Applying a voltage across the electrodes rotates the LC so they are perpendicular to the glass, destroying the twist and thereby reversing the second polarizer's effect. Intermediate voltages produce intermediate transmittances and, hence, grayscale. Reflective TN-LCD have a reflective surface at the rear and only a front polarizer, but work basically the same way. The electrodes can be segmented or arranged in a matrix.

Most passive-matrix LCD today use a twist ranging from 180 to 270°, called a "supertwisted nematic" (STN) configuration, which increases the number of rows that can be addressed but reduces viewing angle, response speed, and grayscale. Active-matrix LCD (AMLCD) provide the best viewing angle, speed, and grayscale, but are more costly. Other important types of LCD today are ferroelectrics (FLC), which are bi-stable and respond to control voltages very quickly but have difficulty producing grayscale, polymer-dispersed (PDLC) and nematic curvilinear aligned phase (NCAP) LCD, both of which vary between transparent and light-scattering states and use no polarizers but have trouble producing high contrasts, and polymer stabilized cholesteric texture (PSCT) LCD, which are bi-stable and therefore do not produce grayscale readily.

### 4.3. Digital Micromirror Display (DMD)

A DMD is an addressable array of tiny mirrors, each of which is mounted on a flexible stalk. Each mirror can be rotated very quickly (e.g. 10 ms) to either of two orientations, thereby deflecting incoming light out to a projection lens or an absorber. Thus, each mirror is a pixel that can be either on or off. Grayscale is produced by temporal multiplexing; dimming is accomplished by modulating the illumination. Color is produced usually by temporal integration using a filter wheel, but triple-DMD systems using addition and dual-DMD hybrids are also available. DMD do not lend themselves to direct-view applications because of the need for a projection lens; furthermore, it is impractical at the moment to make them bigger than a few square centimeters in area. They are used widely in projectors, though, and head-mounted applications are also possible.

## 5. OPTICAL DISPLAY SYSTEMS

### 5.1. Projectors and Screens

Projection displays use one or more lenses to enlarge and focus an image of an internal display onto a screen. Rear projectors usually incorporate the screen into the housing and therefore tend to be bulky, whereas front projectors use a separate screen and are therefore more compact but require the screen to be distant from the projector. Rear projection screens are usually treated to reduce reflections from room lighting and incorporate lenses to produce a non-Lambertian light distribution that reduces the image's viewing angle but provides greater luminance within the intended angle — a feature called "gain." Front projection screens can also have gain.

Some projectors use CRT to create an image; others, termed "light valves," use non-emissive displays. The least expensive, lightest, and most compact design for a color projector uses a single display and lens system; field-sequential color is most common in self-contained versions, whereas spatial integration is the norm in LCD overlays for overhead projectors (subtractive overlays have also been produced). Greater luminance can be obtained, though, by using three displays to produce separate RGB images that are combined additively; this design is more common in self-contained systems. Most triple-display projectors use dichroic beamsplitters, which reflect one band of wavelengths while passing the rest, to add the images internally and, in light-valve projectors, to first separate white light into R, G, and B so each component can illuminate the appropriate display. Most CRT projectors, however, use a separate projection lens for each CRT and add the RGB images at the screen.

The most important human factors issues for projectors are resolution, luminance, contrast, optical distortion, viewing angle, size, and weight.

### 5.2. Head-up Display (HUD)

As the name implies, a HUD is intended to permit the operator of a vehicle (originally an aircraft and, more recently, an automobile) to view a display without looking down toward the instrument panel. It does this by superimposing a display image on the outside world. The original intent was only to assist weapon aiming, but another advantage has been recognized since: The virtual image is normally at or near optical infinity, so the operator does not have to refocus when switching from viewing the world scene to viewing the display.

A HUD consists of three major components: (1) an image source (e.g. a CRT); (2) optics to produce a virtual image of the image source; and (3) an optical combiner, which combines the virtual image with the directly viewed exterior world scene. There are basically two types of HUD: those that produce a real exit pupil and those that do not. If the HUD uses relay optics, it has a real exit pupil, which means at least one eye must be within the exit pupil to see any part of the display. If the HUD uses simple magnifying optics, it does not have a real exit pupil and eye position is less critical, although at least one eye must still be positioned within a limited volume of space to see the entire display.

The most important human factors issues for HUD are the combiner ratio (which affects the relative luminances of the symbology and outside world scene), viewing location, exit pupil size or viewing volume, symbology size, type, luminance, and color, stray reflections, and the size and shape of the field of view.

### 5.3. Head- and Helmet-mounted Display (HMD)

An HMD consists typically of three major components: (1) a miniature display; (2) optics to convey a virtual image of the display to the eye(s); and (3) some (usually adjustable) means of mounting the display and optics to the head. The optics may or may not include a combiner: If the HMD is meant to provide an image that can be superimposed on the external world scene, the optics must include a see-through combiner; if the HMD is for "immersion" only, a combiner is unnecessary and the optics are easier to design. Most HMD use miniature CRT or LCD, which must have very high resolution (in terms of cycles/mm or pixels/mm) to produce a satisfactory image after magnification by the optical system. One exception is the "virtual retinal display," which produces an image by scanning a modulated laser beam directly across the retina in much the same way a CRT scans an electron beam across a phosphor screen. Another exception scans an image of an LED array back and forth across the retina.

