Radiol Phys Technol (2013) 6:35–41
DOI 10.1007/s12194-012-0166-1

Basic concepts and development of an all-purpose computer interface for ROC/FROC observer study

Junji Shiraishi · Daisuke Fukuoka · Takeshi Hara · Hiroyuki Abe

Received: 14 March 2012 / Revised: 16 June 2012 / Accepted: 18 June 2012 / Published online: 5 July 2012

Abstract In this study, we initially investigated various aspects of requirements for a computer interface employed in receiver operating characteristic (ROC) and free-response ROC (FROC) observer studies which involve digital images and ratings obtained by observers (radiologists). Secondly, by taking into account these aspects, an all-purpose computer interface utilized for these observer performance studies was developed. Basically, the observer studies can be classified into three paradigms, such as one rating for one case without an identification of a signal location, one rating for one case with an identification of a signal location, and multiple ratings for one case with identification of signal locations. For these paradigms, display modes on the computer interface can be used for single/multiple views of a static image, continuous viewing with cascade images (i.e., CT, MRI), and dynamic viewing of movies (i.e., DSA, ultrasound). Various functions on these display modes, which include windowing (contrast/level), magnifications, and annotations, are needed to be selected by an experimenter corresponding to the purpose of the research. In addition, the rules of judgment for distinguishing between true positives and false positives are an important factor for estimating diagnostic accuracy in an observer study. We developed a computer interface which runs on a Windows operating system by taking into account all aspects required for various observer studies. This computer interface requires experimenters to have sufficient knowledge about ROC/FROC observer studies, but allows its use for any purpose of the observer studies. This computer interface will be distributed publicly in the near future.

Keywords Receiver operating characteristic (ROC) analysis · FROC analysis · Observer study · Computer interface

J. Shiraishi (✉)
Faculty of Life Sciences, Kumamoto University, 4-24-1 Kuhonji Chuo-ku, Kumamoto, Kumamoto 862-0976, Japan
e-mail: j2s@kumamoto-u.ac.jp

D. Fukuoka
Technology Education, Faculty of Education, Gifu University, 1-1 Yanagido, Gifu, Gifu 501-1193, Japan

T. Hara
Division of Regeneration and Advanced Medical Sciences, Department of Intelligent Image Information, Graduate School of Medicine, Gifu University, 1-1 Yanagido, Gifu, Gifu 501-1193, Japan

H. Abe
Department of Radiology, The University of Chicago, 5841 South Maryland Avenue MC2026, Chicago, IL 60637, USA

1 Introduction

Receiver operating characteristic (ROC) analysis has been introduced in medical fields for evaluation of the diagnostic accuracy of imaging modalities and/or various imaging procedures [1–3]. Because diagnostic accuracy was estimated by use of ROC curves and/or the area under the ROC curve (AUC), which were obtained from the data of ROC observer studies, the statistical theory for estimating ROC curves has been a major topic among researchers for investigating ROC analysis [4–7]. At present, a number of types of software for ROC [3, 8], including locational ROC (LROC) [9], free-response ROC (FROC) [10], alternative FROC (AFROC) [11], and jackknife AFROC [12] (JAFROC), are available for use. However, even though the ROC/FROC curve-fitting programs and software can be available to researchers, the software does not care about any biases due to incorrect experimental design of observer studies. Therefore, no valuable conclusion would be obtained if an experimental design was wrong and thus insufficient data would be provided. For example, one bias due to the reading order of cases in an observer study can produce incorrect experimental data if the order was determined incorrectly by not taking into account reading order effects [13]. Although a number of ROC observer studies have been performed, there was no standard for performing such studies precisely because of variations in the subjects of research such as diseases, types of modalities, and the tasks required for radiologists [3, 8, 13].

In this study, we aimed to develop an all-purpose computer interface for ROC/FROC observer study to assist researchers who plan to perform observer studies, but who are not familiar with designing an experiment for such study. We initially investigated various aspects of the requirements for a computer interface utilized in ROC and free-response ROC (FROC) observer studies which involve digital images and ratings obtained by observers (radiologists). Secondly, by taking into account these aspects, we developed an all-purpose computer interface to be utilized for these observer performance studies.

