Accurate assessment and prediction of noise in clinical CT images

Xiaoyu Tian
Carl E. Ravin Advanced Imaging Laboratories, Department of Biomedical Engineering, Department of Radiology, Duke University, Durham, North Carolina 27705

Ehsan Samei
Carl E. Ravin Advanced Imaging Laboratories, Department of Radiology, Medical Physics Graduate Program, Departments of Physics and Biomedical Engineering, Duke University Medical Center, Durham, North Carolina 27705

(Received 28 June 2015; revised 6 December 2015; accepted for publication 10 December 2015; published 31 December 2015)

Purpose: The objectives of this study were (a) to devise a technique for measuring quantum noise in clinical body computed tomography (CT) images and (b) to develop a model for predicting that noise with high accuracy.

Methods: The study included 83 clinical image sets at two dose levels (clinical and 50% reduced dose levels). The quantum noise in clinical images was measured by subtracting sequential slices and filtering out edges. Noise was then measured in the resultant uniform area. The noise measurement technique was validated using 17 clinical image cases and a turkey phantom. With a validated method to measure noise in clinical images, this noise was predicted by establishing the correlation between water-equivalent diameter (Dw) and noise in a variable-sized phantom and ascribing a noise level to the patient based on Dw estimated from CT image. The accuracy of this prediction model was validated using 66 clinical image sets.

Results: The error in noise measurement was within 1.5 HU across two reconstruction algorithms. In terms of noise prediction, across the 83 clinical image sets, the average discrepancies between predicted and measured noise were 6.9% and 6.6% for adaptive statistical iterative reconstruction and filtered back projection reconstruction, respectively.

Conclusions: This study proposed a practically applicable method to assess quantum noise in clinical images. The image-based measurement technique enables automatic quality control monitoring of image noise in clinical practice. Further, a phantom-based model can accurately predict quantum noise level in patient images. The prediction model can be used to quantitatively optimize individual protocol to achieve targeted noise level in clinical images. © 2016 American Association of Physicists in Medicine. [http://dx.doi.org/10.1118/1.4938588]

Key words: medical imaging, computed tomography, image quality, noise, image-based

1. INTRODUCTION

Substantial technical improvements have led to the expanding use of computed tomography (CT) in clinical body examinations.1 CT has become an indispensable imaging modality for the diagnosis of a wide array of diseases in both pediatric and adult populations.2,3 However, given the perceived risk associated with its utilization, there is a need to balance its benefit (achieving accurate diagnoses) against radiation dose.4,5 For that objective, it is crucial to have accurate and relevant metrics of image quality. Moreover, if one would be able to accurately predict the image quality of an exam before the exam is initiated, it would be feasible to personalize the exam by adjusting the scanning parameters to achieve a desired level of image quality.5

The key objective for image quality assessment should be its quantification in clinical images; that is the only characterization of image quality that clinically matters as it is most directly related to actual quality of the clinical image. In actual practice, however, complex and variable anatomical structures complicate the quantitative assessment of image quality in CT images. As a result, in our current quality metrology, the image quality of clinical CT images is not measured. Rather generally uniform and structurally simple phantoms are used to characterize only the expected image quality of a CT system.7 The current phantom-based and system-based methods serve a beneficial function in ascertaining the minimum amount of radiation exposure necessary to achieve a reference level of quality.6 But a reference quality and the actual quality in clinical images may differ. However, if one would be able to accurately estimate the image quality in clinical images, it would be possible to draw validated correspondence between phantom-based measurements and expected clinical image quality as a function of patient size and scanner attributes; it would be possible to measure the image quality from a phantom and apply the results to accurately predict the image quality level in clinical images.

In the present study, we propose a method to measure image quality of clinical CT images. As a first step in our long-term goal, the focus of this work was stochastic noise. An image-based method was developed to measure noise in...
clinical CT images. The accuracy of method was ascertained using repeated scans. With a validated estimate of stochastic noise in clinical images, a methodology was then developed to draw correspondence between clinical and phantom-based measurements. The accuracy of the methodology was validated in terms of the agreement between the measured and predicted noise values in a database of clinical chest and abdominal exams.

