



ELSEVIER

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

## Physica Medica

journal homepage: <http://www.physicamedica.com>

## Original Paper

## Objective assessment of low contrast detectability in computed tomography with Channelized Hotelling Observer

Damien Racine <sup>a,b,\*</sup>, Alexandre H. Ba <sup>a</sup>, Julien G. Ott <sup>a</sup>, François O. Bochud <sup>a</sup>, Francis R. Verdun <sup>a</sup><sup>a</sup> Institute of Radiation Physics, Lausanne University Hospital, 1 Rue du Grand-Pré, 1007 Lausanne, Switzerland<sup>b</sup> University Joseph Fourier Grenoble I, 621 Avenue Centrale, 38041 Saint-Martin-d'Hères, France

## ARTICLE INFO

## Article history:

Received 27 August 2014

Received in revised form 24 August 2015

Accepted 23 September 2015

Available online 26 October 2015

## Keywords:

Computed tomography (CT)

Channelized Hotelling Observer model

Iterative reconstruction

Low contrast detectability

## ABSTRACT

**Purpose:** Iterative algorithms introduce new challenges in the field of image quality assessment. The purpose of this study is to use a mathematical model to evaluate objectively the low contrast detectability in CT. **Materials and methods:** A QRM 401 phantom containing 5 and 8 mm diameter spheres with a contrast level of 10 and 20 HU was used. The images were acquired at 120 kV with CTDI<sub>vol</sub> equal to 5, 10, 15, 20 mGy and reconstructed using the filtered back-projection (FBP), adaptive statistical iterative reconstruction 50% (ASIR 50%) and model-based iterative reconstruction (MBIR) algorithms. The model observer used is the Channelized Hotelling Observer (CHO). The channels are dense difference of Gaussian channels (D-DOG). The CHO performances were compared to the outcomes of six human observers having performed four alternative forced choice (4-AFC) tests.

**Results:** For the same CTDI<sub>vol</sub> level and according to CHO model, the MBIR algorithm gives the higher detectability index. The outcomes of human observers and results of CHO are highly correlated whatever the dose levels, the signals considered and the algorithms used when some noise is added to the CHO model. The Pearson coefficient between the human observers and the CHO is 0.93 for FBP and 0.98 for MBIR.

**Conclusion:** The human observers' performances can be predicted by the CHO model. This opens the way for proposing, in parallel to the standard dose report, the level of low contrast detectability expected. The introduction of iterative reconstruction requires such an approach to ensure that dose reduction does not impair diagnostics.

© 2015 Associazione Italiana di Fisica Medica. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

Computed Tomography (CT) represents about 10% of all radiological examinations in Switzerland but about 70% of the collective effective dose [1]. Since no dose limit is applicable for patients, a first attempt to optimize patient exposure in radiology was the introduction of diagnostic reference levels (DRL) [2]. This concept makes it possible to identify situations in which dose level is beyond the examinations' common practices [3]. Despite this, a focus restricted to dose alone is highly insufficient in a framework of optimization between the radiological risk and image information. A surrogate to assess image information is the measurement of physical metrics such as image noise, spatial resolution, and contrast-to-noise ratio. However, these concepts are only valid for

linear systems or algorithms. The introduction of iterative reconstructions into CT introduces new challenges in the field of image quality assessment since most of the standard metrics are no longer applicable [4–6]. State-of-the art medical image quality assessment takes another approach by defining image quality as how well the desired information for a given task can be extracted from an image [7]. Simple binary tasks, such as discrimination between the presence and absence of pathology among a given population, are usually characterized by the use of Receiver Operating Curve (ROC) studies [8,9]. Unfortunately these studies are time-consuming [10] and difficult to implement in practice. Therefore, it is necessary to develop tools such as model observers [11,12] which make it possible to quantify image quality using a similar paradigm but in much simpler ways [13,14].

It has been shown that mathematical model observers, such as the Non Pre-whitening model with Eye-filter (NPWE) or Channelized Hotelling Observer (CHO) [15], can predict the capacity of human vision to detect low contrast targets. The advantage of this approach is that it enables testing the whole imaging chain [13] but it requires a substantial amount of data to be statistically robust.

\* Corresponding author. Institute of Radiation Physics, Lausanne University Hospital, 1 Rue du Grand-Pré, 1007 Lausanne, Switzerland. Tel.: +41 21 31 49 249; fax: +41 21 314 8299.

E-mail address: [Damien.Racine@chuv.ch](mailto:Damien.Racine@chuv.ch) (D. Racine).

The purpose of this work was to investigate if the approach of image quality assessment by means of the CHO model observer could be applicable in routine practice in order to facilitate a control of image quality while reducing patient exposure. Thus, our work used a limited number of acquisitions with a specifically designed phantom that allowed the sampling of several realizations per slice. The use of such a methodology could then produce an image quality indicator that could be analyzed with the standard dose report. We also compared the results of the CHO model used with the outcome of human observers while varying several acquisition and image reconstruction parameters on an abdominal phantom.

## Materials and methods

### Data acquisitions

The QRM 401 phantom (Moehrendorf, Germany; see Fig. 1) was imaged on a GE 750 HD scanner (GE Healthcare, USA). The phantom is made of muscle, liver, spleen and bone (vertebrae) tissue equivalents at 120 kV. Two dedicated moduli (moduli A and B) can be inserted into the phantom shell. Modulus A includes spheres of various diameters: 8, 6, 5, 4 and 3 mm; each size having a contrast of 10 and 20 HU relative to the background at 120 kV. This phantom enables the assessment of in-plane and axial low-contrast detectability. In this study we only considered the in plane low contrast detectability for two sphere diameters: 5 and 8 mm. The spheres of each contrast level were positioned within the same slice position. Modulus B is homogeneous and was used to produce images free from low contrast target.