2 Requirements for computer interface utilized in observer study

Basically, all ROC/FROC observer studies can be classified into three paradigms, namely, one rating for one case without identification of a signal location (i.e., for ROC study), one rating for one case with identification of a signal location (i.e., for LROC study), and multiple ratings for one case with identification of signal locations (i.e., for FROC, AFROC, and JAFROC study). For these three paradigms, common requirements for the computer interface are (1) to display images (either static or dynamic) on single/multiple monitors, (2) to provide to observers additional information such as vital signs, age/gender, and marks obtained by computer-aided diagnosis (CAD), (3) to obtain ratings for cases from observers, (4) to record ratings and reading times, (5) to allow modification of the reading order of images, (6) to allow users to distinguish true positives from false positives by use of specified rules and the gold standard, (7) to output rating data with specific formats, and (8) to provide and obtain informed consent from observers electrically. Therefore, one computer interface can treat all three paradigms for ROC/FROC observer study, if each function for the requirements can be adjustable corresponding to the method used in each paradigm.

The static images include those of computed radiography (CR), digital radiography (DR), each slice image of CT and MR that can be displayed continuously in cine mode, dynamic images including movie files of digital subtraction images (DSA), and ultrasonography (US). Except for movie files in MPEG and AVI formats, a majority of medical images would be distributed to the picture archiving and communication system (PACS) in a standard DICOM format. Therefore, the computer interface must treat images in DICOM format. For viewing of images on the monitor, the computer interface needs to have various functions, which include windowing (contrast/level), magnification, and annotations. Basically, these functions will be selected by an experimenter corresponding to the purpose of the research.

We assumed that ratings for ROC/FROC observer study should be obtained in a continuous manner, because it is nearly impossible to obtain a reliable estimate of the AUC by use of ordinal 5 category scales [8]. For rating of the possibility for the presence of an abnormal lesion with its location, a rating scale needs to appear close to the point marked as a location of an abnormal lesion. If a location of the rating scale is fixed and observers are required to move their mouse to this rating scale every time, this will create a huge time loss for observers. In the experimental design of an observer study, the experimenter should try to reduce useless time and useless procedures for the observers as much as he/she can.

Rating data during an observer test should be recorded in real time. If the data were deleted due to computer or human error, it is impossible to recover the motivation of the observer for again reading cases with high attention.

The reading order of cases should be controlled to keep reproducibility in the observer studies. Therefore, a randomized function for determining the reading order is not necessary for the computer interface. For running the computer interface, a data sheet for reading of cases, which includes the image file name, directory names of image files, image matrix size, image pixel size, reading order of the cases, and other information would be prepared in advance.

The rules of judgment for distinguishing between true positives and false positives are an important factor for estimating diagnostic accuracy in an observer study. When the observer study is performed as a detection task with identification of the location of an abnormal lesion, the distance between a mark pointed out by an observer and the location of the center of an actual lesion is used as a measure for classification. If two or more responses were made within a limited region (i.e., diffuse lung disease) and all responses correctly identified an abnormal lesion, all responses were considered as true positives in our software. However, for ROC/FROC analysis, they needed to be gathered and to be considered as a single true-positive for one abnormal lesion. The rule for this classification would

Table 1 An example of a case-setting file used for the computer interface

Image file L[a, c]	Image file R[b, c]	Truth[d]	Modality name[e]	Case number[f]	Treatment number[g]	Comment left[h]	Comment right[h]
C:\Image\case01L.dcm		P	Without	1	1	Original	
C:\Image\case01L.dcm	C:\Image\case01R.dcm	P	With	1	1	Original	With CAD
C:\Image\case02L.dcm		N	Without	2	1	Original	
C:\Image\case02L.dcm	C:\Image\case02R.dcm	N	With	2	1	Original	With CAD
C:\Image\case03L.dcm		N	Without	3	1	Original	
C:\Image\case03L.dcm	C:\Image\case03R.dcm	N	With	3	1	Original	With CAD

The file needs to be created in CSV format

[a] Image file name which is displayed on the left monitor

[b] Image file name which is displayed on the right monitor

[c] Image formats available in this software are BMP, JPEG, TIFF, and DCM (DICOM). For multi images (i.e., CT, MR) input directory name where all images are archived: Ex. \Folder name*.dcm ("*.dcm" meant all DICOM images archived in the directory "Folder name.")