2. MATERIALS AND METHODS

This study was conducted with institutional review board (IRB) approval and in compliance with the Health Insurance Portability and Accountability Act.

2.A. Clinical cases

The study used 83 deidentified clinical cases consisting of ~35,000 image slices including 20 chest image sets, 52 abdomen image sets, and 11 chest–abdomen–pelvic image sets. All images were retrospectively selected from a previously conducted IRB approved clinical trial. The images were acquired using a commercial CT scanner (Discovery CT 750 HD; GE Healthcare, Waukesha, WI) at two radiation dose levels: a clinical dose level and a 50% reduced dose level (Fig. 1). Two reconstruction algorithms, filtered back projection (FBP) and adaptive statistical iterative reconstruction (ASiR), were employed. The detailed scanning parameters are listed in Table I.

The 83 datasets were divided into two groups. Seventeen datasets were randomly selected to validate the noise measurement technique, while the remaining 66 datasets were used to validate the noise prediction method.

2.B. Measurement of quantum noise in clinical images

An experimental methodology was developed to measure stochastic noise in clinical images (Fig. 2). The main focus of this measurement technique was to separate the quantum noise from the anatomical signals. First, adjacent image slices were subtracted from one another to remove the majority of the anatomy that tended to be correlated between slices. An edge enhancement filter (sobel filter) was then applied to the subtracted images to identify the remaining anatomical features (MATLAB, The Mathworks, Inc., Natick, MA). The default threshold level (twice of the average gradient across the image) for the edge filter in the MATLAB program was used. A noise measurement program was developed to loop through the image and extract all the available ROIs (the ones that do not contain edge pixels). On average, 30,000 ROIs (30×30 pixels) were extracted from each image slice. The ROI size was selected to be large enough for sufficient pixel sampling, but small enough to minimize the effects of nonuniformities in CT numbers (e.g., beam hardening effect). The pixel standard deviations within the ROIs were divided by \(\sqrt{2}\) to account for the added noise due to subtraction. The mode number of these noise values was calculated to yield a single noise for each image slice.
The accuracy of the proposed measurement technique was validated using three methods. The first method used a large frozen turkey as a biologically representative phantom (Fig. 3). The scanning protocols were tabulated in Table I. We performed each scan twice and subtracted the repeated images to obtain a true noise map. The true noise values were compared with those obtained by the subtraction-filtering technique, and the discrepancies between the two were regarded as the error in the noise measurement technique.

The measured noise values were second validated against the expected physical relationship between the quantum noise and photon fluence in clinical images. For quantum noise limited systems (e.g., CT systems operated at typical dose levels), the quantum noise is expected to be inversely proportional to the square root of mAs. Our clinical images included imaging of the same patient at two dose levels. Since the dose ratio between two sets was known, a certain noise ratio between image sets would be expected. The discrepancy between the derived and measured noise ratio can be regarded as an indicator of error in the noise measurement technique. To quantify noise measurement accuracy based on such principle, we applied the noise measurement technique to 17 clinical sets containing 7000 image slices randomly chosen from the 83 sets, including 5 chest image sets, 6 abdomen image sets, and 6 chest–abdomen–pelvic image sets. With images acquired at two radiation dose levels, the noise of the standard dose images was extrapolated from the noise in the low dose images based on the noise-dose relationship. The percentage differences between the measured and extrapolated noise in the standard dose images were quantified as the error in the noise measurement.

As a third validation of the noise metric, we designed an observer study to assess the correlation between the noise metric and image noise level perceived by the observers. The aforementioned 17 clinical sets were included in the observer study with average sizes ranging from 24 to 32 cm. Ten image slices were selected from each clinical case. Four observers participated in the observer study. Each observer was instructed to place four ROIs in the uniform area of the image. The noise level for a given image slice was determined as the average value of the standard deviation of the four ROIs. Such noise values were compared with the noise magnitude measured using the subtraction-filtering method. The Pearson correlation coefficient between the two datasets was determined. The average differences were further assessed.

