

63

Display Optimization and Human Factors

Elizabeth A. Krupinski

CONTENTS
63.1 Introduction ........................................................................................................ 1275
63.2 What Is the Human Visual System Capable Of? .............................................. 1275
63.3 Visual Search ...................................................................................................... 1276
63.4 Visual Search and Reader Error ....................................................................... 1277
63.5 Display Optimization and Reader Error ............................................................ 1278
63.6 General Environment Considerations ............................................................... 1278
63.7 Impact on Radiologists: Physical Injuries, Fatigue, and Diagnostic Performance 1279
63.8 Conclusions ....................................................................................................... 1280
References ............................................................................................................... 1280

63.1 Introduction

There are many steps in the imaging chain before a clinical radiology case (i.e., images) makes it to the radiologist’s work list and computer workstation, including acquisition, transformation from a “raw” image to “for presentation” image, perhaps some compression, depending on the modality and application, or possibly some transformation from a series of individual images into a fully reconstructed 3D rendered volumetric image. The images can be archived and transmitted from one institution to another for interpretation or other purposes. No matter what steps precede, the final clinical step is the interpretation by the radiologist or other qualified clinician (e.g., an obstetrician interpreting a fetal ultrasound) of the image data. Currently, nearly all radiographic images are displayed on and interpreted from some sort of electronic digital display. Other chapters have already reviewed the more technical parameters of medical image displays, including a technology overview, quality control and the DICOM GSDF (Digital Imaging and Communications in Medicine Grayscale Standard Display Function) (see Section I, Chapter 64). Although the DICOM GSDF takes the perceptual capabilities of the human observer into account, a broader discussion of the role of the human observer in display optimization and how observers interact with displays is the focus of this chapter.

63.2 What Is the Human Visual System Capable Of?

The human eye–brain system is complex, but there are a few key aspects that are relevant to the optimization of displays for medical image interpretation. The eyes are the starting place, since radiology is essentially a visual task. The eyes are specialized sensory organs designed to receive and process light photons (see Oyster [1999] for a thorough explanation of the eye and its function). These photons pass through the cornea and pupil, then are focused by the lens of the retina. Each of these components impacts the perception of image data.

The pupil is important, as it controls the amount of light entering the eye, contracting in bright light and dilating in low light, in a process called adaptation. Adaptation (aka, dark adaptation) is important, as it impacts contrast sensitivity or the ability to detect differences in scene/image luminance (or color) corresponding to different objects in the scene. In radiology, dark adaptation is influenced by the overall ambient light settings, the display luminance, and the image on the display. There is significant evidence (Brennan et al. 2007) that ambient room light settings do impact lesion detection accuracy—they cannot be set too high or too low.

It is recommended (Norweck et al. 2013) that ambient room lights be set to 20–40 lux. If possible, lighting should be indirect (e.g., wall-mounted), rather than direct (e.g., ceiling), and adjustable. This will help with dark adaptation, reduce reflections on the display monitor and reduce fatigue. For primary interpretation, the maximum luminance should be at least 350 cd/m<sup>2</sup> (for mammography 420 cd/m<sup>2</sup>) and for secondary displays 250 cd/m<sup>2</sup>. As the term adaptation implies, it does take time for the eyes to adjust to a given lighting environment and, on average, it takes 10–15 minutes to adequately adapt, so radiologists should consider this before starting a reading session. Although there have been no studies to date using digital displays and clinical images to study the impact of dark adaptation on lesion detection performance, Baxter et al. (1982) did conduct a psychophysical study and computer simulation and the results support the “common knowledge” belief that viewing conditions and adaptation levels do impact performance, especially for low contrast targets.

The cornea helps protect the eye and also serves as the initial point in the focusing process. As light hits the cornea, it is
bent on its way to the lens, which is the focusing system of the eye. The task of the lens is to focus light on the retina where the photoreceptors (rods and cones) are located. To do this the curvature of the lens is changed via the ciliary muscles and parasympathetic pathways in a process called ocular accommodation. Accommodation is accompanied by vergence, a slightly faster process in which the eyes are oriented properly to a given point of focus in space. Both of these are impacted by the distance between the eye and the point in space—in this case the monitor displaying the radiographic image. For most people the ideal viewing distance is about 18 inches, but it will vary as a function of the individual, their vision and whether corrected or not. Corrective lenses are often required (especially with age), as the physical properties of the lens often deviate from the ideal. Myopia (near sightedness) occurs when the lens focuses the eyes too far in front of the retina; while far sightedness occurs when the lens focuses too far behind the retina. Since vision is so important to radiologists, it is highly recommended that they get annual eye exams and corrective lenses as necessary.