Acquisitions were performed at 120 kVp in helical mode (0.984 as pitch factor). In order to get  $CTDI_{vol}$  values of 5, 10, 15 and 20 mGy the X-ray tube rotation time was kept constant (1 s) while varying the tube current. Indicated  $CTDI_{vol}$  values were verified as described in the IEC 60601-2-44. The X-ray beam collimation geometry was  $64 \times 0.625$  mm (leading to a total X-ray collimation at isocenter of 40 mm) and the reconstructed scan FOV was  $320 \times 320$  mm using a  $512 \times 512$  matrix size. Slices were reconstructed with a nominal thickness of 2.5 mm and a slice interval of 2.5 mm. The reconstruction filter used was the standard “body” filter provided by the manufacturer.

Images were reconstructed in the axial plane with three algorithms: filtered back-projection (FBP), and two iterative algorithms: Adaptive Statistical Iterative Reconstruction 50% (ASIR 50%) and the model-based iterative reconstruction (VEO 2.0) [16–20]. ASIR 50% was chosen since it corresponds to the option that is used for standard abdominal acquisitions in our center.

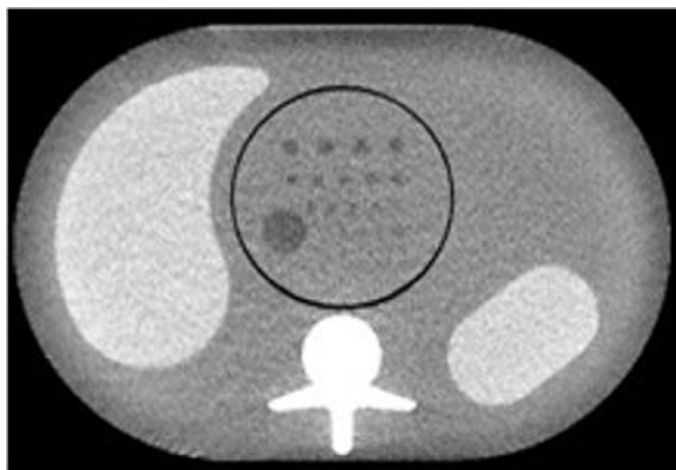


Figure 1. QRM 401 phantom acquired with MBIR at 20 mGy.

We investigated 48 different categories (3 reconstruction algorithms  $\times$  4 dose levels  $\times$  2 signal sizes  $\times$  2 contrast levels). The phantom with modulus A was (see Fig. 1) positioned at the isocenter of the CT unit and scanned ten times for each dose level, without changing its position between acquisitions. The phantom with modulus B was scanned only once for each dose level without changing its position between acquisitions.

### Generating signal-absent and signal-present images

The program was implemented with the Python programming language. The first step performed by the software was the automatic production of ROIs. For that, the vertebra which represents the reference point was searched in the central slice. Using the coordinates of this reference point, a relative reference frame was created in the image and the ROIs were created automatically based on pre-established coordinates (derived from the technical plan of the phantom). The vertebra was used as a reference because it is the most contrasted material present in the phantom, ensuring that the template matching method is robust enough even at low dose levels. For each acquisition, 4 signal-centered ROIs per signal size/contrast combination ( $22 \times 22$  pixels; pixel size of 0.625 mm) were automatically extracted from the images. Signal-absent ROIs were extracted on images of Modulus B using the same in-plane (x,y) coordinates as the images obtained with Modulus A. However, the signal-absent ROIs were extracted in successive slices whereas the signal-present ROIs were extracted at a unique longitudinal position. In the following section ROIs will be called signal-present images if the signal is present in the ROIs or signal-absent images if the signal is absent. The sample consists of 40 signal-present cases (see Fig. 2) and 520 signal-absent cases (see Fig. 3) for each category. The same images were used for model and human observers.

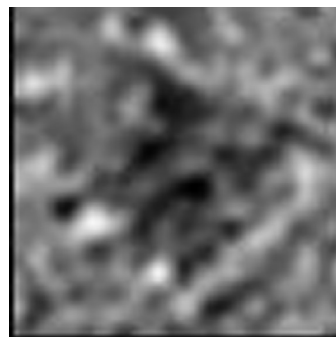


Figure 2. Signal-present image (sphere of 8 mm/20 HU).

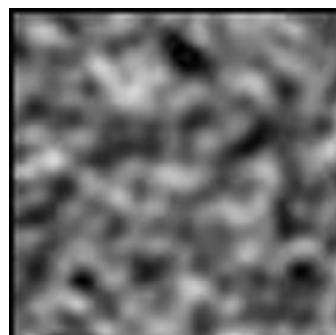


Figure 3. Signal-absent image.

### Human observer study

Six medical physicist students rated the images. These naive human observers (non radiologist) conducted four alternative forced choice (4-AFC) tests (see Fig. 4). The images were displayed on a Siemens SMM 21200P screen in conformity with DICOM 3.14 and AAPM TG18 standards [21]. The reading of images was performed in a room with an ambient light level of about 10 lux. Observers were at a distance of about 50 cm from the screen. No time limit was imposed on the observer to make their decision. During a 4-AFC study with 4 independent image combinations, three images with signal-absent and one with signal-present images were displayed and the observer was requested to identify which image contained the signal. To avoid any bias, the signal is randomly assigned to one of the 4 positions. The images were magnified to  $128 \times 128$  pixels using a bilinear interpolation to appear at their actual size on the screen. The display window level and window width were adjusted to produce the best signal and then kept constant.

Each observer tested 30 images per category for a total of 1440 images. The test was divided into three sessions (480 trials per session) in order to minimize the loss of concentration; each session interval was at least 24 hours. The first session began with a training session with 52 images acquired at high doses ( $CTDI_{vol} = 35$  and  $50$  mGy). During this session the result of each trial, “correct” or “incorrect”, was shown to the observers after they replied.

For each category, at the end of the test each observer obtained a percentage of correct responses (PC). This metric represents

the probability of correctly identifying the image containing the signal, and the higher the PC, the better the performance. PC was then converted into detectability to enable comparison between the performances of model observers with those of the human observers using Eq. (1) [15].

$$PC = \int d\lambda \phi(\lambda - d_{4AFC}) \Phi(\lambda)^3 \quad (1)$$

where  $\Phi$  is the cumulative Gaussian distribution function and  $d_{4AFC}$  represents the detectability obtained from a 4-AFC test performed by the human observers. The value of  $d_{4AFC}$  can be found using specific tabulated values [22–24].

