Digital Image Quality: A Clinical Perspective

John A. Carrino, MD

What is medical image quality? To answer appropriately from a clinical perspective, one should first review the objectives of medical imaging. These can be broadly categorized as the ability to extract information (signal) from within living organisms (noise) and to provide spatially discrete mapping (anatomical information). In this sense, the goal of imaging is to act as a surrogate record of morphology and physiology.

What makes an image diagnostically useful? To answer, one should review the diagnostic process and chain of events from image generation to diagnosis (Fig. 3.1). To begin with, some type of energy source (eg, X ray, ultrasound, radiofrequency, or gamma ray) is used to expose a subject to develop a signal pattern. An imaging system detector (eg, imaging plate, transducer, coil, or photomultiplier tube) records the signal pattern. The image then must be constructed in either an analog or digital fashion. This record then must be displayed, which, again, may be in an analog or digital medium. An observer then views the images. The quanta of light that come off an image are detected on the retina of the eye. The retina contains an active neural network and performs functional differentiation. Therefore, processing begins at the retinal level. The sensory pathways and the visual cortex further process this information. Interpretation also relies greatly on nosology, the knowledge of normal anatomy, imaging techniques and artifacts, and disease processes—part of the reason that radiology training takes at least 4 years. A diagnosis is then generated and reflects a complex synthesis of all the components.

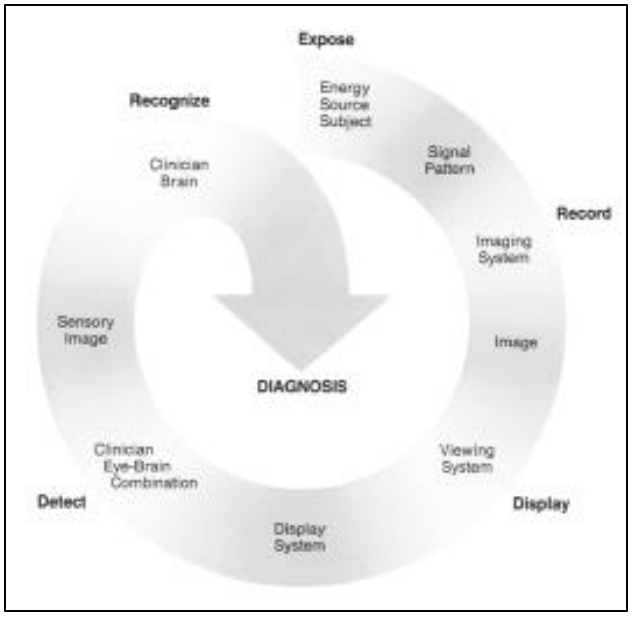

Figure 3.1 Conceptualization of the diagnostic process. Reproduced with permission from (*7*).

In the electronic environment (ie, a picture archiving and communication system [PACS]), image acquisition and display are uncoupled, and this paradigm may be extended to image quality assessment. Therefore, one should consider acquisition and display as separate image quality areas. Several quality measures may be used in assessing acquisition (Table 3.1). Some are specific to certain types of modalities, such as the detective quantum efficiency (DQE) measure for radiography. Others, such as the signal-to-noise ratio (SNR), may be more broadly applied. For example,

TABLE 3.1
Acquisition Quality Measure

- Modulation transfer function
- Detective quantum efficiency
- Signal-to-noise ratio
- Limiting spatial resolution
- Noise power spectrum
- Noise equivalent quanta
- Contrast (photometric) resolution
- Exposure dynamic range
- Uniformity
- Artifacts

human performance predicted by SNR shows 30%–50% efficiency for noise-limited tasks and ~50% efficiency for statistically defined signals and backgrounds. Efficiency is defined as the square of the ratio of human observer performance to ideal observer performance. However, there is no single acquisition quality measure that can sufficiently or completely characterize an imaging system or, more important, define the ability to make a diagnosis. Therefore, metrics based on arbitrary properties of an image are not sufficient predictors of human performance for clinically relevant visual detection tasks.

ACQUISITION

On the acquisition side, the most common clinical modalities are radiography, fluoroscopy, computed tomography (CT), ultrasonography, magnetic resonance (MR) imaging, and nuclear medicine imaging.

