Unicentre
CH-1015 Lausanne
http://serval.unil.ch

Year : 2016

Iterative reconstruction in CT: Using mathematical model observers to determine low dose images' trustworthiness

OTT Julien

OTT Julien, 2016, Iterative reconstruction in CT: Using mathematical model observers to determine low dose images' trustworthiness

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_4E7D5D1CA1472

Droits d'auteur
L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright
The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Département de radiologie médicale - CHUV
Institut de radiophysique (IRA)

Iterative reconstruction in CT: Using mathematical model observers to determine low dose images' trustworthiness

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Julien OTT

Master ès Sciences en physique, Université Joseph Fourier de Grenoble

Jury

Prof. Francine Behar-Cohen, Président (CHUV-UNIL)
Prof. Francis R. Verdun, Directeur de thèse (CHUV-UNIL)
Prof. François Gudinchet, Co-directeur (CHUV-UNIL)
Dr. Rémy Klausz, Expert (GE, Buc)
Dr. Albert Lisbona, Expert (Centre René Gauducheau, Nantes)
MSc Sue Edyvean, Expert (Public Health, England)

Lausanne 2016

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président · e	Madame Prof. Francine **Behar-Cohen**
Directeur · rice de thèse	Monsieur Prof. Francis **Verdun**
Co-directeur · rice	Monsieur Prof. François **Gudinchet**
Experts · es	Monsieur Dr Rémy **Klausz**
	Madame Dr Sue **Edyvean**
	Monsieur Dr Albert **Lisbona**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Julien Ott

Master ès Sciences en physique de l' Université Joseph Fourier, Grenoble, France

intitulée

Iterative reconstruction in CT : Using mathematical model observers to determine low dose images'trustworthiness

Lausanne, le 13 janvier 2016

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Francine Behar-Cohen

"Ach, dass der Mensch doch durchsichtig wäre wie eine Qualle und dass man den Sitz seiner Leiden schauen könnte!"

Wilhelm C. Röntgen

Remerciements

Mener à bien une thèse de doctorat s'est révélé être un travail exigeant au cours duquel j'ai eu la chance de bénéficier du soutien de nombreuses personnes. J'aimerais donc par ces quelques lignes prendre le temps de remercier, à la fois celles et ceux qui ont participé directement à ce travail mais également toutes les personnes qui m'ont soutenu durant cette période.

Je tiens tout d'abord à remercier mon directeur de thèse, le Pr. Francis R. Verdun pour m'avoir accueilli dans son groupe à l'IRA et m'avoir témoigné sa confiance pour la réalisation de ce projet. Ses remarques judicieuses ont été d'une grande aide tout au long de mon travail et la liberté dont j'ai disposé au cours de mes recherches ainsi que les moyens qu'il a su mettre en œuvre pour développer ce projet ont sans aucun doute contribué à son aboutissement dans les meilleures conditions possibles. Mes remerciements sont également adressés à mon co-directeur de thèse, le Pr. François Gudinchet pour ses explications sur les aspects médicaux en rapport avec ce travail.

Je tiens aussi à remercier profondément le Dr. Fabio Becce pour nos diverses collaborations. Travailler avec Fabio s'est révélé particulièrement enrichissant. Son efficacité, ses conseils avisés, ses idées toujours pertinentes et novatrices ainsi que le temps et la quantité de travail qu'il a consacré à tous nos projets communs ont apporté une grande plus-value au travail accompli.

Mes remerciements sont également destinés à toutes les personnes qui ont été co-auteur de mes publications. Parmi eux, j'adresse à Pascal Monnin et au Pr. François Bochud un remerciement particulier. A Pascal pour ses conseils et nos discussions concernant l'évaluation de la qualité d'images via les méthodes usuelles en traitement du signal et à François pour ses explications et son expertise sur tous les sujets relatifs aux modèles d'observateurs. Merci également à ces deux amoureux de la montagne pour toutes nos discussions à ce sujet.

Je voudrais également remercier mes collègues de bureau, plus particulièrement Georg, Alexandre et Damien. Pour leur input scientifique à mes travaux, mais également pour les moments passés en dehors du bureau. Qu'il s'agisse de course à pied, de vélo ou d'autres sorties ensemble, tout cela m'a permis de profiter d'une atmosphère plus qu'appréciable durant les heures de travail.

Merci également à Julie Bize, Christel Elandoy, Silvano Gnesin et Nick Ryckx, mes collègues du GIM pour leur bonne humeur au quotidien et l'ambiance de travail agréable. Un remerciement particulier est à adresser à Christel, pour le temps passé lors des acquisitions d'images, ainsi qu'à Nick, d'une part pour nos collaborations professionnelles (toujours constructives) et d'autre part pour nos collaborations extra-professionnelles (non moins constructives, quoique… !) en compagnie de notre cercle d'amis communs.

Je remercie également tous mes collègues de l'IRA qui ont contribué à rendre agréable l'ambiance au travail durant ces trois dernières années (un remerciement particulier en ce sens va à Andréas pour son assiduité lors de nos séances de course à pied à la mi-journée), ainsi que toute l'équipe du secrétariat pour leur soutien logistique sans faille durant mon travail.

Finalement, j'aimerais remercier toute ma famille pour leur soutien durant ces trois années de thèse et plus généralement durant toute ma scolarité. Merci à ma mère de m'avoir enseigné et transmis la détermination et ténacité dont elle savait faire preuve pour mener à bien tout ce qu'elle entreprenait. Merci à mon père pour son support sans faille et permanent en toutes circonstances. Merci à mes grands parents pour leurs encouragements, leurs pensées positives et leur affection à mon égard. Merci à mes frères pour les moments passés lors de mes retours occasionnels à la maison. Merci également à Andrée, ainsi qu'à Frédéric pour la sympathie dont ils ont toujours su faire preuve, ainsi qu'à Roland pour s'être toujours dévoué à l'aspect logistique au cours de mon parcours. Je remercie également Jean-François et Karine pour les nombreux weekends savoyards à base de gastronomie et de cyclisme (qu'il soit pratiqué le dimanche matin ou suivi à la télé le dimanche après-midi). Enfin merci à Marie, son soutien, sa bonne humeur et sa gentillesse au quotidien ont su m'aider à affronter les étapes jusqu'au bout de ce défi qu'est un travail de thèse.

Ce travail a été partiellement financé par le Fonds National Suisse (FNS) de la recherche scientifique (No. 320030-140995).

Table of contents

Résumé......2
Abstract......3
List of acronyms......4
1. Introduction......7
1.1 Dose reduction for CT procedures: A challenging task......7
1.2 Radiation dose estimation......10
1.3 A major advance in CT: Iterative reconstruction......11
1.4 Usual methods for image quality assessment in CT......18
1.4.1 Clinical approach......*18*
1.4.2 Physical approach......*18*
1.4.3 Limitations of these image quality assessment methods......*22*
1.5 Statistical decision theory using model observers......23
1.6 Goals of the study......28
2. PhD thesis milestones......29
2.1 Clinical approach: Applying existing methods to improve the use of IR in the clinic......29
2.1.1 Optimisation of IR strength levels for clinical cervical spine images......*29*
2.1.2 Use of ASIR for dose reduction in hip clinical images......*30*
2.1.3 Using iterative reconstruction techniques for optimising radiation dose and image quality in musculoskeletal CT......*31*
2.2 Physical approach: Going further by developing fully adapted physical methods in order to analyse IR images......32
2.2.1 Update of the NPW model observer: Specific assessment of ASIR and MBIR algorithms......*32*
2.2.2 Assessment of low-contrast detectability in CT using image space model observers......*33*
2.2.3 Image quality assessment in different reconstruction planes using an updated NPWE model observer......*34*
3. Papers......35
3.1 Clinical approach......35
3.2 Physical approach......62
4. Conclusions and perspectives......144
References......146
Appendix: Details of FBP reconstruction......152
Curriculum Vitae......154

Résumé

La tomodensitométrie (TDM) est une technique d'imagerie pour laquelle l'intérêt n'a cessé de croitre depuis son apparition au début des années 70. De nos jours, l'utilisation de cette technique est devenue incontournable, grâce entre autres à sa capacité à produire des images diagnostiques de haute qualité. Toutefois, et en dépit d'un bénéfice indiscutable sur la prise en charge des patients, l'augmentation importante du nombre d'examens TDM pratiqués soulève des questions sur l'effet potentiellement dangereux des rayonnements ionisants sur la population. Parmi ces effets néfastes, l'induction de cancers liés à l'exposition aux rayonnements ionisants reste l'un des risques majeurs. Afin que le rapport bénéfice-risques reste favorable au patient il est donc nécessaire de s'assurer que la dose délivrée permette de formuler le bon diagnostic tout en évitant d'avoir recours à des images dont la qualité est inutilement élevée. Ce processus d'optimisation, qui est une préoccupation importante pour les patients adultes, doit même devenir une priorité lorsque l'on examine des enfants ou des adolescents, en particulier lors d'études de suivi requérant plusieurs examens tout au long de leur vie. Enfants et jeunes adultes sont en effet beaucoup plus sensibles aux radiations du fait de leur métabolisme plus rapide que celui des adultes. De plus, les probabilités des évènements auxquels ils s'exposent sont également plus grandes du fait de leur plus longue espérance de vie.
L'introduction des algorithmes de reconstruction itératifs, conçus pour réduire l'exposition des patients, est certainement l'une des plus grandes avancées en TDM, mais elle s'accompagne de certaines difficultés en ce qui concerne l'évaluation de la qualité des images produites.

Le but de ce travail est de mettre en place une stratégie pour investiguer le potentiel des algorithmes itératifs vis-à-vis de la réduction de dose sans pour autant compromettre la qualité du diagnostic. La difficulté de cette tâche réside principalement dans le fait de disposer d'une méthode visant à évaluer la qualité d'image de façon pertinente d'un point de vue clinique.

La première étape a consisté à caractériser la qualité d'image lors d'examen musculo-squelettique. Ce travail a été réalisé en étroite collaboration avec des radiologues pour s'assurer un choix pertinent de critères de qualité d'image. Une attention particulière a été portée au bruit et à la résolution des images reconstruites à l'aide d'algorithmes itératifs. L'analyse de ces paramètres a permis aux radiologues d'adapter leurs protocoles grâce à une possible estimation de la perte de qualité d'image liée à la réduction de dose. Notre travail nous a également permis d'investiguer la diminution de la détectabilité à bas contraste associée à une diminution de la dose ; difficulté majeure lorsque l'on pratique un examen dans la région abdominale. Sachant que des alternatives à la façon standard de caractériser la qualité d'image (métriques de l'espace Fourier) devaient être utilisées, nous nous sommes appuyés sur l'utilisation de modèles d'observateurs mathématiques. Nos paramètres expérimentaux ont ensuite permis de déterminer le type de modèle à utiliser. Les modèles idéaux ont été utilisés pour caractériser la qualité d'image lorsque des paramètres purement physiques concernant la détectabilité du signal devaient être estimés alors que les modèles anthropomorphes ont été utilisés dans des contextes cliniques où les résultats devaient être comparés à ceux d'observateurs humain, tirant profit des propriétés de ce type de modèles.

Cette étude a confirmé que l'utilisation de modèles d'observateurs permettait d'évaluer la qualité d'image en utilisant une approche basée sur la tâche à effectuer, permettant ainsi d'établir un lien entre les physiciens médicaux et les radiologues. Nous avons également montré que les reconstructions itératives ont le potentiel de réduire la dose sans altérer la qualité du diagnostic. Parmi les différentes reconstructions itératives, celles de type « model-based » sont celles qui offrent le plus grand potentiel d'optimisation, puisque les images produites grâce à cette modalité conduisent à un diagnostic exact même lors d'acquisitions à très basse dose. Ce travail a également permis de clarifier le rôle du physicien médical en TDM: Les métriques standards restent utiles pour évaluer la conformité d'un appareil aux requis légaux, mais l'utilisation de modèles d'observateurs est inévitable pour optimiser les protocoles d'imagerie.

Abstract

Computed tomography (CT) is an imaging technique in which interest has been quickly growing since it began to be used in the 1970s. Today, it has become an extensively used modality because of its ability to produce accurate diagnostic images. However, even if a direct benefit to patient healthcare is attributed to CT, the dramatic increase in the number of CT examinations performed has raised concerns about the potential negative effects of ionising radiation on the population. Among those negative effects, one of the major risks remaining is the development of cancers associated with exposure to diagnostic X-ray procedures. In order to ensure that the benefits-risk ratio still remains in favour of the patient, it is necessary to make sure that the delivered dose leads to the proper diagnosis without producing unnecessarily high-quality images. This optimisation scheme is already an important concern for adult patients, but it must become an even greater priority when examinations are performed on children or young adults, in particular with follow-up studies which require several CT procedures over the patient's life. Indeed, children and young adults are more sensitive to radiation due to their faster metabolism. In addition, harmful consequences have a higher probability to occur because of a younger patient's longer life expectancy.

The recent introduction of iterative reconstruction algorithms, which were designed to substantially reduce dose, is certainly a major achievement in CT evolution, but it has also created difficulties in the quality assessment of the images produced using those algorithms.

The goal of the present work was to propose a strategy to investigate the potential of iterative reconstructions to reduce dose without compromising the ability to answer the diagnostic questions. The major difficulty entails disposing a clinically relevant way to estimate image quality.

To ensure the choice of pertinent image quality criteria this work was continuously performed in close collaboration with radiologists. The work began by tackling the way to characterise image quality when dealing with musculo-skeletal examinations. We focused, in particular, on image noise and spatial resolution behaviours when iterative image reconstruction was used. The analyses of the physical parameters allowed radiologists to adapt their image acquisition and reconstruction protocols while knowing what loss of image quality to expect. This work also dealt with the loss of low-contrast detectability associated with dose reduction, something which is a major concern when dealing with patient dose reduction in abdominal investigations. Knowing that alternative ways had to be used to assess image quality rather than classical Fourier-space metrics, we focused on the use of mathematical model observers. Our experimental parameters determined the type of model to use. Ideal model observers were applied to characterise image quality when purely objective results about the signal detectability were researched, whereas anthropomorphic model observers were used in a more clinical context, when the results had to be compared with the eye of a radiologist thus taking advantage of their incorporation of human visual system elements.

This work confirmed that the use of model observers makes it possible to assess image quality using a task-based approach, which, in turn, establishes a bridge between medical physicists and radiologists. It also demonstrated that statistical iterative reconstructions have the potential to reduce the delivered dose without impairing the quality of the diagnosis. Among the different types of iterative reconstructions, model-based ones offer the greatest potential, since images produced using this modality can still lead to an accurate diagnosis even when acquired at very low dose.

This work has clarified the role of medical physicists when dealing with CT imaging. The use of the standard metrics used in the field of CT imaging remains quite important when dealing with the assessment of unit compliance to legal requirements, but the use of a model observer is the way to go when dealing with the optimisation of the imaging protocols.

Keywords: Computed tomography, image quality, dose reduction, iterative reconstruction, model observer

List of acronyms

AAPM	American Association of Physicists in Medicine
ADMIRE	Advanced Model-Based Iterative Reconstruction
AIDR	Adaptive Iterative Dose Reduction
AIR	Algebraic Iterative Reconstruction
ASIR	Adaptive Statistical Iterative Reconstruction
ATCM	Automated Tube Current Modulation
AUC	Area Under the Curve
BKE	Background-Known-Exactly
CHO	Channelized Hotelling Observer
CNR	Contrast-to-Noise Ratio
CSF	Contrast Sensitivity Function
CT	Computed Tomography
CTDI	Computed Tomography Dose Index
$CTDI_{vol}$	Volume CTDI
$CTDI_{w}$	Weighted CTDI
DDoG	Dense Difference of Gaussian
DLP	Dose Length Product
DQE	Detective Quantum Efficiency
DRL	Dose Reference Level
DECT	Dual Energy CT
E	Effective Dose
FBP	Filtered Back-Projection
FNF	False Negative Fraction
FOM	Figure Of Merit
FPF	False Positive Fraction
FST	Fourier Slice Theorem
FT	Fourier Transform
HO	Hotelling Observer
HU	Hounsfield Unit
ICRP	International Commission on Radiological Protection
IMR	Iterative Model Reconstruction
IR	Iterative Reconstruction
IRIS	Iterative Reconstruction in Image Space
LCD	Low Contrast Detectability
LG	Laguerre-Gauss
LNT	Linear No-Threshold
LSF	Line Spread Function
(M)-AFC	(M)-Alternative Forced-Choice
MAP	Maximum A Posteriori
MBIR	Model-Based Iterative Reconstruction
MDCT	Multi-Detector CT
MPR	Multi-Plane image Reformatting
MRI	Magnetic Resonance Imaging

MTF	Modulation Transfer Function
NEQ	Noise-Equivalent Quanta
NPS	Noise Power Spectrum (also known as Wiener Spectrum)
NPW	NonPreWhitening matched filter
NPWE	NonPreWhitening matched with Eye filter
PC	Percent of Correct responses
PMMA	PolyMethyl MethAcrylate
PSF	Point Spread Function
PW	PreWhitening matched filter
ROC	Receiver Operating Characteristics
ROI	Region Of Interest
SAFIRE	Sinogram AFfirmed Iterative REconstruction
SD	Standard Deviation
SDNR	Signal Difference-to-Noise Ratio
SIR	Statistical Iterative Reconstruction
SKE	Signal-Known-Exactly
SNR	Signal-to-Noise Ratio
SSDE	Size-Specific Dose Estimate
TNF	True Negative Fraction
TPF	True Positive Fraction
TTF	Target Transfer Function
VGA	Visual Grading Analysis

1. Introduction

1.1 Dose reduction for CT procedures: A challenging task

Computed tomography (CT) has undergone a fast technological development since its introduction in the early 70's, leading to a substantial increase in the number of CT units and procedures performed. This imaging technique is now used so extensively that it may become a victim of its own success if special attention is not paid in terms of its impact on population exposure. Indeed, data show that the number of examinations performed over the last few decades is continuously increasing, reaching for example 62 million CT exams in 2008 in the US [Brenner, 2008]. In Switzerland, successive surveys have indicated that between 1998 and 2008, the amount of radiation delivered by CT went from 28% to 68% of the dose delivered for medical reasons [Aroua, 2002; Samara, 2012]. This number even reached 71% in 2013 [Le Coultre, 2016]. In parallel, the number of CT examinations in Switzerland also increased significantly over this same time period, going from 3.4% to 9.6% of the total radiological procedures. Despite this situation, it is important to remember that the increase of CT indications has also improved patient care and significantly reduced the number of more invasive procedures such as those involving fluoroscopy. However, to ensure that the benefits-risk ratio still remains in favour of the patient, it is necessary to make sure that the dose delivered leads to the proper diagnosis without producing unnecessary high-quality images or images with insufficient conspicuity that could impair the diagnosis.

