Comparison and validation of noise magnitude estimation methods from patient CT images

F. Ria, T. B. Smith, E. Abadi, J. B. Solomon, E. Samei

Purpose

Image quality estimation in CT is crucial for technology assessment, procedure optimization, and overall radiological benefit evaluation, with noise magnitude playing a key role. Over the years, several methods have been proposed to estimate noise surrogates *in vivo*. The most accurate approach is to assess ensemble noise by scanning a patient multiple times and sampling each pixel noise within the ensemble of images, an ethically undoable repeated imaging process. Such impasse can be surmounted using Virtual Imaging Trials (VITs) that use computer-based simulations to simulate clinically realistic scenarios. The purpose of this study was to compare two different noise magnitude estimation methods with the ensemble noise measured in a VIT population.

Methods

This study included a set of 47 XCAT-phantom repeated chest exams acquired virtually using a scanner-specific simulator (DukeSim) modeling a commercial scanner geometry, reconstructed with FBP and IR algorithms. Noise magnitudes were calculated in soft tissues (GNI) and air surrounding the patient (AIRn), applying [-300,100]HU and HU<-900 thresholds, respectively. Furthermore, for each pixel in GNI threshold, the ensemble noise magnitudes in soft tissues (En) were calculated across images. Noise magnitude from different methods were compared in terms of percentage difference with correspondent En median values.

Results

For FBP reconstructed images, median En was 30.6 HU; median GNI was 40.1 HU (+31%) and median AIRn was 25.1 HU (-18%). For IR images, median En was 19.5 HU; median GNI was 25.1 HU (+29%) and median AIRn was 18.8 HU (-4%).

Conclusion

Compared to ensemble noise, GNI overestimates the tissue noise by about 30%, while AIRn underestimates it by 4 to 18%, depending on the reconstruction used. These differences may be applied as adjustment or calibration factors to the related noise estimation methods to most closely represent clinical results. However, air noise cannot be assumed to represent tissue noise.

Innovation/Impact

Noise magnitude is one of the main indicator that can be used as a surrogate for the quality assessment of the diagnostic image properties. Several methods have been introduced over the years to measure noise both in phantoms and in patients: different approaches rely on different strategies to segment the images, apply HU thresholds, and define regions of interest in which the noise magnitude is calculated. As a consequence, it is impossible to compare results obtained with different methods. This scenario is further complicated by the dose reduction techniques introduced in modern CT, such Automated Tube Current Modulation and Iterative Reconstruction algorithms (IR) that have changed the traditional statistical relationships between radiation dose and image quality.

Inconsistency in noise magnitude estimations can negatively affect the overall evaluation of the radiological procedure, impacting the assessment of the technology, as well as justification and optimization. Therefore, it is crucial to compare each methods with the results obtained using a standard accurate approach, as described in this work. The analysis compared the noise magnitude values obtained in patients (*in vivo*) using two different methodologies (GNI and AIRn) with the measurements performed in an ensemble of virtually generated images, which is considered to be the best noise surrogate ethically obtainable. Data showed that air noise underestimates ensamble tissue noise by varying amounts, depending on the reconstruction used. The spatial noise further overestimates the noise compared to ensemble noise. The results can be used as a calibration to better predict clinical results *in vivo* improving consistency across different methodology assessing noise magnitude in clinical operations.

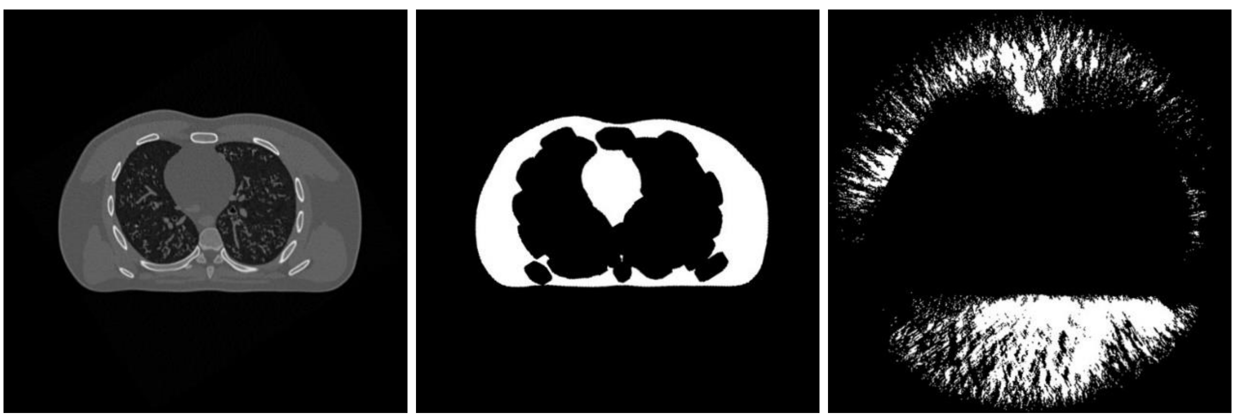

Figure 1. From left to right: example of one image considered in the study reconstructed with FBP; when GNI threshold is applied; when AIR threshold is applied.