Radiography traditionally has been analog, using a film–screen combination to capture the signal from X rays that have been transmitted through a body part. Digital mechanisms in radiography come in two broad categories: computed radiography (CR) and direct radiography (DR). CR uses a photostimulable phosphor imaging plate (IP) and a screen analogous to a film–screen system. The latent image is created on the IP by light emitted from the screen after exposure to X rays. The latent image from the IP is read with a laser light through a processor to create the optical densities. DR is the direct conversion of electrons into signal and does not use a processor. It is based on electronics such as charged-coupled devices (CCDs), thin-film transistors (TFTs), and selenium or silicon detectors.

CT and MR imaging are inherently digital, with the signal being created by complex mathematical transforms generated by computers. CT imaging is a transmission modality using X rays as the energy source. Rather than generating a projection image (ie, a "shadowgram") from a single wide-area exposure, a cross-section is constructed by backprojections, based on data from a collimated X-ray beam that rotates around the patient. MR imaging is a type of emission tomography in which protons are excited by radiofrequency energy and "sing" back via echoes (eg, spin echo, gradient echo, etc.) as they return to the relaxation state. Ultrasound and nuclear medicine are also captured by digital devices but have traditionally been displayed on an analog medium (film). However, because of their lower spatial resolution and smaller memory requirements, they are target areas for early implementation of PACS. Traditionally, the focus for modality acquisition quality has been on engineering better design and construction for imaging systems.

Display

Several media are used for image display, including film, paper, cathode ray tube (CRT), and flat-panel detectors. Flat-panel detectors for medical imaging may be plasma display panels (PDPs) or liquid crystal display (LCDs). Each of these display media has been used for medical imaging, and it is likely that each will continue to have a role in some fashion. It is unlikely that digital display panels will completely replace paper or film for all applications in the near future.

Artifacts

Artifacts are produced by the acquisition device and/or subject factors. Postprocessing may be performed to decrease artifacts, but there is often a penalty of increased noise within the image. Although many observers learn to recognize artifacts, the need to negotiate around numerous artifacts leads to reduced observer confidence. Task-specific interference may also affect the diagnosis. Image variability is derived from two main sources. System variability is related to image acquisition (eg, "quantum" noise), and object variability is mostly related to "anatomical" noise. Obviously, normal anatomy is an essential consideration for detecting structural abnormalities. The three main areas in which image quality affects observer performance are the search process, false positives (false signal), and masking (camouflaging). In addition, difficulties in detecting subtle abnormalities often arise because of the psychovisual elements that influence the detection process and the nature of decision making.

Light and Display Quality

When discussing image quality, it is important to review the various measures of light involved in clinical diagnostic imaging. Illuminance is the measure of light in a room (light falling on a surface) and is typically measured in foot-candles, candela per square meter per steradian (sr), or lux (1 foot-candle = ~10.764 lux; and 1 lux = 1 candela/m^2/sr). Luminance is a measure of light emitted from a surface into a small solid angle (usually over about 6°). It is measured in nits, candela per square meter or foot-lamberts (ft-l) (1 ft-l = ~10.764/pi nits; 1 nit = 1 candela/m^2). More important then the absolute values are the relative values that exist between different displays. The luminance of a typical viewbox will range from 1000 to 5000 nits,

depending on whether it is low intensity, high intensity, or used for mammography applications. The luminance of CRT monitors employed for medical imaging may range from 75 to 600 nits but more typically is in the 75–100 nits range. Therefore, an order of magnitude separates the luminance available from a viewbox from that available on a CRT monitor. The implications become quite apparent when attempting to view images in different ambient light conditions.

THE HUMAN VISUAL SYSTEM

It is also important to consider certain aspects of the human visual system (HVS) and determine whether softcopy images are suitable for viewing with the human eye. Important characteristics of the HVS that should be considered are contrast sensitivity (to both spatial resolution and temporal resolution) and dynamic range. Contrast sensitivity and spatial resolution depend on a threshold contrast, with enough of a difference in luminance from the baseline luminance to perceive a discrete structure. The probability that a standard observer can perceive an object also depends on the test pattern size, shape, frequency content, and any background noise (structured or unstructured). Contrast sensitivity and temporal resolution are largely related to the refresh rate of the monitor screen. The eye becomes quite sensitive to refresh rates that are <30 Hz (cycles/sec) but also perceives flicker up to ~60 Hz. However, this is not usually an issue, because most monitors can produce refresh rates >70 Hz and thus have "flicker-free" displays.

