

A Psychophysical Comparison of Two Methods for Adaptive Histogram Equalization

John B. Zimmerman, Steve B. Cousins, Karin M. Hartzell, Mark E. Frisse, and Michael G. Kahn

Adaptive histogram equalization (AHE) is a method for adaptive contrast enhancement of digital images. It is an automatic, reproducible method for the simultaneous viewing of contrast within a digital image with a large dynamic range. Recent experiments have shown that in specific cases, there is no significant difference in the ability of AHE and linear intensity windowing to display gray-scale contrast. More recently, a variant of AHE which limits the allowed contrast enhancement of the image has been proposed. This contrast-limited adaptive histogram equalization (CLAHE) produces images in which the noise content of an image is not excessively enhanced, but in which sufficient contrast is provided for the visualization of structures within the image. Images processed with CLAHE have a more natural appearance and facilitate the comparison of different areas of an image. However, the reduced contrast enhancement of CLAHE may hinder the ability of an observer to detect the presence of some significant gray-scale contrast. In this report, a psychophysical observer experiment was performed to determine if there is a significant difference in the ability of AHE and CLAHE to depict gray-scale contrast. Observers were presented with computed tomography (CT) images of the chest processed with AHE and CLAHE. Subtle artificial lesions were introduced into some images. The observers were asked to rate their confidence regarding the presence of the lesions; this rating-scale data was analyzed using receiver operating characteristic (ROC) curve techniques. These ROC curves were compared for significant differences in the observers' performances. In this report, no difference was found in the abilities of AHE and CLAHE to depict contrast information.

© 1989 by W.B. Saunders Company.

KEY WORDS: Adaptive histogram equalization, contrast limiting, psychophysics, contrast-enhancement.

From the Electronic Radiology Laboratories, Mallinckrodt Institute of Radiology, Washington University, St Louis and the Department of Internal Medicine, Washington University, St Louis.

This work supported in part by Philips Medical Systems, Inc.

Address reprint requests to John B. Zimmerman, PhD, Assistant Professor, Department of Computer Science, Washington University, St Louis, MO 63130.

© 1989 by W.B. Saunders Company.
0897-1889/89/0202-0006\$03.00/0

A CONTINUING PROBLEM in the display of medical images is the low dynamic range of electronic display devices. Contemporary imaging modalities such as computed tomography (CT), magnetic resonance imaging (MRI), and digital radiography may contain ten to 12 bits of significant information; a typical electronic display such as a cathode ray tube (CRT) screen may be able to represent less than seven bits of information (90-100 just noticeable differences). The problem then is to display all the information of interest to the physician by the use of some appropriate method of contrast enhancement, while avoiding the introduction of artifacts into the image and, preferably, reducing the need for manual interaction with the image by the physician.

Typically, the necessary contrast enhancement has been accomplished by the linear rescaling of a subrange of the intensity values in the image to occupy the full display range of the electronic device. This method, linear intensity windowing (often called center/width windowing), requires both manual intervention to produce optimal results and the use of several intensity subranges to depict all the contrast of interest in the image. Adaptive histogram equalization (AHE), an alternative to linear intensity windowing, was developed independently by Ketcham et al.,¹ Hummel,² and Pizer et al.³ AHE is an automatic, reproducible, contrast enhancement technique that simultaneously depicts most of the contrast of interest and shows good performance in both formal evaluations and preliminary clinical trials.⁴ In a previous report,⁵ the authors demonstrated that for some cases of subtle contrast detection in a signal-known-exactly experiment, the ability of AHE to depict luminous contrast is not significantly different from that of linear intensity windowing.

One disadvantage of AHE is that the amount of contrast enhancement that it performs is so great that in relatively homogeneous areas or areas of low signal (eg, the background of CT images), the noise component becomes very prominent. The resulting appearance is unattractive

and distracting to physicians attempting to perform diagnosis. Furthermore, in certain images, such as digital chest radiographs, the amount of enhancement produced by AHE may cause some breakup of the objects within the image; continuity is lost among the various organs. To address these difficulties, Pizer has proposed a variant of AHE called contrast-limited adaptive histogram equalization (CLAHE).⁶ In this method, the amount of contrast enhancement that can be produced within a local area of the image is limited by a parameter that may be adjusted to accommodate different image types. The resulting images are more natural in appearance than those produced by AHE. Preliminary assessment of CLAHE by radiologists at the Mallinckrodt Institute of Radiology in St Louis indicates that it is preferred over AHE for processing high-resolution, high-contrast digital radiographs, such as those generated by the Philips Computed Radiography (PCR) unit (Philips Medical Systems, Inc, Shelton, CT).