### Model observer: Channelized Hotelling Observer

Model observers are mathematical models based on the statistical decision theory to estimate the detection performance of ideal or human observers. In this study a linear anthropomorphic CHO model observer was chosen. The decision variable which is the outcome of the model is given by the dot product between the template  $\mathbf{w}$  and the reconstructed image  $\mathbf{g}_i$  ( $i = 0$  or  $i = 1$  respectively represents signal-absent or signal-present hypothesis), expressed as an  $N \times 1$  column vector (see Fig. 5) [7,13].

$$\lambda_i = \mathbf{w}^T \cdot \mathbf{g}_i \quad (2)$$

The CHO model is a derivative of the Hotelling Observer (HO) which is too computing expensive to be used in practice [9]. To reduce the dimensionality of HO and take advantage of the spatial selectivity behavior of the human visual system, the image is first passed through  $J$  channels, where  $J$  is significantly lower than  $N$  [25]. The channel output, a scalar  $\mathbf{v}_i$ , is obtained by the dot product between the channel  $\mathbf{u}_j$  and the image  $\mathbf{g}$ .

$$\mathbf{v}_i = (\mathbf{u}_j)^T \mathbf{g} \quad (3)$$

Thus,  $\mathbf{U}$ , the matrix representation of the channel filters, is an  $N \times J$  matrix where each column is one of channel  $\mathbf{u}_j$  [26].

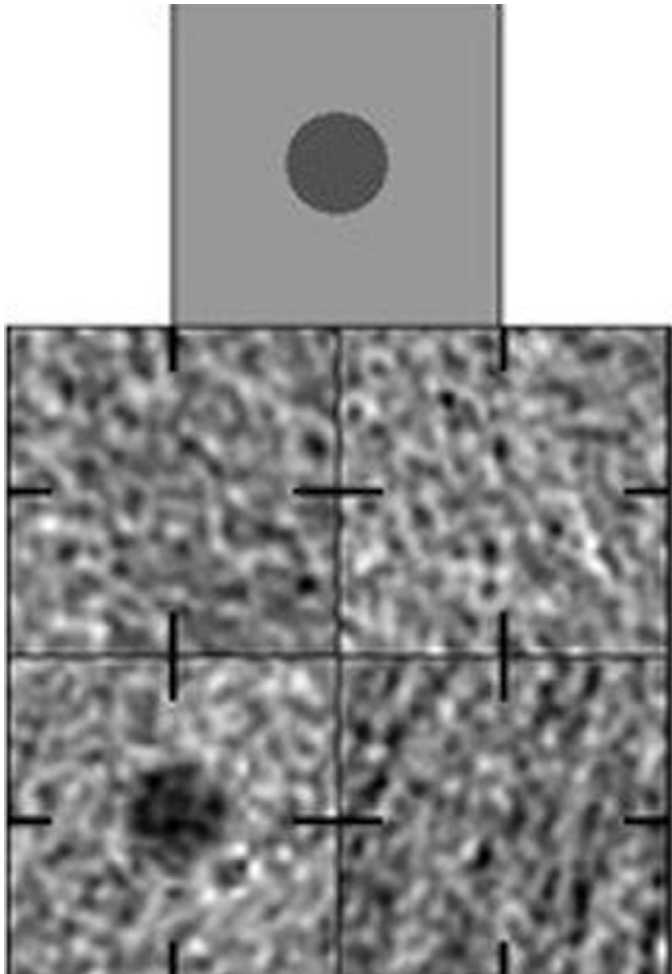


Figure 4. Example of test 4 AFC.

$$g(x, y) = \begin{bmatrix} g_{0,0} & g_{0,1} & \cdots & g_{0,m} \\ g_{1,0} & g_{1,1} & \cdots & g_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n,0} & g_{n,1} & \cdots & g_{n,m} \end{bmatrix} \rightarrow \mathbf{g} = \begin{bmatrix} g_{0,0} \\ g_{0,1} \\ g_{0,m} \\ g_{1,0} \\ g_{1,1} \\ \vdots \\ g_{n,m} \end{bmatrix}$$

$$w(x, y) = \begin{bmatrix} w_{0,0} & w_{0,1} & \cdots & w_{0,m} \\ w_{1,0} & w_{1,1} & \cdots & w_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,0} & w_{n,1} & \cdots & w_{n,m} \end{bmatrix} \rightarrow \mathbf{w} = \begin{bmatrix} w_{0,0} \\ w_{0,1} \\ w_{0,m} \\ w_{1,0} \\ w_{1,1} \\ \vdots \\ w_{n,m} \end{bmatrix}$$

Figure 5. Image and template as column vectors.

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_j]$$

For the type of targets to be detected in this study the CHO model, using the dense of difference of Gaussian (D-DOG) channels type, is recognized as a good model of human vision, and this is even with a limited number of 10 channels enabling a drastic reduction of the images required to compute the model observer outcome [27,28].

In this model, the radial profile of each frequency is given by the following formula:

$$C_j(\rho) = e^{-\frac{1}{2}\left(\frac{\rho}{Q\sigma_j}\right)^2} - e^{-\frac{1}{2}\left(\frac{\rho}{\sigma_j}\right)^2} \quad (4)$$

where  $\rho$  is the spatial frequency and  $\sigma_j$  is the standard deviation of each channel. Each  $\sigma_j$  value is given by  $\sigma_j = \sigma_0 \alpha^{j-1}$ . Factor  $Q$  is the bandwidth of the filter. Generally the parameters used are:  $\sigma_0 = 0.005$ ,  $\alpha = 1.4$  and  $Q = 1.67$  values from [27].

The template of the resulting covariance matrix was calculated from all images containing no signal according to:

$$\mathbf{w}_{\text{CHO}} = (\mathbf{K}_{v/n})^{-1} \langle \mathbf{v}_s - \mathbf{v}_n \rangle \quad (5)$$

where  $(\mathbf{K}_{v/n}) = \mathbf{U}^T \mathbf{K}_{v/n} \mathbf{U}$ .

