

The Importance of Perception Research in Medical Imaging

Elizabeth A. Krupinski

The goal of this paper is to provide the reader with an introduction to the importance of perception research in medical imaging. It is well known that radiologists' performance is not perfect: they make both false positive and false negative decisions, both of which can impact on patient care and treatment. Some of these errors can be attributed to technical difficulties such as underexposing a plain film X-ray image. Such technical explanations cannot, however, account for all the errors that are made; missed lesions are often found in retrospect. These errors can be attributed to perceptual and/or cognitive factors. The study of why these perceptual and cognitive errors occur and what steps can be taken to ameliorate them is a relatively small but growing area in the field of medical imaging. Understanding the capabilities of the human visual system with respect to medical imaging is becoming even more important as we make the transition from the traditional film-based display to soft-copy monitor viewing of medical images. If we understand what the human visual system is capable of, we can tailor the display of medical information to take advantage of these perceptual capacities.

Key words: medical image perception, observer performance, errors, expertise, eye-tracking, soft-copy displays

INTRODUCTION

MEDICAL IMAGE PERCEPTION RESEARCH began to burgeon soon after World War II when a study¹ designed to determine what viewing method was best for detecting tuberculosis in chest images found a high degree (sometimes >30%) of inter-observer and intra-observer variation. A series of subsequent studies²⁻⁴ confirmed that such variation was not uncommon. Until these early studies were conducted, it was generally thought that all radiologists tended to agree with each other and that an individual radiologist would be consistent in his/her diagnoses over time. Neither was true. Radiologists varied in their diagnostic decisions, and they often varied a lot. Early research was aimed at quantifying these differences and discovering why radiologists differ in their diagnostic decisions. From this line of study grew interest in discovering what types of diagnostic errors radiologists make, how often they occur, what causes them, and what we can do to decrease

error rates and variation among radiologists in their interpretations. Even today, there are studies^{5,6} being done to investigate rates of disagreement among radiologists and why disagreement occurs.

In 1997, a new society, the Medical Image Perception Society (MIPS), was formed to promote research and education in medical image perception and provide a forum for the discussion of perceptual, psychophysical, and cognitive issues by radiologists and basic scientists. In 1998, a similar society, the Nippon Medical Image Perception Society (NMIPS), was formed in Japan. NMIPS held its Second Annual Meeting in Matsuyama, Japan, on January 28, 2000. Both societies share a number of basic goals and areas of interest. Recently, the Medical Image Perception Society (MIPS) outlined some of these basic goals for perception research.^{7,8} Broadly, they include (1) mathematical modeling of the detection of discrete abnormalities in noise-limited images, (2) understanding how observers find discrete abnormalities in images when their locations are unknown, (3) understanding how knowledge and experience affect the recognition and detection of abnormalities, (4) developing perceptually based standards for image quality, (5) developing computer-aided perception tools, and (6) developing quantitative methods for describing natural images and for measuring

Received March 14, revision accepted June 12, 2000.

Department of Radiology, University of Arizona, USA

Reprint requests to Elizabeth A. Krupinski, Ph.D., Department of Radiology, University of Arizona, 1609 N. Warren Bldg. 211, Tucson, AZ 85724, USA.



Fig. 1. Example of a typical eye-position pattern generated by a radiologist searching a bone image for fractures. The small circles represent fixations or locations where the eye lands. The lines represent the sequence in which the fixations were generated. The large circle on the wrist indicates the location of a fracture.

human detection and recognition performance. The goal of the present paper is to describe some of the image perception research being conducted and to demonstrate how perception research in general benefits clinical radiology. It is hoped that it will help stimulate more research in this exciting and important area.