The retina, as noted, contains the rods and cones. There are light-sensitive cells that receive the light photons, undergo a chemical transformation, and transmit electrical energy that acts on nerve cells that transmit signals to the visual cortex and related brain systems. There are about 115 million rods, mostly in the periphery of the retina. The rods are responsible for lower resolution and motion perception and are mostly associated with peripheral vision. As motion detectors, their temporal response is faster than that of the cones, allowing for rapid detection and subsequent orientation of the eyes to changes in the periphery. As such, they are often thought of as “bug detectors.” In radiology, this system can come into play, for example when scrolling through a series of chest CT slices. Pulmonary nodules (Figure 63.1) tend to “pop” in and out of surrounding anatomy as they are scrolled through, catching the attention of the observer, causing them to slow down, focus on the “pop out” location and decide if what was detected is indeed a nodule. This “pop out” effect has garnered a significant amount of attention in the psychology literature since it was first described by Treisman (1985) as being a pre-attentive processing step in visual perception.

There are about 6.5 million cones, which are located in the central or foveal region of the retina. These photoreceptors are responsible for fine, high-resolution detail perception. Basically, this foveal region can be thought of as the direction of gaze or focus of attention. Foveal vision is limited to about a 2.5 degree radius (with parafoveal vision extending a bit further out, but not as far as peripheral vision). What this means practically is that, in order to bring image details into focus, the eyes need to move around the image in a process called visual search.

Visual acuity and contrast, or differences in color and brightness that allow you to distinguish between objects and their background, are intertwined. Contrast sensitivity has been found to peak in the mid-spatial frequency range, around 3–5 cycles/degree. For radiologists, this means that low contrast lesions (e.g., pulmonary nodules, breast masses, subtle fractures) can often go undetected, especially when viewing conditions are not optimal. Contrast sensitivity is measured using tests developed with grating patterns (alternating black and white lines, where the average luminance remains the same but contrast between the light and dark areas differ) to establish contrast levels perceptible by the human eye. This discrimination is described in terms of cycles per degree or grating frequency, as noted above.

The other key function of the cones is color perception. Hence, there are actually three types of cones, each sensitive to a particular range of wavelengths. Thus, there are S, M, and L cones, sensitive to short, medium, and long wavelengths, respectively (or blue, green, and red). In radiology, color vision is not especially important as images are predominantly grayscale (thus, being a color blind radiologist is not a problem!). In other clinical specialties, however, like pathology, dermatology, and ophthalmology, color vision can be quite important and display calibration for optimal (i.e., relative to the color capabilities of the human visual system) rendering and interpretation of color images is critical (Badano et al. 2015). To date, however, there is no single accepted color display calibration protocol like the DICOM GSDF in medical imaging and, surprisingly, the results on whether color calibration or quality impacts diagnostic performance is mixed (Krupinski et al. 2012; Campbell et al. 2015; Kimpe et al. 2016).

63.3 Visual Search

As noted above, due to the limited field of high-resolution foveal vision, the eyes need to move around the scene (image on a monitor) in order to bring details into focus, making it possible to detect subtle abnormalities embedded in the complicated and often camouflaging context of the anatomic background (often referred to as anatomic or structured noise). This limited field of high-resolution foveal vision is often referred to as the “useful visual field” and can be thought of as a “cone” of vision, as shown in Figure 63.2. Peripheral vision still processes information around the image outside of this cone, but not at the level of detail the useful visual field is capable of.
Interestingly, radiologists are capable of detecting quite a bit in very short views of radiographic images. One of the seminal studies in medical image perception (Kundel and Nodine 1975) had 10 radiologists view 20 chest radiographs, half normal and half abnormal, tachistoscopically for 0.2 seconds versus unlimited viewing time. Surprisingly, they found that in the 0.2 second flash view detection accuracy was 70% (97% in the unlimited search). This study pretty much introduced the concept of the “visual or global gist” into the medical image perception field.

Carmody et al. (1980) followed this up and illustrated just how much and how fast targets can be detected in chest radiographs. They had radiologists view a series of chest X-ray images with small pulmonary nodules using very fast or tachistoscopic presentations of the images (to prevent visual search). When viewed directly for 300 milliseconds, 85% of the nodule targets were detected. When the targets were off center by as little as 5 degrees from the axis of gaze, performance dropped by half, illustrating the critical role of foveal vision in lesion detection, but also showing how good peripheral vision can be at detecting small, subtle lesions—performance did not drop to zero. Similar studies have verified these findings and explored underlying mechanisms (Pietrzyk et al. 2014).