Dynamic range is quite interesting. Across the entire visible spectrum, the absolute range for the HVS is >6 orders of magnitude. However, for clinical imaging work, the range is much reduced. For grayscale, about 60–90 just noticeable differences (JNDs) are available to the HVS, and, for color, >500 JNDs are available. Monochrome CRTs have spatial resolution in the range of a 1–2-k matrix. The contrast ratio (maximum luminance over minimum luminance) is variable and can range from 300 to 600 nits. This is because the maximum luminance varies close to 600 nits and the minimum approaches the scotopic region of the HVS. Veiling glare is related to light reflected off the face plate from internally generated sources and unequivocally impairs diagnostic ability. Glare also comes from outside sources that appear as reflections in the face plate (specular reflections). The photometric resolution for most displays is 8–12 bits but also depends on the gamma (characteristic) curve of the specific monitor.

As discussed in other chapters of this primer, the dynamic range of a display is quite important. Some confusion exists about the differences between brightness and contrast versus window and level. These are different but can influence the same attributes of an image. They are basically separated at the digital-to-analog conversion (DAC) process. The pre-DAC component, in which the native image will be manipulated, is digital. The post-DAC component is the analog video signal to the CRT display. The pre-DAC component permits the selection of a range (window and level) of tonal information to be presented, whereas the post-DAC component is for setting the display's dynamic response (brightness and contrast).

DICOM GRAYSCALE DISPLAY FUNCTION

The Digital Imaging and Communications in Medicine (DICOM) grayscale display function (GSDF) determines how the CRT dynamic range can be mapped to the HVS so that images appear similar, irrespective of display media. Although monochrome CRT monitors are suitable for HVS viewing of softcopy images, color CRT monitors have different attributes, complicating their ability to render diagnostic quality images. A color monitor uses three beams (red, green, and blue [RGB]), corresponding to the phosphor types. A shadow mask absorbs a substantial amount of the electrons and thus reduces overall luminance. The RGB pixels must be summated to produce a white intensity, and the shadow mask dictates the pixel density. A monochrome CRT usually has a single phosphor, and the optics dictate the pixel density. The trade-off for using a color CRT instead of a monochrome CRT is as follows. The shadow mask puts limitations on luminance, contrast, and resolution. The maximum luminance is usually in the range of 80–100 nits. The veiling glare is increased, which results in poorer small detail. A look-up table (LUT) is needed to optimize a grayscale rendition on color CRT monitors, and the maximum spatial resolution is nominally limited to 1280 x 1024 pixels to allow appropriate dot pitch and refresh rate and to minimize aliasing and other artifacts.

IMAGE QUALITY STANDARDS

Image quality standards should be considered in two components: optimization and consistency. Optimization is subjective (ie, "image quality is in the eye of the beholder"). The Medical Image Perception Society is working on perceptually based standards for image quality.

Consistency is an important concept and assures that images appear similar irrespective of display media. This is addressed by DICOM services, which include the GSDF, presentation LUT, and presentation state storage. In addition, the American Association of Physicists in Medicine (AAPM) has a task group (TG18) that is addressing the assessment for display performance of medical imaging systems. The AAPM TG18 scope is to cover monochrome CRTs and flat-panel displays. The goals are to provide standardization, performance criteria, education, communication, and professional development.

The DICOM GSDF (part 14 of the standard) facilitates having an image appear similar across dissimilar displays or on the same or similar displays at different times (a solution for the consistency issue). It relies on the concept of perceptual linearization. This is the uniform mapping of presentation (P) values into density and luminance, based on the Barton model of human contrast sensitivity, and of JNDs. Perceptual linearization allows for a mathematically defined relationship between the input value to a device and its output value. The process results in a uniform perceptual change for any given input. The grayscale softcopy presentation-state storage takes the presentation LUT and maps it into P space. Presentation-state storage supports the storage of parameters related to the way a specific image was displayed and the manipulations (if any) applied by the person interpreting that image. It is a mechanism to communicate intended presentation states for softcopy and to store/record a specific presentation state.