[d] Truth (gold standard) of images displayed: For ROC, Use "P or Positive" for positive cases and "N or Negative" for negative cases. For LROC, Use "P-location1" for positive cases and "N-location1" for negative cases. For FROC, Use "P-locations" for positive cases and "L-locations" for negative cases

[e] Modality names for comparison: Ex. "without CAD versus with CAD", "CT versus MR", and "CR versus SF"

[f] Case number or patient number. The same case needs to have the same case number

[g] When ROC/FROC study is performed for two or more treatments in one reading session, use treatment number to classify rating data into each treatment

[h] Comments shown in an image displayed on left or right monitor. Use this function for checking cases and reading order in advance

be determined by taking into account the characteristics of abnormal lesions such as an average size and shapes.

Because a variety of ROC/FROC software are available and the statistical backgrounds are already established, the computer interface does not need to include any functions for ROC/FROC curve fitting and also a statistical procedure for estimating statistical significance of a difference between two conditions which were obtained by the observer study. Instead, it is very convenient for experimenters, who would use the computer interface, to include functions to output rating data in specific formats which can be used as an input file for specific software without any modifications.

In the observer study, informed consent from the observers should be obtained because their rating data and the results obtained from the observer study were personal information for each observer. Moreover, his/her diagnostic performance can be used for judging his/her diagnostic skill which often relates to her/his salary directly. To minimize the time for an observer study, it is convenient to include a specific function for providing and obtaining informed consent in the computer interface.

3 Computer interface for ROC/FROC observer study

For developing the computer interface, we assumed that this interface would be used only for viewing of digital images, including digitized analog images, and for rating of these images by a human observer. The computer interface was developed by use of Windows Visual Studio 2008. This computer interface is able to run on any PC with Windows operating system XP or a later version.

In order to use one computer interface for three ROC/FROC paradigms, we used a case-setting file for changing display modes as well as rating methods. Table 1 shows one example of a case-setting file for the computer interface we developed. This setting file needed to be prepared for one reading session of an ROC/FROC observer study; one row of the file indicates one reading for one case with one or two images. We adopted a maximum of two views for one case as a default. If only one view for one reading was required, the second column was left blank. For viewing static images, which includes JPEG, BMP, TIFF, and DCM (DICOM) formats, an image file name for each case was entered first and (if necessary) a second column with a certain reading order which should be determined in advance by taking into account a reading order effect. For viewing of stacked images such as CT and MRI (DCM format only), an image directory name of the stack images was entered instead of an image file name. In addition, for viewing movies in AVI and MPEG formats, movie file names were used instead of image file names. For DICOM-format images with various kinds of data compression, such as "Lossy JPEG", "Lossless JPEG", and "Run length encoding (RLE)", DICOM Toolkit (DCMTK)[1] needs to be installed in advance.

[1] Available at http://dicom.offis.de/dcmtk.php.en.