ERRORS IN INTERPRETATION

General estimates suggest that, overall, there is about a 20-30% miss rate (false negatives) in radiology with a 2-15% false positive rate.^{6,9,10} The question arises as to why these errors occur. Generally, the false positives are easier to understand than the false negatives. Overlaying anatomic structures often mimic disease entities. A vessel on end or an unclear nipple marking might easily be mistaken for a nodule. False negatives, however, are harder to understand, especially when the missed lesions can be seen in retrospect. Using eye-position recording techniques, it has been found that false negatives can be classified into three categories based on how long they are fixated or dwelled on.¹¹ Some lesions are missed because they are never looked at with high-resolution foveal vision (search errors). Some are looked at, but not long enough to detect or recognize any suspicious lesion features (recognition errors). Finally, there are those that are looked at for long periods of time (often as long as lesions that are looked at and reported), but are still missed (decision errors). This hierarchy of types of errors based on visual dwell patterns

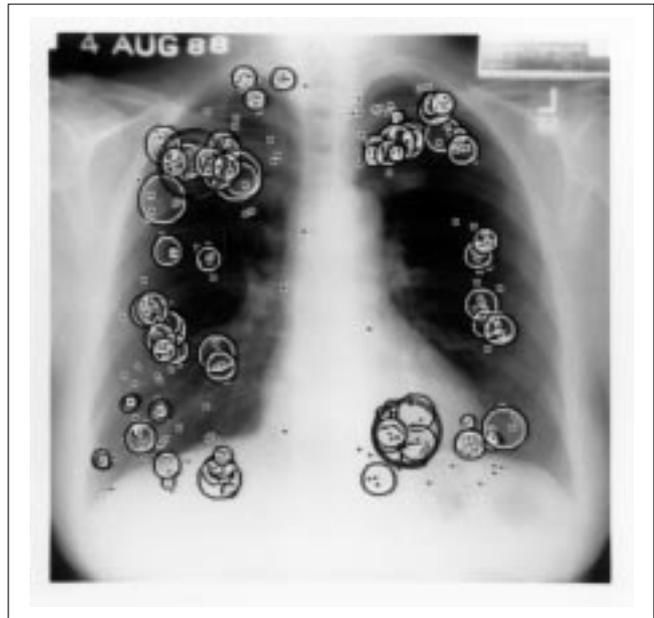


Fig. 2. Example of a typical eye-position pattern generated by a radiologist searching a chest image for pulmonary nodules. The small dots represent fixations or locations where the eye lands. The larger circles represent clusters of fixations. Clusters that have a cumulative dwell of ≥ 1000 msec can be circled and fed back to the radiologist for a second look. Examples of feedback circles are the dark circle in the upper right quadrant of the lung and the dark circle behind the heart. Feedback of this sort resulted in an increase in ROC Az performance of 16%.

has been observed in chest images,^{11,12} bone images,¹³ and mammograms.¹⁴ Figure 1 shows a typical example of an eye-position pattern generated by a radiologist searching a bone image for fractures. It seems that for recognition and decision errors the visual system has perceived some perturbation in the image, prolonging visual dwell, but the features detected do not cohere into a recognizable object and so are not cognitively perceived as a potential lesion.

Given the finding that a fairly large number (about two-thirds) of missed lesions (false-negatives) receive prolonged perceptual attention and processing, the question naturally arose as to whether this phenomenon could be exploited somehow to improve lesion detection and recognition. One idea was to improve these error rates by using perceptually-based feedback.¹⁵ Basically, one could record the eye-position of the radiologist as he/she searches the image for lesions, and then use this data to find out where the radiologist spent a good deal of time (>1000 msec) examining the image. Since missed lesions tend to be associated with these long dwells, one could then circle these long dwell areas (Fig. 2) and show them to the radiologist for a second look. Use of such a visual feedback system resulted in a 16% increase in

observer performance for radiologists looking for pulmonary nodules, compared to just showing the image again without any dwell locations indicated! The true-positive rate increased and the false-positive rate decreased, indicating a true improvement in performance rather than just a shift along the ROC (Receiver Operating Characteristic) curve.