### 2.C. Noise prediction

With a validated method to measure quantum noise in clinical images, a phantom-based approach was developed to predict the noise based on patient attenuation. First, recognizing that noise has a strong dependence on patient size, noise measurements were performed using a custom-designed phantom [Mercury phantom 2.0, Duke University, Durham, NC (Fig. 4)].

The phantom was composed of four cylindrical sections with varying sizes (16, 23, 30, and 37 cm in diameter),

---

**Table I.** Scan parameters used in this study.

| Parameter               | Clinical images | Turkey phantom | Mercury phantom |
|-------------------------|-----------------|----------------|-----------------|
| kVp                     | 120             | 120            | 120             |
| CTDI<sub>vol</sub> (mGy) | 3.5–20.3        | 1.4, 4.3, 8.6  | 12, 15, 18, 21 |
| Pitch                   | 1.375           | 1.0            | 1.375           |
| Collimation (mm)        | 40 (64×0.625)   | 40 (64×0.625)  | 40 (64×0.625)   |
| Slice thickness (mm)    | 0.625           | 5              | 0.625           |
| Reconstruction algorithm| FBP, ASiR<sup>a</sup> 50% | FBP, ASiR 50% | FBP, ASiR 50%  |
| Convolution kernel      | Standard        | Standard       | Standard        |
| Tube current modulation | On              | Off            | Off             |
| Noise index (if applicable) | 9.9–21.5      | None           | None            |

<sup>a</sup>Filtered back projection.

<sup>b</sup>Adaptive statistical iterative reconstruction.

---

**Fig. 2.** The process to measure quantum noise in clinical images. (a) Example of a clinical chest image. (b) Noise image obtained by subtracting two adjacent image slices. (c) The remaining anatomical region on the noise image identified by the edge filter.
with interconnecting cone-shaped sections. The phantom sizes represented the size distribution of clinical patients. The phantom was scanned on GE Discovery 750 HD scanner at four dose levels (12, 15, 18, and 21 mGy) and reconstructed with ASiR 50% and FBP algorithms. The noise was measured in terms of average pixel standard deviation within multiple ROIs in each diameter section.

To establish the relationship between quantum noise and attenuation, a water-equivalent diameter, denoted as Dw, was estimated for each image slice. A water-equivalent area of the scanned object was first estimated as $\sum (\text{HU}/1000 + 1)\alpha$, in which HU denotes the CT number for each pixel and $\alpha$ denotes the pixel area. The calculation was summarized across all the pixels within the image field of view. Dw was then estimated as the diameter of the water-equivalent area. For the patients that exceed the scan FOV, the water-equivalent diameter cannot be accurately quantified from the CT image. The alternative method is to quantify water-equivalent diameter from topogram images.

A regression analysis was then established (R, version 2. 14.1) between noise and phantom water-equivalent diameter as

$$\sigma = B \sqrt{\exp(\mu Dw)},$$

where Dw denotes the water-equivalent diameter of the object estimated for each image slice. $\mu$ denotes the attenuation coefficient of water at the effective energy level of the CT spectrum, and $B$ is a CT protocol-dependent parameter affected by scanner model, bowtie filter, kVp, dose levels, pitch, reconstruction algorithm, and reconstruction kernel.

Fig. 3. Example CT slice of the turkey phantom acquired at 4.3 mGy CTDIvol reconstructed using FBP (a) and ASiR (b) reconstruction. The reference noise map obtained by subtracting the repeated image sets for FBP (c) and ASiR (d).