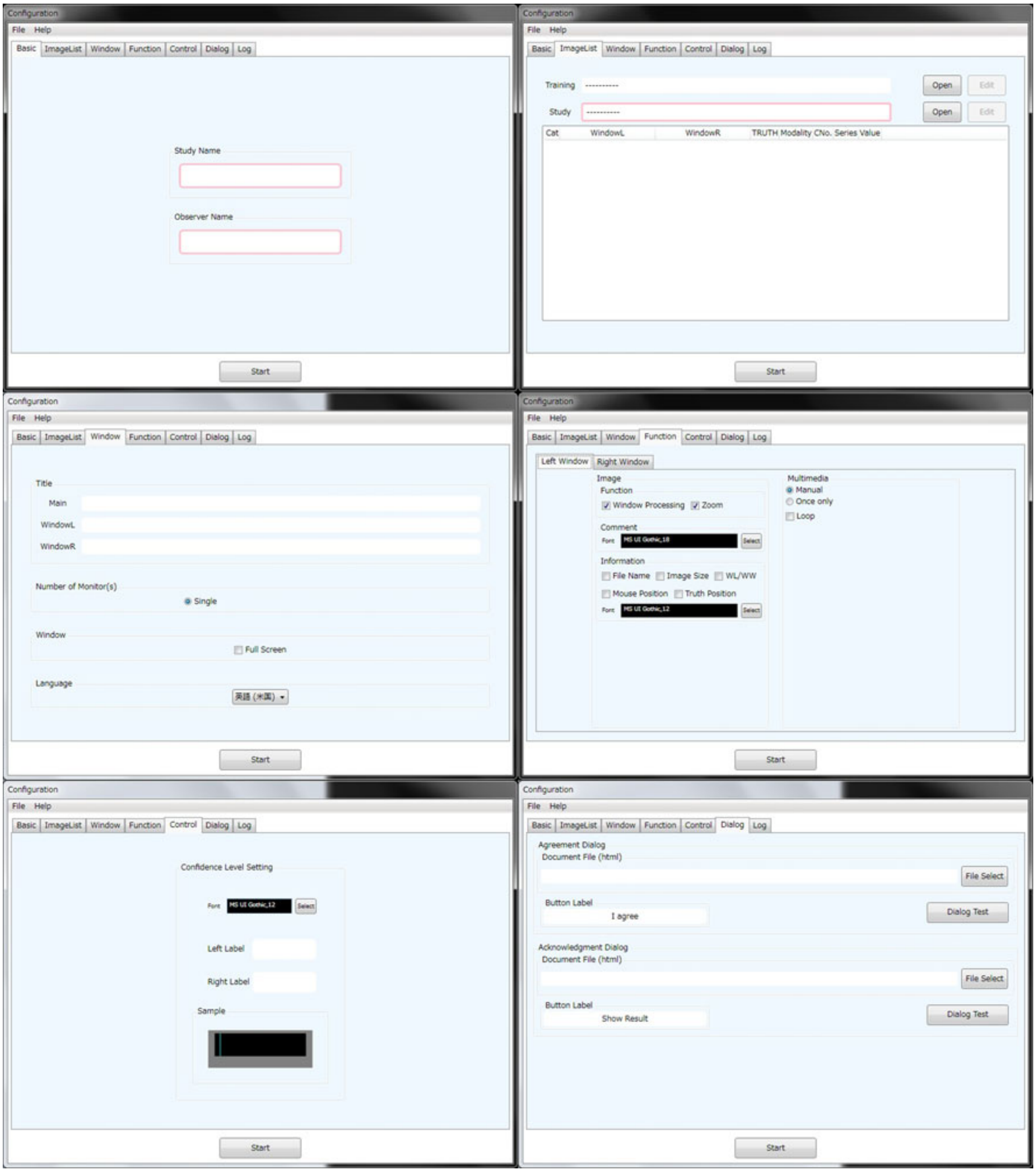

Fig. 1 The six configuration windows for managing various items and parameters in the computer interface. *Left upper* for input study name and observer name, *right upper* for reading case-setting files for training and testing, *left middle* for input titles displayed at the top of monitor and for selecting the number of monitors, *right middle* for selecting image display functions and information on the screen, *left bottom* for input captions on the both ends of a rating bar, *right bottom* for setting dialogs for agreement of informed consent and acknowledgment that appear at the end of all ratings

The third column of the setting file classifies the observer study into three ROC/FROC paradigms. When an ROC paradigm which requires an observer to identify one signal for one case without its location is employed for the observer study, a truth of the case (positive or negative) is entered in this column. For LROC study requiring one signal for one case with its location, a code of "P-location1" or "N-location1" is used. For FROC study with multiple signals, another code of "P-locations" or "N-locations" is used. By using these codes, the computer interface distinguishes the paradigms automatically.

For sequential rating in an ROC/FROC study (i.e., without and with computer aid), each case requires two rows for two ratings. The first row is for viewing and initial rating of a case only with an image on the left viewer, and the second row is for viewing and sequential rating of a