Patient's major risk associated with exposure to diagnostic X-ray procedures is the development of cancers. However, this consequence cannot currently be demonstrated for low-dose range radiation exposure, which makes this assertion a highly debated question and means that the principle of caution must be applied. One common approach to quantify the number of excess cancers due to X-ray medical procedures involves estimating the cumulated absorbed dose in the most exposed organs. Then, one assesses how many of those patients developed a tumour over a given period of time and compares the results with a control population [Berrington de González, 2004]. Using this method, researchers managed to highlight that *"if 10,000 children under the age of 10 each received one CT scan, this would be associated in a single additional patient developing a brain tumour or leukaemia over the next 10 years"* [NHS, 2012]. Despite the fact that the results obtained are in good agreement with the risk factors proposed by the international commission on radiological protection (ICRP), several questions remain for that particular study. First, the delay between the exposures and the development of cancer seems quite short in comparison to the more commonly expected delays. Another weakness of the study is the choice of the populations. A population of children who require a brain CT is different, from a health status point of view, than the general child population. Finally, the precision with which the absorbed dose was measured was limited. This study certainly underscores the fact that caution should be applied with a strict justification and optimisation of the procedure, but a direct association between CT and cancer induction should not be made.
At the moment there are two major trends: one that uses new elements that seem to confirm the validity of the linear no-threshold (LNT) model also in the low-dose range (inferior to 50 mGy) [Pearce 2012]. The other that claims that sparing dose *"never constitutes a logical reason to avoid an imaging procedure that is otherwise medically indicated"* [Durand, 2011].

In such a context, the most reasonable approach involves reinforcing the justification as well as the optimisation principle. The justification of the examination is a medical task and will be assessed in the framework of clinical audits. The optimisation process is a complex task involving manufacturers, radiologists, radiographers and medical physicists. It relies on the fact that a radiological image is considered a test to help patient management and should therefore convey the maximum amount of useful information while minimising patient exposure. If the quantification of patient dose is relatively easy, it is not the case for the image quality level—the very thing upon which we rely to provide a successful and reliable diagnosis. The aim of this work was to investigate possible options when dealing with image quality within the framework of patient exposure optimisation.

Manufacturers already started the first step of the optimisation process a few years ago, when they began to develop strategies and devices to reduce patient exposure as much as possible while providing an image quality level compatible with diagnostic requirements. Among the major technological advances worth mentioning is the introduction of the automated tube current modulation (ATCM) [Kaza, 2014]. This is of major importance since, unlike plain analogue film radiography where excessively high radiation exposures result in reduced image quality or even a non-readable image, digital radiography or CT image quality does not deteriorate at high exposures; on the contrary, images become even easier to read. Some tube current modulations can even take into account the fact that during the acquisition the X-ray will expose particularly radiosensitive organs such as the eye lenses, the thyroid, or the breast (e.g. organ based tube current modulation). However, if the idea seems rather enticing the optimal positioning of the sensitive organs remains a major issue in order to fully benefit from the technique [Taylor, 2015]. Recently, X-ray tube high-voltage modulation (kV modulation) has been proposed as a strategy to further reduce the radiation dose when dealing with imaging of high Z materials, taking advantage of the benefit of the photoelectric effect [Suh, 2013]. Much progress has also been made concerning not only the efficiency of the detector, but the penumbra of the beam. Some advances in terms of the X-ray beam collimation have also been proposed with a real-time beam tracking during scanning in order to minimise the beam collimation and size and thus reduce unused patient exposure at ends of helical scanning (so called: adaptive collimation) [Deak, 2009].

The way information is extracted from the data has also been drastically improved by the introduction of iterative reconstruction (IR). Users are now able to produce readable images at very low-dose levels [den Harder, 2015; Naoum, 2015; Padole, 2015]. However, image content remains, up to a certain level, dose dependent and one should remember that drastic dose reduction particularly impairs the detection of low-contrast structures [Schindera, 2013].

The next step of the optimisation process concerns the way the devices are used by radiographers. Indeed, if a CT unit, when it complies with international standards, is a relatively safe piece of equipment, the way it is used on patients might lead to undesired effects [Imanishi, 2005; Rehani, 2015]. To limit over-exposure as much as possible, some manufacturers have proposed software that checks whether the settings of the protocol could lead to acute tissue reactions or present major errors [Howard, 2014]. On top of this, manufacturers offer training sessions to ensure that all the technological solutions available on the unit are optimally used.

Finally, the last necessary step of the optimisation process concerns radiologists and physicists and should be done with the clinical applications in mind. The assessment of the unit for generic protocols is generally performed by medical physicists who should be involved in setting quality assurance programs [EURATOM 2013, IAEA. 2014]. Compliance of the units with legal requirements is made with standard test objects where image quality criteria are far from clinically relevant tasks [Verdun 2015]. When willing to optimise actual clinical protocols, as for radiography, we should use task-based methodologies. This can be done by using mathematical model observers and standard quality assurance phantoms as shown by several groups. [Hernandez-Giron, 2011, 2014; Kofler, 2015; Samei 2015]. To get closer to the clinical world, one can use anthropomorphic phantoms and either human or mathematical model observers [Yu, 2013; Tseng, 2014; Zhang, 2014; Ott, 2015 a; Verdun & Racine, 2015]. Finally, methods of assessing the adequacy of image quality with the diagnosis should be done in patient studies [Miéville, 2013; Smedby, 2013; Zarb, 2015]. In spite of their difficulty to implant they remain necessary when drastic changes in dose and data processing are introduced.

Having briefly described patient dose and image quality while considering image quality at the first stage, other initiatives have used patient dose instead. Dose quantities are a lot easier to assess than image quality, which should be related to diagnostic tasks. The introduction of dose reference levels (DRLs) that rely on an automatic and permanent update of the recommendations for the practice has already improved the situation by helping to reduce the spread of the practice. But these values are still poorly related to a specific diagnostic task. They are often considered to be limits, which is not the case, and concepts like the introduction of ”target dose” values or “achievable dose” meant to reduce patient dose as much as possible might be counterproductive in the future. Image quality should remain a main priority of the optimisation process [Rehani, 2015]. Tailoring image quality to diagnostic requirements increases the number of CT protocols. If this process is important in the context of radiation protection, it introduces a serious bias because of the lack of standardisation. One of the first efforts to standardise CT acquisitions was applied quite successfully in the paediatric population by colour-coding the protocols, but there is still a way to go [Frush, 2002; Singh, 2011]. Nevertheless, many initiatives are focused on protocol standardisation meant to enable an optimal use of the software solutions and making it possible to compare between centres in terms of patient exposure [Kofler, 2014; Escalon, 2015; MacGregor, 2015]. This is certainly a quality criterion, but image information should not be forgotten. In addition, optimisation should not only consider radiation risk but other risks such as the one associated with the use of contrast media [Noda, 2015].

On the whole, thanks to technological breakthroughs in CT, significant progress has been achieved in terms of improving the optimisation process (i.e. conveying the maximum amount of useful information in the image while minimising patient exposure). Amongst this progress, the recent introduction of IR algorithms is certainly a major achievement since they helped to bring about substantial dose reductions. However, the difficulty of this approach remains in image quality assessment due to image texture modifications and signal non-linearity which are inherent to those algorithms. All this shows that efforts still have to be made in order to optimise image quality in CT.
Traditionally, the role of the medical physicist was to check the compliance of the CT unit with legal requirements or technical characteristics claimed by manufacturers. Tools adapted from signal processing theory were used satisfactorily. However, this approach was limited when dealing with the actual image quality/patient dose optimisation in the clinical world. Our work demonstrates that the use of model observers enables the medical physicists to enter the clinical world. Much progress

remains to be made in order to improve the outcomes of model observers to clinical requirements but even in their present infancy stage they have enabled radiologists to adapt their practice.

1.2 Radiation dose estimation

Today, CT devices rely on two metrics in order to quantify the radiation dose delivered by the unit during an examination.

The first is called “volume CT dose index” ($CTDI_{vol}$), expressed in mGy. It is calculated using the CT dose index (CTDI), which corresponds to the integral of the dose profile along a line perpendicular to the tomographic plane divided by the product of the nominal tomographic section thickness and the number of tomograms produced in a single scan. CTDI measurements at the centre and the periphery of a standardised phantom (16- and 32-cm-diameter polymethyl methacrylate (PMMA) phantoms) are then averaged to produce the weighted CTDI ($CTDI_w$). Finally, in order to take the effect of the pitch on the radiation dose into account, $CTDI_w$ is divided by the pitch value, yielding the $CTDI_{vol}$. This quantity can be related to the image noise level of reconstructed slices.
The dose-length product (DLP) which corresponds to the $CTDI_{vol}$ multiplied by the scan length expressed in mGy·cm [AAPM 96, 2008] is the other quantity used to quantify the amount of radiation delivered by the unit.

CTDI and DLP do however only give a measure of the radiation amount delivered by the unit and do not correspond to the dose received by the patient. Thus, in order to represent the stochastic health effects of ionizing radiation on the human body, the quantity called effective dose is used.
Effective dose itself is based on the equivalent dose, which represents the physical quantity of absorbed dose, and also taking into account the biological effectiveness of the radiation, which is dependent on the radiation type and energy. Then, the effective dose is calculated using the tissue-weighted sum of the equivalent doses in all specified tissues and organs of the human body. [ICRP 103, 2007]

However, Monte Carlo simulations have shown that the $CTDI_{vol}$ concept has important limitations since it could lead to an underestimation of the dose for two reasons: The length of the CTDI phantom which is quite small (14 cm) and therefore produces less scatter than in actual patients, and the length of the ion chamber (10 cm) that does not collect the total amount of scattered radiation [Dixon, 2003]. Moreover, $CTDI_{vol}$ only provides information about CT radiation output without taking patient morphology into account. Thus, this metric can be seen as a measure of the radiation produced by the device, but not as a patient dose estimator.

To better estimate the local exposure of patients, a new metric called “size-specific dose estimate” (SSDE) has been proposed [AAPM report 204]. This metric relies on the use of conversion factors that are based on patient size measurements.

Eventually, the best way to link CT dose indicators to the organ dose is to compute an effective dose (E) using the tissue weighting factors [ICRP 103, 2007]. However, and despite the previously stated weaknesses of the CTDI, only this metric and not SSDE should be used to compute an effective dose to the organs. Indeed, the tissue weighting factors were estimated for adult patients only. Also, effective dose is a value that is used to predict statistical risk and was not meant to be applied to a single individual. Thus, Monte Carlo simulations remain the best way to proceed when it’s necessary to perform a risk assessment of a specific organ dose on a single individual.

1.3 A major advance in CT: Iterative reconstruction

Analytical methods for CT reconstruction

Until recently, image reconstruction in CT was performed using **analytical methods**. Among them, the filtered back projection (FBP) is the most commonly encountered. This method consists of resolving an inverse problem: the projections of an object are acquired under various angles between 0 and π radians (This step is in fact the computation of the so-called sinogram of the image) before being used to return the image of the investigated object. An illustration of the process is represented in Figure 1.

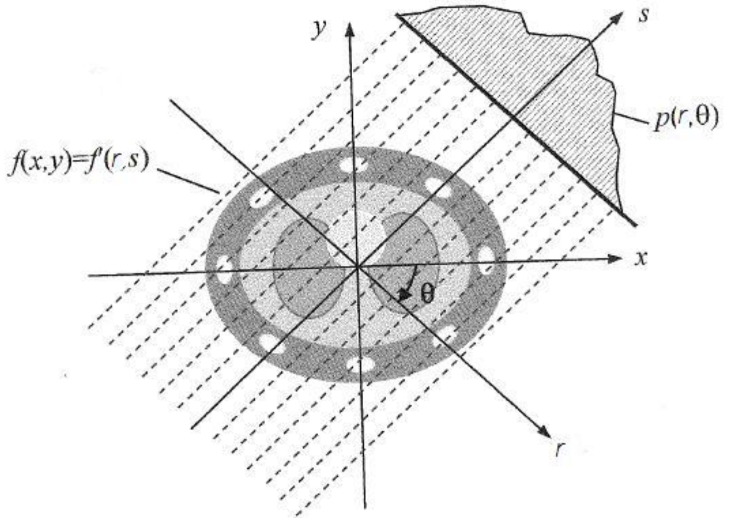


Figure 1: Acquisition of the projections $p(r,\theta)$ of an object $f(x,y)$ (Extracted from [Hsieh, 2009])

It can be easily shown that the image $f(x,y)$ can then be computed using Eq. 1 (Extensive details on its derivation are given in annex):

$$f(x,y) = \underbrace{\int_0^{\pi} \underbrace{\int_{-\infty}^{+\infty} \underbrace{\underbrace{TF[p(\rho,\theta)]}_{TF[Projection]} \, |\rho|}_{Filtering} \, e^{2i\pi\rho r} d\rho}_{TF^{-1}[filtered\, projection]} \; d\theta}_{Backprojection} \tag{1}$$

In Eq. 1 $TF[p(\rho,\theta)]$ is the Fourier Transform of the sinogram. The inner integral is the inverse Fourier transform of the quantity $TF[p(\rho,\theta)]\,|\rho|$, which corresponds to a Fourier-transformed projection filtered by a function whose frequency domain response is $|\rho|$. This function is called a ramp filter. Finally, using the inverse Fourier transform to go back into the spatial domain and project the result is what creates the desired image. This explains why this algorithm is called filtered back projection. The process can be summarised according to Figure 2:

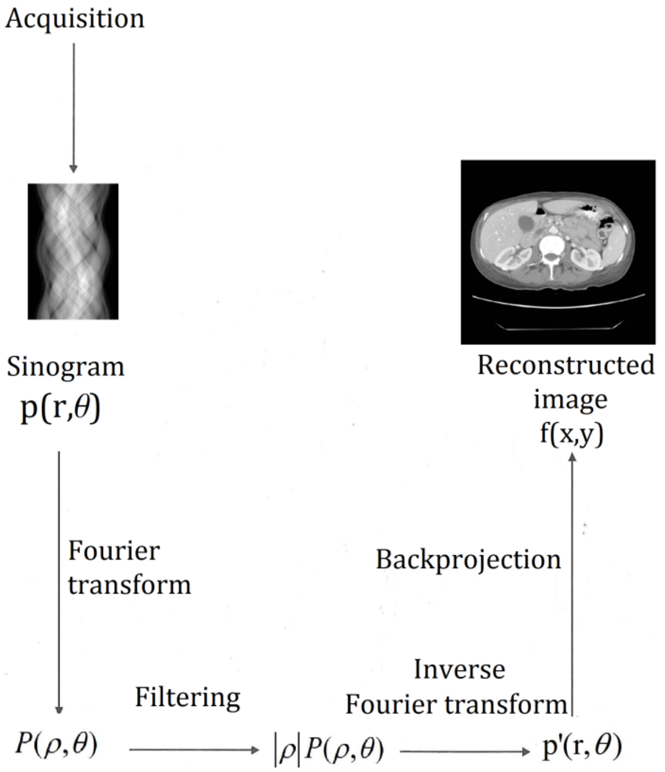


Figure 2: Explanation of the FBP algorithm

Eq. 1 is, however, only true for a parallel X-ray beam and some adjustments are made in practice to take the effect of the fan beam into account. Moreover, other filters beyond the classical ramp are also implemented in practice during the filtering step (Extensive details can be found in Chapter 3 of Hsieh's book [Hsieh, 2009]). The main problem of this approach is that the Fourier plane may be well covered by the sampling points which are expressed in the polar coordinate system, but switching to the image space necessitates a resampling of the point in the Cartesian coordinates and this resampling step is responsible for information loss. Also using a finite number of projections implies the non uniqueness of the solution, meaning that an infinite number of projections (that is to say an infinite number of angles θ) should be used so that Eq. 1 can be mathematically true. Also, the solution is highly unstable, meaning that small differences over the projections could lead to very different reconstructed images. Finally, FBP does not take into account the variations of the number of photons used to image our object; these variations are understood through Poisson distribution. Thus, and because of Poisson's distribution properties, the noise in FBP reconstructed images will always be inversely proportional to the square root of the dose.

Iterative methods for CT reconstruction

In order to overcome some of these limitations, an alternative approach for CT reconstruction is also possible, namely the **iterative methods**. This approach involves a discretisation of the problem. Indeed, instead of resolving the problem $f(x,y) = \int_0^{\pi} p'(r,\theta)d\theta$ which requires strong mathematical hypothesis, it is possible to proceed as follows:

Assuming that no scatter or beam hardening effect is encountered, the Beer-Lambert law states that an X-ray beam with an intensity I_0 going through an object with an attenuation coefficient *attenu(x,y,z)* and along a path of length *L* will come out with an intensity I_{out}. That is to say:

$$I_{out} = I_0 e^{-\int_L attenu(x,y,z)dl} \tag{2}$$

Since the object we are trying to image is represented on a digital system, the image can be discretised. Applying that assumption to our problem, which is summarised in Fig. 3, Eq. 2 becomes:

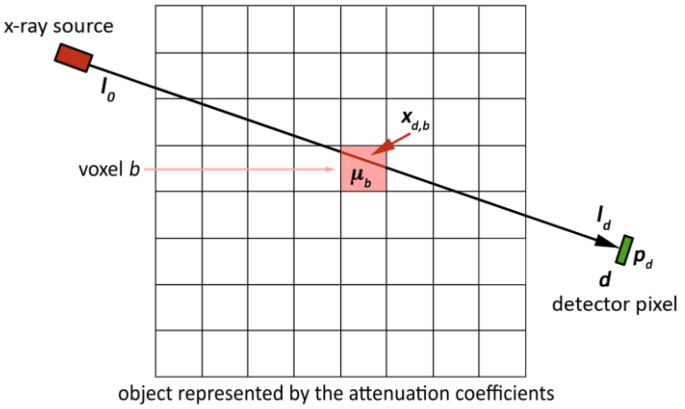
$$I_d = I_0 e^{-\sum_{b=1}^{B} \mu_b \times x_{d,b}} \tag{3}$$

$$\Leftrightarrow \qquad Ln\frac{I_0}{I_d} = p_d = \sum_{b=1}^{B} \mu_{b\times} x_{d,b} \tag{4}$$

$ln\frac{I_0}{I_d}$ is the quantity measured by the detector, in other words the projection measurements of our object. Thus, and using matrix notation, Eq. 4 becomes:

$$\boldsymbol{p} = \boldsymbol{X\mu} \tag{5}$$

The vector ***p*** contains the acquired projections.
The vector ***μ*** contains the properties of the object we are trying to estimate.
The matrix ***X*** is the projecting operator. Each coefficient $X_{d,b}$ of this DxB size matrix represents the intersection length between the beam reaching the detector number d and the voxel b. It is therefore related to the probability that an event emitted in the pixel b be detected in the detector cell d.



b	voxel object, 1 < *b* < *B*, with *B* the number of voxel composing the object
μ_b	attenuation coefficient of voxel *b*
$x_{d,b}$	interaction length in the voxel *b*
d	detector cell, 1 < *d* < *D*, with *D* the number of cell of the detector
p_d	the projection measurements in the detector cell *d*

Figure 3: X-ray Beam emitted by the source and absorbed by the detector (Extracted from [Miéville PhD, 2012])

Using this formulation of the problem, the image reconstruction can be done in resolving D equations of B unknown. However, the dimensions of the problem rapidly lead to a large size matrix. Therefore, performing an inversion of ***X*** to solve Eq. 5 is not possible and iterative methods have to be employed to reconstruct the image. Those iterative reconstruction techniques are divided into two kinds: **algebraic methods** and **statistical methods**.