The GSDF calibration for monitors may be performed manually, automatically, or remotely. The first step is to measure the characteristic curve of the display system and the ambient light level. The second step is to compare this with a standardized display system that has the GSDF. The third step is to compensate the monitor and to measure the quality of the standardization in steps. Studies have shown that visual searching using a monitor with a linearized display (DICOM GSDF) is more efficient and has greater diagnostic accuracy than with a nonlinearized softcopy display. In addition, the linearized display offers decreased total viewing time, less fixation clusters, and decreased dwell time for all types of decision making (true negatives, false negatives, true positives, and false positives). These data support the use of DICOM GSDF in clinical practice and present an area in which vendors have yet to accommodate in PACS deployments.

TABLE 3.2
Softcopy Display Quality Measures

- Spatial resolution
- Contrast (photometric) resolution
- Luminance
- Modular transfer function
- Veiling glare
- Noise (spatial and temporal)
- Flicker
- Display size
- Display uniformity

Viewing Conditions

As stated previously, considerations for viewing conditions are paramount given the differences in luminance between softcopy displays and viewboxes. Illuminance is measured in lux. Direct sunlight produces ~10^5 lux, and an overcast day produces ~10^3 lux. Twilight produces ~10^1 lux, and moonlight produces ~10^{-1} lux. A typical office room is 75–100 lux, and a radiology reading room is on the order of 2–25 lux. This is important, because reading rooms will need to have less ambient light and more indirect lighting sources to produce suitable viewing environments for softcopy displays.

Several important aspects affect the ways in which ambient light can reflect on the softcopy display. Reflected light, which may be specular or diffuse, can show up in the face plate. Ambient light also decreases the perceived image contrast. Antireflective coating can be employed, but this decreases the luminance of the monitor and typically is worse for CRTs than for flat panels. The effect is most pronounced in environments such as the operating room and outpatient clinics, where ambient light is not at as low as that in a diagnostic radiology reading room or cannot be easily controlled. A flat-panel display may perform more favorably, given its ability to produce higher luminance than CRTs.

Another issue with luminance is whether absolute or relative values are more important. Contrast resolution may be the predominant factor mitigating the need for very bright monitors. Higher luminance monitors do have a physiological effect. Pupil diameter is decreased with higher luminance displays, which decreases overexcitation and improves recovery time

of retinal photoreceptors. This phenomenon, coupled with evidence that detection of low-to-moderate contrast target detection is improved with smaller pupil size, favors high-luminance displays. There is also evidence showing increased fatigue with lower luminance displays.

Observer Performance

Observers are a key aspect of the imaging quality "chain." Psychophysics is the discipline that encompasses physics and psychology. The methodology for assessing human performance is important because of both inter- and intraobserver variability.

Receiver operating characteristic (ROC) curves are accepted as the most complete way to quantify and report accuracy in two group classification tasks. Radiologists achieve different combinations of sensitivity and specificity by altering the "threshold" between normal and abnormal and by using more lax or strict criteria. ROC analysis is suited to isolating "decision threshold" effects from inherent differences in diagnostic accuracy. The area under the curve (AUC), also denoted by A_z, is a popular summary measure of diagnostic accuracy. The AUC corresponds to the probability of a pair of independent diseased and nondiseased measurement values being in the correct order (ie, the probability of disease is greater than the probability of no disease or vice versa, depending on the location on the curve). Alternatively, one may compare the partial AUC that exists between two fixed a priori values for specificities between two diagnostic modalities or permutations thereof.

Other summary accuracy measures include confidence intervals and bands, point of intersection, and optimal intercept. A horizontal or vertical confidence interval of the true-positive rate (TPR) for a specified false-positive rate (FPR) or vice versa may be constructed. The point of intersection is the ROC point at which sensitivity and specificity are equal and believed to reflect test accuracy. The optimal intercept is a line tangent to the ROC curve with the optimal slope. ROC studies are time consuming and costly, because they require a large number of human observations, especially if the number of possible conditions is large. In addition, several practical and important issues for ROC analyses are unresolved, but solutions are being pursued.

As in other imaging analyses, ROC studies may suffer from verification bias and the imperfect criterion ("gold standard") bias. Verification bias occurs when not all subjects who undergo an imaging procedure receive the criterion standard verification. Typically, patients with negative imaging are less likely to undergo a confirmatory procedure (eg, patients with negative knee MR imaging are unlikely to have arthroscopy). A so-called gold standard may be somewhat tarnished and not entirely accurate for comparison. Again, using the knee as an example, an arthroscopist may not visualize all aspects of both menisci, depending on the number and types of portals used. Also, areas such as the subarticular bone and extra-articular soft tissues are not well evaluated with arthroscopy but are exquisitely evaluated with MR imaging. Ideally, to provide an unbiased estimator for a test's accuracy, the disease status of each patient needs to be determined independently from the imaging result.