Since CLAHE deliberately limits the amount of contrast enhancement performed on an image, the detectability of subtle contrast in CLAHE images may be reduced relative to that in AHE images. Before further developing CLAHE as a contrast enhancement method for digital radiography, it is necessary to determine whether the use of CLAHE results in a reduction in diagnostic performance. As a first step in this determination, the authors have undertaken formal psychophysical observer studies to measure the relative detectability of subtle contrast features when AHE and CLAHE are used. Observers were shown CT scans of the chest, processed with either AHE or CLAHE, into some of which subtle artificial lesions were introduced. The observers were asked to assess their confidence that the artificial lesions were present. Their relative performance was measured to see whether there was a difference in their ability to detect the contrast features when using the two different enhancement modalities.

AHE AND CLAHE

AHE is a locally-adaptive contrast enhancement algorithm that is sensitive to the local information content of the image. In the AHE algorithm, the enhancement at a given pixel in the image is a function of the intensity values

immediately surrounding the pixel. The basic method is illustrated in Fig 1. For each pixel in the image, a region centered about the pixel, called its contextual region, is assigned. The intensity values in that region are used to calculate a histogram equalization mapping which is then applied to the pixel in question. The result is an image in which the mapping applied to each pixel is different and is adaptive to the local distribution of pixel intensities rather than to the global information content of the image. In practice, this technique produces an image in which different objects whose intensity values lie in different subranges of intensity values are simultaneously visible.

This algorithm, based upon calculating a separate histogram equalization mapping for each pixel in the image, works well but is extremely time-consuming to calculate; the enhancement of a 512×512 image requires about two hours to compute on a Sun 3 computer (Sun Microsystems Inc, Mountain View, CA). An alternative algorithm proposed by Pizer reduces the amount of computation by choosing a number of sample pixels, typically 16 to 64 pixels arranged on a regular grid. The correct histogram equalization mapping is calculated for each sample pixel; the mapping for any given pixel is then derived as a

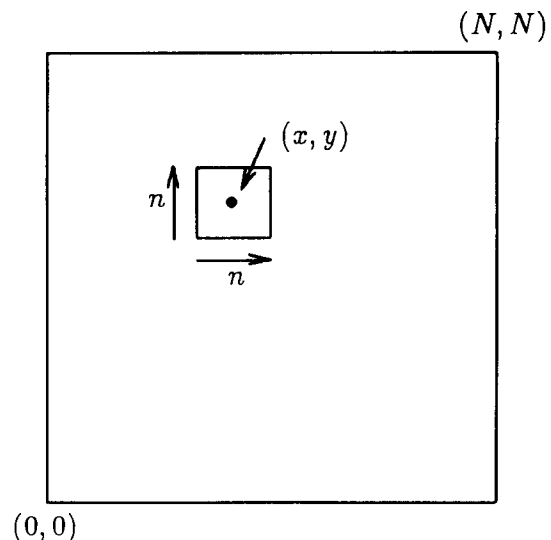


Fig 1. Adaptive histogram equalization. The contextual region shown is a square of $n \times n$ pixels about a pixel at location (x, y) . The pixels in this region are used to calculate a histogram equalization mapping which is then applied to the pixel at (x, y) .

bilinear interpolation of the correct mappings at the four sample pixels that are nearest to it. This interpolated version of AHE gives results that are very close to the uninterpolated version. A full description of this method may be found in Pizer et al.⁶ Interpolated AHE allows 512×512 medical images to be processed in tens of seconds on a general purpose minicomputer or a few seconds on many special-purpose image processing systems.

CLAHE differs from AHE in that it places restrictions on the histogram equalization mapping. For small contextual regions, it is often the case that the range of pixel intensities is limited. When histogram equalization is performed, this small intensity range will be expanded to the full intensity range of the image; this results in a large contrast enhancement and is the fundamental reason why AHE is effective as a contrast enhancement algorithm. This effect may be seen by calculating the cumulative distribution function (CDF) of the pixel intensities within the contextual region. The CDF (the mapping which is used to equalize the histogram) will have large values for its first derivative (i.e., the slope of the CDF at some intensities may be quite steep), indicating intensity ranges which have strong contrast enhancement. CLAHE controls the amount of contrast enhancement in the image by restricting the maximum number of pixels in each bin of the histogram and thus the slope of the CDF. A single parameter, the clipping level, specifies the number of pixels in a histogram bin (Fig 2). For 512×512 CT images with a total intensity range of about 2,048 Hounsfield units and a 64×64 pixel contextual region, a typical regional histogram will contain about 1,024 bins with a median occupancy of about 15 to 20 pixels/bin. Preliminary experiments indicate that a clipping level of 20 pixels/bin produces results essentially identical to AHE, while with a clipping level of five, the amount of contrast enhancement is severely restricted. These studies have shown that a clipping level of 10 pixels/bin is appropriate for such CT scans, giving images that have considerable contrast enhancement, but which retain a more natural appearance than images processed with AHE. The results of applying AHE and CLAHE to a CT scan are shown in Fig 3. Figure 3A shows the original image, while Figs 3B and 3C are the result of