Algebraic iterative reconstructions (AIR) work by iteratively resolving the system $\boldsymbol{p} = \boldsymbol{X\mu}$. They rely on projection methods to determine the values of the elements belonging to the vector $\boldsymbol{\mu}$ [Buvat, 2011]. While AIR have been employed in the first CT prototypes [Herman, 1980], their high computational cost have led people to abandon them in favour of analytical methods like FBP despite its previously stated drawbacks.

Statistical iterative reconstruction techniques (SIR) work by using a cost function based on statistical models; this function tries to obtain the most likely image using the measured projections [Fessler, 2009]. The introduction of these new CT statistical iterative solutions made it possible to obtain CT images that are still usable for a diagnosis but with a substantial dose reduction [Schindera, 2011].
Among statistic IR, two other subsets can be determined, namely the **hybrid IR algorithms** and the **fully statistical iterative reconstruction algorithms**.

Hybrid algorithms are called this because they combine analytical and iterative reconstruction in different proportions. Several approaches are possible for this. The first entails generating the image with the FBP method before applying iterative methods in the image space to de-noise and improve image properties. The second approach consists of performing the iterative de-noising in the sinogram space before using FBP to reconstruct the image. The proportion in which analytical (that is to say FBP in this case) and iterative methods are combined can be selected by the user. Finally, it is also possible to combine image and sinogram de-noising methods for the reconstruction. These various approaches have facilitated remarkable advances in terms of image quality over the last decade, but they also have their own drawbacks, including edge blurring or resolution loss.

In opposition to hybrid algorithms, fully statistical methods rely entirely on the principle of maximising a cost function. In this class of IR belong the **model-based iterative reconstructions (MBIR)** which rely on a probabilistic view of CT reconstruction based on CT physics. To do this, their statistical cost function includes X-ray physics and a model of the CT optics. They constitute the most complex and advanced reconstruction methods used today since they strongly reduce noise and artefacts and even enhance spatial resolution [McCollough, 2012]. The following paragraph summarises the approach used by these algorithms to perform image reconstruction. Extensive details on formula derivation are voluntarily omitted since they have already been treated by numerous authors. The reader interested in the details of the subject will find them in Thibault's work [Thibault, 2007].

The first step, as always, is to acquire the projection of the object on the CT detectors. Since X-ray intensity is known to follow a Poisson law, we can reasonably assume that the same is true for the projection measurements. Then, the ith intensity measurement on the detector will be expressed as:

$$I_{d_i} \sim Poisson\left\{ I_{0_i} e^{-\overline{p_i}} \right\} \tag{6}$$

$\overline{p_i}$ is the noiseless projection, that is to say the ideal line integral of the image along an X-ray beam. It is computed based on the true attenuation coefficients μ.

The cost function can then be expressed using the Bayesian framework of the maximum *a posteriori* (MAP) estimate:

$$\underset{\substack{\text{Attenuation values}\\ \text{estimate}}}{\boldsymbol{\mu}} = \arg\max_{\mu} \underbrace{P(\boldsymbol{\mu} \mid \boldsymbol{p})}_{\substack{\text{Posterior}\\ \text{probability}}} \tag{7}$$

$$\Leftrightarrow \qquad \boldsymbol{\mu} = \arg\max_{\mu} \left\{ \underbrace{Ln\left(P(\boldsymbol{p} \mid \boldsymbol{\mu})\right)}_{Log\ likelihood} + \underbrace{Ln\left(P(\boldsymbol{\mu})\right)}_{Log\ prior\ probability} \right\} \qquad (8)$$

Then, using Eq. 3 and adding a vector ***n*** representing the random fluctuations of the system (such as photon and electric noise) we can model the transformation performed by the CT unit using Eq. 9:

$$\boldsymbol{p} = \boldsymbol{X}\boldsymbol{\mu} + \boldsymbol{n} \qquad (9)$$

Substituting Eq. 9 into Eq. 8 introduces the system's physics into the reconstruction algorithm.
Finally, it is possible to rewrite the log-likelihood term in this new equation using a second order Taylor development. This yields the statistical cost function of the MBIR algorithm:

$$\boldsymbol{\mu} = \arg\min_{\mu} \left\{ \frac{1}{2}(\boldsymbol{p} - \mathbf{X}\boldsymbol{\mu})^T \mathbf{D}(\boldsymbol{p} - \mathbf{X}\boldsymbol{\mu}) + U(\boldsymbol{\mu}) \right\} \qquad (10)$$

In this equation, ***X*** models the CT system. It embeds elements to model CT optics and X-ray physics.
D is a diagonal matrix whose coefficients are used to model the noise.
*U(**μ**)* is a scalar equal to the sum of log prior probability with a constant. It is called the penalty function because it is used to penalise local differences between voxels. Its role is basically an object model.

Eq. 10 is then solved iteratively using approaches like iterative coordinate descent for example to obtain the desired image.

The use of MBIR has recently been made possible mostly thanks to hardware and computer developments which helped to manage the huge computational cost inherent in those techniques [Hsieh, 2013].

More details on the different IR techniques as well as their classifications and way to work can be found in Geyer's work [Geyer, 2015]. Figure 4 also gives a comprehensive summary of the different types of reconstruction algorithms.

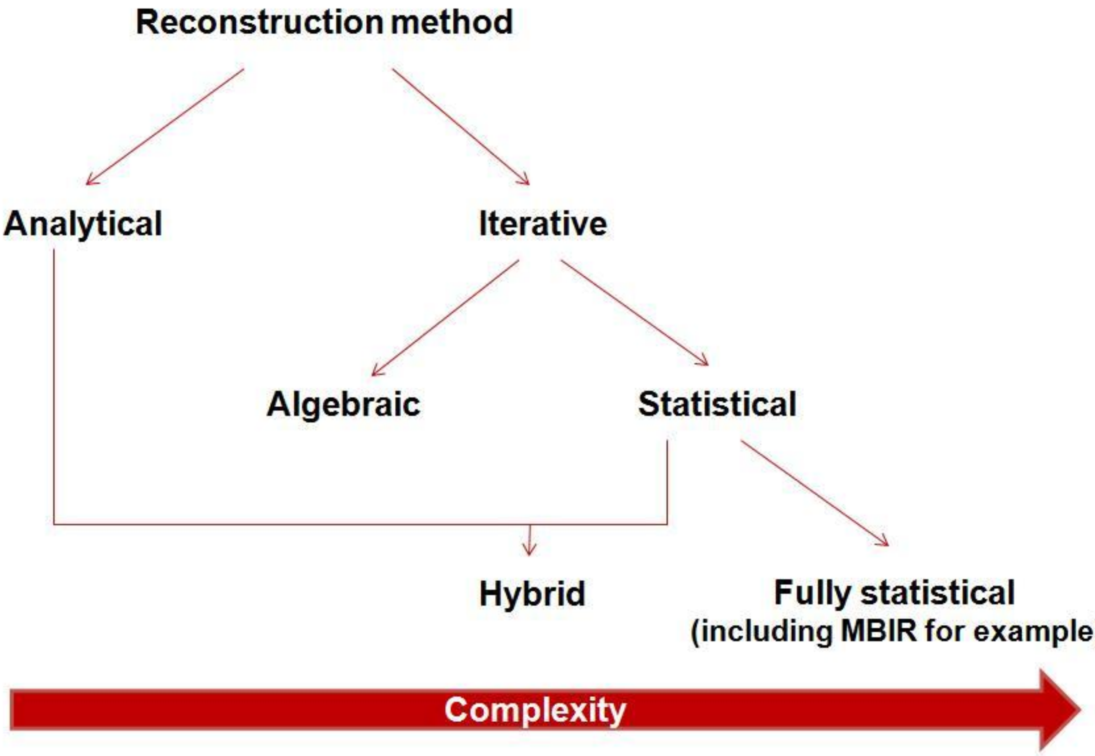


Figure 4: A summary of the classification of different reconstruction algorithm types

IR strategies of the different CT vendors

The very first statistical iterative reconstruction commercially available for clinical CT devices was introduced by General Electric Healthcare (GE, Milwaukee, WI, USA) in 2008 under the name adaptive statistical iterative reconstruction (ASIR$^{(TM)}$). Then, in 2010, Siemens (Erlangen, Germany) released its algorithm called iterative reconstruction in image space (IRIS$^{(TM)}$). One year later, Philips commercialised its own iterative method under the name iDose$^{4\ (TM)}$. The same year, Toshiba (Tochigi, Japan) also released its adaptive iterative dose reduction (AIDR$^{(TM)}$).

The next big step in the field of IR happened in 2012, when GE released its second iterative reconstruction named VEO$^{(TM)}$, and which was also the first commercially available MBIR. The same year, Siemens introduced its first raw-data-based iterative solution, called sinogram affirmed iterative reconstruction (Safire$^{(TM)}$), as a successor of IRIS.

Recently, Philips and Siemens also released their own MBIR, respectively called iterative model reconstruction (IMR) and advanced model-based iterative reconstruction (ADMIRE). At the same time, GE launched its latest reconstruction algorithm called ASIR-V, whose aim is to offer image quality comparable to Veo but with a smaller reconstruction time. For this, this MBIR algorithm **Table 1** shows the existing reconstruction techniques of the four manufacturers and details to which type of reconstruction algorithm each belong. The classification of each algorithm was done using Figure 4 and the work done by Patino et al. [Patino, 2015].

Table 1: Classification of the different commercially available IR techniques

Manufacturer	Reconstruction technique	Algorithm type	Iteration domain	Strength levels or FBP vs IR proportion	Reconstruction time compared to FBP
GE	ASIR	Hybrid	Image and sinogram space	Adjustable FBP percentage (from 0 to 100% with 10% intervals)	Comparable to FBP
	Veo	MBIR	Image and sinogram space	No strength level	Substantially higher than FBP
	ASIR-V	MBIR	Image and sinogram space	Adjustable FBP percentage (from 0 to 100% with 10% intervals)	Slightly higher than FBP
Philips	iDose^4	Hybrid	Image and sinogram space	7 strength levels	Comparable to FBP
	IMR	MBIR	Image and sinogram space		Slightly higher than FBP
Siemens	IRIS	Hybrid	Image space	No strength level	Comparable to FBP
	Safire	Hybrid	Image and sinogram space	5 strength levels	Comparable to FBP
	ADMIRE	MBIR	Image and sinogram space	5 strength levels	Slightly higher than FBP
Toshiba	AIDR	Hybrid	Image and sinogram space	3 strength levels	Comparable to FBP
	FIRST	MBIR	Image and sinogram space	Not known yet	Comparable to FBP

1.4 Usual methods for image quality assessment in CT

These new techniques all raised related questions regarding their potential, their benefits, and their use in clinical routine. This last point in particular remains critical. Indeed, IR algorithms can produce images that tend to have a very unusual *"plastic appearance"* compared to the classical FBP images [Hara, 2009]. Therefore, the radiological community legitimately asks itself the question to know "if *the appearance of those images can affect the conspicuity of patient diagnostic".* Correctly answering the relevant clinical question is fundamental to the optimisation process, since it will guarantee correct patient management.

An approach used by medical physicists to tackle this problem involves developing and using tools like physics metrics to determine the amount of information that can be found in the image, improving radiologist's confidence.

Otherwise, image quality assessment can also be done using a clinical approach which often entails asking a trained reader to use the images in order to perform a task.

1.4.1 Clinical approach

Clinical approaches, like the use of the visual grading analysis (VGA) or receiver operating characteristics (ROC) methods could be seen as a kind of gold standard with which to evaluate diagnostic image quality because they take the whole imaging chain into account. This method usually involves visually evaluating the appearance of image parts using a several levels scale to rate the visibility of the structures.

1.4.2 Physical approach

Conventional image quality metrics: use of signal processing theory

Image signal and image noise are key parameters in image quality assessment. In the ideal and linear case, image signal (*S*) is directly linked to the detected number of photons *N*, while the noise (*σ*) may be seen as the pixel's stochastic fluctuation around their mean value. The photons are distributed according to Poisson's law, meaning that the quantity σ is equal to $\sqrt{N}$. The ratio of these two quantities yields the **signal-to-noise ratio (SNR)**, expressed as:

$$SNR \propto \frac{S}{\sigma} = \frac{N}{\sqrt{N}} = \sqrt{N} \qquad (11)$$

The number of quanta N_{true} that really contribute to image formation in the device is called **noise-equivalent quanta (NEQ)**. We can thus write:

$$SNR^2_{real} = N_{real} = NEQ \qquad (12)$$

In an ideal device, each quantum is counted by the detector and contributes to the image. By estimating the ratio between this ideal photon number and the number of photons actually impinging the image detector we can compute the **detective quantum efficiency (DQE)** which characterises the efficiency of the detector as follows:

$$DQE = \frac{SNR_{real}^2}{SNR_{true}^2} = \frac{NEQ}{N_{true}} \tag{13}$$

Another global image quality index is the **contrast-to-noise ratio (*CNR*)**, defined as follows:

$$CNR = \frac{I_{object} - I_{Background}}{\sigma} \tag{14}$$

Fourier-based image quality metrics

Image quality can also be estimated in the Fourier space. In this case, the spatial resolution is assessed by estimating one of the three available "spread" functions and then converting it into the Fourier space.

The **point spread function (PSF)** corresponds to the impulse response of a system, that is to say the response of the system to a Dirac input ($\delta(x, y)$). It has the following interesting property: $I(x, y)$, which is that the image of an input object denoted by $f(x, y)$ can be expressed as the convolution product between this input object and the point spread function:

$$I(x, y) = \iint f(x - x', y - y') PSF(x', y')\, dx'\, dy' \tag{15}$$

The **line spread function (LSF)** is the response of the system to a straight line ($\delta(x)$). Thus, the relationship between the LSF and the PSF can be derived from Eq. 15, in which the input function is replaced by the equation of a straight line, yielding:

$$LSF(x) = \int_{-\infty}^{+\infty} PSF(x, y)\, dy \tag{16}$$

Besides those two metrics, it is also possible to estimate the resolution using the **edge spread function (ESF)**, that is to say the response of the device to an edge. An edge can be mathematically approached by the Heaviside function. Injecting it into Eq. 15 and using the properties of this function yields:

$$LSF(x) = \frac{\partial ESF(x)}{\partial x} \tag{17}$$

Hence, PSF, LSF and ESF are all related to each other and it is possible to use one of those in the frequency space to perform resolution estimation.
The resulting metric is called **modulation transfer function (MTF)** and can be defined as:

$$MTF(u, v) \overset{\text{def}}{=} \frac{|FT\{PSF(u, v)\}|}{|FT\{PSF(0,0)\}|} \tag{18}$$

Using the convolution product properties together with Eq. 15 and 18 shows that:

$$FT[Image] = FT[Input\ object] \times MTF \qquad \textbf{(19)}$$

This last equation clearly shows that the MTF expresses the modifications of the object's frequency spectrum by the system, offering an objective spatial resolution estimator (Figure 5). It also has the advantage of requiring a simple product to estimate the final image of the object, instead of a complex convolution when working with an image-space spread function. We will not discuss the ways to compute this metric, since lots of methods were already extensively investigated by several authors ([Nickoloff 1985; Boone 2001; Judy 1976; Nakaya, 2012]).

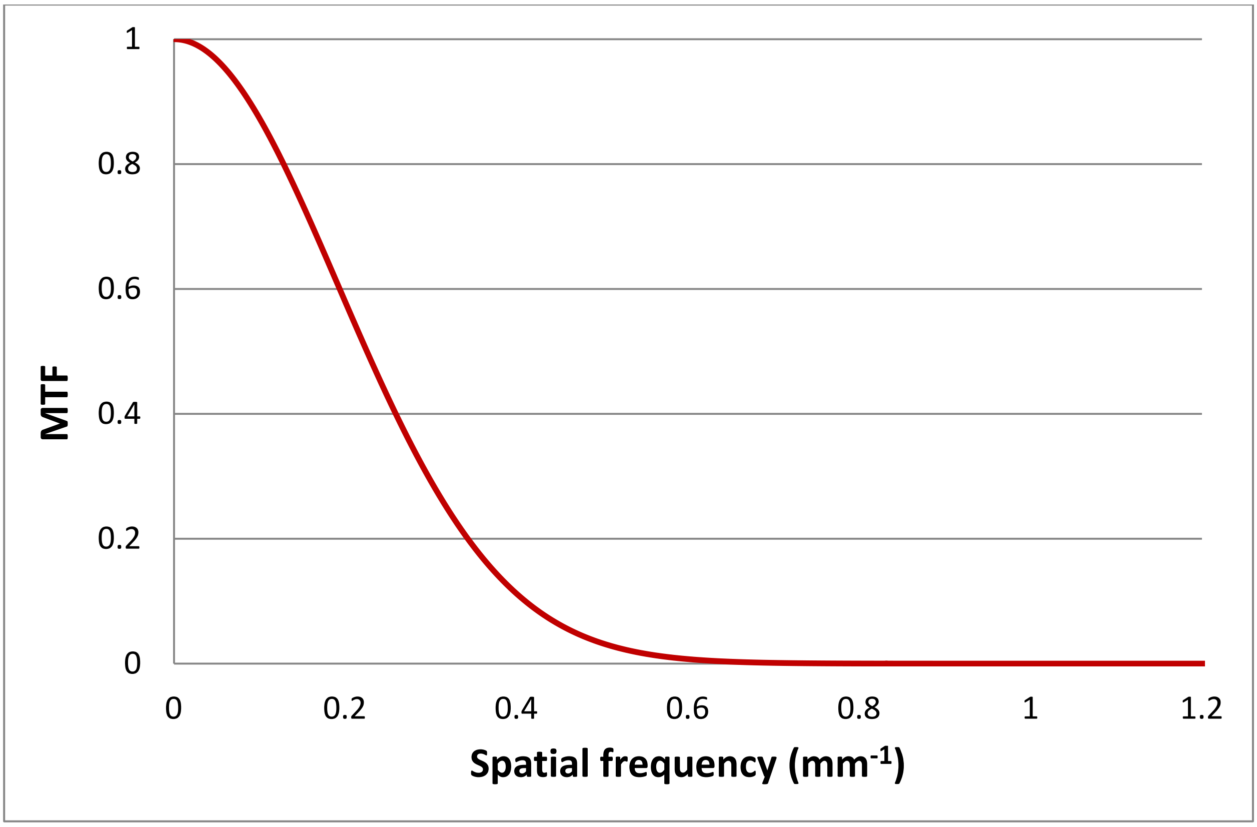


Figure 5: Example of a 1 dimensional MTF curve

As with resolution, and of equal importance for SNR transfer, image noise can also be estimated in the frequency space. Assuming noise stationarity, **the noise power spectrum (NPS)** is a frequency metric which gives a complete description of the noise by providing its amplitude over the entire frequency range of the image.