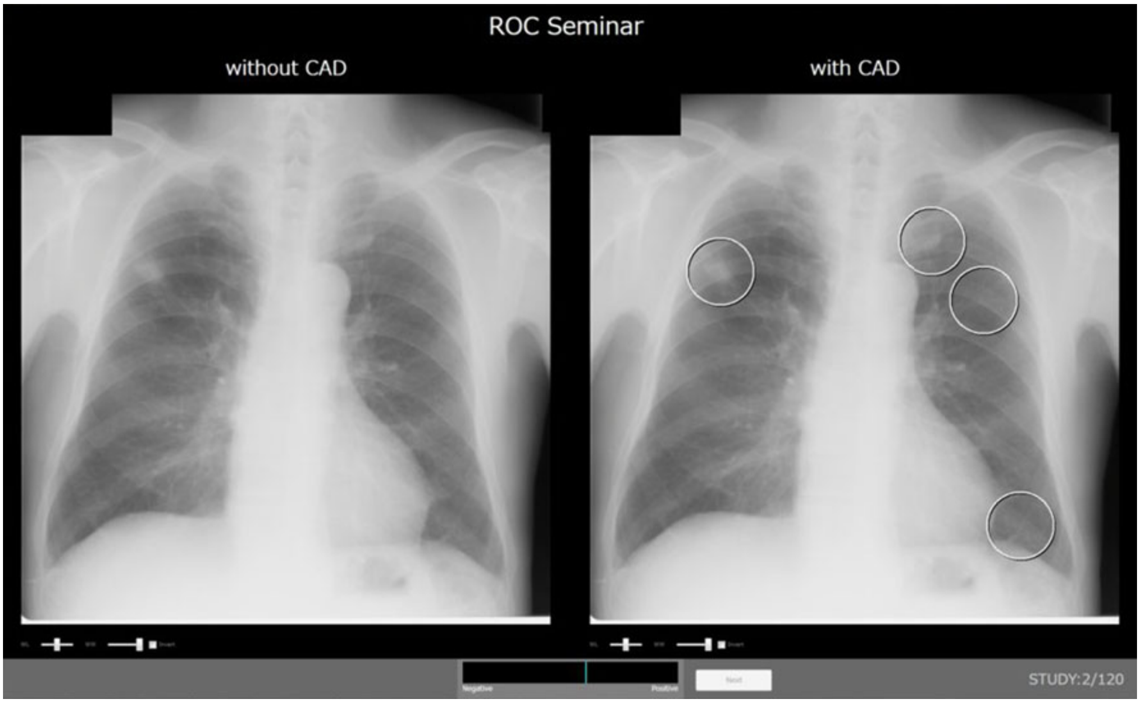

Fig. 2 Screen shot of the computer interface for viewing two images in the ROC observer study. *Left window* displays an original chest radiograph, and the *right* displays that with CAD marks (*circle*) for lung nodule candidates. The observer could adjust windowing by use of *slide bars* located at the *left bottom* of each image

case with two images on both views. In this type of observer study, two modality names compared in this ROC/FROC study are "without" and "with", and they are entered in the fourth column. Case numbers in the fifth column will be used for checking the correspondence between the modalities and for producing an output file with the same order of cases for each modality.

To set various parameters, we adopted configuration windows in the computer interface. These configuration windows allowed us to use a number of common functions easily for the three ROC/FROC paradigms. Figure 1 shows six configuration windows in our computer interface. They are used for managing the following items: (1) study and observer names, (2) setting file names for training and testing cases, (3) a main title and subtitles for each view, (4) the number of monitors used in the study, (5) languages (English or Japanese) for display, (6) functions for windowing and magnification, (7) information displayed on the monitor such as file name, image matrix size, window width/window level, coordinates of a mouse, and the location of truth, (8) multimedia mode for movies, (9) annotations for a rating bar, and (10) dialog settings for showing informed consent and a message for acknowledgment.

Figure 2 shows a screen shot of our computer interface which was used for viewing of two images to evaluate the clinical utility of CAD in the ROC observer study. A black bar at the bottom of the screen is a rating bar with a line which moves left and right within the bar corresponding to the location of a mouse. In the ROC observer study, the observer was not allowed to move to the next case without rating a present case.

Figure 3 shows how the observer rated his/her confidence level by use of a mouse in the FROC paradigm. Once the observer clicks a mouse on the point where he/she believes that there is an abnormal lesion, a dialog with rating bar appears close to the point clicked. Then, after the observer decides on his/her confidence level on the rating bar, the dialog will disappear, and a circle mark with the rating score will be indicated on the point clicked. For stack images for CT or MRI, a slide bar for moving between slices is indicated at the bottom of the screen, as shown in Fig. 4.

Although we did not plan to include any ROC curve-fitting estimations except for a simple line graph of an ROC curve, our computer interface could produce experimental results with specific formats required for ROC (Rockit [14]) and JAFROC (JAFROC [15]) software, which is available publicly. Thus, the users of this computer interface can use their data directly on that ROC/FROC software.