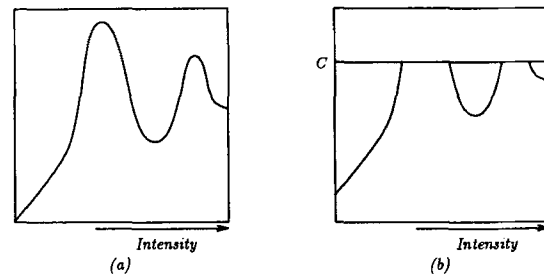


Fig 2. Histogram clipping. (a) shows the histogram of a contextual region before clipping. In (b), each bin of the histogram with more than a certain number of pixels has been clipped and the excess pixels uniformly distributed to the other bins such that after the distribution, no bin contains more than C pixels, where C is the clipping level.

applying AHE and CLAHE respectively. The clipping level for the CLAHE image was 10 pixels/bin.

EXPERIMENTAL PROTOCOL

The purpose of these experiments was to determine if the use of CLAHE results in a reduction in the ability of observers to detect subtle features in an image when compared with AHE. The objective is to evaluate the two contrast enhancement methodologies using observer tasks that are as clinically realistic as possible while still allowing the results to be analyzed by well-understood statistical methods. In a previous experiment, we compared the relative difference in the ability of intensity windowing and AHE to depict subtle contrast differences⁵; our current work retains the experimental methodology used earlier. A full description of that methodology is given in the reference cited; the following is an overview of the experiment and a description of the differences between the current experimental design and that of the earlier work.

Overview of the Experiment

A receiver operating characteristic (ROC) rating experiment was performed to compare the ability of AHE and CLAHE to depict luminance contrast. In this experiment, a set of test images was prepared from normal CT images of the chest. In each normal image, four sites were chosen for the insertion of simulated lesions, two sites in the lungs and two in the mediastinum. These sites were chosen with the assistance of an experienced radiologist. For each field (lungs and mediastinum), two simulated lesions were prepared. The simulated lesions were two-dimensional Gaussian spots that varied in linear size and gray-scale intensity. The linear sizes were chosen as appropriate for the given field (lungs or mediastinum). No attempt was made to account for the disarrangement of normal structure at the insertion site, but this effect was considered small due to the restricted linear sizes of the simulated lesions. The intensity of each lesion was chosen such that when the lesion was inserted into an image it was very subtle. Preliminary experiments allowed us to estimate, through the use of a simple formula, the intensity needed in order for the simulated lesions to appear approxi-

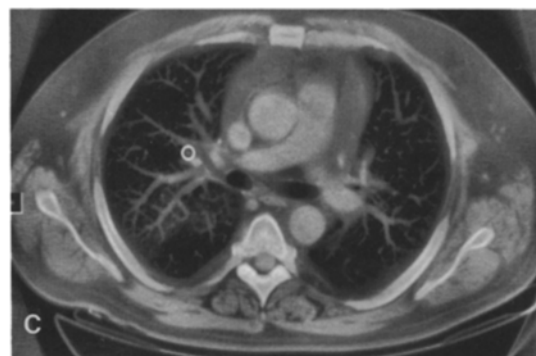
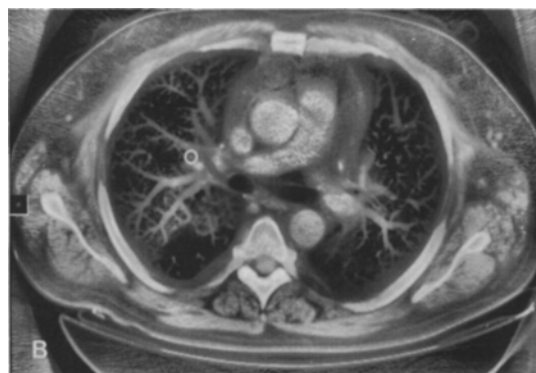
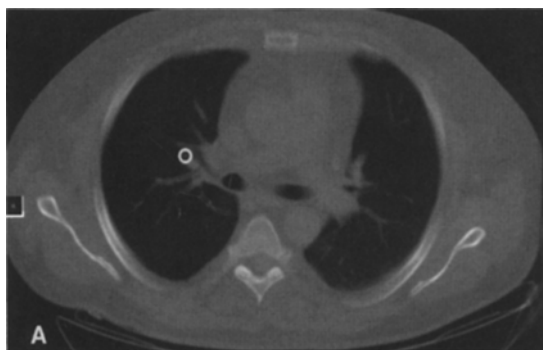


