Reprinted from SPIE Vol. 535—Application of Optical Instrumentation in Medicine XIII
Medical Image Production, Processing, and Display

Detectability of lesions of various sizes on CT images

Philip F. Judy and Richard G. Swensson

Department of Radiology, Brigham and Women's Hospital and
Harvard Medical School, Boston, MA 02115

Abstract

This experiment studied the detectability of lesions superimposed at specified locations on CT images of water phantoms. Each set of images contained 120 pedestals (a constant CT number added to circular areas of the water-phantom images), 60 of which also had a circular lesion profile of a given size and contrast superimposed on their centers. Lesion detectability (estimated from observers' ROC curves) was measured for 24 separate sets of images, representing four different levels of contrast for each of six lesion sizes (3.0 to 16.0 mm in diameter). Direct calculations of each lesion's "signal-to-noise ratio" (SNR), from the physical CT values in each set of images, predicted its detectability for observers, and there was no systematic change in the linear relation between detectability and lesion SNR as a function of lesion size. The levels of contrast required to produce either a given level of lesion SNR or a given level of detectability were approximately linear functions of the lesion's diameter on a log-log plot, and parallel to the contrast-detail curve that we measured independently with the same image format.

The slope of these curves (about -1.0) is consistent with those measured by other investigators, and is inconsistent with the signal detection predictions that assume an ideal ramp function for the CT noise power spectrum. These results suggest that the ideal noise power spectrum may not be realized in actual CT scanners, and that the actual deviations at low spatial frequencies affect the detectability of large lesions.

Introduction

Whether or not contrast-detail (CD) curves predict the usefulness of CT images has been a controversial topic [1] and still remains so [2]. A CD curve plots the degree of "contrast" (e.g., difference in CT number) required to satisfy some visibility "threshold" as a function of the feature's size (e.g., circular diameter). While those controversies involving the technical issues of CT imaging have generally been resolved, there remains controversy about the proper interpretation of empirical CD curves obtained from observers' subjective judgments of feature visibility. When plotted on log-log coordinates, these measured CD curves have a slope of about -1.0 [3], which is inconsistent with the slope of -3/2 predicted from theoretical, signal-detection formulations [4]. The procedures used to measure CD curves have been criticized [2], because there is no assurance that observers' contrast "thresholds" remain "constant" (imply a constant detection capability) at all feature sizes.

The signal detection theory of observer performance assumes that pixel values at the image locations which may contain a feature are used to calculate a variable that determines the observer's detection decision. The expected value of this decision variable would differ, depending on whether or not the feature were actually present at the given location. Even neglecting any variability that may be added by the human observer, however, noise on the physical image will produce variability in the distributions of any physical quantity calculated from the pixel values. The feature's contrast will determine the degree of separation in the distributions of a physical decision variable, calculated for image locations with and without actual features, relative to the variations produced by image noise. Since an increase in the contrast will increase the relative separation between these distributions, it will increase the accuracy of detection decisions based upon this physical variable.

A statistical index that is useful in summarizing the expected performance of a detector is its signal-to-noise ratio (SNR) for a specified detection task. The SNR is defined as a difference in mean values of the decision variable for the two situations to be differentiated (usually assumed to be normally distributed with equal variance), expressed in units of its standard deviation. In such cases, the SNR completely specifies the ROC curve, the function that relates the conditional probabilities of "true-positive" and "false-positive" decisions (given the presence or absence of the feature) over changes in the detection criterion. The decision variable for a human observer can be measured only indirectly, however, from the observer's decisions or the measured ROC curve for a particular detection task. The fitted parameters of a measured ROC curve can be used to estimate an index of "feature detectability" (analogous to an "observer SNR") that does not depend on the observer's choice of any particular detection criterion.