4 Discussion

The computer interface for ROC/FROC study is very important for reducing the time of analysis and for

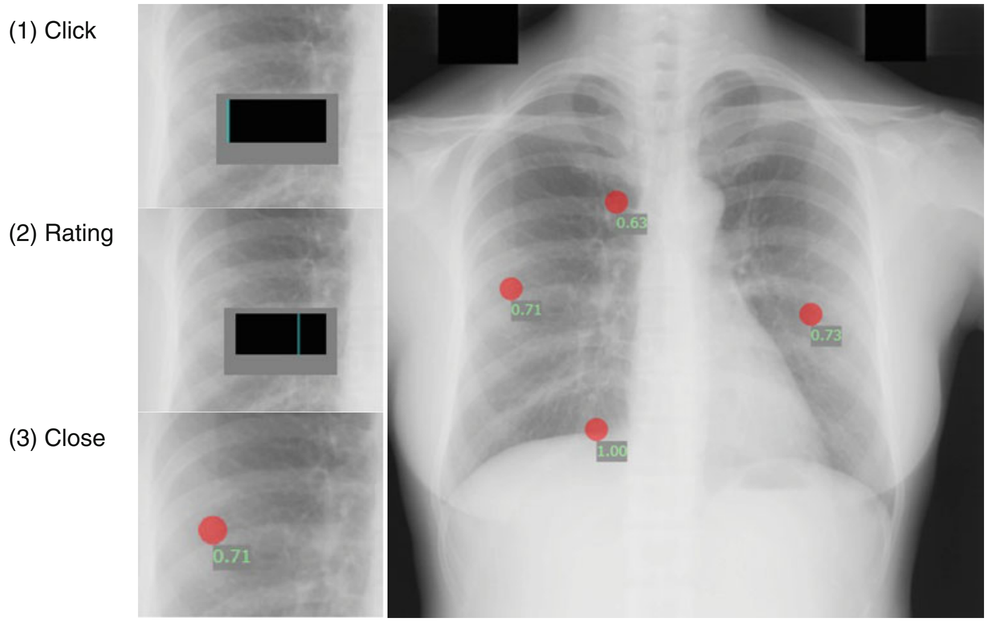

Fig. 3 Screen shot of the computer interface for rating an abnormal lesion with its location in an FROC paradigm. When the observer clicked at a point where he/she believed that there was a possible lung nodule (1) a small window with a rating bar appeared close to the point, (2) the observer determined a confidence level for the presence of a lung nodule by clicking at the rating bar, and (3) the small window disappeared and a *red circle* for indicating the observer's mark with a rating score was displayed

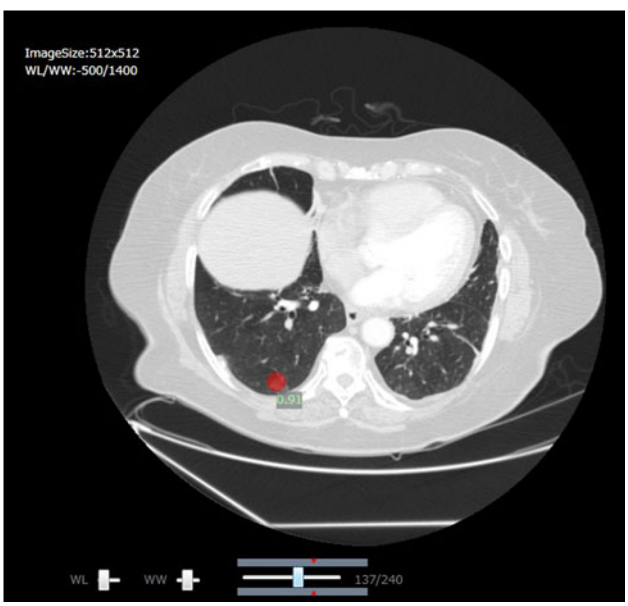

Fig. 4 Screen shot of the computer interface for viewing CT images with stack mode. All slice images could be viewed with cine mode by use of a *slide bar* located at the *bottom* of images. Ratings with location pointed out were made with the same manner as that for FROC study shown in Fig. 3. When the observer marked a possible lesion on the slice image, the location of the slice image among a series of stacked slice images was shown with a *red mark* at the *top* and *bottom* of the *slide bar*