$\mathbf{K}_{v/n}$  represents the covariance matrix computed from channelized images containing no signal. In Eq. (5)  $\mathbf{v}_s$  represents the vector containing the data of the signal images as seen through the channels, and  $\mathbf{v}_n$  represents the vector containing the data of the signal-absent as seen through the channels [15,28,29].

The decision variable of the CHO model is obtained by combining Eq. (5) and Eq. (2). However, with the CHO model the decision variable is not computed with the images but with the channelized images ( $v_i$ ):

$$\lambda_{\text{CHO}} = \mathbf{w}_{\text{CHO}}^T \mathbf{v}_i \quad (6)$$

#### Internal noise

Model observers like CHO with well suited channels are more efficient than human observers for simple detection tasks such as Signal Known Exactly/Background Known Exactly (SKE/BKE). To adjust the detection outcomes of model observers to human observers it is necessary to add some internal noise,  $\varepsilon$ , to the model observer formula as shown by Eq. (7) [30].

$$\lambda_{\text{noisy}} = \lambda + \varepsilon \quad (7)$$

Internal noise  $\varepsilon$  is added to the decision variable  $\lambda$ , with a probability proportional to the standard deviation of the distribution of the decision variable from the signal absent images [30].

$$\varepsilon = \alpha \times \sigma_{\text{bg}} \times \xi \quad (8)$$

where  $\alpha$  is the weighting factor and  $\xi$  is a random number generated between  $-1$  and  $1$ ,  $\sigma_{\text{bg}}$  is the standard deviation of the distribution of the decision variable of signal-absent.

In this study, the coefficient  $\alpha$  was calibrated using images containing the signal 8 mm/10 HU at 15 mGy reconstructed with FBP and VEO.  $\alpha$  was varied between 0 and 10 iteratively. The value that minimized the difference between the model observers and the human observer outcomes for each algorithm was then selected.

#### Assessment of the model outcomes

For each category, 560 decision variables were calculated (520 from signal-absent images and 40 from signal-present images). The ROC curves were then generated from pairs of TPF and FPF and the

area under the curve was calculated by the trapezoidal method using 100 points.

Concerning the uncertainties of the results the average and standard deviation of the area under the curve (AUC) are obtained using the bootstrap method [31]; in our study the error bars represent plus or minus one standard deviation (68% for a Gaussian distribution). In order to estimate the mean and the standard deviation, the bootstrap was made using 1000 iterations for each category, and for each iteration, 520 signal-absent images and 40 signal-present images were randomly selected and replaced. Finally, to compare the performance of the CHO and the human performance, the AUC and its uncertainties were converted into detectability index ( $d_A$ ) using Eq. (9) to be used as a figure of merit [15]. For our calculations, the maximum value was set to 6 whereas theoretically, detectability varies between 0 and infinity. Obviously, the higher is the index value, the better is the signal visibility.

$$d_A = 2\Phi^{-1}(2\text{AUC} - 1) \quad (9)$$

where  $\Phi$  is the normal cumulative distribution function.

$$\Phi(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-y^2} dy \quad (10)$$

## Results

#### Comparison of FBP, ASIR 50% and MBIR

In this section, algorithms FBP, ASIR 50% and MBIR are compared with the CHO model without internal noise implemented with a D-DOG. Figure 6 shows the performance of the CHO model without noise addition for the 3 algorithms. FBP is taken as the reference algorithm. The detectability index,  $d_A$ , varies from 0.67 to 6; i.e. it varies from almost no detection to a perfect detection. The results for ASIR 50% are comparable to the ones of FBP whatever the size and contrast tested. ASIR 50% is however slightly better than FBP when the dose reaches a certain level. For signals 5 mm/10 HU, the performances between the algorithms are very similar. When the signal is hardly detectable MBIR does not improve the performance. Nevertheless, with 8 mm/10 HU or 8 mm/20 HU, MBIR improves the detectability resulting in better outcomes whatever the dose when compared to ASIR 50% and FBP. The detectability index,  $d_A$ , generally increases with dose reaching a plateau; this plateau is reached faster with MBIR algorithm than with ASIR 50% or FBP.

#### Internal noise calibration

Figure 7 shows the variation of  $d_A$  as a function of  $\alpha$  at a dose level of 15 mGy for the sphere 8 mm/10 HU reconstructed with FBP. As expected, the higher the  $\alpha$ , the lower the  $d_A$ . From these data it appears that  $\alpha$  set to 3.15 provided a good match between the CHO and human observers for the image of this category. We decided to take that value for the other categories. The calibration was also performed with the algorithm VEO using the category 8 mm/10 HU at 10 mGy (see Fig. 8). The resulting alpha coefficient is 3.6 which is quite similar to the previous one but it enables a better comparison.

#### Correlation of performances between model observers and human observers

All the detectability indexes obtained using the CHO model were compared with the 4-AFC results for images reconstructed using the algorithms MBIR and FBP, only since ASIR 50% led to comparable results