Fig. 4. Schematic diagram (a), photograph (b), and scout image (c) of the phantom (Mercury 2.0) used in this study (Ref. 8).
2.D. Validation of noise prediction model

The noise prediction model was validated using the 66 clinical patient images. The validation was performed across patient sizes and anatomical regions. To predict noise, the patient water-equivalent diameter was first estimated from CT images using the method described above. Since the clinical images were acquired using tube current modulation (TCM), each slice had a different mAs value. The noise was further predicted by substituting the tube current levels and water-equivalent diameter into Eq. (1) to derive the corresponding noise magnitude level. As a reference standard, the noise in clinical images was further measured using our validated subtraction-filtering technique described previously. The accuracy was characterized in terms of the discrepancy between the predicted and the measured noise values.

3. RESULTS

3.A. Validation of noise measurement technique

3.A.1. Turkey phantom images

The average absolute error in noise measurement technique when applied to images of the turkey phantom was 1.44 (6.5%), 1.12 (7.4%), and 1.51 HU (9.3%) for FBP reconstructions at three dose levels. For ASiR, the corresponding average errors were 1.37 (7.1%), 1.26 (8.6%), and 1.47 HU (9.8%), respectively. The maximum error ranges from 10.3% to 17.4% across different reconstruction algorithms, dose levels, and anatomical regions.

3.A.2. Clinical patient images

Figures 5 and 6 illustrate the accuracy of noise measurement technique in clinical image sets reconstructed with FBP. Figure 5 shows the histogram of percentage and absolute error of the noise measurement. The error exhibited a largely Gaussian distribution with an average error of 1.63 HU, 5.4% for the 5000 images. Figure 6 shows one example of a chest–abdomen–pelvic case with the extrapolated noise (red points) agreeing closely with measured ones (blue points).

3.A.3. Validation against observer studies

Figure 7 illustrates the relationship between automated and observer-based noise measurements for 17 clinical image datasets. The average difference between the automated and observer-based noise measurements was 0.95 HU (4.3%), which was less than the intraobserver variability of 1.7 HU (5.3%). A better agreement was found in the abdominal images than in the thoracic images [0.89 (3.9%) and 1.4 HU (4.7%) for the abdominal and thoracic images, respectively]. The average Pearson correlation coefficient between the two datasets was 0.998.

3.B. Noise prediction technique

Figures 8 and 9 illustrate prediction accuracy for 66 clinical images. Figure 8 shows six examples of noise predictions. The predicted noise generally agreed well with measured noise across reconstruction algorithms and patient anatomies (chest, abdominal, and pelvic regions). Figure 9 summarizes
prediction accuracy for all clinical sets, with each data point representing the average noise value of one patient. The predicted noise showed a linear relationship with the measured noise, with unity slopes and \( R^2 \) of 0.98 and 0.97 for the FBP and ASiR 50\% reconstructions, respectively. The corresponding average differences between the predicted and measured noise in 66 clinical sets were 6.6\% and 6.9\%. The average maximum error across 66 clinical patient datasets was 14.3\% and 14.1\% for FBP and ASiR 50\% reconstructions.

4. DISCUSSION

Characterizing CT image quality is essential to the optimization and design of CT protocols, the evaluation of CT algorithms and systems, and improved diagnostic accuracy. However, current approaches for image quality measurements are largely based on uniform phantoms and simple tasks. As an initial step to move image quality evaluations toward more clinically meaningful measurements relevant to actual clinical images, this study proposed a practical method to measure quantum noise in clinical images and use that to design a validated model for predicting image noise for individual patients.

The noise measurement technique of this program alone can provide crucial information in a clinical operation. Similar to the current dose monitoring program, this technique can essentially enable monitoring noise in clinical images. One may improve the system performance by identifying and investigating cases with noise levels over or below a desired range, enforcing protocol-specific reference noise level, and establishing consistent system performance across scanners. Such a noise management strategy was implemented by Larson et al. to achieve the targeted noise range based on phantom measurements.\(^6\) Our approach offers the added advantage of being based on the actual measured attributes of clinical images.

