Subjective and Objective Assessment of Image Quality—A Comparison

Walter F. Good, David Gur, John H. Feist, F. Leland Thaete, Carl R. Fuhrman, Cynthia A. Britton, and B. Simon Slasky

Forced-choice just noticeable difference (JND) studies are extremely sensitive to image quality variations that are below the threshold at which the differences are apparent to or definable by the observer. Paired comparisons of 4K and 2K laser-printed posteroanterior chest images consistently demonstrated that although images are viewed as comparable by radiologists, when forced to choose the better ("sharper") image, they actually select the higher-resolution images in 83% of the paired observations. We conclude that small differences in image quality may be detectable even in image sets which are considered to be comparable by subjective assessments.

KEY WORDS: just noticeable difference (JND), observer performance, analog film digitization.

IT HAS BEEN ESTIMATED that high-quality chest radiographs are roughly equivalent to 8,000 × 10,000 pixels in each image. In the digital world, there are clear advantages to reducing image matrices to 2,000 × 2,500 pixels or less. In most instances, all visible abnormalities on conventional analog films can be shown on laser-printed films with significantly reduced matrix size.[1] If the differences between images are not noticeable, one can hypothesize that diagnostic performance is not affected; hence, the reduced matrix can be used as the gold standard for digital acquisition (eg, computed radiography systems) or digitization of film for primary diagnosis purposes, or for comparisons with compressed images.

Contrary to this logic, observer performance has been shown to be marginally affected when high-quality conventional images are digitized and laser printed with matrix sizes as large as 4,000 × 5,000 pixels.[1-4] In an attempt to better understand detectability of extremely small image quality differences, we performed the following just noticeable difference (JND) study.

METHOD

Twelve high-quality posteroanterior chest images were digitized at 4,000 × 5,000 × 12-bit matrices and were laser printed with the same matrix after adjusting the lookup table to enable generation of images which would match the original conventional films as closely as possible (within 0.1 optical density or better). The film digitizer used in the study is an experimental digitizer (Eastman Kodak, Rochester, NY). Details of its performance characteristics are described elsewhere.[5] The images were then reduced by pixel averaging and filtering using a Gaussian kernel[6] to a matrix of 2,000 × 2,500 × 12 bits and laser printed (after cubic spline interpolation and lookup table adjustment) onto a full-size film to match the size and look (brightness and contrast) of the 4,000 × 5,000-pixel images.

Five experienced radiologists were presented with 12 pairs of laser-printed images displayed side-by-side in random order. The radiologists were told that in each pair there was a higher resolution, "sharper" image, and they were asked to select that image. To avoid being affected by dominant eye acuity differences or display quality (viewbox), the radiologists were asked to swap the images in each pair (right/left) several times before making a selection. At the end of each series, each observer was asked to estimate his/her accuracy (number of correct responses) in selecting the sharper images.

RESULTS

All participating radiologists indicated that the quality of the presented images was comparable, and they did not believe that "information content" varied between the two images of each pair. Four of five radiologists indicated that they felt like "flipping a coin" when selecting the "sharper" image. The observers' estimated number of correctly selected images and

From the Department of Radiology, University of Pittsburgh, Pittsburgh, PA.
Supported in part by Grant Nos. PO1 CA43684 and RO1 CA51248 from the National Cancer Institute.
Address reprint requests to David Gur, ScD, A449 Scaife Hall, Department of Radiology, University of Pittsburgh, Pittsburgh, PA 15261-0001.

0897-1889/94/0702-0014$03.00/0

Table 1. Observer-Estimated and Actual Number of Correctly Selected 4K Images as the Better or "Sharper" Image

Reader	Actual Number of Correct Selections	Observer-Estimated Number of Correct Selections
1	11	8
2	12	9
3	9	7
4	8	8
5	10	8

Note: The total set included 12 images.

their actual performance is given in Table 1. It should be noted that all participating radiologists correctly selected more than six images (50%), and four of five estimated their performance to be lower than the actual results. Despite the small sample size, these results are statistically significant ($P < .01$).

One radiologist was asked to repeat the study after 6 weeks to assess whether the results were repeatable. The results of this repeated experiment were similar to the original observation, although different errors (incorrectly selected images) were made during the second experiment.

DISCUSSION

The results of this study indicate that extremely small differences in image quality may be detectable even when radiologists believe that images are comparable. The study also raises questions about the use of 2,000 × 2,500 image matrices as the gold standard or the starting point for data compression studies. In an environment where many factors affect detection performance in small increments, the acceptance of 2,000 × 2,500-pixel images as the "noncompressed" gold standard may already reduce high-frequency information that is important for optimal detection of certain abnormalities.[1-3]

Trade-offs are often made in radiology for a variety of reasons (eg, reduced exposure, reduced cost, improved access). It is not clear whether 2,000 × 2,500-pixel images are or are not the optimal selection for primary diagnosis in a digital environment. However, *a priori*, systems driven selection may be the wrong approach to optimal performance in a complicated multifactorial environment. The study also shows the high sensitivity (small sample size requirements) of the JND methodology for the detection of extremely small differences in image quality. Although our results were significant in this limited-scope study, we wish to caution against their broad generalization. Additional investigations are clearly needed in this regard.

REFERENCES

1. Slasky BS, Gur D, Good WF, et al: Receiver operating characteristic analysis of chest image interpretation with conventional, laser-printed, and high-resolution workstation images. Radiology 174:775-780, 1990

2. Cox GG, Cook LT, McMillan JH, et al: Chest radiography: Comparison of high-resolution digital displays with conventional and digital film. Radiology 176:771-776, 1990

3. Fajardo LL, Hillman BJ, Pond GD, et al: Detection of pneumothorax: Comparison of digital and conventional chest imaging. Am J Roentgenol 152:475-480, 1989

4. Kundel HL: How much spatial resolution is enough? A meta-analysis of observer performance studies comparing plain films and digital hard copy. Proc SPIE 1899:86-89, 1993

5. Whiting BR, Muka E, Kocher TE, et al: High-resolution, high-performance radiographic film scanner. Proc SPIE 1231:295-305, 1990 .

6. Shaw CC, Herron J, Gur D: Pixel averaging versus digitization using larger apertures—A comparison of the spatial resolution properties. Proc SPIE 1651:128-133, 1992