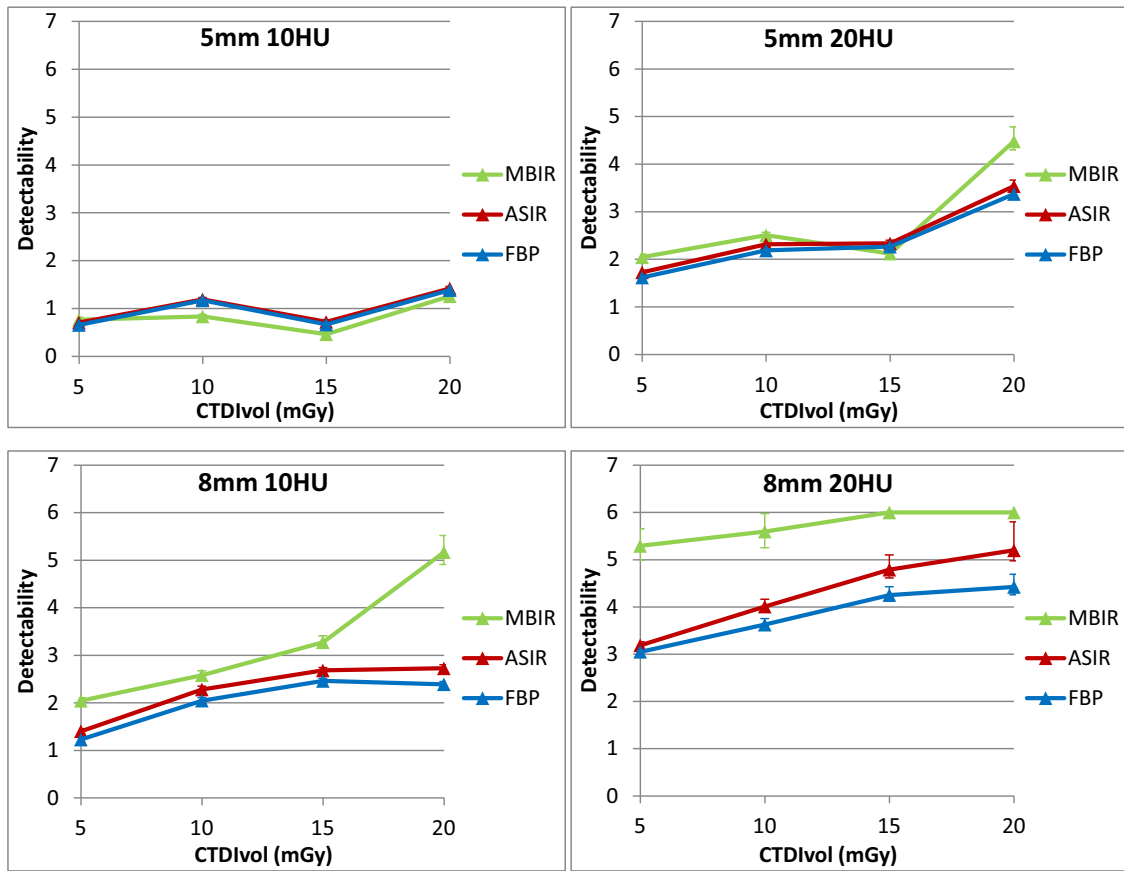


Figure 6. Comparison between FBP, ASIR 50% and MBIR algorithms.

than FBP. For each dose level, performances were strongly correlated except for the 8 mm/20 HU at 5 mGy. The Pearson coefficient is 0.911 for the 5 mm sphere FBP-reconstructed, 0.948 for 8 mm sphere FBP-reconstructed, 0.971 for the 5 mm sphere MBIR-reconstructed and 0.983 for 8 mm sphere MBIR-reconstructed. Error bars for the model observer represent plus or minus one standard deviation uncertainty obtained

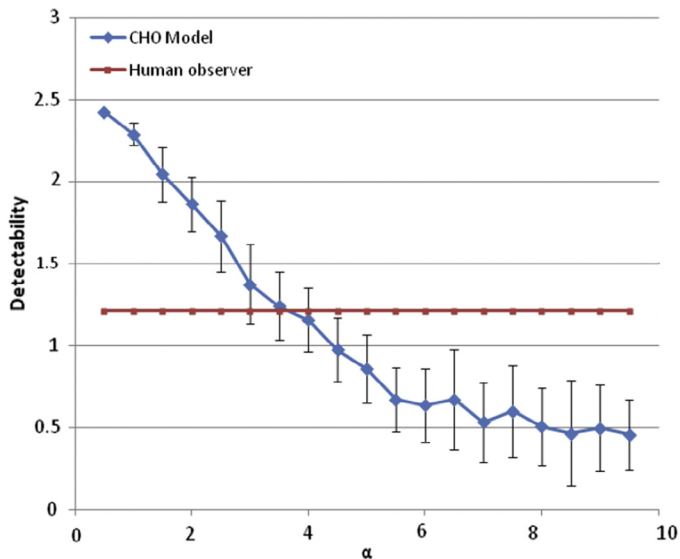


Figure 7. Internal noise calibration for an 8 mm and 10 HU sphere reconstructed with FBP algorithm.

from 200 internal noise realizations. For human observers, uncertainty is also plus or minus one standard deviation for responses recorded during the 4-AFC experiments.

All human observer results show a dose dependent increase in the detectability index. This is also the case when the size and contrast increase. For 8 mm/20 HU the increase in dose is not associated

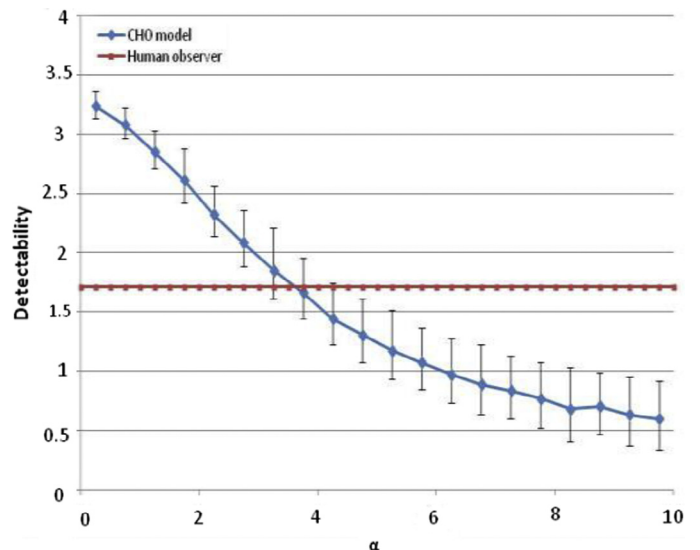


Figure 8. Internal noise calibration for an 8 mm and 10 HU sphere reconstructed with MBIR algorithm.