The generic task of lesion detection is complicated by several factors. One needs to locate an abnormality within a complex pattern of anatomical details, relying on a visual search and expertise. Considerations for evaluating imaging systems include the task, the observer, and the image properties. Imaging tasks may be categorized as detection, comparison, staging, classification, and estimation according to Kundel. In a detection task, the goal is to characterize an image as normal or abnormal. Comparison involves identifying a change in appearance or determining same–different status. Staging is determining the anatomical location and extent of a lesion. Classification is making an assignment to a predefined class of objects. Estimation encompasses quantifying the size, shape, or intensity of an object. The ideal observer is a model that describes the performance of an optimum decision maker with respect to a specific task. Model observers are computer algorithms that attempt to predict human visual performance in noisy images. One goal is to develop model observers that will become the desired metrics for optimizing image quality. The key components required for the use of model observers for medical image quality evaluation are: the visual task, signal and image backgrounds, figure of merit, the model observer itself, and comparison to the human observer. This complex interaction is made more difficult by the number and variety of model observers and an assortment of methodologies available to determine performance.

Search is a critical function performed by an observer. Search involves a target, prior knowledge about the target-to-background relationship, and selective

attention focused on the task. Decision-centered models for visual perception (eg, Gregory–Rock model) are based on observer hypothesis testing. The global-focal search model (Kundel and Nodine) follows from the decision-centered model. It specifies that the initial activity is a preliminary analysis of the total retinal image and that specific analyses of selected areas ensue.

Expertise is a desirable attribute for an observer. It may be considered the ability to acquire and use contextual knowledge. Medical imaging experts develop stored knowledge representations of anatomical maps (or templates) that are known as schema. Radiologists are characterized by using a searching strategy that is based on heuristics. This strategy is neither a random process (as a layperson might attempt) nor a rote, exhaustive viewing of all areas on an image (as a computer program would perform). Expert observers first look for perturbations in the schema (this is the global component of the global-focal search model). Visual scrutiny is then applied to these "focal" areas of perturbations. Object segmentation and feature extraction are performed, and then the accumulated attributes are compared against a "diagnostic list" until a decision is made as to whether the identified perturbation is considered a lesion. Expertise is dependent on the amount of mentored practice (ie, a residency training program) and the number of cases read (ie, increasing the repertoire of schema available).

Compression

A few words should be mentioned with regard to data compression and the potential impact on image quality. Data compression is possible because information in an image is not random. Compressed images are a more compact form of representation. The benefits of a smaller file size are obvious, in that it reduces archiving requirements and increases the data transfer rate. However, the influence of compression on image quality is uncertain and the subject of some controversy. Some evidence suggests that, at low levels of compression, the noise in the image is preferentially compressed and, thus, this serves as a filtering mechanism for noise reduction. However, the ultimate question is impact on diagnostic yield. Compression parameters include compression ratio, computation time, and image quality (degradation). The compression ratio is expressed as the ratio of the file size of the original to the compressed image (eg, 2:1 or 100:1). Lossless ratios are at about 2:1 or 3:1. Lossy techniques can achieve any compression ratio but offer less fidelity at higher ratios. Computation time is also involved in encoding and decoding the compression. Once images are transferred, they must be uncompressed for viewing, but these times usually are insignificant. The three key steps to image compression (transformation, quantization, and encoding) have been outlined in a white paper by Erickson.