A detector's SNR depends on the feature to be detected, on the physical characteristics of the images (e.g., resolution and noise), and on the details of the calculation performed. Wagner [5] has shown that many proposed indices of image quality can be derived by a formulation that assumes a particular feature to be detected and a specific calculation to be performed upon the physical images. This same formulation can be used to predict the form of the CD curve for the human observer, by assuming that the observer decides that a feature has sufficient contrast to be detected when it yields some (minimum constant value of) SNR for a specified calculation on the images. Essentially, this assumes that: 1) the observer's threshold judgments and detection decisions both depend on the same specified physical calculation of the image information and 2) the observer's contrast "thresholds" imply a constant level of detectability for the features of all sizes.

A popular assumption, with some empirical support [6,7], is that the human observer performs a "matched-filter" calculation--i.e., a cross-correlation of the feature's expected feature profile with the pixel values at a given location on the CT image. Wagner [4] showed that the CD curve for such a matched-filter observer (using a constant SNR criterion) would be predicted to have a slope of -3/2 on a log-log scale, assuming a theoretically ideal ramp function for the noise power spectrum of the CT image. There are several reasons why this signal-detection formulation may fail to predict the shape of CD curves measured by observers' contrast "threshold" judgments:

1. The theoretical characterization of the image properties may not be correct for actual CT images.

2. The assumed (matched-filter) calculation may not model the information that the human observer obtains from the CT image.

3. The observer's subjective judgments of contrast may imply systematic changes in the feature's detectability, as a function of the feature size.

Our experiment measured a feature's detectability (from observers' ROC curves) and its SNR (from a realized matched-filter calculation), as a function of the feature's size and contrast on the CT image. These measured values were used to generate CD curves that represent features of equal detectability and features of equal SNR, for comparison with both the "ideal" CD curves and those obtained from observers' subjective judgments of feature visibility.

Method

The image format consisted of seven circular pedestals, arranged in a hexagonal array. These pedestals were created on the CT images of water phantoms from the Technicare Delta 2020 scanner, by adding a constant value of 50 to the pixel CT values within areas of 64 mm in diameter. The six outer pedestals (at a constant radius of the circular water phantom) were used for presenting the features to be detected or judged by observers, while the central pedestal on each image always contained a high-contrast example of the feature as a reference for its size and location. The images were photographed onto transparencies for the visual stimuli of the experiment, using a display window of 120 CT numbers centered at the pedestal mean value (50).

The visual stimuli for the measurements of feature detectability were 24 different sets of images. Each set consisted of 120 separate image pedestals, 60 of which contained a darker circular feature of a specified size and contrast superimposed at the center. There were six different feature sizes (3.0 to 16.1 mm in diameter) and four levels of contrast for each size, selected (from pretest data) to span a useful range of detectability values. Six experienced observers read all 24 sets of images, using a six-category scale to rate the likelihood that each image pedestal contained the specified feature. The observers were instructed to vary their viewing distance as they judged each image pedestal. For convenience, a minification lens was used to extend the effective viewing distance (by a factor of 3) without a change in position.

An index of feature detectability (d_ε) was estimated from a two-parameter fit [8,9] to the ROC curve generated from each observer's ratings of the 120 pedestals in a given set of images. This particular index of detectability specifies the operating point on the observer's fitted ROC curve where the two types of detection errors are equiprobable, and it has an interpretation similar to that of the SNR for a physical calculation [10].

The SNR for the matched-filter calculation was determined directly from each of the 24 sets of images. This calculation uses the realized values of the cross-correlation between the image pixel values and the feature's expected profile on the 120 separate pedestals. These direct estimates of the feature SNR include the effects of the discrete gray-scale and pixel representations, and any unique effects of the particular samples of noisy images within each set.

The visual stimuli for the direct measurements of observers' CD curves were sets of three images, one set for each of 17 feature sizes (2.5 to 16.1 mm in diameter). All 18 pedestals in each set of images contained a feature of the specified size, but superimposed at a systematically decreasing level of contrast. Preliminary data were used to select an appropriate range of contrast values for each feature size; the feature contrast decreased by the factor 1.2 between each successive pair of the 18 pedestals. Four observers judged the 17 sets of images, reporting the last pedestal of each set in which the feature could still be seen.