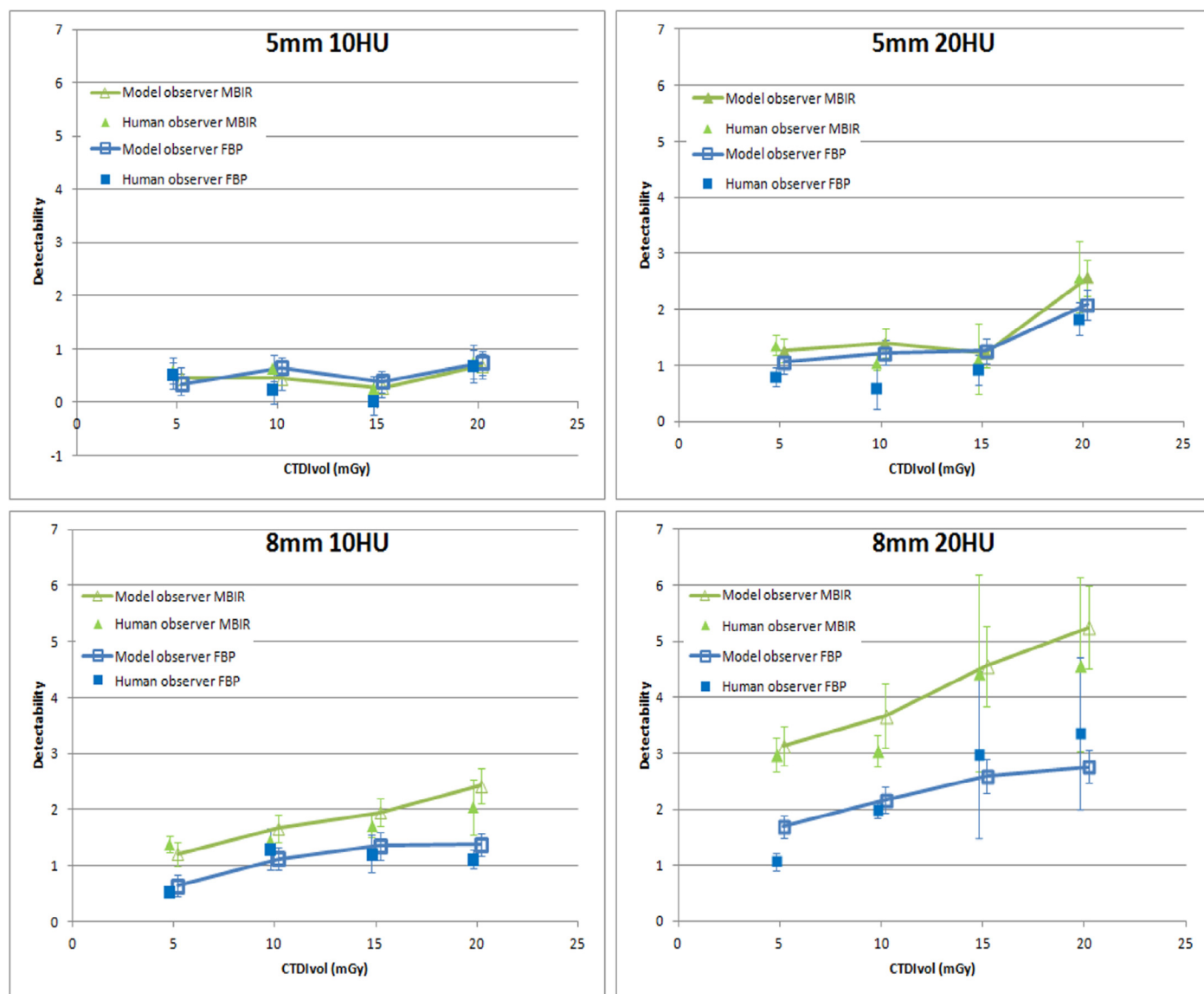


Figure 9. Performance comparison between the CHO model and human observers.

with a detectability index benefit since human observer outcomes are already very good (see Fig. 9).

To compare the performances between CHO (internal noise added) and humans, the Bland–Altman plot was used. Its ordinate is the difference between the values obtained by the two types of observers and its abscissa is the average value of the detectability index obtained by the two types of observers. Each point represents a class and the dotted lines represent plus or minus two standard deviations (95% for a Gaussian distribution). All the points are in the range plus or minus two standard deviations (see Fig. 10). Internal noise provides a good agreement between the responses of human observers and model observers. Also there is no relationship between the deviations from the mean and increased detectability.

## Discussion

Currently, patient radiation protection is a major issue and there is a tendency to significantly reduce dose without paying much attention on the potential loss of low contrast detectability. If this parameter was controlled by CNR measurements in the past (using FBP and keeping the same reconstruction kernel) the introduction

of iterative reconstruction algorithms does not allow such an approach anymore. One way to ensure the detection of a low contrasted lesion is to try to mimic human detection with mathematical model observers such as the CHO model used in this study. Our results show that the trends of low contrast detection provided with the CHO model used are compatible with human observers. However, without the addition of internal noise the CHO model outperforms human outcomes. It is therefore necessary to tune the internal noise of the CHO model to obtain a good correlation between the responses. We have shown that under our conditions a unique additional noise setting gave satisfactory results whatever the sizes, contrasts and dose involved. Thus, our method makes it possible to link a dose level to low contrast detectability performances. This information should improve the way optimization between image quality and patient exposure is balanced.

According to manufacturers, iterative reconstructions enable a drastic dose reduction without major loss of image quality. Our results show that in terms of low contrast detectability caution must be exercised in particular with the iterative reconstruction of the first generation tested (ASIR 50%) in spite of having chosen a percentage recommended to get an image quality improvement without