Three principle methodologies guide evaluation of irreversible compression: numerical analysis, subjective evaluation, and diagnostic accuracy. Numerical analysis may be performed on the pixel values before and after compression. Subjective observer evaluation may be performed focusing on the aesthetic acceptability and estimated (not actual) diagnostic value. Objective measurement of diagnostic accuracy is performed by use of blinded or masked evaluations versus a criterion or "gold standard" reference. However, there are problems with each of these methods. The quantitative methods (mean square error, peak SNR) do not correlate well with observer opinion. Subjective techniques have two basic assumptions. First, if a radiologist rates one rendition superior to another, then it is easier (more pleasing) to look at. Second, this ease of viewing also makes it easier to provide a correct diagnosis. Unfortunately, a pleasing image may have more to do with familiarity (ie, training) than true image quality superiority, although observers tend to equate aesthetics with quality. Subjective evaluation by an expert observer is appealing. However, "visually lossless" is not equal to "diagnostically lossless," and subjective preference is not equal to diagnostic capability. Therefore, subjective methodologies have problems with validity. Although observer performance studies using ROC curves are more objective (more valid), they are also problematic, in that many differential diagnoses exist for a specific imaging study of a specific body part and all these diagnostic possibilities are rarely exhaustively tested. Therefore, observer performance studies have problems with generalizability. Inherent intra- and interobserver variability can be substantial. The perceptual and cognitive factors may outweigh differences in performance based on the "technical" factors that were originally the confirmatory hypothesis to be tested, thus reducing the power of these types of studies. Although no single metric exists for diagnostic yield, the combination of subjective and objective (or at least more objective) assessments used in numerous different studies and settings strengthens support for the use of compression. Current evidence indicates compression ratios on the order of 10:1 are suitable for most types of examinations.

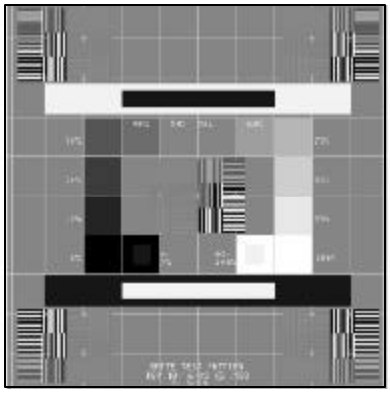

Figure 3.2 Society of Motion Picture and Television Engineers (SMPTE) test pattern.

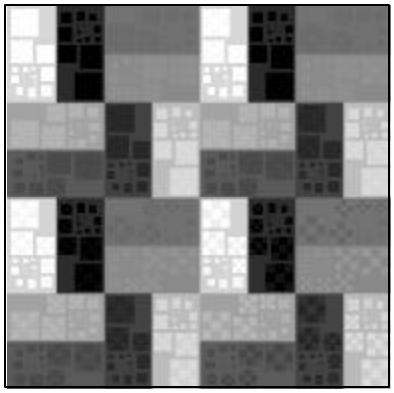

Figure 3.3 The Briggs test pattern.

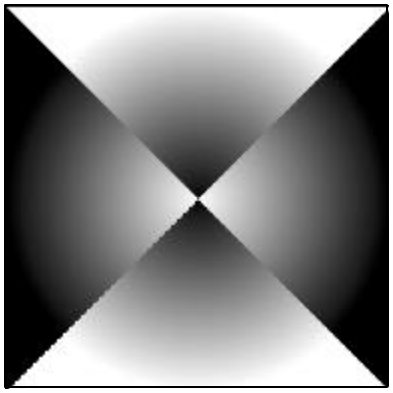

Figure 3.4 Continuous grayscale monitor test pattern. Reproduced, with permission, from Dr. Philip F. Judy from the Brigham and Women's Hospital Medical Image Perception Lab, Boston, MA.

Issues about development, implementation, and deployment affect the use of proprietary compression, with associated risks in both support and life span of equipment. This becomes especially problematic for long-term archives (several military installations have large archive platters that are no longer supported by the vendor and include proprietary compression). This may compromise interoperability. Interoper-ability risks are minimized by using DICOM. In a DICOM network, there is negotiation of a transfer syntax, and both receiving and sending devices will work if they support the same compression algorithm. Remember that uncompressed baseline is guaranteed in DICOM. However, for DICOM media there is an a priori decision required if compression is used with no negotiation mechanism. Therefore, compression in an archive can be used to reduce media costs and, by employing standard compression schemes, the migration risk is reduced.

It is commonly believed that not all the information in an imaging study or data set is equally utilized in all situations and by all types of observers. These are the context-specific issues for which the main question becomes: how much "quality" is enough? Or, how poor an image can you tolerate for your purposes? Some observers try to segment or stratify needed image quality versus intended use or user. In this paradigm, there are several "classes" of users. Some may separate intended use into the following categories: diagnosis, review, electronic medical record (archive), education (students, house officers, and patients), and procedure guidance ("road mapping"). The main argument against this stratification concept is that it is very difficult to know a priori the intended use for the image. In addition, studies on observer performance show that expert observers (eg, radiologists) fair better with relatively poor-quality images then do nonradiologists in clinical situations. There is no doubt that many nonradiologists use images in a fashion much like radiologists. That is, crucial diagnostic information is extracted and therapeutic decisions are made based on that interpretation, mandating that full fidelity images are available to these observers.