Fig 3. Images processed with AHE and CLAHE. The image is a 512×384 CT scan. (A) The original image. (B) The image processed with AHE. (C) The image processed with CLAHE. In (B) and (C), the interpolated version of the algorithm was used with a 64×64 pixel square contextual region. Each image has a simulated lesion inserted in the lung field; the site of insertion is emphasized by the white circle. This circle does not appear in the images shown to observers. The same lesion is inset into the left side of the image. The intensity of the lesion has been exaggerated by a factor of five for illustrative purposes. The cross hairs used in the observer experiment are not shown.

mately one just noticeable difference (JND) brighter than their background.⁵ The characteristics of the lesions are given in Table 1. The height factors in Table 1 were used to modify the predicted intensity so that, for example, lesion 0 in the lungs was 0.85 of the intensity predicted by the formula (ie, approximately 0.85 JND).

Preparation of the Image Test Set

A set of 24 normal CT scans of the chest was selected from five different patients with the assistance of an experienced radiologist. The images were taken from a Technicare 2060 scanner (Technicare, Inc, Cleveland, OH); their resolution was 512×512 pixels. Adjacent slices were avoided in order to minimize the memorization of normal structure. From each normal image, eight additional images were generated, each image having one of the lesions inserted at one of the selected sites. From each of these images, two more images

were prepared by processing the given image with AHE and CLAHE. The complete test set consisted of the normal images with AHE applied, the normal images with the lesions inserted and AHE applied, the normal images processed with CLAHE, and the normal images with the lesions inserted processed with CLAHE. In order that there be equal numbers of images with and without the artificial lesions, multiple copies of the processed images without lesions inserted were included; hence, for each image with a lesion inserted there was a companion image with no lesion. The result was a test set of 768 images, which was then presented in random order to five observers.

In the processing of the images, both AHE and CLAHE used a contextual region size of 64×64 pixels and the pixel intensity mappings were calculated using the interpolated algorithm previously described. For CLAHE, the necessary clipping level parameter value was obtained by examining ten CT scans of the chest processed using different values for the clipping level. This study indicated that a clipping level of 10 pixels/bin was appropriate; all images processed with CLAHE in the formal study used a clipping level of 10 pixels/bin.

Table 1. Parameters for the Simulated Lesions

Lesion Number	Width (σ)	Height Factor
Lungs		
0	1.41	0.85
1	1.41	1.15
Mediastinum		
0	2.00	0.85
1	2.00	1.15

The widths correspond to the standard deviation of the two-dimensional Gaussians, given in pixels. The height factor multiplies the intensity which is predicted to result in a one JND difference between the lesion and its background (see text).

Selection of Observers

The five observers consisted of two internists, experienced in the reading of CT scans, and three laymen. The observers were trained in the observing techniques and were allowed two practice runs of 64 images each; the first of these runs was conducted with feedback to the observers to allow them to calibrate their perceptions against the appearance of the images. The results from these test runs were not included in the final data.

Experimental Apparatus and Layout

The images in the test set were presented to the observers on a Barco CDCT 5351 color CRT display (Barco Video & Communications NV, Kortrijk, Belgium) attached to a Pixar Image Computer (Pixar, San Rafael, CA). The images were displayed at a resolution of 512×512 pixels using the full screen display size of 35×35 cm; the range of gray-scale intensities on the monitor varied from 0 to 21.3 foot-lamberts. Before beginning the experiments, the monitor was carefully calibrated and the gray-scale intensity mapping was linearized using the procedures described by Johnston et al.⁷ It was found that the perceived dynamic range of the monitor was 78.8 JNDs.

The observers were seated before the monitor at a distance of approximately 70 cm; no attempt was made to constrain their movements. All room lights were extinguished except for an ambient illumination of 2.5 lux.

Presentation of the Images

The images were presented to the observers in 12 sessions of approximately 25 minutes each. For each image in the test set, the observers were shown the image on the CRT monitor; inset into one side of the test image was a replica of the lesion which may have been inserted into that image. This replica was prescaled to approximate the appearance of the lesion after processing. Cross hairs were placed on the image at the site location of the prospective lesion. The observers were able to turn the cross hairs on and off at will, but could not affect its location in the image. Typical images are shown in Fig 3; in these images, the inset replica lesion is at the left edge of the image and the small white circle shows one of the sites used for placement of the simulated lesions into the lung field. The intensity of the lesion has been exaggerated in these photographs for demonstration purposes. The cross hairs are not shown.