In order to compute the NPS of an image, it is necessary to acquire homogeneous CT images and select regions of interest (ROI) in this stack. The 2D *NPS* can then be computed as:

$$NPS_{2D}(f_x, f_y) = \frac{\Delta_x \Delta_y}{L_x L_y} \frac{1}{N_{ROI}} \sum_{i=1}^{N_{ROI}} |FT_{2D}\{ROI_i(x,y) - \overline{ROI_i}\}|^2 \qquad \textbf{(20)}$$

Where Δ_x, Δ_y are the pixel sizes in the x and y dimension, $\mathrm{L_x\, L_y}$ are the ROI's lengths (in pixel) for both dimensions, $\mathrm{N_{ROI}}$ is the number of ROIs used in the average operation and $\overline{\mathrm{ROI_i}}$ is the mean pixel value of the ith ROI.

Commonly, the 2D NPS is averaged along a 1D radial frequency using the equation $\mathrm{f_r} = \sqrt{\mathrm{f_x^2 + f_y^2}}$ (Figure 6). More details on NPS computing can be found in ICRU report 87 [ICRU 87, 2012].

In the end, the NPS characterises the noise texture, thus giving a more complete description of noise than the simple pixel's standard deviation (SD). Moreover, information about the pixel's SD can still be retrieved with knowledge of the Wiener spectrum. Indeed, the Parseval theorem ensures that the total energy is obtained by summing the contribution of the different harmonics and that its value does not depend on the chosen space (image or frequency space). Since the *NPS* is a spectral decomposition of noise over frequencies, we have:

$$\sigma^2 = \iint NPS_{2D}(f_x, f_y)\, df_x df_y \tag{21}$$

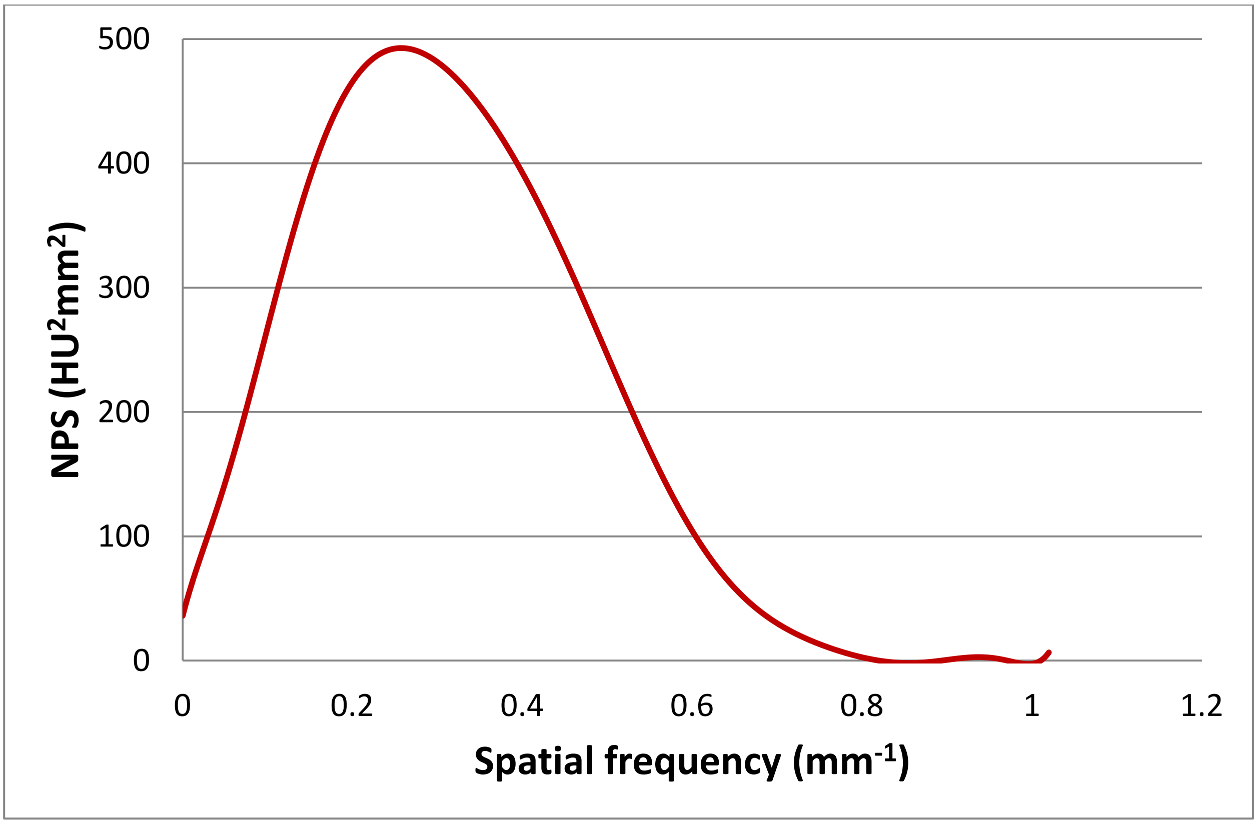


Figure 6: Example of a radially averaged NPS obtained with a standard convolution kernel

As explained before, MTF shows how well the signal frequencies are transferred through an imaging system, that is to say it exhibits the signal response of a system at a given spatial frequency. Similar to the spatial domain, the ratio of signal (i.e. MTF) and noise (i.e. NPS) yields the SNR (cf. Eq. 11 and 12). The NEQ in the frequency space can therefore be calculated as:

$$NEQ(f) = SNR^2_{Real}(f) = \frac{a^2 MTF^2_{1D}(f)}{NPS_{1D}(f)} \tag{22}$$

Where a^2 is the mean pixel value squared.
Using Eq. 13, the DQE in the frequency space can therefore be estimated by:

$$DQE(f) = \frac{SNR_{Real}^2(f)}{SNR_{Ideal}^2(f)} = \frac{a^2 MTF_{1D}^2(f)}{N_{Ideal}\, NPS_{1D}(f)} \tag{23}$$

1.4.3 Limitations of these image quality assessment methods

Both methods presented above are, however, subject to important limitations.

VGA and ROC studies exhibited in the clinical approach are very time consuming and repetitious for radiologists. Therefore, obtaining sufficient statistics often reveals itself to be complicated. Besides, reader subjectivity makes it difficult to produce a repeatable final result.

Fourier metrics also show some limitations since the reconstructed images should respect the **linearity** and **shift-invariance assumptions**. The shift-invariance hypothesis entails supposing that the device's response remains the same no matter where the measure is done on the image. The assumption of linearity entails stating that the output signal remains the same within the Hounsfield units (HU) response range of the imaging system (Clinical CT scanners usually have a linear range from -200 to +200 HU [ICRU 87, 2012]). If those assumptions are approximately satisfied for CT images reconstructed with FBP algorithms and a standard reconstruction kernel, the extent to which they hold for IR algorithms is less certain [Pineda, 2008].

In order for a MTF to be reliable, the linearity hypothesis needs to be fulfilled. Several authors have already highlighted the non-linearity properties of IR algorithms, which manifests itself as contrast dependency of the resolution [Richard, 2012]. Consequently, estimating the MTF with a high Z material that produces an impulse function can give a signal outside this range, which in practice often leads to a resolution overestimation [ICRU 87, 2012].

If image noise is not stationary, the NPS will not provide a complete description of it and the whole covariance matrix is needed for a full characterisation. However, if applied with care - for example working with small ROIs, extracted from a restricted region of the image - the NPS can be applied to both conventionally (i.e. FBP based) and iteratively reconstructed images.

Finally, these metrics are hardly linked with the actual tasks of the radiologists that should, among other tasks, detect, localise and classify image abnormalities.

These elements highlight the need to adapt existing metrics for IR algorithms. In this context, several authors have already dedicated many efforts to modify existing methods so that they fulfil the previously stated requirements [Richard, 2012; Brunner, 2013]. Those modified metrics can then be integrated as basic elements in the development of more complex models like model observer, whose role is to model the human visual system [Brunner, 2013; Ott, 2014; Solomon 2015]. This approach based on those models has been considered by some authors to be an efficient alternative to the other methods stated here [Vaishnav, 2014].

1.5 Statistical decision theory using model observers

Quantitative measurement methods based on **statistical decision theory** rely on four elements: the task (1), the signal and background properties (2), the observer (3) and the figure of merit (FOM) (4). In other words, their principle consists of measuring the performance of an observer conducting a task of clinical interest. The task may be the detection of a signal into a noisy background (classification in signal-present or signal-absent category), the observer is the person or algorithm performing the task (in the case where we use algorithms we will refer to them as model observers) and the FOM measures how well the observer performed the task.

Model Observers

Model observers could be classified into two different categories, which are **the image discrimination models** and the **models for detection in the noise** [Beutel, 2000].

Image discrimination models were developed in order to predict human visual detection of signals superimposed on a uniform or grating background. Their interest is, however, limited in the field of medical imaging. Indeed, they were designed to predict human detection of a signal superimposed on identical backgrounds, whereas radiological images display various types of backgrounds.

Models for detection in noise constitute the second category. They typically include the calculation of a distribution of statistical decision variables (often referred to as $\lambda(\boldsymbol{g})$) which is a linear function of the image data. In the vector notation of images this can be written as an inner product of a template $\boldsymbol{w}$ and an image $\boldsymbol{g}$:

$$\lambda(\boldsymbol{g}) = \boldsymbol{w}^T \boldsymbol{g} \qquad (\mathbf{24})$$

There are different types of linear model observers which all work based on Eq. 24, with the difference remaining in the expression of the template.

ICRU report 54 [ICRU 54, 1996] suggests that those methods should be used to assess diagnostic image quality in nonlinear and non-stationary domains. Indeed, a radiologist looking at an image takes a decision based on the sensitivity and specificity of the diagnosis and the costs associated with the different outcomes. Related to this, the principle on which statistical decision theory is based entails making the best possible choice provided a given amount of information. In this framework, the question of knowing how well the system provides the observer with information related to a given task rises naturally. Also, if it is understood that the SNR constitutes a key component of assessing this parameter, the choice of which linear observer model should be used is more critical. Indeed, they exhibit different properties making a given model more adapted to particular tasks and less adapted to others. However, a simple way around this is possible if we acknowledge that the imaging process can be divided in two different stages: 1) the data acquisition for which the X-ray detector captures the radiant image and conveys the information, and 2) the processing and display of the image (a step in which the observer can perform an analysis) [ICRU 54, 1996]. This differentiation in the imaging process is fortunate since each stage of the imaging process has to be assessed using a different type of observer.

The **ideal observer** utilises all statistical information available regarding the task in order to maximise the task performance as measured by Bayes risk or some other related measures of performance [He, 2013]. It corresponds to a measure of the task-related information transmitted by the system,

and can therefore be used to analyse the first stage of the imaging process. This kind of observer is usually applied in a signal-known-exactly/background-known-exactly (SKE/BKE) paradigm (signal and background are both completely specified and all *a-priori* information is known by the model). The only thing to determine then remains whether the analysed image has to be classified in signal-present or signal-absent category.
The ideal linear model observer is the Hotelling observer (HO) [Barrett, 1993]. This model works in two steps. The first is called prewhitening or noise decorrelation and entails removing noise correlations in the signal and background images using the inverse of the covariance matrix. Covariance matrix of signal-absent and signal-present images will be referred to as $\boldsymbol{K_n}$, n being a value equal to 0 or 1 and referring to the signal-absent or signal present hypothesis, respectively. The second step involves applying a filter matched to the expected signal, allowing the model to compute a metric used to proceed to the image classification (signal-absent image ($\boldsymbol{g_0}$) or signal-present image ($\boldsymbol{g_1}$)) [ICRU 54, 1996]. If the noise statistics are the same in both signal-absent and signal-present images, then signal-present and signal-absent images have no significantly different correlation matrix and HO is the prewhitening matched filter (PW) model observer. If the image noise is white, then the prewhitening step is not required and the HO is identical to the non-prewhitening matched filter (NPW) model observer. Detailed explanations of these correspondences based on their template's expression are given in Figure 7.
A difficulty often encountered with the implementation of the HO remains the size of the covariance matrix. An image with the size NxN will indeed produce a N^2xN^2 covariance matrix, which will be almost impossible to inverse (typical values for N often being around 128 pixels in our work). A process relying on the use of channels (which are basically frequency filters) is therefore employed to reduce the size of the covariance matrix. Using a set of 10 channels for example, will then produce a 10x10 covariance matrix (which we will refer to as $\boldsymbol{K_{vn}}$**)** and that can be inverted more easily. When using this channelization process, it is of note that the signal-absent and signal-present images we are using also have to be filtered through the channels (they will then be referred to as v_0 and v_1**).** The HO model observer therefore becomes the channelized Hotelling observer (CHO) [Myers, 1987]. Extensive details of this process are given in Volume 1 of the Handbook of Medical imaging [Beutel, 2000] or Barrett's work [Barrett, 2004].

The characterisation of the observer's ability to extract the information from the image corresponds to the assessment of the second step of the image formation (display and processing of the image). This step is usually assessed with **anthropomorphic model observers**. These models are more likely to approximate the performances of human observers thanks to the incorporation of some anthropomorphic features. Among these models, the non-prewhitening matched with eye filter (NPWE) works similarly to the NPW, but also integrates a filter called the contrast sensitivity function (CSF) which reproduces the sensitivity of the human eye to the frequency of some patterns [Burgess, 1994]. CHO can also belong to anthropomorphic model observers, provided an anthropomorphic set of channels be used, like for example the dense difference of Gaussians channel (DDoG) set [Abbey and Barrett, 2001]. Since our objective was to be able to tell the radiologist if the diagnostic information could be extracted from images produced with IR, we naturally relied on a model closely aligned with human observer (that is to say anthropomorphic models like NPWE or CHO) in most of our studies. Extensive details on NPWE and CHO models can therefore be found in the material and methods section of subpart 3.2.2. or in Barrett's work [Barrett, 2004] as well as. Figure 7 provides a comprehensive overview of the overall classifications of all previously discussed linear model observers with their template.

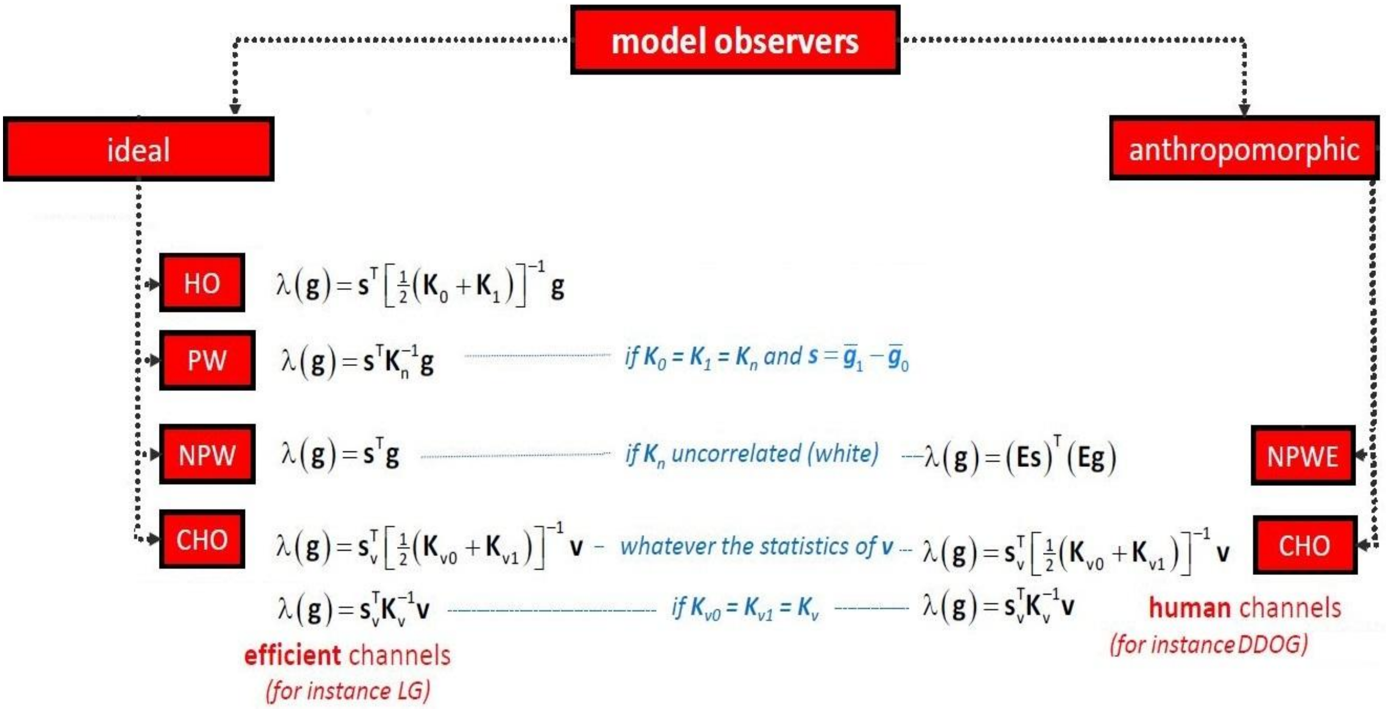

model observers
ideal
anthropomorphic
HO
$\lambda(g) = s^T\left[\frac{1}{2}(K_0 + K_1)\right]^{-1} g$
PW
$\lambda(g) = s^T K_n^{-1} g$
if $K_0 = K_1 = K_n$ and $s = \bar{g}_1 - \bar{g}_0$
NPW
$\lambda(g) = s^T g$
if K_n uncorrelated (white)
$\lambda(g) = (Es)^T (Eg)$
NPWE
CHO
$\lambda(g) = s_v^T\left[\frac{1}{2}(K_{v0} + K_{v1})\right]^{-1} v$
whatever the statistics of v
$\lambda(g) = s_v^T\left[\frac{1}{2}(K_{v0} + K_{v1})\right]^{-1} v$
CHO
$\lambda(g) = s_v^T K_v^{-1} v$
if $K_{v0} = K_{v1} = K_v$
$\lambda(g) = s_v^T K_v^{-1} v$
human channels
(for instance DDOG)
efficient channels
(for instance LG)

Figure 7: Classifications of some linear model observers with their scalar distribution calculation.