In order to study how radiologists move this useful visual field around an image in search of potential abnormalities, eye-tracking studies have been conducted. Figure 63.3 shows an example of a typical pattern of a radiologist searching for pulmonary nodules in a chest X-ray image. Each dot represents where the eye (useful visual field) has landed and is spending time processing visual input data (referred to as a fixation). The lines between the dots represent the paths between fixations as the eye moves around the image (known as saccades). Note that, because of the size of the useful visual field and since peripheral vision is also in play, the entire image need not be fixated with foveal vision. The nature of the search pattern also changes as a function of the type of image (Kundel et al. 1978; Krupinski 1996; Krupinski and Lund 1997; Drew et al. 2013; Rubin et al. 2015) and the nature of the task (e.g., general screening versus searching for a likely target based on the clinical history).

The idea to study how radiologists search images stems from seminal work in the 1940s. A series of studies was conducted to determine what the best technique was for tuberculosis screening (i.e., four radiographic and fluoroscopic techniques) (Birkelo et al. 1947; Garland 1949). It was thought that one imaging technique would readily stand out as the best, but there was so much intra-observer and inter-observer variation that no clear winner in terms of optimal technique could be found. Later studies verified this large amount of reader variation in studies that tasked radiologists with something as seemingly straightforward as describing the physical characteristics of radiographic shadows (Newell et al. 1954).

### 63.4 Visual Search and Reader Error

The studies noted above not only documented the first accounts of reader variability in radiographic image interpretation, but, as a result, also raised the problem of errors in radiology, which is still a major concern today (Brady et al. 2012; Pinto et al. 2012; Lee et al. 2013; Bruno et al. 2015), with estimates as high as 30% (false negatives and false positives) in some areas.

There are obviously many reasons why an abnormality might be missed by a radiologist, but, after the early studies after World War II first noting errors, Tuddenham and Calvert (1961) were one of the first to suggest that errors (misses) may due to inadequate searches and large inter-observer variability in search strategies. They did not actually record search patterns with any
eye-tracking technology, but rather asked radiologists to use a spotlight source with variable diameter to move around the images using a diameter no larger than they thought necessary, to indicate where they were looking. Thomas and Lansdown (1963) probably conducted the first actual eye-tracking study in 1963.

The majority of the early work in eye-tracking, however, was conducted by Kundel and Nodine and their various associates through the early 1990s. After that, eye-tracking equipment became more widely and commercially available (and affordable) and other labs started engaging in this area of research. One of the key contributions of Kundel et al. (1978) was using eye-position recording to classify types of omission errors.

Using data from radiologists looking for pulmonary nodules in chest images, Kundel et al. (1978) classified omission errors into three categories, based on visual dwell times. About one third of these errors fall into each category. Search errors occur when the radiologist never fixes the abnormality with foveal vision. It is presumed that he/she does not process abnormality information with high-resolution vision, so it is not detected or recognized. It should be noted that some abnormalities are obvious enough to be detected with peripheral vision and reported without ever directly fixating them.

The second category is known as recognition errors. In this case, abnormalities are fixated with foveal vision, but not for very long. It is thought that there is, therefore, inadequate processing time, reducing the likelihood that relevant features are detected or recognized. Decision errors occur when abnormalities are fixated for long periods of time, but the radiologist either does not consciously recognize it as a true abnormality or actively dismisses it.

There is significant literature on the nature of visual search and errors and how they differ as a function of level of expertise (Nodine and Mello-Thoms 2010; Drew et al. 2013; Kelly et al. 2016). There is, however, little evidence that display optimization or image quality significantly impact diagnostic performance differentially for experts versus novices. In brief, it is worth noting that, in radiology and other visual specialties like pathology (Mello-Thoms et al. 2012; Krupinski et al. 2013), as expertise develops, visual search patterns become more efficient—readers tend to find abnormalities faster, need to spend less time looking at them with foveal vision to decide if there are indeed lesions, and return to them less often in searches than those with less experience or those in training.

### 63.5 Display Optimization and Reader Error

As is clear from the chapter on the DICOM GSDF, display optimization, at least from the perspective of calibration and taking into account the perceptual capabilities of the human visual system, does impact diagnostic performance. The DICOM GSDF was developed utilizing Barten’s (1999) model of the contrast sensitivity of the human visual system and, in fact, use of the DICOM GSDF has been shown to yield better diagnostic accuracy and search efficiency compared to other calibration options such as the Society of Motion Picture and Television Engineers (SMPTE) pattern (Krupinski and Roehrig 2000).