Further studies^{16,17} were designed to determine exactly why putting a physical circle around the long-dwell areas improved performance. It was found that the presence of the circle actually focuses attention on the lesion area, increasing the frequency and accuracy with which the lesion is actually fixated directly with foveal vision. These results were quite exciting, but it was difficult to translate the procedure into clinical reading of radiographs. Recording eye-position entails wearing the recording apparatus on the radiologist's head and having a visor or monocle in front of the eye(s). After about one hour in the system, the radiologist tends to get fatigued, making use of the system in the clinic rather improbable. However, the finding that pointing out specific image locations improves performance has not been abandoned. Many of the computer-aided detection (CAD) schemes currently being developed also rely on pointing out suspicious image locations to the radiologist for further evaluation.¹⁸⁻²¹ It is quite possible that CAD improves lesion detection performance for the same reasons that perceptual feedback did. CAD may help focus perceptual and attentional resources better than the unaided radiologist can do on his/her own. CAD also helps the radiologist detect lesions in locations they never even looked at (search errors), but it has not been perfected yet. There are still lesions that the radiologist can detect that the computer cannot.²² Perceptual feedback may still hold promise in the future, perhaps using remote recording devices instead of head-mounted systems in conjunction with CAD systems.

EXPERTISE

What makes a "good" or "expert" radiologist? Are there quantifiable differences that we can observe that distinguish the novice from the expert? The most obvious answer to these questions is that the "expert" radiologist generally has better diagnostic accuracy than the novice and that is what makes them an expert. The less obvious answer is actually another question: Why is the expert better than the novice? Or in a more practical vein, can we predict who will make a "good or expert" radiologist? Perhaps there is something special in the way a radiologist approaches and searches an image that sets them apart from other clinicians and predisposes them to a successful career in radiology. On the other hand,

maybe it's all in the training and everything a radiologist needs to know can be learned during residency. More likely it is a combination of these two plus some other factors.

How can we study what makes the "expert" radiologist different from the average radiologist or the novice resident? Again, eye-position recording can help. Some studies have been done that demonstrate specific and predictable differences between novices and experts in terms of their perceptual search behaviors.²³⁻²⁵ We have found that experts tend to find lesions earlier in search than novices, tend to have different fixation and dwell patterns, and tend to have much more efficient search strategies than novices.²⁶ In general, they seem to be able to perceive and process more information, more quickly than novices or those with less expertise, and this generally translates directly into better diagnostic performance. These findings are, however, not all that unexpected. The expert radiologist has been exposed to more images and has had much more practice searching these images than the novice. One would expect their search to be more efficient and accurate.

What is harder to predict, is whether the radiologist is somehow better at searching images in general than other clinicians or lay persons. Are they somehow predisposed to be better searchers and that is why they go into radiology and become expert at searching for lesions in complex radiographic images? There have been a number of studies^{27,28} conducted to determine if there are specific perceptual tasks that radiologists are better at than lay persons or other clinicians. If they are better, then maybe specific tests could be designed to pre-screen people before they enter into the radiology residency, only taking those who are likely to be expert searchers. The results of such studies have been mixed, and little has been found that sets the radiologist apart from others in terms of perceptual abilities.

As an example of such a study, Nodine and Krupinski²⁹ conducted two separate studies comparing radiologists and lay persons searching complicated picture scenes for hidden targets (e.g., Waldo in the "Where's Waldo" children's book). Figure 3 shows a small portion of a typical "Where's Waldo" scene with Waldo partially obscured by another figure, much like a lung nodule can be obscured by an overlying rib.

The tasks were very similar to reading an X-ray image and searching for lesions because the targets of search are embedded in a complicated background that also must be searched and interpreted in order to understand the full import of the scene – much like a tumor in a chest radiograph. Both of these studies recorded eye position of the radiologist and lay searchers. Detection performance for these non-medical targets in artistic



Fig. 3. Typical example of a “Where’s Waldo” picture-book scene. This is only a small portion of a much larger scene presented to the observers in the study. The actual pictures were presented in full color. Waldo (circled in this illustration only) is partially obscured by another figure in this example, much like a lung nodule can be obscured by an overlying rib. Reprinted with permission (©Handford M. *Where’s Waldo? The Magnificent Poster Book*. New York, NY: Little Brown, 1991).

drawings did not differ between radiologists and lay persons. There were however, some differences in the visual search patterns that did differentiate the two groups of observers. On average, the radiologists spent more time overall searching the images for the targets. They also tended to fixate the targets earlier in search than did the lay persons, even though they did not always report the target as having been found on this first fixation. These results suggest that even though the radiologists may have been more careful or thorough searchers (i.e., spending more time than lay persons), they really were no better in terms of absolute detection performance than the lay persons. It can be concluded that the visual search and analysis skills that the radiologist exhibits when searching medical images do not transfer to a more general search task of searching art scenes for embedded figures. In other words, radiologists do not seem to possess special search skills that predispose them to being successful at searching radiographs. Much of what they learn is through repeated exposure to and practice reading radiographic images.