In Figure 7, the template's expressions are given in the image space domain, but these calculations can also be done in the Fourier space domain. To do this, Fourier metrics like the one detailed in subsection 1.4.2 are traditionally used to compute a SNR (correspondences between formulas in image and those in the Fourier space are given in Chapter 10 of the Handbook of Medical Imaging [Beutel, 2000]).

Human observers

The gold standard for the assessment of image quality is human observer performance. Therefore, people often undergo alternative forced choice (AFC) tests in order to produce an indicator of their performance. This indicator is directly linked to the image quality level. Figure 8 provides an example of a 4-AFC test, in which a human observer has to select the signal-present image in a bunch of 3 signal-absent images and one signal-present image. Human observers undergo this trial several times, using randomly selected signal-absent and signal-present-images and leading to a percentage of correct responses (PC). This PC is then compared with the FOM yielded by the model observer.

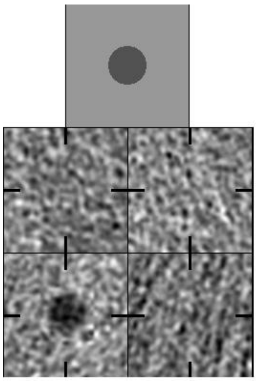


Figure 8: Example of a 4-AFC test

Some usual FOM and their relationships together

As we have seen, the model observers used yield a distribution of the decision variable λ(***g***)***.*** Using this statistical distribution, it is possible to construct a curve displaying the false positive fraction (FPF), true positive fraction (TPF), false negative fraction (FNF) and true negative fraction (TNF) (cf. Figure 9). Thanks to the ROC theory [Metz, 2006], a ROC curve can then be plotted by varying the decision threshold *t.* This curve displays the sensitivity versus (1-specificity), and the area under the curve (AUC) can then be used as an FOM. An area equal to 1 indicates a perfect detectability whereas an area of 0.5 would indicate a guessing observer.

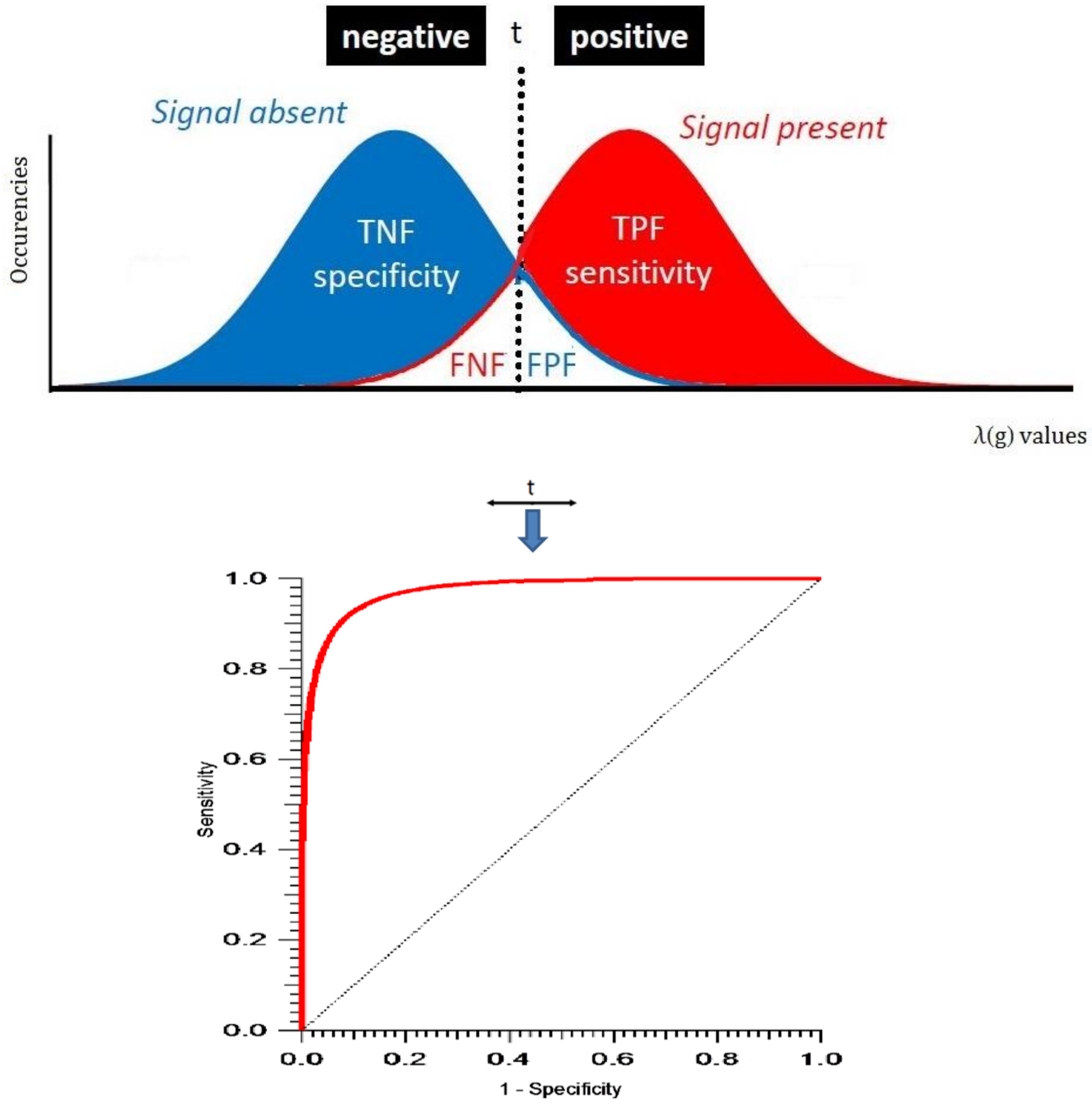


Figure 9: Plot of the repartition of the scalar response λ(***g***) and the subsequent ROC curve.

Model observer and human results can also be expressed in terms of detectability index (often referred to as *d')* instead of scalar values distribution (λ(***g***)), PC or AUC as FOM.

Correspondences between d', AUC and PC exist and make it possible to compare human and model observer results using the same FOM. Figure 10 gives an overview of the ways to perform the conversions. On this figure, $\phi(\lambda)$ represents the distribution of the λ and $\Theta(\lambda)$ represents the cumulative of this distribution. Extensive details on the formula and their demonstration can be found in the Handbook of Medical Imaging [Beutel, 2000] as well as ICRU report 54 [ICRU 54, 1996].

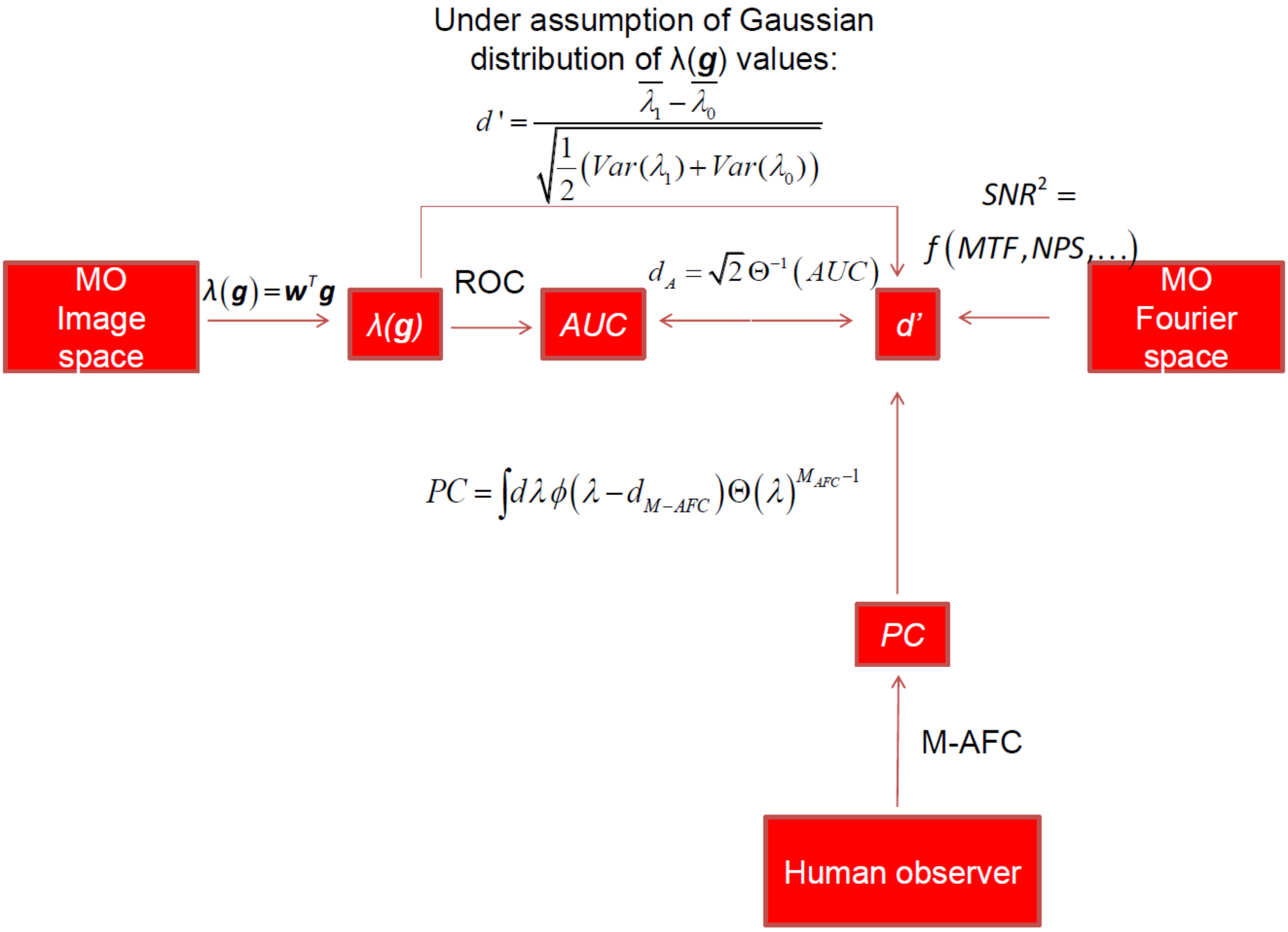

Under assumption of Gaussian distribution of λ(**g**) values:
$d' = \frac{\overline{\lambda_1} - \overline{\lambda_0}}{\sqrt{\frac{1}{2}(Var(\lambda_1) + Var(\lambda_0))}}$
$SNR^2 =$
$f(MTF, NPS, \dots)$
MO Image space
$\lambda(\mathbf{g}) = \mathbf{w}^T \mathbf{g}$
λ(g)
ROC
AUC
$d_A = \sqrt{2}\,\Theta^{-1}(AUC)$
d'
MO Fourier space
$PC = \int d\lambda\, \phi(\lambda - d_{M-AFC})\, \Theta(\lambda)^{M_{AFC}-1}$
PC
M-AFC
Human observer

Figure 10: Relationships between different FOM.

All this shows that quantitative measurement methods based on statistical decision theory can overcome some problems posed by traditional image quality assessment metrics and are well adapted to assess how well the desired information can be extracted from the image.

1.6 Goals of the study

CT technology recently underwent a fast evolution, including e.g. the introduction of IR algorithms. Since then, it is difficult for medical physicists to objectively assess image quality. Indeed, in order to try to understand the perceived effect of IR on clinical images (in particular some changes of image texture), medical physicists used NPS where signal variations remained compatible with the hypothesis of linear signals [Miéville, 2011]. If these kinds of metrics have revealed themselves to be very useful in establishing a link between a physical metric and radiologist perception, it was limited to a texture analysis without being able to assess the performance of structure detection.

Our goal was therefore to propose a strategy to ensure that dose reductions applied when using the latest IR techniques remain compatible with radiologist requirements. In short, we wanted to investigate by how much one could reduce patient exposure with these new tools. We used two successive approaches to tackle this problem: a clinical approach followed by a more physical approach.

The first step of our work was a continuation of the work previously performed and entailed applying the existing tools like Fourier metrics in the context of radiological studies. We focused in particular on image texture and spatial resolution. This allowed us to strengthen our communication with radiologists and may help them in the choice of the best option (choice of IR strength level for example) when dealing with the use of IR.

Then, we focused on the use of special image quality tools (mathematical model observers), making it possible to predict the detection of simple structures in a homogeneous background.

One of these model observers, (the NPW model observer), was applied to characterise image quality when dealing with the detection of high-contrast structures. Since detection per se is not of major concern (easy task) we focused on analysing the way spatial frequency were transferred as a function of image noise and image contrast using a recently introduced metric: Target Transfer Function (TTF). Note that the use of this TTF in a standard model observer allowed us to establish a link with clinical data.

Then we investigated a more challenging problem—taking care of the low-contrast detectability (LCD) which requires the use of more elaborated model observers, namely anthropomorphic model observers.

Eventually, we used the same tools to investigate the changes in image quality that occur when radiologists switch reconstruction plane.

These project’s milestones are summarised in **Chapter 2**.

2. PhD thesis milestones

The peer-review papers of **Chapter 3** compose the core of this thesis, which was developed around two milestones:

A *"Clinical approach",* in which we investigated how the physicist could help the radiologist with the integration of IR techniques into the clinical routine, using the existing tools at our disposal.

A *"Physical approach",* which was exclusively based on phantoms and entailed developing and testing fully IR-adapted tools in order to perform an efficient optimisation scheme.

2.1 Clinical approach: Applying existing methods to improve the use of IR in the clinic

Classical Fourier and image space metrics like MTF, NPS or CNR and SNR have been recognised as being valuable FOM for evaluating and comparing system properties or for improving and optimising system performances. Since those tools were already widely used, understood and studied, we relied on them to help the radiologists in parameter selection when acquiring patient images with the different IR solutions available.

2.1.1 Optimisation of IR strength levels for clinical cervical spine images

The very first part of our work consisted of helping our radiologist determine the optimal IR strength level to use when performing a low-dose multi-detector CT (MDCT) of the cervical spine [Omoumi, 2014]. The study was conducted on clinical CT images of the cervical spine, reconstructed using FBP and IR (Siemens SAFIRE, strength levels 1 to 5). Images were acquired at a dose level classically used in the clinic for those kinds protocols and reconstructed using both standard and bone convolution kernels. Our task was to measure image quality with physical metrics like NPS (Figure 11), SNR and contrast-to-noise ratio (CNR). Those physical parameters associated with ratings of the overall diagnostic image quality by radiologists allowed us to determine the strength levels needed to perform an accurate diagnosis in that particular clinical context.

The results showed that the optimal strength level of IR in the context of low-dose cervical spine MDCT could be radically different depending on the anatomical structure to be analysed. Indeed, we witnessed that an increase of IR strength levels led to an overall noise reduction as well as an improvement of both SNR and CNR. This suggests at first sight, that high IR levels will produce higher image quality. But this trend revealed itself to be true only for some anatomical structures like intervertebral discs, the content of neural foramina and dural sac. The radiologist's analysis showed that other structures like soft tissues were more visible when using low IR levels.

With this study we were able to advise radiologists on the optimal IR strength level to use depending on the anatomical structure to be analysed. We also witnessed a first limitation of our image quality assessment method. Indeed, focusing only on noise reduction by increasing strength level of IR can lead to a drop in image conspicuity. This observation highlights the fact that the way spatial frequencies of the signal are transferred also needs to be analysed if we are willing to fully understand the change of appearance of the anatomical structures.

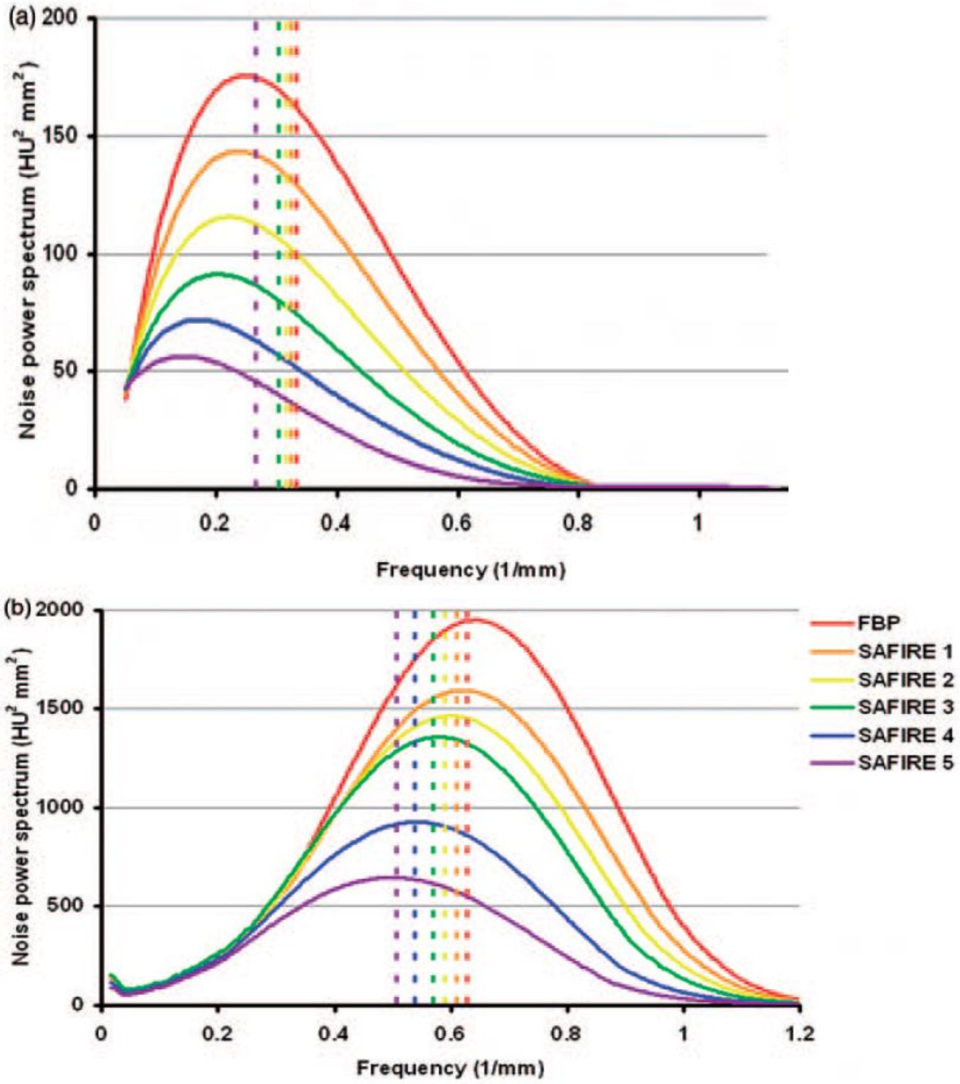


Figure 11: NPS for FBP and different IR levels, using standard (a) and bone (b) convolution kernels. A clear diminishment in the noise amplitude is visible when switching from FBP to IRs.