Having a strong affirmation that we can measure noise in clinical images, this study further established a validated correspondence between phantom-based measurements and clinical image noise level, which can be used for protocol optimization. As an example, consider an adult male patient who undergoes an abdominopelvic CT examination. Assuming a targeted noise level of 15 HU and an average 26 cm water-equivalent diameter, the imaging protocol can be prospectively defined for the GE 750 HD scanner as 120 kVp tube voltage, 361 mAs, 1.375 pitch, 0.625 mm slice thickness, 40 mm collimation, and FBP reconstruction to yield the desired noise. One may further develop an optimized TCM profile to provide a consistent level of noise in the images.

It should be noted that the measured noise remains relatively constant for different threshold levels of the edge filter (difference less than 3.4\% across 5000 image slices). This is due to the fact that the mode number of multiple measurements was calculated as the final noise result. With the increasing of threshold level, there may be more inaccurate noise measurement due to the existence of anatomy. However, the mode number of the multiple ensembles remains the same.

Prediction accuracy was found to be generally higher in the abdominal regions than in the lung regions of our cases. We suspect that this might be due to two factors: first, the lung region contains an air cavity surrounded by soft tissue. The prediction model equates the attenuation to a centrally located circular area, neglecting the effect of the bowtie filter on noise. Bowtie filters preferentially attenuate the x-ray on the periphery. As a result, at a peripheral location in the lung with water-equivalent diameter Dw, the noise is higher than that measured from the center of a water phantom with the same Dw. Further, the noise measurement in the lung region is more challenging since the uniform areas that can be used for noise measurement are relatively small.

The introduction of iterative reconstruction algorithms has posed a challenge to CT image quality methodology. More specifically, measurements based on simple phantoms may not represent corresponding image quality in clinical images due to the nonlinear and shift-variant properties of some iterative reconstruction algorithms. Our approach for noise measurements for ASiR provides good assessment and prediction results for quantum noise. For other reconstruction algorithms, where noise property may be dependent on anatomy and background, further efforts should be devoted to ensuring that the phantom-based measurements can accurately predict the noise observed in actual clinical images.
In this study, we assumed no statistical correlation between the subsequent CT slices; the noise map was determined by subtracting the two adjacent slices and dividing by a factor of $\sqrt{2}$ to account for error propagation. However, the two subsequent CT image slices may be correlated. To estimate the potential error induced by such correlation, a validation study was performed using an anthropomorphic phantom (ATOM, Adult Male, models 701, Computerized Imaging Reference Systems) scanned under a clinical abdominopelvic protocol (120 kVp, 1.375 pitch, 12 mGy CTDIvol, 0.625 mm slice thickness, FBP reconstruction, standard kernel). The phantom was first scanned twice and the two datasets were subtracted from each other to obtain “gold-standard” measurements of the noise magnitude. We further used one image dataset and subtracted the adjacent slices to obtain the noise magnitude measurements that may contain errors induced by slice correlation. The average discrepancy between the two noise measurements is less than 1% across the dataset. Therefore, the effect of potential correlation on noise measurement is generally negligible.

This study was based on a CT system from one manufacturer and a small number of scanning parameters. Future work would include CT scanners from other manufactures and a broader range of protocols and reconstruction algorithms.
Besides, in this study, noise is modeled as a function of multiple variables including the scanner model, bowtie filter, kVp, dose levels, pitch, reconstruction algorithm, and reconstruction kernel. In future study, we aim to reduce the number of dependent variables in the noise model. Furthermore, this study mainly focused on the prediction of quantum noise and assumed the effect of electronic noise is negligible. Electronic noise generally has a minor effect for CT images acquired under standard dose levels. However, it becomes more significant for low-dose protocols since the level of quantum noise may become comparable to the electronic noise magnitude.14 Future study should include a measurement and prediction model for electronic noise, especially for CT images acquired with low-dose protocols (such as lung cancer screening).

5. CONCLUSION

In this study, we proposed a practical technique to measure quantum noise in clinical images. The image-based measurement technique enables one to effectively track and monitor clinical image performance. Furthermore, this study involved the development of a phantom-based model to predict the level of quantum noise in clinical images. Such a model allows one to have the knowledge of the typical noise level in a CT image before it is acquired. Such knowledge can be used to individualize the CT protocol by adaptively adjusting the scanning parameter in advance of a CT exam to achieve a target image quality while minimizing the radiation dose.