DISPLAY QUALITY MEASURES

Similar to acquisition, there are several proposed metrics for display image quality (Table 3.2). Spatial resolution was once considered the "Holy Grail" of medical imaging. However, each new modality developed since (within the notable exception of mammography, which is actually a form of radiography) has much less spatial resolution. "Advanced" imaging modalities, such as CT, MR, and ultrasound, exploit different tissue contrast resolution along with multiplanar cross-sectional capabilities to improve diagnostic accuracy. With respect to electronic displays, radiography has been the rate-limiting (or resolution limiting) modality for softcopy viewing. However, data are accumulating to support the idea that display of radiographic images is more dependent on contrast resolution than spatial resolution. This translates into the ability to use 1K (1024 x 1024) monitors without significantly impairing diagnostic performance, so long as the luminace is maintained (>75ft-1).

Quality Assurance and Quality Control

In terms of operationalizing image quality practices, it is prudent to review quality assurance (QA) and quality control (QC). These two terms are often interchanged but have come to have specific and distinct meanings. QA is typically used to refer to the broad spectrum of management practices for comprehensive, network-wide quality maintenance. The electronic environment requires new areas of skills. The goal is to detect slowly changing parameters before they become problematic or clinically significant. QC is typically regarded as a subset of QA, specifically dealing with tests of equipment function.

CRT monitors need two areas of QC: luminance and sharpness. Luminance calibration is performed using a photometer. The focus (sharpness) is assessed using test patterns. Several state-of-the-art monitors incorporate automatic luminance calibration. However, the need for assuring sharpness of the display remains. This is particularly true as the monitor ages and electron beam current is increased to compensate for the decreased light-emission efficiency. An adequate assessment usually requires multiple measurements at multiple locations, including the center and four corners.

Several test patterns are available, and four will be described briefly to provide the reader a sample of the various types. (1) The most widely used is the Society of Motion Picture and Television Engineers (SMPTE) test pattern (Fig. 3.2), which is specified for image size and bit depth. The background is specified at 50% of the maximum digital driving level. The display function includes an 11-step grayscale tablet ranging from 0% to 100%. Sharpness is assessed using sets of resolution bars. Geometry and image artifact are noted by looking at the grid lines. (2) The Briggs test pattern (Fig. 3.3) was developed by Stewart Briggs of Boeing Aerospace and can also be used for medical imaging. Each target consists of a series of 17 checkerboards arranged in order of decreasing size. The target is scored depending on the visibility of the checkerboards on each target according to a specified set of instructions. An average score or an adjusted score is then used. The Briggs method is considered to be somewhat subjective. (3) A continuous grayscale pattern is available from the Brigham and Women's Hospital Medical Image Perception Lab. When the continuous grayscale pattern is displayed (Fig. 3.4), it should have no concentric ring-like features. If these are present, then this is a global overall indicator that the monitor is not performing within specifications. (4) The Mayo Clinic has developed a four-alternative force-choice (4-AFC) contrast-detailed target. For this display, an observer is presented with three sets of eight test images. Each image consists of an 8 x 4 array of the limited areas, each containing a target in one of four locations (Fig. 3.5). The task is for the observer to record the location of each target. From this, a maximum threshold contrast is calculated at the 15% and 85% levels. The 4-AFC method is believed to be more objective than the other methods but is more time consuming (30–45 minutes). The AAPM TG18 has developed specific test patterns for medical imaging. This standard is fully described in chapter 7.