Thus, the observers knew both the approximate appearance of the target and the exact site at which it might appear, a signal-known-exactly task. The task of the observers was to rate their confidence that the simulated lesion was present on a scale of 0 to 4, with 0 indicating that the lesion almost certainly was not present and 4 indicating that the lesion was probably present. The exact definitions of each rating category, as made known to the observers, are given in Table 2. The observers were encouraged to make use of all five rating categories, using a 0 response when they were most certain the lesion was not present and a 4 when they were most certain that it was present.

Table 2. Rating Criteria for the ROC Response Categories

Rating	Confidence
0	Definitely not present
1	Probably not present
2	Possibly not present
3	Possibly present
4	Probably present

Analysis of the Results

The task presented to the observers was one of luminance contrast detection; the data constituted the results of a rating scale experiment from which a ROC curve could be derived, giving ordered pairs of true positive v false positive responses for any desired level of observer confidence.⁸ Each combination of lesion type, lesion site, and image field (lungs or mediastinum) constituted one experiment; thus there were eight results for each observer. The data was analyzed using the CORROC program developed by Metz and his collaborators for correlated ROC data.⁹ The areas, A_z , under the ROC curves and their standard deviations were calculated and used to compare the two modalities for a given site and lesion; the modalities were assumed to have no difference in their ability to depict luminance contrast if the difference in the areas of the two ROC curves was less than a preset criterion. The results were evaluated for two discrimination criteria, differences of 1.5 and 2.5 standard deviations in the areas under the ROC curves.

RESULTS

In Tables 3 and 4, the results for the five observers are summarized. Each column shows the number of experiments for which each observer expressed a preference of modality. Observers B and C are the physicians; observers A, D, and E are laymen. The full results are given in Tables 5 to 9. In these tables, the headings of each column describe the parameters of a given

Table 3. Summary of Results for the Unpooled Rating Scale Data

Field	C (1.5 σ)	C (2.5 σ)	A (1.5 σ)	A (2.5 σ)	No Preference
Observer A					
Lungs	0	0	1	0	3
Mediastinum	1	0	0	0	3
Observer B					
Lungs	0	0	0	0	4
Mediastinum	0	0	0	0	4
Observer C					
Lungs	1	0	0	0	3
Mediastinum	0	0	1	0	3
Observer D					
Lungs	0	0	0	0	4
Mediastinum	0	0	0	0	4
Observer E					
Lungs	0	0	0	0	4
Mediastinum	0	0	0	0	4

Each column gives the preferred enhancement method, CLAHE (C) or AHE (A), at the shown discrimination level. For each field, lungs or mediastinum, there were four results, corresponding to two sites within the field \times 2 possible artificial lesions.

Table 4. Summary of Results After Pooling of Data Across Lesion Sites

Field	C (1.5 σ)	C (2.5 σ)	A (1.5 σ)	A (2.5 σ)	No Preference
Observer A					
Lungs	0	0	0	0	2
Mediastinum	0	0	0	0	2
Observer B					
Lungs	0	0	0	0	2
Mediastinum	0	0	0	0	2
Observer C					
Lungs	1	0	1	0	0
Mediastinum	0	0	1	0	1
Observer D					
Lungs	0	0	0	0	2
Mediastinum	0	0	0	0	2
Observer E					
Lungs	0	0	0	0	2
Mediastinum	1	0	0	0	1

The columns are as given in Table 3. Two results are given for each image field, corresponding to the two different lesion types as given in Table 1.

experiment (lesion type, site, and processing modality), the integrated area under the resulting ROC curve, and the standard deviation of the area. The standard deviations shown have been

corrected for the correlation of the data by the CORROC program. There is no data pooling in these results. The sixth column shows the number of standard deviations, n_σ , by which the areas of the two ROC curves differ. The seventh column is the two-tailed P value; it represents the confidence with which the null hypothesis, in this case that the two ROC curves have the same area, can be rejected. A small value of P indicates that the two areas being compared are unlikely to have arisen from the same underlying distribution. The final column indicates which, if either, method was found preferable. Those studies marked with a "C" showed a preference for CLAHE with a difference in standard deviation of at least 1.5σ ; those marked with an "A" showed a preference for AHE at the same level. The results were evaluated at two levels; if the entry has no asterisk, the areas under the ROC curves differ by more than 1.5 standard deviations but less than 2.5 standard deviations; if the asterisk is present, the areas differ by more than 2.5 standard deviations. To allow ready comparison, the data is arranged so that experiments with similar parameters (field, lesion type, lesion site, etc) are adjacent with only the contrast enhancement modality different. The values of