ACKNOWLEDGMENTS

The authors would like to thank James Winslow, Olav Christianson, and Joshua M. Wilson for their ample help with noise measurement experiments. The authors are also thankful to Dr. Xiang Li for stimulating discussions throughout the evolution of this work.

---

Fig. 9. Accuracy of noise prediction model for 66 clinical image cases for the FBP (a) and ASiR 50% (b) reconstructed images.

---

14Author to whom correspondence should be addressed. Electronic mail: xt3@duke.edu
1C. H. McCollough, L. Guimaraes, and J. G. Fletcher, “In defense of body CT”, AJR, Am. J. Roentgenol. 193, 28–39 (2009).
2A. S. Raja, K. J. Mortele, R. Hansol, A. D. Sodickson, R. Zane, and R. Khorasani, “Abdominal imaging utilization in the emergency department: Trends over two decades,” Int. J. Emerg. Med. 4, 19 (2011).
3D. C. Levin, V. M. Ruo, L. Parker, A. J. Frangos, and J. H. Sunshine, “Recent trends in utilization rates of abdominal imaging: The relative roles of radiologists and nonradiologist physicians,” J. Am. Coll. Radiol. 5, 744–747 (2008).
4M. K. Kalra, M. M. Maher, T. L. Toth, L. M. Hamberg, M. A. Blake, J. A. Shepard, and S. Saini, “Strategies for CT radiation dose optimization,” Radiology 230, 619–628 (2004).
5L. Yu, X. Liu, S. Leng, J. M. Kofler, J. C. Ramirez-Giraldo, M. Qu, J. Christner, J. G. Fletcher, and C. H. McCollough, “Radiation dose reduction in computed tomography: Techniques and future perspective,” Imaging Med. 1, 65–84 (2009).
6D. B. Larson, L. L. Wang, D. J. Podberesky, and M. J. Goske, “System for verifiable CT radiation dose optimization based on image quality. Part I. Optimization model,” Radiology 269, 167–176 (2013).
7C. H. McCollough, M. R. Bruesewitz, M. F. McNitt-Gray, K. Busch, T. Ruckdeschel, J. T. Payne, J. A. Brink, and R. K. Zeman, “The phantom portion of the American College of Radiology (ACR) computed tomography (CT) accreditation program: Practical tips, artifact examples, and pitfalls to avoid,” Med. Phys. 31, 2423–2442 (2004).
8J. M. Wilson, O. I. Christianson, S. Richard, and E. Samei, “A methodology for image quality evaluation of advanced CT systems,” Med. Phys. 40, 031908 (9pp.) (2013).
9X. Li and E. Samei, “Comparison of patient size-based methods for estimating quantum noise in CT images of the lung,” Med. Phys. 36, 541–546 (2009).
10J. Menke, “Comparison of different body size parameters for individual dose adaptation in body CT of adults,” Radiology 236, 565–571 (2005).
11X. Tian, X. Li, W. P. Segars, E. K. Paulson, D. P. Frush, and E. Samei, “Pediatric chest and abdominopelvic CT: Organ dose estimation based on 42 patient models,” Radiology 270, 535–547 (2014).
12P. Sahbaee, W. P. Segars, and E. Samei, “Patient-based estimation of organ dose for a population of 58 adult patients across 13 protocol categories,” Med. Phys. 41, 072104 (12pp.) (2014).
13X. Tian, X. Li, W. P. Segars, E. K. Paulson, D. P. Frush, and E. Samei, “Dose coefficients in pediatric and adult abdominopelvic CT based on 100 patient models,” Phys. Med. Biol. 58, 8755–8768 (2013).
14X. Duan, J. Wang, S. Leng, B. Schmidt, T. Allmendinger, C. H. McCollough, and K. Grant, “Electronic noise in CT detectors: Impact on image noise and artifacts,” AJR 201, W626–W632 (2013).