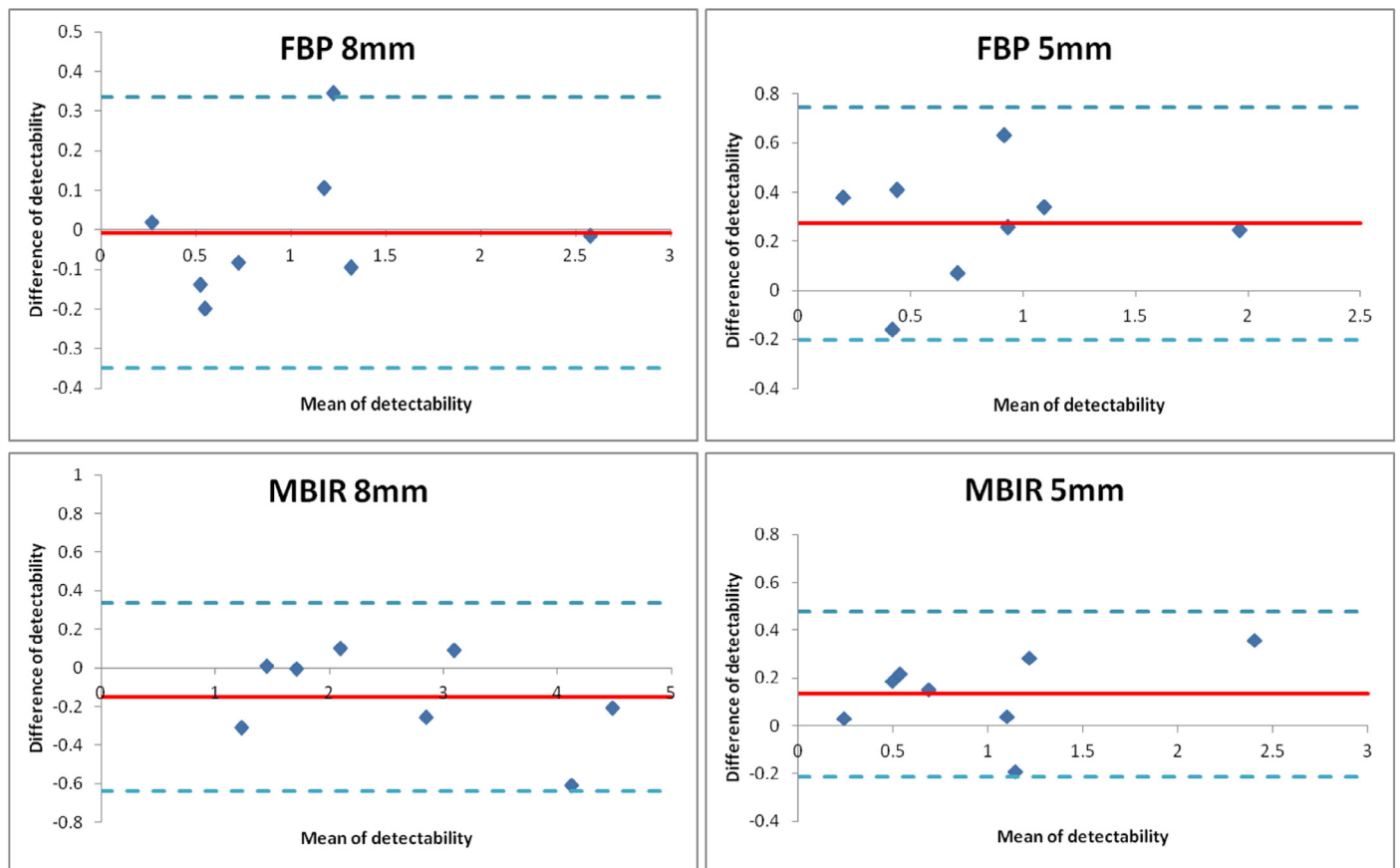


Figure 10. Bland–Altman plot of detectability index difference between human and model observer.

major image texture changes [32]. Concerning the model based iterative reconstruction (MBIR), in spite of being very computing extensive, low contrast detectability cannot be recovered at a low dose and for a very low contrast level. However, after a certain increase in dose the use of MBIR leads to much better results in terms of low contrast detectability than FBP or ASIR 50%. This kind of information is important when willing to lower patient exposure.

One limitation of our study when willing to calibrate the mathematical model observer with human results is the design of the phantom. It enables getting four spheres of a given size and contrast level per acquisition which is an advantage, but these spheres are very close to each other which require the use of a pixel interpolation to get a reasonable image size to be presented to the human observer. For such calibration purposes one should avoid placing several spheres within the same slice in order to generate large ROIs compatible with the suggestions of Yu [29] who proposed an ROI size of 4–5 times the size of the signal. Moreover the CT iterative reconstruction is not shift invariant, so to have a maximum of spheres in the minimum of space the phantom used in this project provides 4 spherical ROIs per acquisition. Unfortunately for this compromise, which is nonetheless an advantage in terms of being able to use this protocol to evaluate clinical protocols or CT units, it was necessary to create the signals close to each other which might also introduce some correlations from one signal affecting the values of nearby signals.

Finally, our results apply in a simple situation in comparison to the actual environment. The background images are homogeneous and the task is quite basic. However, we have been able to demonstrate that dose reduction must be introduced while keeping in mind that the detection of low contrast structures might be lost.

In such a situation some kind of information should be displayed on the unit to inform the radiologists about the type of low contrast sphere they will not be able to detect.

## Conclusion

CHO model coupled to D-DOG channels can be used to predict human observer performance for a 4AFC even with a limited number of acquisitions compatible with routine quality control measurements in order to assess the low contrast detectability of the acquisition for the FBP and iterative algorithms. From our results, we can conclude that the model based iterative generation algorithms (MBIR) offer superior image quality than FBP or ASIR 50% at equivalent dose. Thus, MBIR certainly offers a potential for dose reduction.

A CHO model, such as the one used in our study, could be used in routine to qualify the image quality of any given acquisition protocol. The results provided are easy to present and can be well understood by radiologists and radiographers. Finally, the use of such model observers appears to be necessary to avoid dose reduction that would significantly impair low contrast detectability.

## Acknowledgments

I would like to thank Christel Elandoy for his active participation in acquisitions, but also Ivan Diaz, Archonteia Kyrouti and Nick Ryckx for their participation in detection studies and Georg Kropat for his valuable statistical advice.

This work was partly supported by a grant from the Swiss National Science Foundation (SNSF) [No. 320030-140995].