2.1.2 Use of ASIR for dose reduction in hip clinical images

The second part of our work aimed at assessing the effects of the ASIR algorithm on image quality in hip CT arthrography, as well as evaluating its potential for reducing radiation dose [Tobalem & Dugert, 2014]. Patients were randomised into three different protocols: one with a regular dose ($CTDI_{vol}$ 38.4 mGy) and two with a reduced dose ($CTDI_{vol}$, 24.6 or 15.4 mGy). Images were reconstructed using FBP and four increasing percentages of ASIR (30%, 50%, 70%, and 90%). We measured MTF, NPS (Figure 12) and CNR in order to assess image quality objectively. At the same time, two radiologists independently evaluated several anatomical structures and image quality parameters using a VGA.

As expected, the reduction of patient exposure was associated with a significant increased of image noise and a significant decrease of CNR. At the same time, we found that increasing the percentage of ASIR produced a significant reduction in noise as well as an increase of the CNR. In addition, we noted an enhancement of overall image diagnostic quality when using higher ASIR levels (up to a limit of 50% ASIR). Eventually, we determined a dose reduction that maintained image quality. Our work showed that the use of up to 50% ASIR in hip CT arthrography helps to reduce radiation dose by approximately 35–60%, while maintaining diagnostic image quality.

Compared to the previous one, this study incorporated several evolutions, including an estimation of the potential of the method in order to reduce the dose as well as the assessment of signal frequencies transfer in addition to noise characterisation. This step was, however, carried out using the MTF, and this metric should be used with caution in the context of IR, since it depends on signal amplitude, due to IR non-linearity. An alternative would be to measure a true MTF, for example, using true sine wave signals with different amplitudes. As this method is hardly achievable, alternative methods should be proposed, like the TTF, which is described in the next part.

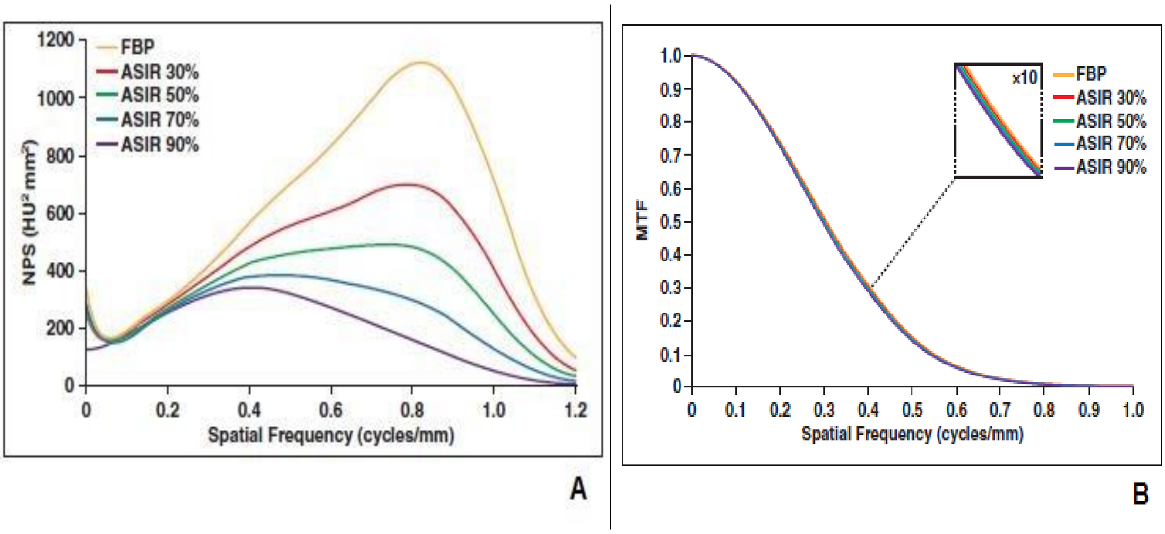


Figure 12: Strong noise reduction is observed when switching from FBP to growing levels of IR (A). No changes in resolution as estimated by MTF are observed when switching from FBP to IRs (B).

2.1.3 Using iterative reconstruction techniques for optimising radiation dose and image quality in musculoskeletal CT

This work [Omoumi, 2015] is a review based on the use of IR techniques in the context of musculoskeletal imaging. It basically summarises the progress that IR brought to the field regarding the dose/image quality compromise.

2.2 Physical approach: Going further by developing fully adapted physical methods in order to analyse IR images

The second part of this project is composed of works relying on physical methods in order to help the radiologist determine whether or not diagnostic information is present when reducing the dose with IR algorithms.

The physical methods we used involve a first step of adapting Fourier metrics to IR algorithm requirements for image quality assessment. Then, integration of those modified metrics in a Fourier space model observer (NPW, see Figure 7) so that high-contrast low-size detection could be performed. On top of that, the use of image space model observers for LCD also constitutes a reliable tool (CHO, see Fig. 7).

2.2.1 Update of the NPW model observer: Specific assessment of ASIR and MBIR algorithms

Our previous work showed that image quality could be well estimated with objective physical metrics, and that extrapolation for dose reduction could even be made based on that. However, a review of the literature revealed a need to use metrics which take the specificity of IR algorithms into account [Richard et al. 2012]. Indeed, it was shown that computing the MTF, for example, poses a problem when dealing with IR algorithms, such as ASIR or MBIR. Given that the TTF had already shown it could accurately express the system resolution even with non-linear algorithms, we decided to replace the standard MTF by the TTF in our characterisations [Richard, 2012; Brunner, 2013]. It was estimated using a custom-made phantom containing cylindrical inserts made of three different materials (Teflon, PMMA and Plexiglas) and surrounded by water (Figure 13). The contrast differences between the inserts and the water were plotted for each acquisition condition, leading to three different ESFs. Then, mathematical transformations as described in part 1.4.2 were performed on those three ESFs, leading to the TTF curves (Figure 14).

As expected, the first results showed the image contrast and noise levels to be dependent on the TTF for both ASIR and MBIR, revealing their nonlinear behaviour. In order to go one step further in our image characterisation we chose to integrate those results into a model observer. In this study, we only focused on the objective assessment of the tested reconstruction techniques using adapted metrics and models. No subjective assessment was then performed by the radiologists in order to be compared to the models' results. Given this experimental paradigm the choice of an ideal model observer was well adapted. Thus, we used the NPW model observer in the Fourier space to objectively characterise the evolution of its figure of merit in various acquisition conditions. The NPW model observer usually requires the use of the MTF, but since IR assessment is performed, we replaced it by the TTF. Our final results showed an enhancement of detectability index when switching from FBP to ASIR to MBIR. Also, IR algorithms greatly improved image quality in low-dose conditions. Based on our results, we determined that the use of MBIR could lead to further dose reduction in several clinical applications.

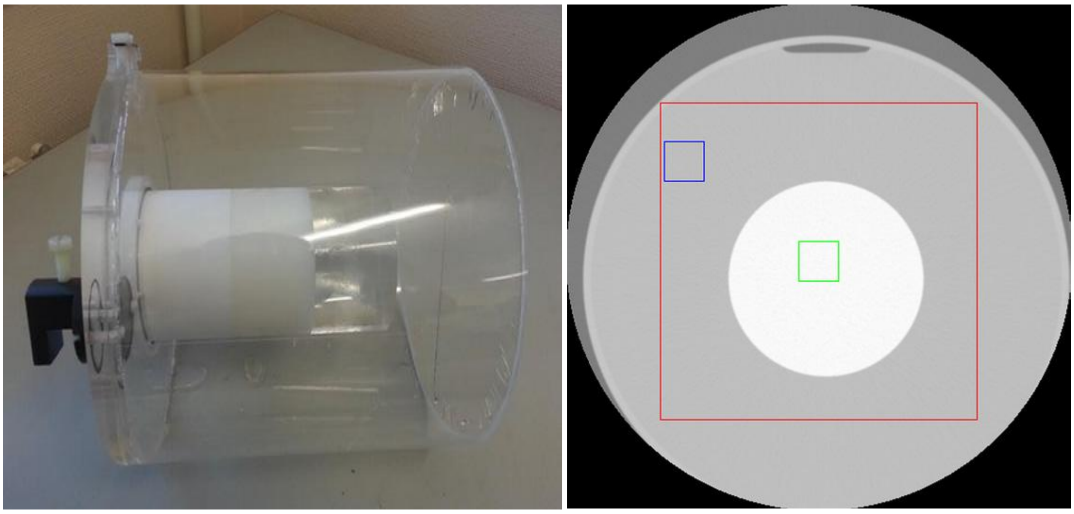


Figure 13: The home-made phantom and a CT image of a slice with the ROIs used for TTF computation.

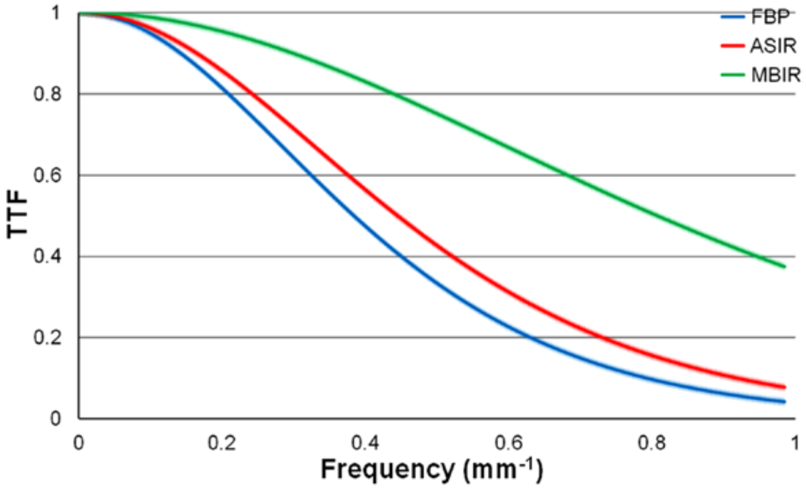


Figure 14: Resolution estimation through the TTF. Clear enhancement is observed with IR, in particular with MBIR. These kinds of changes could not be observed when using MTF.

2.2.2 Assessment of low-contrast detectability in CT using image space model observers

Objective physical metrics, (performed by medical physicists) as well as clinical assessments (performed by radiologists) represent two standard and well-described ways to assess image quality in CT. However, the biggest problem of those methods remains that radiologists and physicists do not dispose of a common method to work on the topic. Anthropomorphic model observers could represent an efficient solution to that problem, since one of their main properties is to attempt to mimic the human visual system by integrating some of their properties. The goal of these studies [Ott, 2015 a; Ott, 2015 b] was to find out if some selected anthropomorphic model observers (NPWE and CHO in this case) are adapted to match human detection in order to propose a fast and reliable methodology when dealing with the assessments of low-contrast structures in CT medical images.

For that, a modified QRM chest phantom containing spheres of 5 and 8 mm at contrast levels of 10 and 20 HU under 120 kVp will be used. Images of the phantom were acquired at several CTDI$_{vol}$ before being reconstructed using FBP, ASIR 40% and MBIR in the first study, and ADMIRE at two strength levels in the second one. At the same time, these images were also assessed by human observers undergoing 4-AFC test, in order to establish a set of reference data. These data were compared with the results obtained from CHO and NPWE model observers (Figure 15). Results of both model observers were obtained by working in the image space, since our experimental paradigm (low-contrast target phantom) did not allow us to work in the Fourier space. Eventually, the impact of the acquisition conditions as well as reconstruction methods were investigated in the study, and a potential dose benefit was estimated.

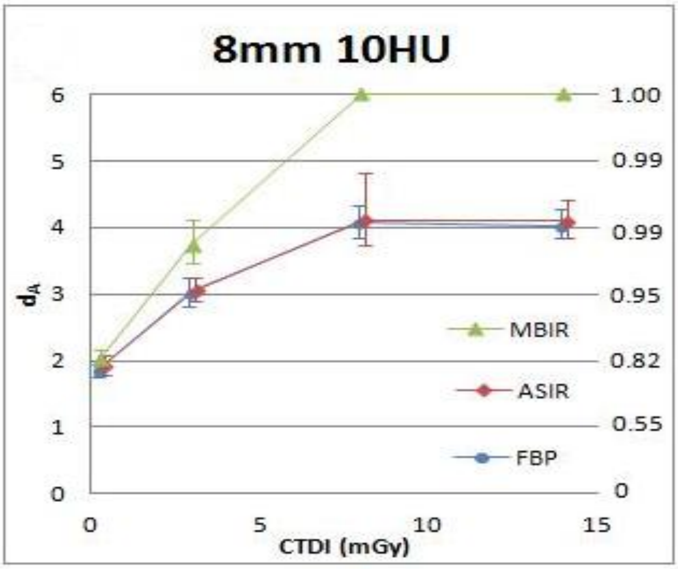


Figure 15: Results for the CHO model observer with the three reconstruction techniques for a target of 8 mm and 10 HU of contrast. (Detectability values are on the left y axis and corresponding values in term of probability of detection are represented on the right y axis).

2.2.3 Image quality assessment in different reconstruction planes using an updated NPWE model observer

This work [Ott, 2015 c] studies the impact of the choice of the reconstruction plane on image quality when using several IR types. We relied on an updated NPWE model observer in the Fourier space to assess image quality objectively. The updates to this model observer were similar to those already performed in [Ott, 2014], that is to say use of the TTF instead of MTF in order to take IR nonlinearity into account. However, we used a NPWE rather than an NPW model observer in this study in order to integrate some properties of the human visual system into our model. This was done because of the clinical paradigm we worked with (simulation of cartilage lesions) and could allow us to compare the results obtained with our model to the radiologists' impression in the future. The choice to work in the Fourier space was motivated by our experimental paradigm in which we used high-contrast targets, allowing the computation of elements to estimate the detectability in the Fourier space.
We observed an increase of spatial resolution and image noise in both the coronal and sagittal plane compared to the axial plane, at least when using an MBIR like Veo 2.0. On top of that, the Veo 3.0 algorithm was also tested. With this new MBIR, the image model and noise model are improved to make the noise covariance more isotropic in all three dimensions and further increase the image quality. Promising results were obtained in terms of image quality were obtained in all reconstruction plane, since the SNR obtained with the 3.0 version surpassed those of the 2.0 version in all planes.

3. Papers

The following papers are enclosed in this section. They are the development of the main points listed in **Chapter 2**.

3.1 Clinical approach

- [**Omoumi, 2014**] P. Omoumi, F. R. Verdun, Y. B. Salah, B. C. Berg , F. E. Lecouvet, J. Malghem, J. G. Ott, R. Meuli, F. Becce, 2014 Low-dose multidetector computed tomography of the cervical spine: optimization of iterative reconstruction strength levels. Acta Radiologica 55(3) pp. 335-344.

- [**Tobalem & Dugert, 2014**] F. Tobalem, E. Dugert, F. R. Verdun, V. Dunet, J. G. Ott, H. A. Rudiger, S. Cherix, R. Meuli, F. Becce 2014 MDCT Arthrography of the Hip: Value of the Adaptive Statistical Iterative Reconstruction Technique and potential for Radiation Dose Reduction. AJR **203(6)**:W665-73

- [**Omoumi, 2015**] P. Omoumi, F. R. Verdun, J. G. Ott, D. Racine, F. Becce 2015 Optimization of radiation dose and image quality in musculoskeletal CT: Emphasis on iterative reconstruction techniques (part 1). Semin Musculoskelet Radiol **19:** 415-21

Original Article

Acta Radiologica
0(0) 1–10
! The Foundation Acta Radiologica 2013
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0284185113494981
acr.sagepub.com
SAGE

Low-dose multidetector computed tomography of the cervical spine: optimization of iterative reconstruction strength levels

Patrick Omoumi[1], Francis R Verdun[2], Yosr Ben Salah[1],
Bruno C Vande Berg[1], Frederic E Lecouvet[1],
Jacques Malghem[1], Julien G Ott[2], Reto Meuli[3] and Fabio Becce[3]

Abstract

Background: Iterative reconstruction (IR) techniques reduce image noise in multidetector computed tomography (MDCT) imaging. They can therefore be used to reduce radiation dose while maintaining diagnostic image quality nearly constant. However, CT manufacturers offer several strength levels of IR to choose from.

Purpose: To determine the optimal strength level of IR in low-dose MDCT of the cervical spine.

Material and Methods: Thirty consecutive patients investigated by low-dose cervical spine MDCT were prospectively studied. Raw data were reconstructed using filtered back-projection and sinogram-affirmed IR (SAFIRE, strength levels 1 to 5) techniques. Image noise, signal-to-noise (SNR) and contrast-to-noise (CNR) ratios were measured at C3–C4 and C6–C7 levels. Two radiologists independently and blindly evaluated various anatomical structures (both dense and soft tissues) using a 4-point scale. They also rated the overall diagnostic image quality using a 10-point scale.

Results: As IR strength levels increased, image noise decreased linearly, while SNR and CNR both increased linearly at C3–C4 and C6–C7 levels ($P < 0.001$). For the intervertebral discs, the content of neural foramina and dural sac, and for the ligaments, subjective image quality scores increased linearly with increasing IR strength level ($P ::: 0.03$). Conversely, for the soft tissues and trabecular bone, the scores decreased linearly with increasing IR strength level ($P < 0.001$). Finally, the overall diagnostic image quality scores increased linearly with increasing IR strength level ($P < 0.001$).

Conclusion: The optimal strength level of IR in low-dose cervical spine MDCT depends on the anatomical structure to be analyzed. For the intervertebral discs and the content of neural foramina, high strength levels of IR are recommended.