FILMLESS RADIOLOGY

Radiologists have been looking at film images ever since Roentgen first discovered X-rays and obtained an image of his wife’s hand. However, since the 1980’s we have been making the leap into the digital world and

increasingly radiologists are viewing images on computer monitors. What soon became evident as this transition from film to monitors began to take place, was that the image on the monitor was not the same as the image on film. Many perceptual and ergonomic issues started to arise when film use started to decrease. Compared with the traditional method of viewing film on a light box, monitors are less bright, have less spatial resolution, have less contrast (dynamic range), and have a limited viewing area. These are all problems that need to be addressed, and those in medical image perception research have begun to look at these issues. In the long run, we must ensure that switching to a different viewing medium will not negatively affect diagnostic accuracy. It also must not significantly affect workflow. If it takes too much time to adapt to and use a new type of workstation or viewing system, radiologists are not likely to make the transition easily or quickly.

There are a number of things that we have discovered in terms of optimal monitor luminance, tone scale, and interface design that draw on perceptual factors and impact the clinical reading environment. In terms of diagnostic accuracy, it has been shown that in most cases performance is about the same with film and monitor viewing. However, there are some characteristics of the monitor that can affect diagnostic performance, making it imperative to examine some of these features when considering which monitor to buy for teleradiology and/or PACS (Picture Archiving and Communication Systems) applications in the clinical setting. For example, performance was found to be better with a high luminance (140 ftL) monitor than a relatively low luminance (80 ftL) monitor.³⁰ It has also been shown that performance is better with a perceptually linearized display curve (e.g., the Barten curve) than with a non-linearized curve (e.g., using the SMPTE pattern to calibrate the monitor).³¹ The Barten curve was developed to try to match monitor output (in terms of gray levels) to the perceptual capabilities of the human visual system, and in fact performance is indeed better when the display system is perceptually linearized.

Eye-position recording has also proven to be a useful tool in understanding how radiologists view images on film versus a monitor. Looking at the same display features of luminance and tone scale, perceptual results were found that fell nicely in line with the performance data. For the high luminance and perceptually linearized monitors, total search time was shorter, time to first fixate the lesion was shorter, and dwell times for all decisions (true and false, positive and negative) were shorter than for the lower luminance and non-perceptually linearized monitors. Recording radiologists’ eye-position as they navigate between the radiographic images displayed on

the monitor and the menu on the monitor (i.e., image processing icons, image management tools) has also been quite informative. For example,³² evaluation of one workstation found that radiologists spent 20% of their search time looking at the menu rather than the X-ray image that they were supposed to be diagnosing! This has significant implications for workstation design and how much extraneous information should be displayed on the monitor in addition to the image being diagnosed. Increasingly it is being found that the best interfaces are the simple, uncluttered ones that need very little training to be able to use.

The advent of digital display of radiographs also makes possible true clinical use of computer-aided detection (CAD) schemes. The goal of CAD is similar to the goal of perceptual feedback discussed earlier: to give the radiologist another look at an image with potential lesion locations pointed out. Instead of using eye-position information, however, CAD uses a variety of image processing algorithms to detect (and sometimes classify) probable lesion sites. How the radiologist will use CAD information in the clinic and how it will affect their diagnostic performance is becoming a topic of quite a bit of interest. It has been shown that although CAD is becoming very good, there are still lesions that the radiologist is able to find that the computer misses.²² If this continues to be the case, radiologists must learn to use CAD as a supplement to their perceptual search strategies; they cannot eliminate their own search of the image and rely on CAD to detect 100% of the lesions. The radiologist also needs to continue to search the whole image to better be able to decide if a suspicious region pointed out by CAD is a true lesion or a false positive.