Table 5. Results for Observer A

Type	Site	Proc	Area	σ	n_σ	P Value	Preference
Lungs							
0	0	C	0.8569	0.0561	0.6136	.5395	None
0	0	A	0.8242	0.0629	0.6136	.5395	
1	0	C	0.8004	0.0657	-2.3696	.0178	
1	0	A	0.9386	0.0417	-2.3696	.0178	A
0	1	C	0.9169	0.0412	1.0348	.3008	None
0	1	A	0.8671	0.0542	1.0348	.3008	
1	1	C	0.9396	0.0339	1.1960	.2317	None
1	1	A	0.8901	0.0482	1.1960	.2317	
Mediastinum							
0	0	C	0.8074	0.0660	-0.3615	.7177	None
0	0	A	0.8370	0.0615	-0.3615	.7177	
1	0	C	0.9027	0.0483	1.7081	.0876	C
1	0	A	0.7924	0.0682	1.7081	.0876	
0	1	C	0.7562	0.0762	-0.4145	.6785	None
0	1	A	0.7951	0.0808	-0.4145	.6785	
1	1	C	0.8010	0.0676	-0.7880	.4307	None
1	1	A	0.8455	0.0594	-0.7880	.4307	

The columns indicate the lesion type, site, and processing method (Proc); the area under the ROC curve and its standard deviation; the number of corrected standard deviations by which the areas differ; the two-tailed P value; and the result. In the results column, the letter indicates the preferred method (C = CLAHE, A = AHE) at the 1.5σ level; the addition of an asterisk indicates that the results were significant at the 2.5σ level. No data pooling was performed.

Table 6. Results for Observer B. No Data Pooling

Observer B							
Type	Site	Proc	Area	σ	n_σ	P Value	Preference
Lungs							
0	0	C	0.6649	0.0841	0.9582	.3380	None
0	0	A	0.5701	0.0985	0.9582	.3380	
1	0	C	0.7259	0.0755	-1.2705	.2039	None
1	0	A	0.8458	0.0629	-1.2705	.2039	
0	1	C	0.6091	0.0884	-0.7744	.4387	None
0	1	A	0.6840	0.0825	-0.7744	.4387	
1	1	C	0.6420	0.0844	-0.1853	.8530	None
1	1	A	0.6621	0.0852	-0.1853	.8530	
Mediastinum							
0	0	C	0.7345	0.0781	-0.3199	.7490	None
0	0	A	0.7665	0.0731	-0.3199	.7490	
1	0	C	0.7690	0.0757	-1.4030	.1606	None
1	0	A	0.8686	0.0575	-1.4030	.1606	
0	1	C	0.6460	0.0870	-0.2099	.8338	None
0	1	A	0.6694	0.0872	-0.2099	.8338	
1	1	C	0.6666	0.0876	-0.5361	.5919	None
1	1	A	0.7179	0.0799	-0.5361	.5919	

1.5 and 2.5 standard deviations correspond to two-tailed P values of 0.1336 and 0.0124 respectively.

Unpooled Results

The results shown in Table 3 indicate that in very few cases was one modality preferred over the other. Of the 40 experimental results, only four showed a preference at the 1.5σ level, and none showed a preference at the 2.5σ level. Three

observers had no results which expressed a preference. Of the four results which showed a preference, two favored AHE and two favored CLAE; in the two different fields (lungs and mediastinum), the results are also evenly split. It is obvious that the performance of the two physicians (Observers B and C) corresponded well with that of the three laymen in this respect.

In the full results (Tables 5 to 9), the areas under the ROC curves, corresponding to the

Table 7. Results for Observer C. No Data Pooling

Observer C							
Type	Site	Proc	Area	σ	n_σ	P Value	Preference
Lungs							
0	0	C	0.7378	0.0747	1.3610	.1735	None
0	0	A	0.6146	0.0840	1.3610	.1735	
1	0	C	0.6923	0.0804	-0.9696	.3323	None
1	0	A	0.7689	0.0748	-0.9696	.3323	
0	1	C	0.7985	0.0699	1.8880	.0590	C
0	1	A	0.6223	0.0880	1.8880	.0590	
1	1	C	0.6785	0.0819	-1.2118	.2256	None
1	1	A	0.7711	0.0707	-1.2118	.2256	
Mediastinum							
0	0	C	0.7949	0.0731	0.2834	.7769	None
0	0	A	0.7689	0.0733	0.2834	.7769	
1	0	C	0.7715	0.0766	-0.9464	.3440	None
1	0	A	0.8493	0.0618	-0.9464	.3440	
0	1	C	0.6152	0.0881	-0.8928	.3720	None
0	1	A	0.7016	0.0810	-0.8928	.3720	
1	1	C	0.5767	0.0841	-2.0014	.0454	
1	1	A	0.7197	0.0779	-2.0014	.0454	A