Keywords

Multidetector computed tomography, filtered back-projection, iterative reconstruction, image quality, radiation dose, cervical spine

Date received: 5 March 2013; accepted: 31 May 2013

Introduction

Multidetector computed tomography (MDCT) of the cervical spine is generally the first-line imaging examination following trauma (1,2). It is also used to assess spondylosis and disc herniation (3–5). Although magnetic resonance imaging (MRI) is the examination of reference to evaluate cervical myelopathy, MDCT has proven to be as accurate in case of neural foraminal stenosis, particularly for distinguishing soft disc herniations from osteophytes (3,4). Moreover, CT can be performed when MRI is contraindicated, complicated (e.g. due to claustrophobia) and/or inconclusive

---

[1] Department of Radiology, Cliniques Universitaires Saint-Luc, Université Catholique Louvain, Brussels, Belgium

[2] Institute of Radiation Physics, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland

[3] Department of Diagnostic and Interventional Radiology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland

Corresponding author:
Fabio Becce, Department of Diagnostic and Interventional Radiology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Rue du Bugnon 46, 1011 Lausanne, Switzerland.
Email: fabio.becce@chuv.ch

(e.g. because of motion artifacts). However, it is still associated with a substantial radiation dose, particularly in the spine (5,6).

Among the tools developed by CT manufacturers to limit or reduce radiation dose (7), iterative reconstruction (IR) techniques have only recently been introduced in routine clinical practice, thanks to significant advances in computational power (8). Unlike traditional filtered back-projection (FBP) algorithms, these techniques are based on a correction loop during the image reconstruction process (8,9). By reducing image noise without altering CT attenuation values, IR methods can increase both signal-to-noise (SNR) and contrast-to-noise (CNR) ratios. Such a gain in objective image quality can be used to reduce radiation dose, while maintaining an acceptable diagnostic image quality. Dose reductions of approximately 50% and 40% have indeed been recently reported in MDCT of the lumbar (10) and cervical (11) spine, respectively.

Several strength levels of IR are available for each technique depending on the manufacturer (8,9). The latter pragmatically recommend a medium strength level of IR so as not to confuse radiologists (used to read FBP-reconstructed images), since the image appearance becomes ''smoother'' or more ''blotchy'' with increasing IR strength levels (8,9,12). To our knowledge, the optimal strength level of IR has however not yet been established in MDCT of the spine.

Therefore, the goal of our study was to determine the optimal strength level of IR to be applied in low-dose cervical spine MDCT.

Material and Methods

Patients

Over a 3-month period (October to December 2011), 31 consecutive patients investigated by low-dose cervical spine MDCT for refractory cervical pain and/or cervicobrachial neuralgia were prospectively enrolled in this single-center study. One patient was subsequently excluded because of metallic artifacts related to implanted surgical material. Hence, the final study group consisted of 30 patients (19 women, 11 men; mean age, 56.1 years; range, 18–77 years).

This study was approved by the institutional review board, with waiver of patient informed consent.

MDCT protocol

All examinations were performed using a 40-detector row MDCT scanner (Somatom Definition AS; Siemens Healthcare, Erlangen, Germany), without intravenous administration of iodinated contrast material. The patients were positioned supine, with their head first on the table. After obtaining CT scout views, data were acquired from the C3 to T1 vertebrae by applying the following parameters: tube voltage, 120 kVp; reference tube current-time product, 150 mAs; effective tube current-time product, 120–201 mAs using CARE Dose 4D (Siemens Healthcare); detector configuration, 40 x 0.6 mm; pitch, 0.8; gantry rotation time, 1 s.

The raw data were reconstructed using both a traditional FBP and an IR (sinogram-affirmed iterative reconstruction, SAFIRE, strength levels 1 to 5; Siemens Healthcare; Figs. 1 and 2) algorithm. As compared to other image domain-based techniques, SAFIRE uses a noise-modeling algorithm supported by the raw data (i.e. sinogram data) and aiming at reducing noise while maintaining image sharpness (9,13,14). Over the course of up to five iterations, this model estimates the noise content in each image pixel by analyzing the raw data contributing to the pixel, and removes it from the current data-set (13). In a validation loop, the result is compared to the initial data, generating an update image, and added to the previous data-set before the next iteration begins. Each iteration leads to further noise reduction. Besides, the strength level of SAFIRE is not related to the number of iteration loops (14). The image reconstruction parameters were as follows: field-of-view, 12 x 12 cm; section thickness/increment, 0.75/0.75 mm; soft tissue (B41s and I41s for the FBP and IR algorithms, respectively) and bone (B70h and I70h, respectively) convolution kernels. The reconstruction time varied with the strength level of IR, lasting up to 1 min 35 s for SAFIRE 5, compared to <20 s for FBP images.

Noise power spectrum analysis

In order to characterize the noise frequency bandwidth transferred during image reconstruction with FBP (B41s and B70h convolution kernels for soft tissue and bone, respectively) and SAFIRE (I41s and I70h convolution kernels, respectively; strength levels 1 to 5) algorithms, the noise power spectra (NPS) of MDCT images of water-filled phantom were calculated for each reconstruction algorithm, using the same method as described in detail elsewhere (11).

Radiation dose estimations

This MDCT protocol yielded a mean ± standard deviation (SD) $\mathrm{CTDI_{vol}}$ (expressed in the 16-cm diameter CTDI phantom) and dose-length product (DLP) of 22.8 ± 2.6 mGy and 272.9 ± 62.2 mGy x cm, respectively. After multiplying the DLPs by the appropriate conversion coefficient (i.e. $0.0051\ \mathrm{mSv \times mGy^{-1} \times cm^{-1}}$ (15)), the mean ± SD effective dose was estimated to be 1.39 ± 0.32 mSv.

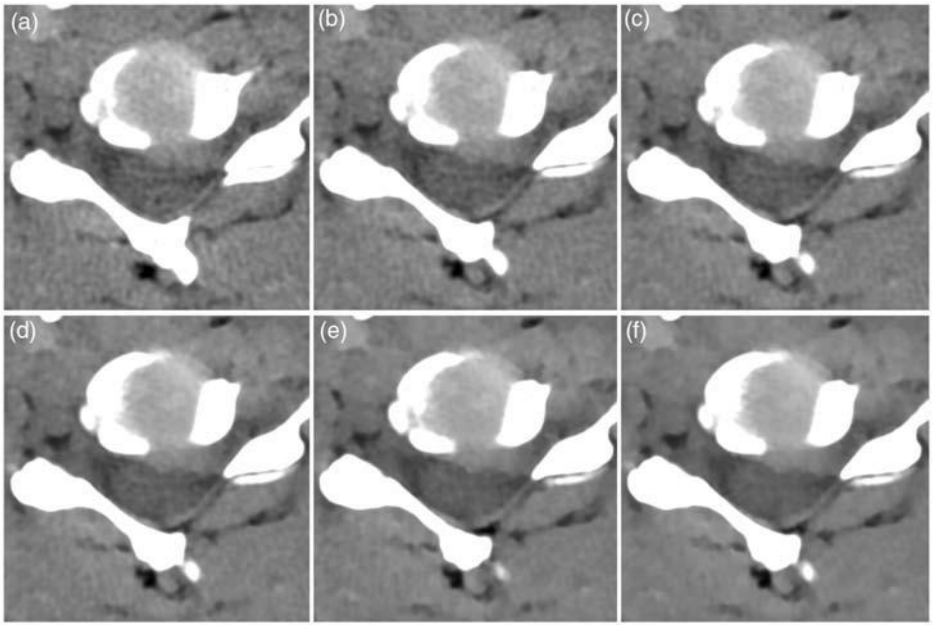


Fig. 1. A 34-year-old woman with left C7 radiculopathy. Axial unenhanced MDCT images (soft tissue convolution kernel; window level/width, 60/300 HU) of the cervical spine at C5–C6 level show a central-left subarticular focal disc protrusion. The images are reconstructed using either a (a) filtered back-projection (FBP) or (b–f) sinogram-affirmed iterative reconstruction (SAFIRE) algorithm with strength levels of (b) 1, (c) 2, (d) 3, (e) 4, and (f) 5, respectively. Note the change in appearance and sharpness of the intervertebral disc and other anatomical structures with increasing strength level of IR.

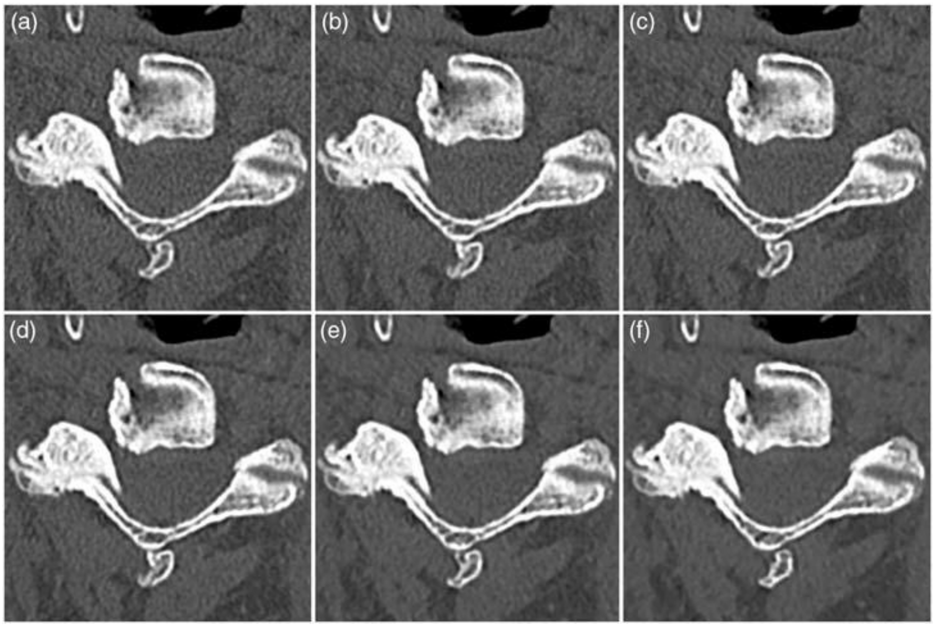


Fig. 2. A 77-year-old man with chronic cervical pain and right C4 radiculopathy. Axial unenhanced MDCT images (bone convolution kernel; window level/width, 400/2000 HU) of the cervical spine at C3–C4 level show a right neural foraminal stenosis caused by hypertrophy of the facet and uncovertebral joints. The images are reconstructed using either a (a) filtered back-projection (FBP) or (b–f) sinogram-affirmed iterative reconstruction (SAFIRE) algorithm with strength levels of (b) 1, (c) 2, (d) 3, (e) 4, and (f) 5, respectively. Note the subtle change in appearance and conspicuity of the cortical and trabecular bone of vertebrae with increasing strength level of IR.

Image analysis

All examinations were independently and blindly reviewed by two musculoskeletal radiologists (Observers 1 and 2, working in two different institutions, and with 4 and 7 years of experience in spine imaging, respectively) on a picture archiving and communication system (PACS) workstation (Carestream Client version 11.3; Carestream Health, Rochester, NY, USA). Before the analysis and in order to get used to the appearance of IR-processed MDCT images, the readers jointly reviewed and rated a selection of 10 examinations performed in the same fashion as for the study but not included in the study group.

Quantitative analysis

The measurements were performed by Observer 1 on MDCT images reconstructed using soft tissue convolution kernels, according to the same method as described in detail elsewhere (11). The regions of interest (ROIs) were directly copied and pasted from the FBP to the five corresponding IR images to be exactly at the same position. The image noise, defined as the SD of the mean CT numbers (CT_n, i.e. Hounsfield unit [HU]) within a ROI, was measured in the posterior paraspinal muscles. The SNR and CNR were subsequently calculated according to the following formulas (16):

- SNR ¼ mean CT_n/SD, within the ROI drawn in the posterior paraspinal muscles
- CNR ¼ � mean CT_n/[(:E SD)/2], within the two ROIs drawn in the intervertebral disc and dural sac, respectively

Qualitative analysis

The sharpness of several anatomical structures (the intervertebral discs, the content of neural foramina and dural sac, the flavum and posterior longitudinal ligaments, the subcutaneous tissue and paraspinal muscles, the cortical and trabecular bone of vertebrae), the presence of streak artifacts, and the overall diagnostic image quality (in terms of diagnostic acceptability) were assessed at both C3–C4 and C6–C7 levels, using the same semi-quantitative scales described in detail elsewhere (11). This image grading system was inspired by the European guidelines on quality criteria for CT (17).

Statistical analysis

The data were processed using a statistical software package (MedCalc version 11.6; MedCalc Software, Mariakierke, Belgium). Univariate one-way repeated measures analyses of variance (ANOVAs) were used to assess the impact of the five IR strength levels on the quantitative and qualitative analysis. A trend analysis was further performed to detect linear or non-linear (i.e. quadratic or cubic) trends. In addition, pairwise comparisons were performed between image quality scores obtained with IR 5 and other IR strength levels. The Bonferroni correction for multiple comparisons was applied to the P values. P values <0.05 were considered to be statistically significant. Inter-observer agreement was evaluated by calculating weighted (linear weighting) kappa coefficients, and interpreted as follows: :::0, poor; 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and >0.81, almost perfect agreement.

Results

Noise power spectrum analysis

(Fig. 3a) shows that the centroids of the frequency range obtained with soft tissue convolution kernels were shifted towards the low spatial frequencies when switching from FBP to IR. Moreover, the higher the strength level of IR, the lower the frequency range centroid was. With our acquisition parameters, the centroid of convolution kernel B41s lied at 0.33 mm^{-1}; it was slightly reduced at 0.31 mm^{-1} (-6%) with IR 3 and reached 0.27 mm^{-1} with IR 5 (-18%). As compared to FBP, the SDs were reduced from approximately 30% and 50% when using IR strength levels 3 and 5, respectively.

(Fig. 3b) presents similar information but when dealing with bone convolution kernels. As expected, the centroids of the frequency range were in a higher spatial frequency range. With the same acquisition parameters, the centroid of convolution kernel B70h lied at 1.63 mm^{-1}; it was slightly reduced at 0.57 mm^{-1} (-10%) with IR 3 and reached 0.50 mm^{-1} with IR 5 (-21%). As compared to FBP, the SDs were also reduced from approximately 30% and 50% when using IR strength levels 3 and 5, respectively.

Quantitative analysis

The results of the quantitative analysis are illustrated in Fig. 4. With increasing strength level of IR, image noise decreased linearly ($P < 0.001$), while the SNR ($P < 0.001$) and CNR ($P < 0.001$) both increased linearly at C3–C4 (Fig. 4a) and C6–C7 (Fig. 4b) levels. The mean ± SD noise was 11.6 ± 2.4, 10.7 ± 2.2, 9.3 ± 1.8, 8.2 ± 1.9, 6.8 ± 1.3, and 5.5 ± 1.1 HU at C3–C4 and 21.0 ± 8.7, 18.8 ± 7.7, 16.9 ± 6.9, 14.7 ± 5.9, 13.2 ± 5.3, and 11.8 ± 4.7 HU at C6–C7 levels, for FBP and IR strength levels 1 to 5, respectively. The mean ± SD SNR was 4.8 ± 1.1, 5.2 ± 1.2, 5.9 ± 1.3, 6.8 ± 1.7,

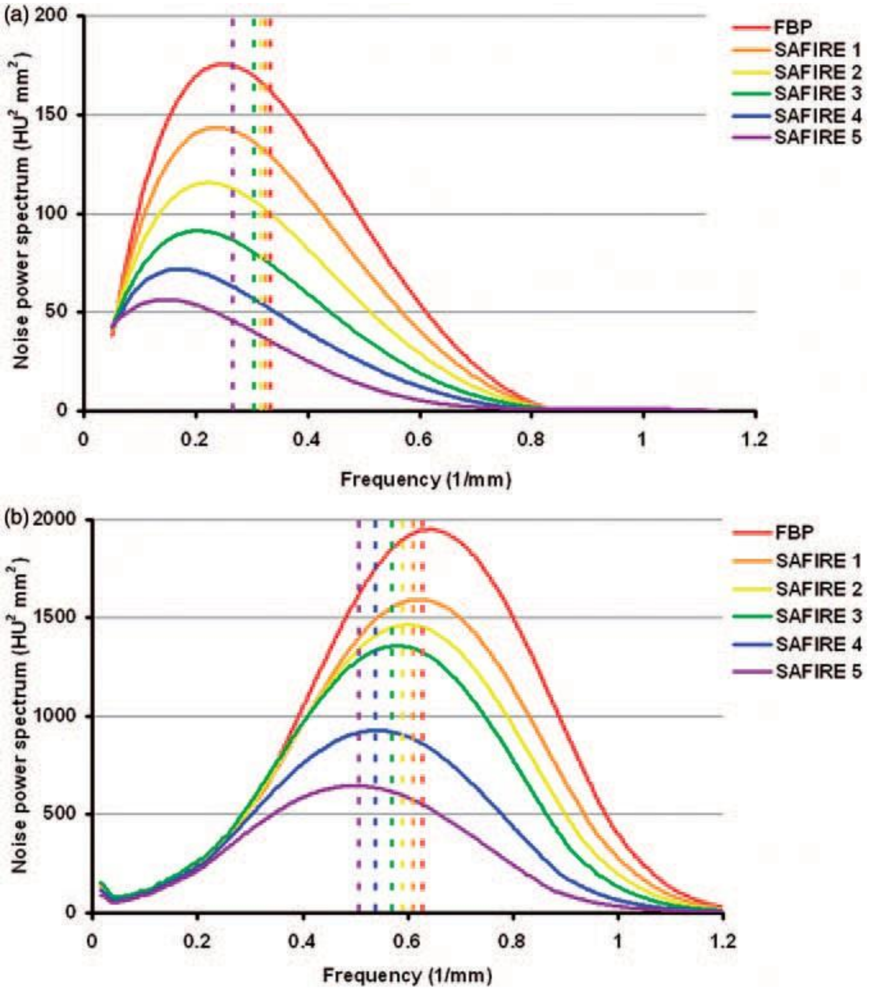


Fig. 3. Line graphs illustrate the impact of filtered back-projection (FBP) and sinogram-affirmed iterative reconstruction (SAFIRE, strength levels 1 to 5) algorithms on the noise power spectra (NPS) of MDCT images reconstructed using a (a) soft tissue and (b) bone convolution kernel. With increasing strength level of IR, the centroids (dotted lines) of the frequency range are shifted towards the low spatial frequencies.