Results and discussion

For each feature size, an increase in contrast (specified feature-pedestal difference in CT numbers) increased both the values of feature detectability (d_ε), measured from observers' ROC curves, and the feature SNR, measured from the matched-filter calculations on the CT images. Figure 1 plots the mean estimates of d_ε against the SNR for each of the 24 combinations of feature size and contrast studied in this experiment.

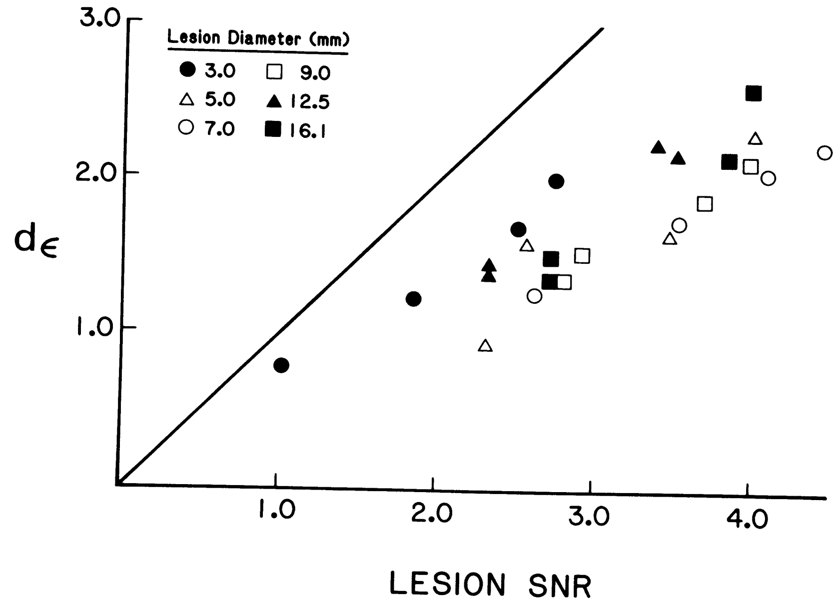

Figure 1. Mean detectability (d_ε) for six observers in 24 different experimental conditions, plotted against the feature (lesion) SNR for the realized matched-filter calculations. The four data points plotted for each feature size represent different levels of specified contrast, the feature-pedestal difference in CT numbers.

The six different regression lines (separately fitted to the four data points for each feature size) did not change systematically with the feature's size. The separate linear regressions were used to estimate the values of contrast required to produce equal levels of detectability for features of the six different sizes.

Figure 2 plots the contrast values required for d_ε=2.0 against the feature size, as a CD curve on a log-log scale. Also plotted in Figure 2 are the contrast values required to produce a feature SNR of 3.0 for the matched-filter calculations. These CD curves for equal feature detectability and equal feature SNR are nearly identical in shape, and both are parallel to the CD curve we derived from the observers' direct judgments of their contrast thresholds. The slopes of all three curves are close to -1.0, consistent with previous direct measurements of CD curves [3] and inconsistent with the theoretical prediction of -3/2, which assumes the noise power spectrum of an ideal CT scanner.