HMD can be monocular, bi-ocular, or binocular, monochrome, multi- or full-color, see-through or non-see-through, adjustable focus or fixed focus, and real exit-pupil forming or non-pupil forming. The most important human factors issues associated with them are weight, comfort, size, luminance, binocular alignment, resolution, optical distortion, optical distance of the image, size and shape of the field of view, exit pupil size or viewing volume, and the distance the eye can be from the optics ("eye relief").

### 5.4. Night Vision Goggle (NVG)

NVG are a special type of HMD. Most NVG are binocular, with each ocular consisting of three major components: (1) the objective lens, which produces an optical image on the input side of an (2) $I^2T$; and (3) an eyepiece lens to produce a virtual image of the $I^2T$'s output. NVG typically include assemblies that permit them to be adjusted and worn on the head. NVG containing second-generation $I^2T$ are most sensitive to wavelengths from 400 to 750 nm; third-generation devices are most sensitive to 650–900 nm.

The most important human factors issues associated with NVG are ocularity (binocular, bi-ocular, or monocular), output luminance, signal to noise ratio, optical distortion, eye relief, gain, adjustability (fore/aft, interpupillary distance, focus, tilt, up/down), weight, size, and comfort.

### 5.5. Three-dimensional (3D) or Stereoscopic Display

"3D" is used sometimes to refer to images seen on a conventional display that include depth cues such as perspective, surface shading and hidden lines. A stereoscopic 3D display, however, provides binocular disparity, so each eye sees a slightly different image that corresponds to what it would see in the real scene. There are several ways to produce binocular disparity. One is to use two independent image sources (e.g. CRT) and an optical system that presents a different image to each eye. Another approach is to use a single display and encode or multiplex the left- and right-eye images so each eye sees only the appropriate one. A common encoding technique, used with projection displays, polarizes the left- and right-eye images oppositely and superimposes them on the screen; the observers wear oppositely polarized lenses, so each eye sees only the appropriate image. A common multiplexing technique, used with direct-view displays, places electronic shutters in front of the eyes and alternates the left- and right-eye images on the display in synchrony with the opening and closing of the shutters.

The most important human factors issues associated with stereoscopic displays are the accuracy of the binocular disparities and hence depth cues, visibility of the images to the wrong eyes ("crosstalk"), frame rate, luminance, color, binocular alignment, and the effects of mismatch and lack of normal covariance between the images' optical distance (which is usually fixed) and the varying vergence cues and requirements produced by the binocular disparities.

## 6. FURTHER READING

For further information about display human factors issues, see Tannas (1985), Travis (1991), and Widdel and Post (1992). For display technology, see Csorba (1985), Tannas (1985), Sherr (1993, 1998), MacDonald and Lowe (1997) and Nelson and Wullert (1997). For optical display systems, see Weintraub and Ensing (1992) and Melzer and Moffitt (1997).

## ACKNOWLEDGMENTS

The authors thank Howard L. Funk (H.L. Funk Consulting), Bernard Lechner (consultant), Jeffrey Ohstrom (Thomson Components and Tubes Corp.) and Lawrence Tannas Jr (Tannas Electronics) for advising us on several important matters, Reginald Daniels and Gurdial Saini (Air Force Research Laboratory) for reviewing the FED and OLED sections, and especially Alan Sobel (Editor of *Journal of the Society for Information Display*) for reviewing the manuscript.

## REFERENCES

CSORBA, I.P., 1985, *Image Tubes* (Indianapolis: Howard W. Sams).

MACDONALD, L.W. and LOWE, A.C., 1997, *Display Systems Design and Applications* (Chichester: Wiley).

MELZER, J.E. and MOFFITT, K. (eds), 1997, *Head-mounted Displays* (New York: McGraw-Hill).

NELSON, T.J. and WULLERT II, J.R., 1997, *Electronic Information Display Technologies* (River Edge: World Scientific).

SHERR, S., 1993, *Electronic Displays* (2nd ed.) (New York: Wiley).

SHERR, S., 1998, *Applications for Electronic Displays* (New York: Wiley).

TANNAS, L.E., Jr (ed.), 1985, *Flat-panel Displays and CRT* (New York: van Nostrand Reinhold).

TRAVIS, D.S., 1991, *Effective Color Displays: Theory and Practice* (San Diego: Academic Press).

WEINTRAUB, D.J. and ENSING, M., 1992, *Human Factors Issues in Head-up Display Design: The Book of HUD* (OH: Wright-Patterson AFB, CSERIAC).

WIDDEL, H. and POST, D.L. (eds), 1992, *Color in Electronic Displays* (New York: Plenum).