Table 8. Results for Observer D. No Data Pooling

Observer D							
Type	Site	Proc	Area	σ	n_e	P Value	Preference
Lungs							
0	0	C	0.7621	0.0709	-1.3473	.1779	None
0	0	A	0.8342	0.0617	-1.3473	.1779	
1	0	C	0.8208	0.0630	-0.2801	.7794	None
1	0	A	0.8383	0.0595	-0.2801	.7794	
0	1	C	0.8702	0.0531	0.1994	.8419	None
0	1	A	0.8597	0.0555	0.1994	.8419	
1	1	C	0.8997	0.0473	-0.4364	.6625	None
1	1	A	0.9202	0.0442	-0.4364	.6625	
Mediastinum							
0	0	C	0.7511	0.0741	-0.7027	.4822	None
0	0	A	0.8090	0.0681	-0.7027	.4822	
1	0	C	0.8717	0.0603	0.1192	.9051	None
1	0	A	0.8628	0.0651	0.1192	.9051	
0	1	C	0.7950	0.0677	0.5056	.6131	None
0	1	A	0.7573	0.0748	0.5056	.6131	
1	1	C	0.8207	0.0657	-0.9614	.3363	None
1	1	A	0.8856	0.0520	-0.9614	.3363	

percent correct that would be obtained in a two-alternative forced-choice experiment, may be compared. Observers A, B, and E (laymen) performed at approximately the same level, with areas A_z in the 0.75 to 0.90 range. Observers B and C (physicians) were consistently lower than the other observers by a slight amount, with areas A_z of approximately 0.6 to 0.8 and a slightly larger standard deviation in their results. Informal discussions suggested that the physi-

cians were more cognizant of the possible presence of normal structure which might masquerade as the lesion. While we attempted to avoid confusing normal structure in the selection of the lesion sites, this effort was not completely satisfactory. We are not inclined to assign any significance to this difference in performance at this time, since this experiment was not designed to compensate for the variance in experience and knowledge of the observers. The observer perfor-

Table 9. Results for Observer E. No Data Pooling

Observer D							
Type	Site	Proc	Area	σ	n_e	P Value	Preference
Lungs							
0	0	C	0.7903	0.0712	-0.6594	.5097	None
0	0	A	0.8300	0.0606	-0.6594	.5097	
1	0	C	0.8693	0.0526	-0.7447	.4565	None
1	0	A	0.9052	0.0489	-0.7447	.4565	
0	1	C	0.8572	0.0567	-0.4113	.6808	None
0	1	A	0.8786	0.0535	-0.4113	.6808	
1	1	C	0.9211	0.0401	0.5636	.5730	None
1	1	A	0.9019	0.0488	0.5636	.5730	
Mediastinum							
0	0	C	0.8711	0.0570	0.9078	.3640	None
0	0	A	0.8099	0.0681	0.9078	.3640	
1	0	C	0.8851	0.0515	0.2947	.7682	None
1	0	A	0.8683	0.0565	0.2947	.7682	
0	1	C	0.8282	0.0638	1.3409	.1799	None
0	1	A	0.7397	0.0759	1.3409	.1799	
1	1	C	0.7905	0.0711	0.1939	.8462	None
1	1	A	0.7794	0.0752	0.1939	.8462	

mances on a relative scale (ie, the preference in modalities) are consistent across all observers.

Pooled Results

Given the good observer performances and relatively small standard deviations in the results, we analyzed the results again after pooling the data across lesion sites. That is, for a particular observer, the data for a given field and lesion type was combined before analysis. The summarized results are shown in Table 4. Again, neither modality is clearly preferred. Out of four possible results for each observer, Observers A, B, and D showed no preferences, Observer E showed one preference, and Observer C showed three preferences. Neither modality was preferred at the 2.5σ level in any of the pooled results. It should be noted that there are more preferences expressed in these data than in the unpooled data: four out of ten. However, these preferences were again evenly split between the two modalities. Examination of the full data shows a lower standard deviation for each area estimate, consistent with the larger number of observations for each experiment due to the data pooling. It would appear that the pooling of the data is statistically valid.