## References

- [1] Samara ET, Aroua A, Bochud FO, Ott B, Theiler T, Treier R, et al. Exposure of the Swiss population by medical x-rays: 2008 review. *Health Phys* 2012;102:263–70. doi:10.1097/HP.0b013e31823513ff.
- [2] Office fédéral de la santé publique OFSP. Notice R-06-06: niveaux de référence diagnostiques en tomodensitométrie 2010.
- [3] Smith-Bindman R. Is computed tomography safe. *N Engl J Med* 2010;363:1–4.
- [4] Silva AC, Lawder HJ, Hara A, Kujak J, Pavlicek W. Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm. *AJR Am J Roentgenol* 2010;194:191–9. doi:10.2214/AJR.09.2953.
- [5] Miéville FA, Gudinchet F, Brunelle F, Bochud FO, Verdun FR. Iterative reconstruction methods in two different MDCT scanners: physical metrics and 4-alternative forced-choice detectability experiments – a phantom approach. *Phys Med* 2013;29:99–110. doi:10.1016/j.ejmp.2011.12.004.
- [6] Vaishnav JY, Jung WC, Popescu LM, Zeng R, Myers KJ. Objective assessment of image quality and dose reduction in CT iterative reconstruction. *Med Phys* 2014;41:071904. doi:10.1118/1.4881148.
- [7] Barrett HH, Myers KJ. *Foundations of image science*. Hoboken, New Jersey: Wiley-Interscience; 2004.
- [8] International Commission on Radiation Units and Measurements. Receiver operating characteristic analysis in medical imaging. ICRU Rep. No. 79, vol. 79. Bethesda (MD): International Commission on Radiation Units and Measurements; 2008.
- [9] Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. *Proc Natl Acad Sci U S A* 1993;90:9758–65.
- [10] Gang GJ, Lee J, Stayman JW, Tward DJ, Zbijewski W, Prince JL, et al. Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance. *Med Phys* 2011;38:1754–68. doi:10.1118/1.3560428.
- [11] He X, Park S. Model observers in medical imaging research. *Theranostics* 2013;3:774–86. doi:10.7150/thno.5138.
- [12] Burgess AE, Jacobson FL, Judy PF. Human observer detection experiments with mammograms and power-law noise. *Med Phys* 2001;28:419–37.
- [13] Hernandez-Giron I, Geleijns J, Calzado A, Veldkamp WJH. Automated assessment of low contrast sensitivity for CT systems using a model observer. *Med Phys* 2011;38:S25–35. doi:10.1118/1.3577757.
- [14] COCIR. CT Manufacturer's Voluntary Commitment Regarding CT Dose to HERCA Working Group 2013.
- [15] Beutel J, Kundel H, Van Metter R. *Handbook of medical imaging*, vol. 1. Physics and psychophysics. Bellingham, Washington: SPIE Press; 2000. p. 595–695.
- [16] Singh S, Kalra MK, Hsieh J, Licato PE, Do S, Pien HH, et al. Abdominal CT: comparison of adaptive statistical iterative and filtered back projection reconstruction techniques. *Radiology* 2010;257:373–83. doi:10.1148/radiol.10092212.
- [17] Hara AK, Paden RG, Silva AC, Kujak JL, Lawder HJ, Pavlicek W. Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study. *AJR Am J Roentgenol* 2009;193:764–71. doi:10.2214/AJR.09.2397.
- [18] Patino M, Fuentes JM, Singh S, Hahn PF, Sahani DV. Iterative reconstruction techniques in abdominopelvic CT: technical concepts and clinical implementation. *AJR Am J Roentgenol* 2015;205:W19–31. doi:10.2214/AJR.14.13402.
- [19] Thibault J-B, Sauer KD, Bouman CA, Hsieh J. A three-dimensional statistical approach to improved image quality for multislice helical CT. *Med Phys* 2007;34:4526–44. doi:10.1118/1.2789499.
- [20] Scheffel H, Stolzmann P, Schlett CL, Engel L-C, Major GP, Károlyi M, et al. Coronary artery plaques: cardiac CT with model-based and adaptive-statistical iterative reconstruction technique. *Eur J Radiol* 2012;81:e363–9. doi:10.1016/j.ejrad.2011.11.051.
- [21] Samei E, Badano A, Chakraborty D, Compton K, Cornelius C, Corrigan K, et al. Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report. *Med Phys* 2005;32:1205–25. doi:10.1118/1.1861159.
- [22] Craven BJ. A table of d' for M-alternative odd-man-out forced-choice procedures. *Percept Psychophys* 1992;51:379–85.
- [23] Hacker MJ, Ratcliff R. A revised table of d' for M-alternative forced choice. *Percept Psychophys* 1979;26:168–70. doi:10.3758/BF03208311.
- [24] Dahlquist G, Björck Å. *Numerical Methods*. Courier Corporation; 2012.
- [25] Myers KJ, Barrett HH. Addition of a channel mechanism to the ideal-observer model. *J Opt Soc Am A* 1987;4:2447–57. doi:10.1364/JOSAA.4.002447.
- [26] Baydush AH, Catarious DM, Lo JY, Floyd CE. Incorporation of a Laguerre–Gauss channelized hotelling observer for false-positive reduction in a mammographic mass CAD system. *J Digit Imaging* 2007;20:196–202. doi:10.1007/s10278-007-9009-8.
- [27] Abbey CK, Barrett HH. Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability. *J Opt Soc Am A* 2001;18:473–88. doi:10.1364/JOSAA.18.000473.
- [28] Tseng H-W, Fan J, Kupinski MA, Sainath P, Hsieh J. Assessing image quality and dose reduction of a new x-ray computed tomography iterative reconstruction algorithm using model observers. *Med Phys* 2014;41:071910. doi:10.1118/1.4881143.
- [29] Yu L, Leng S, Chen L, Kofler JM, Carter RE, McCollough CH. Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: impact of radiation dose and reconstruction algorithms. *Med Phys* 2013;40:041908. doi:10.1118/1.4794498.
- [30] Leng S, Yu L, Zhang Y, Carter R, Toledano AY, McCollough CH. Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain. *Med Phys* 2013;40:081908. doi:10.1118/1.4812430.
- [31] Efron B, Tibshirani RJ. *An introduction to the bootstrap*. Boca Raton, Florida: CRC Press; 1994.
- [32] Schindera ST, Odedra D, Raza SA, Kim TK, Jang H-J, Szucs-Farkas Z, et al. Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? *Radiology* 2013;269:511–18. doi:10.1148/radiol.13122349.