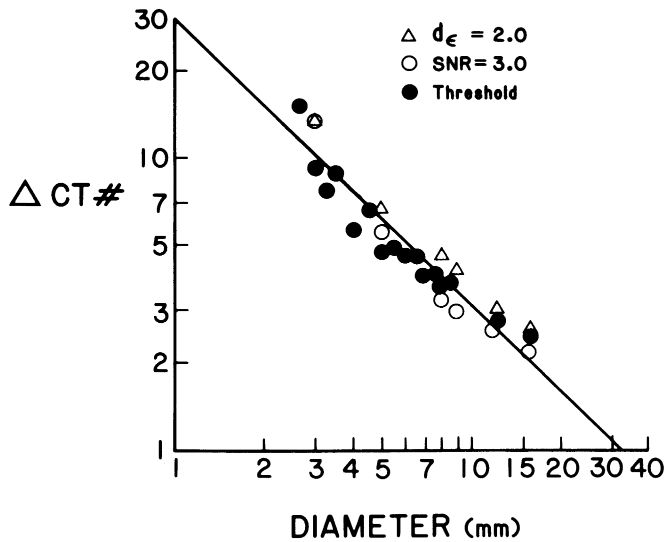

Figure 2. Contrast-detail curves measured by three different procedures. Each CD curve plots the contrast (difference in CT number) required to satisfy a particular type of measurement criterion, as a function of the feature size (on a log-log scale). Solid data points show the CD curve obtained when the measurement criterion was the observer's direct judgment of the contrast threshold for feature visibility (mean contrast thresholds from 4 observers at 17 feature sizes). The two sets of open points are the CD curves estimated for the six different feature sizes studied in the detection experiment. They show the contrast required to achieve an equal feature detectability for observers (open triangles, d_ε=2.0) or an equal SNR for the matched-filter calculations on the realized CT images (open circles, SNR=3.0). The solid line has a slope of -1.0.

Our data demonstrate that the disagreement between measured CD curves and those predicted by the theoretical analysis cannot be explained methodologically, by difficulties in the interpretation of observers' subjective judgments of contrast "thresholds". In addition to replicating previous direct measurements of the CD curve, we obtained a CD curve of the same shape for equally detectable features, estimated from observers' measured ROC curves. Our results also suggest that the source of difficulty is not the image calculation assumed for the human observer, since matched-filter calculations on the physical CT images predicted the variations in feature detectability for our observers.

For the matched-filter calculations on the realized CT images, however, the contrast required to achieve a constant value of SNR did not show the dependence on feature size that is predicted theoretically. Previous measurements of the noise power spectrum in realized CT images [11] have revealed deviations from the ideal spectrum at low spatial frequencies. Our data suggest that these deviations affect the detectability of features and the shape of the CD curve, by increasing the contrast required to achieve a specified level of SNR.

Acknowledgement

The research was supported by PHS grants CA32813 and GM18674.

References

1. Hasegawa, B.H., "Mis-use of Contrast Detail Curves," A.J.R., vol. 135, pp. 199-200. 1980.
2. Goodenough, D.J. and Weaver, K.E., "Factors Related to Low Contrast Resolution in CT scanners," Comput. Radiol., vol. 8, pp. 297-308. 1984.
3. Cohn, G., "Contrast-detail-dose Analysis of Six Different Computed Tomographic Scanners," J. Comput. Asst. Tomogr., vol. 3, pp. 197-203. 1979.
4. Wagner, R.F., Brown, D.G. and Pastel, M.S. "Application of Information Theory to the Assessment of Computed Tomography," Med. Phys., vol. 6, pp. 83-94. 1979.
5. Wagner, R.F., "Decision Theory and the Detail Signal-to-Noise Ratio of Otto Schade," Photog. Sci. Eng., vol. 22, pp. 41-46. 1978.
6. Judy, P.F., Swensson, R.G., Szulc, M., "Lesion Detection and Signal-to-Noise Ratio in CT Images," Med. Phys., vol. 8, pp. 13-23. 1981.
7. Burgess, A. and Ghandeharian, H., "Visual Signal Detection. I. Ability to Use Phase Information," J. Opt. Soc. Am., vol. 1, pp. 900-905. 1984.
8. Dorfman, D.D. and Alf, E., "Maximum-likelihood Estimation of Parameters of Signal-detection Theory and Determination of confidence Intervals--Rating-method Data," J. Math. Psych., vol. 6, pp. 487-496. 1969.
9. Grey, D.R. and Morgan, B.J.T., "Some Aspects of ROC Curve-fitting: Normal and Logistic Models," J. Math. Psych., vol. 9, pp. 128-139. 1972.
10. Sandor, T. and Swensson, R.G., "Evaluation of Observer Performance in Detecting Blood Vessels on Simulated Angiographic Images," Med. Phys., vol. 5, pp. 380-386. 1978.
11. Kijewski, M.F., Judy, P.F., Swensson, R.G., Toltzis, H., "Stimuli for Lesion Detection on CT Images," Med. Phys., vol. 10, pp. 525. 1983.