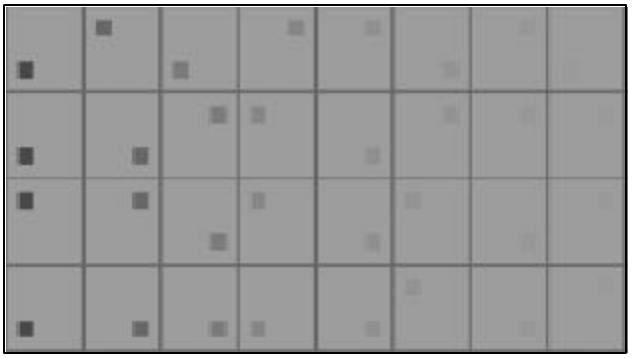

Figure 3.5 The Mayo Clinic four-alternative force-choice (4-AFC) contrast-detailed target. Reproduced, with permission, from Dr. Nicholas Hangiandreou.

CONCLUSION

Many factors influence image quality and the detection of abnormalities in medical images, including acquisition technique, image processing, image presentation, and viewing conditions. Furthermore, observer performance issues should not be ignored as a substantial cause for "quality" variation. In a PACS environment, acquisition and display are uncoupled. As a result, optimization of quality also may be uncoupled and addressed at these two levels. Although each component should be optimized separately, the ultimate "test" is the diagnostic performance of the human observer. Optimizing is best performed with respect to a specific imaging task. The ROC methodology for assessing human performance can be difficult to imple-

ment but will likely maintain a very important role for task-specific questions. Clinical trials are the standard for comparing one or two different therapeutic modalities but are impractical for optimizing a diagnostic modality, as in medical imaging. The results of early (preliminary) studies suggest that model observers can be used with synthetic and real medical images to predict human performance. Model observers are likely to play a greater role, given the limitations and practical difficulties in performing ROC curve studies for every possible diagnosis with every possible modality and for every level of observer.

SELECT REFERENCES

1. Van Metter RL, Beutel J, Kundel HL, eds. *Handbook of Medical Imaging. Volume 1. Physics and Psychophysics.* Bellingham, WA: SPIE Press; 2000.
2. Burgess A. Image quality, the ideal observer, and human performance of radiologic decision tasks. *Acad Radiol.* 1995;2:522–526.
3. Carrino JA, Unkel P, Miller ID, Bowser CL, Freckleton MW, Johnson TG. Large-scale picture archiving and communications system (PACS) implementation. *J Digit Imaging.* 1998;11(suppl 1):3–7.
4. Compton K. Factors affecting CRT display performance: specifying what works. *J Digit Imaging.* 2001;14:92–106.
5. *J Digit Imaging.* 1995;8:21–43. *J Digit Imaging.* 1995;8:21–43.Erickson B. *Irreversible Compression of Medical Images.* White paper. Great Falls, VA: Society for Computer Applications in Radiology; 2000.
6. Hemminger BM, Johnston RE, Rolland JP, Muller KE. Introduction to perceptual linearization of video display systems for medical image perception. *J Digit Imaging.* 1995;8:21–43.
7. Jaffe CC. Medical imaging, vision, and visual psychophysics. Kodak monograph series. *Med Radiogr Photogr* 1984;60.
8. Kundel HL. Medical image perception. *Acad Radiol.* 1995;2(suppl 2):S108–110.
9. Krupinski EA, Roehrig H. The influence of a perceptually linearized display on observer performance and visual search. *Acad Radiol.* 2000;7;8–13.
10. Krupinski EA, Kundel HL, Judy PF, Nodine CF. The Medical Image Perception Society. Key issues for image perception research. *Radiology.* 1998;209:611–612.
11. Lubin J. A human vision model for filmless radiology system design and optimization. *J Digit Imaging.* 1998;11:179.
12. Roehrig H, ed. Image quality assurance for CRT display systems. Part I. *J Digit Imaging.* 1999;12:1–3; Part II. *J Digit Imaging.* 1999;12:49; Part III. *J Digit Imaging.* 1999;12:153–165.
13. Rose A. *Vision: Human and Electronic.* New York, NY: Plenum Press; 1973.

WEB RESOURCES

Society for Computer Applications in Radiology
www.scarnet.org

Medical Image Perception Society
www.radiology.arizona.edu/krupinski/mips/index.html

Dave Clunie's website with many DICOM resources
www.dclunie.com

OTech: Resource for PACS implementation
www.otechimg.com

American Association of Physicists in Medicine TG18
www.deckard.mc.duke.edu/~samei/tg18

National Electrical Manufacturers Association DICOM
www.medical.nema.org/

Video Monitor Test Pattern Tutorials
www.brighamrad.harvard.edu/research/topics/vispercep/tutorial.html