There are a number of other physical display parameters that have been shown to impact diagnostic accuracy and/or visual search, stressing the need to properly optimize them for interpretation of clinical images. It is important to note that, even though these display characteristics have been assessed in the past, it is often necessary to redo them as display technology continues to evolve and change. Most of the early studies compared cathode ray tube (CRT) displays to film (Krupinski and Lund 1997; Hertzberg et al. 1999; Krupinski et al. 1999), then CRTs to LCDs (liquid crystal displays) (Krupinski et al. 2004; Balassy et al. 2005; Saunders et al. 2006), LCDs to iPads (McEntee et al. 2012; Yoshimura et al. 2013; Salati et al. 2015), phones and other portable devices (Szekely et al. 2013; Hirschorn et al. 2014; Kim et al. 2015; Park et al. 2016), monochrome versus color (Krupinski et al. 2008; Okumura et al. 2014) and medical-grade versus commercial-off-the-shelf (COTS) displays (Krupinski 2009; Okumura et al. 2014). Surprisingly, although evaluations of newer technologies often revealed lower diagnostic accuracy and efficiency compared to the status quo, after the technology matured most studies revealed relative equivalence. Thus, today there are many situations in which portable devices and COTS displays are used on a regular basis for primary interpretation.

In particular, key display characteristics that have been shown to lead to better diagnostic accuracy and efficiency include higher luminance (Krupinski et al. 1999; Burs et al. 2007), ambient lights between 25 and 50 lux (Brennan et al. 2007; Hellen-Halme et al. 2008; Liu et al. 2014), orthogonal viewing (Badano et al. 2003; Krupinski et al. 2005) and calibrating to the DICOM GSDF (Krupinski and Roehrig 2000; Leong et al. 2012; McIlgorm et al. 2013). Characteristics that do not seem to significantly impact accuracy and efficiency include 8-bit versus 11-bit data displays (Krupinski et al. 2007), and color versus monochrome high-performance LCD displays (Averbukh et al. 2005; Doyle et al. 2005; Langer et al. 2006; Geijer et al. 2007; Okamura et al. 2014).

### 63.6 General Environment Considerations

The display is clearly only part of the reading environment and its optimization only one piece of the greater picture (Hori et al. 2003; Prabhu et al. 2005; Krupinski and Kallergi 2007). As noted, external factors such as ambient room lights also play a large role in the optimization of the more global reading environment. There are many other factors to consider. For example, the user interface is a critical factor that every radiologist must deal with, and it became quite evident in the early days of picture archiving and communications system (PACS) that there was no one-size-fits-all interface that would suit every radiologist and every task or modality. Clearly radiologists have hanging protocol preferences, but it goes beyond that. Allowing for individually tailored or customized interfaces can actually increase reading efficiency or, at least, the perception of increased efficiency (Jorritsma et al. 2014, 2015). Some of the key aspects of the user interface that have been identified as important to consider when selecting the optimal workstation include hanging protocol and default display options, image processing and analysis tools or functions, reporting options, speed, means of interaction (e.g., mouse, joystick, touch), decision support tools and compatibility with other systems (e.g., is a separate workstation required for 3D rendering options).
Other environmental factors to consider include ambient noise levels (e.g., from computers, colleagues, hallways), airflow and temperature and even cleanliness. A recent study (Duszak et al. 2014) assessed the degree of bacterial contamination of radiology workstations (with toilet seat and doorknob as reference points). They tested dictation microphones (mean = 69.4) and computer mice (mean = 46.1) and found that all were contaminated with bacteria—at significantly higher colony count rates than toilet seats (mean = 10.5) or doorknobs (mean = 14.8)! The majority of the workstations (64%) grew Staphylococcus aureus and 21% grew enteric organisms. Simple swabbing with a commercial antiseptic pad practically eliminated the contamination.

### 63.7 Impact on Radiologists: Physical Injuries, Fatigue, and Diagnostic Performance

Another important consideration that has gained much attention in recent years relates in part to the choice of input device (e.g., mouse, joystick), but also to the more general phenomenon of radiologists sitting in front of computer workstation displays for hours on end interpreting cases. In the film environment, the problem was either minimal or non-existent, as there were no input devices and reading sessions did not last for hours upon hours, as the films on the view box or alternator had to be changed periodically. In a sense, this was forcing the radiologist to get up and do something else for at least a short period of time. The digital reading environment, however, has raised significant concerns regarding physical injuries to the radiologist. For example, Rodrigues et al. (2014) surveyed 148 radiologists (trainees and attendings) on the prevalence of musculoskeletal (MSK) symptoms related to radiology work. They found that 38% reported radiology-associated MSK symptoms, with lower back pain being the most common (41%), followed by shoulder, neck, wrist, eye, hand, and elbow discomfort. Thompson et al. (2014) found that repetitive strain injuries are also common among breast radiologists and they increase as a function of longer work hours.