DISCUSSION AND CONCLUSIONS

In interpreting the results, it is wise to keep the following points in mind. First, the task that the observers performed was very difficult; the need to make the stimulated lesions subtle enough to be ambiguous in this test means that they would be essentially undetectable in clinical practice. Second, the criterion that the areas under the ROC curve differ by 1.5σ is a relatively weak one; such an eventuality might occur by chance one time in eight. Only a small number of results were significant at this level, and none were significant at the more stringent 2.5σ level; those results that did show a preference were evenly divided between AHE and CLAHE. Finally, the performances of AHE and CLAHE seem to be equally good in both the lung and mediastinum fields of the images, indicating that the presence of differing kinds of normal structures are handled equally well by the two modalities.

Thus, the authors assert that in this case, it appears to be valid to use either AHE or CLAHE, provided that the clipping level parameter of CLAHE is carefully chosen for the image type and scanner characteristics. The need to choose

this parameter is an added complication of CLAHE which is unnecessary with AHE, but it may well be worthwhile given the superior subjective image quality in homogeneous regions of CLAHE. Further work is needed to see if this result carries over to digital chest radiography; this is especially important since the subjective image quality of CLAHE in digital chest is preferred by physicians over that of AHE. Experiments will begin shortly in the authors' laboratory on a similar experiment comparing the current results with those obtained using digital chest radiographs.

Finally, the difference in the performances of physicians and laymen for this experiment is not significant as it stands, but may warrant further investigation of the validity of this type of testing. These performances may indicate a lack of realism in emulating the process of diagnosis, rather than suggesting that one can use laymen for this type of study. Even though the task here is one that emphasizes the fundamental contrast detection ability of the eye as opposed to the trained diagnostic response of the physician, it appears that the knowledge of the possible presence of interfering structures in the image may have a noticeable effect on observer performance.

The following is a summary of this report: (1) CLAHE is a contrast enhancement method which offers subjectively superior image quality; (2) In the current work, no difference was found in the ability of CLAHE and AHE to depict subtle contrast in an image; (3) The use of CLAHE rather than AHE may well be valid for the CT modality, provided that the clipping level parameter of CLAHE is carefully chosen; and (4) The extension of these results to other digital imaging modalities requires further investigation.

ACKNOWLEDGMENT

The authors gratefully acknowledge an equipment grant from Philips Medical Systems, Inc. which was provided through their Pegasus Project. We thank Edward V. Staab for supplying the CT image data which the authors used and his material assistance in the preparation of the test set. Steven M. Pizer and R. Eugene Johnston assisted in the experimental design; Steve Pizer also supplied a working implementation of the CLAHE algorithm. John D. Austin helped establish the proper parameters for the CLAHE clipping level. Our thanks also go to R. Gilbert Jost, G. James Blaine, Jerome R. Cox, Jr., and Ronald G. Evens, for their work in establishing the Electronic Radiology Laboratories at the Mallinckrodt Institute of Radiology, without which this work would not have been possible.

REFERENCES

1. Ketcham DJ, Lowe R, Weber W: Real-time enhancement techniques, in *Seminars on Image Processing*. Pacific Grove, CA, Hughes Aircraft Company, 1976, pp 1-6
2. Hummel R: Image Enhancement by Histogram Transformation. *Comput Graphics Image Process* 6:184-195, 1977
3. Pizer SM, Zimmerman JB, Staab EV: Adaptive grey level assignment in CT scan display. *J Comput Assist Tomogr* 8:300-308, 1984
4. Ter Haar Romeny BM, Pizer SM, Zuiderveld K, et al: Recent Developments in AHE, in *Exhibit at 71st Scientific Assessment and Annual Meeting*. Chicago, Radiological Society North America, 1985
5. Zimmerman JB, Pizer SM, Staab EV, et al: An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement. *IEEE Transact Medical Imaging* 7:304-312, 1988
6. Pizer SM, Amburn EP, Austin JD, et al: Adaptive Histogram Equalization and its Variations. *Comput Vision Graphics Image Process* 39:355-368, 1987
7. Johnston RE, Zimmerman JB, Rogers DC, et al: Perceptual Standardization. *Proceedings of the SPIE (PACS III)* 536:44-49, 1985
8. Green DM, Swets JA: *Signal Detection Theory and Psychophysics*. New York, Robert E. Krieger, 1974
9. Metz CE, Wang P-L, Kronman HB: A new approach for testing the significance of differences between ROC curves measured from correlated data, in Deconinck F (ed): *Information Processing in Medical Imaging (Proceedings of the 8th Conference)*. The Hague, Martinus Nijhoff, 1984