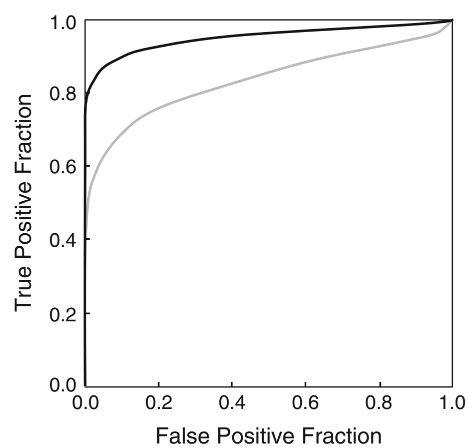

Fig. 5 ROC curves in which rating data were obtained by use of the computer interface developed. This ROC study was performed for evaluating the detectability of image unsharpness due to the movement of an object by use of a diagnostic liquid crystal display (LCD) (*black solid line*) and preview LCD (*gray solid line*)

improving the robustness of data obtained in an observer study. Therefore, special computer interfaces for specific purposes were developed and used by individual institutions where academic computer scientists are available. However, a majority of researchers who tried to perform ROC/FROC observer studies did not have such computer interfaces or the skill to develop them. One purpose in this study was not only developing a computer interface for ROC/FROC observer study, but also diffusing appropriate knowledge for ROC/FROC observer studies.

To maintain easy use for experimenters and high credibility of the data obtained, this computer interface was developed only for the traditional fully crossed design with paired cases and paired observers [8]. In addition, there was no connectivity of the data for two or more reading sessions performed by use of this computer interface. Therefore, if two reading sessions with independent ratings of one observer were performed by this computer interface, the experimenter would need to combine two rating data sets to estimate a statistically significant difference between two ROC curves.

The utility of this computer interface has been tested in ROC seminars held by the Japanese Society of Radiological Technology (JSRT) for teaching basic theory and experimental design of observer studies in the last 2 years. We have modified this computer interface by taking into account a number of opinions provided by seminar participants and our experience obtained at the seminar. As a result, one seminar participant performed one ROC study by use of this computer interface after attending and publishing her research as an original paper [16]. Figure 5 shows the first ROC curves which were obtained by use of our computer interface and published in the *Journal of Radiological Physics and Technology*.

5 Conclusions

We developed a computer interface by taking into account various factors which need to be considered for appropriate ROC/FROC observer studies.

Acknowledgments We gratefully acknowledge support from the Japanese Society of Radiological Technology (JSRT) Research Grant (2020, and 2011). This work is also supported in part by Scientific Research Grant (C), 22611014, 2010.

References

1. Goodenough DJ, Rossmann K, Lusted LB. Radiographic applications of receiver operating characteristic (ROC) curves. Radiology. 1974;110:89–95.
2. Metz CE. ROC methodology in radiologic imaging. Invest Radiol. 1986;21:720–33.
3. ICRU Report 79. Receiver operating characteristic analysis in medical imaging. Oxford University Press, Oxford; J ICRU 2008;8(1).
4. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143:29–36.
5. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8:283–98.
6. Metz CE, Herman BA, Shen J-H. Maximum-likelihood estimation of ROC curves from continuously-distributed data. Stat Med. 1998;17:1033–53.
7. Pesce LL, Metz CE. On the convexity of ROC curves estimated from radiological test results. Acad. Radiol. 2010;256(1):29–31.
8. Shiraishi J, Pesce L, Metz CE, Doi K. Experimental design and data analysis in receiver operating characteristic studies: lessons learned from reports in Radiology from 1997 to 2006. Radiology. 2009;253:822–30.
9. Swensson RG. Unified measurement of observer performance in detecting and localizing target objects on images. Med Phys. 1996;23:1709–25.
10. Chakraborty DP, Breatnach EA, Yester MV, Soto B, Barnes GT, Fraser RG. Digital and conventional chest imaging: a modified ROC study of observer performance using simulated nodules. Radiology. 1986;158:35–9.
11. Chakraborty DP, Winter LH. Free-response methodology: alternative analysis and a new observer-performance experiment. Radiology. 1990;174:873–81.
12. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. Med Phys. 2004;31:2313–30.
13. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. Invest Radiol. 1989;24: 234–45.
14. Metz ROC Software web site in Department of Radiology, The University of Chicago. http://xray.bsd.uchicago.edu/krl/roc_soft6.htm. (Accessed 14 Mar 2012)
15. Dr. Dev's Chakraborty's FROC web site http://www.devchakraborty.com/index.php. (Accessed 14 Mar 2012)
16. Tanaka R, Shiraishi J, Takamori M, Watari C. ROC analysis for evaluating the detectability of image unsharpness due to the patient's movement: phantom study comparing preview and diagnostic LCDs. J Radiol. Technol. 2011;67(7):772–8.