In addition to documenting these issues, there are also a number of solutions provided to help combat them (Hoffmann et al. 2016), including getting seating with adequate lumbar support, wrist supports, setting aside time for movement and stretching exercises and even standing while interpreting cases (Krupinski 2013; Richardson 2014). American Optometric Association. (https://www.aoa.org/documents/infographics/SaveYourVisionMonth2016-1.pdf) is a validated tool to help evaluate where existing “pain points” might exist, thereby providing guidance on what aspects might need to be addressed to optimize the reading environment (Hedge et al. 1999).

Another way that the digital reading environment impacts radiologists is in terms of visual strain (Krupinski and Berbaum 2009; Halpenny et al. 2012; Gerard et al. 2013; Ikushima et al. 2013). Ikushima et al. (2013) found that the type of monitor, ambient lighting conditions and degree and type (contacts versus glasses) were all associated with subjective visual fatigue levels. Krupinski et al. (2010) measured visual accommodation (ability to focus on a point in space or a scene such as a radiograph on a computer monitor) in radiologists and radiology residents (Figure 63.4) and found that accommodation at near distances was significantly worse overall compared to far distances, and was significantly worse after a day of digital reading at all distances. In other words—after a long day of reading clinical images, visual strain was increased to the point where they were actually becoming temporarily myopic and less able to focus (Krupinski et al. 2010).

The critical question, of course, is whether visual strain and subjective feelings of fatigue have any impact on the ultimate issue of importance—diagnostic accuracy (Krupinski and Reiner 2012; Reiner and Krupinski 2012a,b). A number of studies have been conducted in recent years to investigate this issue, with the majority finding that diagnostic performance and efficiency are indeed impacted negatively. It has been shown that visual accommodation, subjective fatigue and detection accuracy for fractures are all negatively impacted after just 8 hours of clinical reading, with diagnostic accuracy decreasing significantly by about 4% (Krupinski et al. 2010). The detection of pulmonary nodules in CT is similarly impacted (Krupinski et al. 2012). In an analysis of six studies (Taylor-Phillips et al. 2015) that were not initially designed to assess the impact fatigue, it was found that there are systematic changes in diagnostic accuracy over just the course of a single laboratory reading session! Across the studies, time taken to interpret each case decreased by 9%–23% as time progressed and sensitivity decreased or specificity increased significantly as well.

Some possible solutions to avoid visual fatigue (and other errors) have also been proposed (Lee et al. 2013), such as automated accommodative training to strengthen the eyes and reduce the tendency towards fatigue-induced myopia. The “20-20-20 rule” is an easy tip to remember to help reduce eye strain and fatigue—every 20 minutes, look towards an object 20 feet away for at least 20 seconds (American Academy of Ophthalmology 2016). Other suggestions include avoiding sitting too close to air heating and cooling systems, using lubricating eye drops frequently, not sitting too close to the monitor, reducing glare as much as possible, humidifying the air if necessary and avoiding dust, mold, and pollen as much as possible. These types of considerations regarding environmental factors become even more important when using mobile devices outside of the reading room environment and especially in outdoors situations.
63.8 Conclusions

The nature of medical image viewing has changed dramatically in the past 30 years in many ways, including the way images are acquired and, hence, displayed (the move from film to digital), the display medium (CRTs, LCDs, iPads, smartphones) and the sheer volume and types of images that require interpretation in a typical day. Changes will continue to occur as new types of image acquisition devices are invented and new types of viewing devices evolve. Stereoscopic displays have never really taken off in radiology, in part due to technology limitations and the need to wear special glasses, but there are numerous advances being made in 3D volumetric imaging and innovative ways to display these images for diagnostic use (Chen et al. 2016; Perhac et al. 2016). Even the basic 2D viewing devices continue to evolve—the smartphone of today will soon be replaced by yet another portable device with capabilities yet to be considered and, just as with the devices that are available today, we must continue to evaluate them in terms of their impact of diagnostic accuracy and efficiency, and optimize them for clinical interpretation by the human observer.

REFERENCES