A recent study again used eye-position recording to study the perceptual strategies of experienced mammographers versus residents reading mammograms with and without CAD information.³³ Significant differences based on level of expertise were found again. The experienced mammographers spent more time doing a more thorough search of the images before they accessed the CAD pointers than did the residents. During the search without CAD, the mammographers also fixated more of the lesions than the residents, so when they did access the CAD pointers they used them more to confirm their suspicions about potential lesions. The residents seemed to use the CAD to guide them to an initial inspection of the potential lesions. With this sort of strategy, the less experienced readers are less likely to discover lesions that the CAD system does not detect. Results like this could have significant implications for how CAD is to be used in the clinical environment. It may be that residents or radiologists not expert in mammography need to be instructed explicitly to

conduct their own thorough search of the image before seeing what CAD has to offer.

CONCLUSIONS

As we move into the new millennium and as radiology completes the transition to a filmless practice, perception issues will not decrease. In fact, the importance of medical image perception research may increase. Technology will allow us to present better and different images to the radiologist. How will they deal perceptually with such things as compressed images, the addition of color, and computer-aided prompts superimposed over the images? All of these new ways of presenting images and image information are currently being researched, and it will not be long before the radiologist is asked to look at these new types of images. Image perception research will help us guide the basic research, will help us to understand how to present this new information to the radiologist, and will teach us how to help improve their diagnostic performance by improving their perception of the images. In conclusion, some insight into how image perception research is conducted has hopefully been gained as well as some insight into how to measure and analyze observer performance and perception. In addition it should also be clear why medical image perception research is important to the practicing radiologist.

ACKNOWLEDGMENTS

This paper is based on a talk presented at the Second Annual Nippon Medical Image Perception Society (NMIPS) held on January 28, 2000, in Matsuyama, Japan. I would like to thank Junpei Ikezoe, M.D. (Professor and Chairman, Department of Radiology, Ehime University School of Medicine) and Takeo Ishigaki, M.D., Ph.D. (Professor and Chairman, Department of Radiology, Nagoya University School of Medicine) for inviting me to speak at the conference and write this paper based on that talk. Some of the work presented in the paper (monitor evaluation) was supported in part by Toshiba Corporation, Tokyo, Japan.

REFERENCES

- 1) Birkelo CC, Chamberlain WE, Phelps PS, Schools PE, Zacks D, Yerushalmy J. Tuberculosis case finding: comparison of effectiveness of various roentgenographic and photofluorographic methods. *JAMA*, 133: 359–366, 1947.
- 2) Garland LH. On the scientific evaluation of diagnostic procedures. *Radiology*, 52: 309–328, 1949.
- 3) Garland LH. On reliability of roentgen survey procedures. *AJR*, 64: 32–41, 1950.

- 4) Cochrane AL, Garland LH. Observer error in interpretation of chest films: international investigation. *Lancet*, 2: 505–509, 1952.
- 5) Siegle RL, Baram EM, Reuter SR, Clarke EA, Lancaster JL, McMahan CA. Rates of disagreement in imaging interpretation in a group of community hospitals. *Acad Radiol*, 5: 148–154, 1998.
- 6) Robinson PJA. Radiology's Achilles' heel: error and variation in the interpretation of the Roentgen image. *Br J Radiol*, 70: 1085–1098, 1997.
- 7) Krupinski EA, Kundel HL, Judy PF, Nodine CF. Key issues for image perception research. *Radiology*, 209: 611–612, 1998.
- 8) Krupinski EA, Kundel HL. Update on long-term goals for medical image perception research. *Acad Radiol*, 5: 629–633, 1998.
- 9) Muhm JR, Miller WE, Fontana RS, Sanderson DR, Uhlenhopp MA. Lung cancer detection during a screening program using four-month chest radiographs. *Radiology*, 148: 609–615, 1983.
- 10) Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. *Radiology*, 184: 613–617, 1992.
- 11) Kundel HL, Nodine CF, Carmody DP. Visual scanning, pattern recognition and decision-making in pulmonary tumor detection. *Invest Radiol*, 13: 175–181, 1978.
- 12) Kundel HL, Nodine CF, Krupinski EA. Searching for lung nodules: visual dwell indicates locations of false-positive and false-negative decisions. *Invest Radiol*, 24: 472–478, 1989.
- 13) Hu CH, Kundel HL, Nodine CF, Krupinski EA, Toto LC. Searching for bone fractures: a comparison with pulmonary nodule search. *Acad Radiol*, 1: 25–32, 1994.
- 14) Krupinski EA. Visual scanning patterns of radiologists searching mammograms. *Acad Radiol*, 3: 137–144, 1996.
- 15) Kundel HL, Nodine CF, Krupinski EA. Computer-displayed eye position as a visual aid to pulmonary nodule interpretation. *Invest Radiol*, 25: 890–896, 1990.
- 16) Krupinski EA, Nodine CF, Kundel HL. Perceptual enhancement of tumor targets in chest x-ray images. *Perception & Performance*, 53: 519–526, 1993.
- 17) Krupinski EA, Nodine CF, Kundel HL. A perceptually based method for enhancing pulmonary nodule recognition. *Invest Radiol*, 28: 289–294, 1993.
- 18) Qian W, Li L, Clarke LP. Image feature extraction for mass detection in digital mammography: influence of wavelet analysis. *Med Phys*, 26: 402–408, 1999.
- 19) Kallergi M, Carney GM, Gaviria J. Evaluating the performance of detection algorithms in digital mammography. *Med Phys*, 26: 267–275, 1999.
- 20) Armato SG, Giger ML, MacMahon H. Computerized analysis of abnormal asymmetry in digital chest radiographs: evaluation of potential utility. *J Dig Imaging*, 12: 34–42, 1999.
- 21) Ashizawa K, Ishida T, MacMahon H, Vyborny CJ, Katsuragawa S, Doi K. Artificial neural networks in chest radiography: application to the differential diagnosis of interstitial lung disease. *Acad Radiol*, 6: 2–9, 1999.
- 22) Krupinski EA, Nishikawa RM. Comparison of eye position versus computer identified microcalcification clusters on mammograms. *Med Phys*, 24: 17–23, 1997.
- 23) Kundel HL, LaFollette P. Visual search patterns and experience with radiological images. *Radiology*, 103: 523–528, 1972.
- 24) Nodine CF, Kundel HL, Lauver SC, Toto LC. Nature of expertise in searching mammograms for breast masses. *Acad Radiol*, 3: 1000–1006, 1996.
- 25) Nodine CF, Kundel HL, Mello-Thoms C, et al. How experience and training influence mammography expertise. *Acad Radiol*, 6: 575–585, 1999.
- 26) Krupinski EA. Visual scanning patterns of radiologists searching mammograms. *Acad Radiol*, 3: 137–144, 1996.
- 27) Bass JC, Chiles C. Visual skill: correlation with detection of solitary pulmonary nodules. *Invest Radiol*, 25: 994–998, 1990.
- 28) Smoker WRK, Berbaum KS, Luebke NH, Jacoby CG. Spatial perception testing in diagnostic radiology. *AJR*, 143: 1105–1109, 1984.
- 29) Nodine CF, Krupinski EA. Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. *Acad Radiol*, 5: 603–612, 1998.
- 30) Krupinski EA, Roehrig H, Furukawa T. Influence of film and monitor display luminance on observer performance and visual search. *Acad Radiol*, 6: 411–418, 1999.
- 31) Krupinski EA, Roehrig H. The influence of a perceptually linearized display on observer performance and visual search. *Acad Radiol*, 7: 8–13, 2000.
- 32) Krupinski EA, Lund PJ. Differences in time to interpretation for evaluation of bone radiographs with monitor and film viewing. *Acad Radiol*, 4: 177–182, 1997.
- 33) Krupinski EA. An eye-movement study on the use of CAD information during mammographic search. Paper presented at the 7th Far West Image Perception Conference, Oct 16–18, 1997, Tucson, AZ.