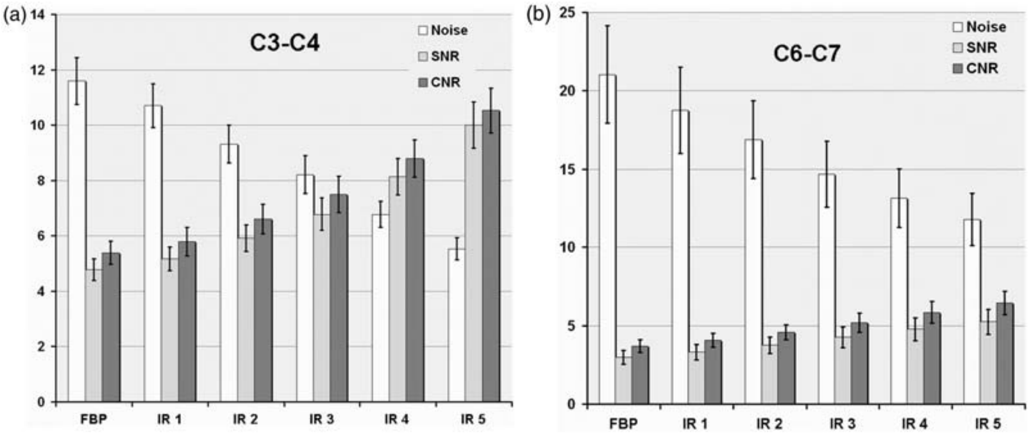


Fig. 4. Bar charts illustrate the impact of filtered back-projection (FBP) and sinogram-affirmed iterative reconstruction (SAFIRE, strength levels 1 to 5) techniques on the image noise (in Hounsfield units [HU]), the signal-to-noise (SNR) and contrast-to-noise (CNR) ratios, at (a) C3–C4 and (b) C6–C7 levels. With increasing strength level of IR, there is a linear decrease in noise and a concomitant linear increase in SNR and CNR, at both levels. Error bars represent 95% confidence intervals.

8.1 ± 1.8, and 10.0 ± 2.3 at C3–C4 and 3.0 ± 1.3, 3.3 ± 1.4, 3.8 ± 1.5, 4.3 ± 1.9, 4.8 ± 2.0, and 5.2 ± 2.2 at C6–C7 levels; while the mean ± SD CNR was 5.4 ± 1.2, 5.8 ± 1.5, 6.6 ± 1.5, 7.5 ± 1.8, 8.8 ± 1.9, and 10.5 ± 2.2 at C3–C4 and 3.7 ± 1.1, 4.1 ± 1.2, 4.6 ± 1.3, 5.2 ± 1.7, 5.9 ± 1.9, and 6.4 ± 2.1 at C6–C7 levels, for FBP and IR strength levels 1 to 5, respectively. For these three variables, all strength levels of IR were significantly different from the lower levels ($P < 0.001$).

Besides, there were no significant change in the mean CT_n (within the ROIs drawn in the posterior paraspinal muscles) obtained with FBP and the five strength levels of IR (P ¼ 0.19), nor a significant difference in the pairwise comparisons ($P > 0.05$).

Qualitative analysis

The results of the qualitative analysis are reported in Table 1. With increasing strength level of IR, a linear trend towards an increase in image quality scores was found for the intervertebral discs (Fig. 5a), the content of neural foramina (Fig. 5b), the content of the dural sac, the ligaments, as well as for the overall diagnostic image quality (Fig. 5e) at C3–C4 and C6–C7 levels. Conversely, there was a linear trend towards a decrease in scores for the soft tissues (Fig. 5c) and trabecular bone (Fig. 5d) at both levels. In contrast, no trend was found for cortical bone either at C3–C4 or C6–C7 level.

With regard to artifacts, no trend was noted either at C3–C4 (Observer 1, P ¼ 0.16; Observer 2, P ¼ 0.33) or C6–C7 (Observer 1, P ¼ 0.77; Observer 2, P ¼ 0.07) level.

Inter-observer agreements were at least moderate for all items, except for the content of the dural sac at C3–C4 level with IR 3 (k ¼ 0.39) and for the soft tissues at C6–C7 level with FBP (k ¼ 0.36), IR 2 (k ¼ 0.32) and IR 3 (k ¼ 0.40) algorithms.

The results of the pairwise comparisons are summarized in Table 2. The image quality scores obtained with IR strength level 5 were significantly higher than with IR 3 or 4 for the intervertebral discs, the content of neural foramina and dural sac, the ligaments, and the overall diagnostic image quality, except for the content of neural foramina at C3–C4 level (Observer 1) and the ligaments at C6–C7 level (Observer 2) for which they were only higher than with IR ::: 2. Conversely, the scores obtained with IR 5 were significantly lower than with IR 3 or 4 for the soft tissues and trabecular bone, except for trabecular bone at C3–C4 level (Observer 2) for which it was lower than with IR 2: 2. No significant difference in image quality scores was found for the content of neural foramina at C3–C4 level (Observer 2), the content of the dural sac at C6–C7 level (Observer 2), and cortical bone (Observers 1 and 2). As regards artifacts, there was no significant difference either at C3–C4 or C6–C7 level (Observers 1 and 2, $P > 0.99$).

Discussion

Our patient-based study demonstrates that the optimal strength level of IR to apply in low-dose cervical spine MDCT depends on the anatomical structure to be analyzed and, therefore, on the clinical context. Except

Table 1. Qualitative image quality scores for filtered back-projection (FBP) and the five strength levels of iterative reconstruction (IR).

		FBP		IR 1		IR 2		IR 3		IR 4		IR 5		P value	
Item	Level	Obs 1	Obs 2	Obs 1	Obs 2	Obs 1	Obs 2	Obs 1	Obs 2	Obs 1	Obs 2	Obs 1	Obs 2	Obs 1	Obs 2
Intervertebral disc (mean)	C3–C4	1.9	1.8	1.8	1.9	2.4	2.3	2.4	2.5	2.9	2.9	3.1	3.1	<0.001	<0.001
	C6–C7	1.6	1.7	2.0	2.1	1.9	2.0	2.2	2.3	2.8	2.9	3.2	3.0	<0.001	<0.001
Content of neural foramina (mean)	C3–C4	1.6	1.8	1.6	1.8	1.7	1.8	2.0	2.0	2.1	2.1	2.1	2.1	<0.001	0.007
	C6–C7	1.3	1.5	1.6	1.7	1.6	1.8	1.7	1.7	2.0	1.9	2.3	2.1	<0.001	<0.001
Content of the dural sac (mean)	C3–C4	1.1	1.3	1.2	1.4	1.2	1.5	1.2	1.4	1.5	1.7	1.6	1.8	<0.001	<0.001
	C6–C7	1.1	1.1	1.3	1.3	1.3	1.4	1.3	1.3	1.4	1.4	1.6	1.5	0.001	0.03
Ligaments (mean)	C3–C4	2.0	2.1	2.0	2.2	2.4	2.5	2.4	2.6	2.9	2.9	3.0	3.0	<0.001	<0.001
	C6–C7	2.2	2.2	2.3	2.4	2.5	2.5	2.6	2.6	3.1	3.0	3.5	3.2	<0.001	<0.001
Soft tissues (mean)	C3–C4	3.6	3.3	3.2	3.0	2.9	3.0	2.0	2.2	1.5	1.7	1.0	1.5	<0.001	<0.001
	C6–C7	3.2	3.1	3.0	2.9	2.6	2.8	2.1	2.2	1.3	1.7	1.0	1.5	<0.001	<0.001
Cortical bone (mean)	C3–C4	3.9	3.9	3.8	3.8	3.9	3.9	3.8	3.8	3.9	3.9	3.9	3.9	0.41	0.43
	C6–C7	3.9	3.9	3.8	3.8	3.8	3.9	3.8	3.8	3.9	3.9	3.9	3.9	0.41	0.43
Trabecular bone (mean)	C3–C4	2.8	2.9	2.5	2.9	2.6	2.8	2.1	2.3	2.2	2.2	1.9	2.0	<0.001	<0.001
	C6–C7	2.7	2.6	2.3	2.4	2.3	2.5	2.2	2.3	2.0	2.0	1.5	1.7	<0.001	<0.001
Overall diagnostic image quality (mean)		4.9	5.1	5.4	5.5	5.9	5.8	6.2	6.3	7.0	6.9	7.2	7.1	<0.001	<0.001

trauma (1,2,18), other regular indications for cervical spine MDCT include chronic cervical pain and/or cervicobrachial neuralgia; the findings to look for being spondylosis, herniated disc disease and/or neural foraminal stenosis (3–5,18,19). In this setting, we found that subjective image quality scores for the intervertebral discs, the content of neural foramina and dural sac, and for the ligaments were all significantly higher with increasing strength level of IR (Table 1 and Fig. 5a and b). However, this increase in IR strength level was also associated with a significant decrease in image quality scores for the paraspinal soft tissues and, to a lesser extent, trabecular bone (Table 1 and Fig. 5c and d). While the scores for trabecular bone remained mostly ''acceptable'', those for the soft tissues became ''unacceptable'' with IR strength levels >3. Consequently, for low-dose MDCT of the cervical spine, we would recommend combining a high strength level of IR for the analysis of the intervertebral discs, neural structures and ligaments, with a medium strength level of IR for the study of trabecular bone. In the context of trauma, we would suggest applying

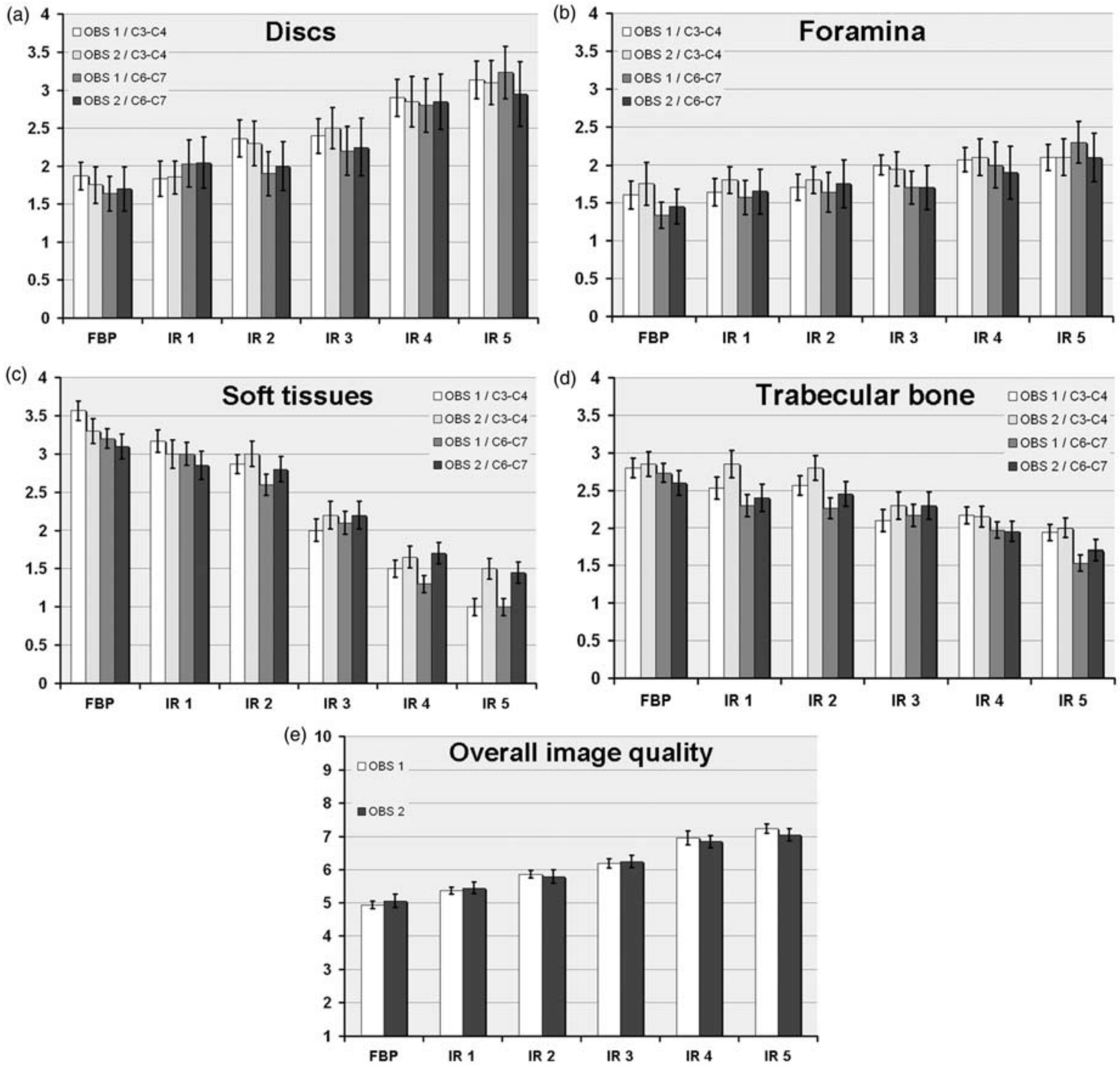


Fig. 5. Bar charts illustrate the impact of filtered back-projection (FBP) and sinogram-affirmed iterative reconstruction (SAFIRE, strength levels 1 to 5) techniques on the four-point image quality scores for (a) the intervertebral discs, (b) the content of neural foramina, (c) the soft tissues, (d) trabecular bone and (e) the overall diagnostic image quality, at C3–C4 and C6–C7 levels. With increasing strength level of IR, there is a linear trend towards an increase in scores for (a) the intervertebral discs, (b) the content of neural foramina and (e) the overall diagnostic image quality, while a linear trend towards a decrease in scores is noted for (c) the soft tissues and (d) trabecular bone, at both levels. Error bars represent 95% confidence intervals; OBS 1, Observer 1; OBS 2, Observer 2.

Table 2. Summary of the pairwise comparisons in the qualitative analysis.

		Observer 1		Observer 2	
Item	Level	Strength level of IR*	P value	Strength level of IR*	P value
Intervertebral disc	C3–C4	5 vs. 4	<0.001	5 vs. 3	0.004
	C6–C7	5 vs. 4	0.004	5 vs. 3	<0.001
Content of neural foramina	C3–C4	5 vs. 2	0.002	No significant difference	2:0.15
	C6–C7	5 vs. 4	0.02	5 vs. 3	0.03
Content of the dural sac	C3–C4	5 vs. 3	0.002	5 vs. 3	0.003
	C6–C7	5 vs. 4	0.01	No significant difference	>0.99
Ligaments	C3–C4	5 vs. 4	<0.001	5 vs. 3	0.001
	C6–C7	5 vs. 4	0.01	5 vs. 2	0.002
Soft tissues	C3–C4	5 vs. 4	<0.001	5 vs. 3	<0.001
	C6–C7	5 vs. 4	0.02	5 vs. 3	<0.001
Cortical bone	C3–C4	No significant difference	>0.99	No significant difference	>0.99
	C6–C7	No significant difference	>0.99	No significant difference	>0.99
Trabecular bone	C3–C4	5 vs. 3	<0.001	5 vs. 2	<0.001
	C6–C7	5 vs. 4	<0.001	5 vs. 3	0.004
Overall diagnostic image quality		5 vs. 4	0.04	5 vs. 3	<0.001

*The IR strength level 5 is compared to other strength levels of IR. For each item, the highest strength level with a statistically significant difference is reported.
IR, iterative reconstruction.

SAFIRE 3 to search for vertebral fractures, and SAFIRE 5 to assess spinal ligament injuries. Based on the results of the current study, we have modified all spine MDCT protocols in our institution and now routinely perform two sets of reconstruction: SAFIRE 3 (with a bone convolution kernel) for bony structures and SAFIRE 5 (with a soft tissue convolution kernel) for all other spinal structures.

Recently, a few different IR techniques have been implemented by CT manufacturers, with the common goal of reducing image noise while preserving spatial resolution (7–14). Each manufacturer has developed its own IR method (adaptive statistical iterative reconstruction, ASIR, and model-based iterative reconstruction, MBIR, for General Electric; iterative reconstruction in image space, IRIS, and SAFIRE for Siemens; iDose for Philips; and adaptive iterative dose reduction, AIDR, for Toshiba) (8,9,20). These algorithms are executed either in the raw data or the image domain, as with SAFIRE (9,13,14). Moreover, most algorithms provide a set of strength levels to choose from, but what determines these levels also depends on the manufacturer. As an example, the strength level of IR may correspond to the number of iterations, or to the proportion of IR and FBP in the blend used to reconstruct the final MDCT images (8,9,12,20). With SAFIRE, the estimations and approximations included in the noise models are set up differently for each strength level: the higher the level chosen, the stronger the noise reduction the model is aiming for (9,13,14,20). According to our quantitative analysis, increasing the strength level of SAFIRE by one should reduce image noise by approximately 10% (Fig. 4).

As expected, we found that the higher the strength level of IR, the lower the noise amplitude was (Fig. 3). Moreover, the noise amplitude was not only reduced but also shifted towards the low-frequency range. This shift might not only smoothen image noise but also low-contrast transitions within MDCT images. When dealing with high-contrast structures, such as in our case, it has been shown that SAFIRE does not influence the frequency content of MDCT images. With such properties, it appears that the use of SAFIRE strength level 5 is optimal for the diagnostic task presented here.

To our knowledge, only two studies addressing the value of IR in MDCT of the spine have been performed so far (10,11). Gervaise et al. reported that a radiation dose reduction of up to 50% was achievable by using such techniques in lumbar spine MDCT, while preserving diagnostic image quality (10). The current study is the first designed to optimize the strength level of IR by separately assessing its impact on the various anatomical structures of the spine, comprising both soft (intervertebral discs, neural structures, ligaments, muscles and fat) and dense (bone) tissues. Indeed, only a few studies have evaluated the impact of varying IR strength levels on image quality (21–24), since most studies on IR used medium strength levels, as recommended by manufacturers (13,14,16). In chest MDCT, Singh et al. found that a percentage of ASIR 30–50% was the optimal strength level to be applied (21). On the other hand, Mie´ville et al. reported that a percentage of ASIR 20–40% was the optimal setting for pediatric cardiac MDCT (22). In contrast, Rampado et al. showed in a phantom study that a percentage of ASIR 50–70% was optimal for MDCT of the chest and upper abdomen (23). Finally, Rapalino et al. recently suggested a high strength level of ASIR (i.e. 2:60%) as the optimal setting in head MDCT, because it reduced image noise and improved low-contrast reso- lution without affecting spatial resolution (24). It is cer- tainly difficult to compare our findings with SAFIRE to these studies on ASIR because, as previously high- lighted, the techniques as well as the definition of IR strength levels vary from one manufacturer to another. However, our results suggest that higher strength levels of IR are recommended for low-dose MDCT of the cervical spine. It is not clear whether these discrepancies are due to the different anatomical structures considered, to the fact that they were all assessed separately, or because our readers were trained with IR-processed images before the analysis. Although some authors mentioned that MDCT images reconstructed with high strength levels of IR may have an ‘‘over-smooth’’ appearance (8,9,12), this study demonstrates that when trained readers were asked to analyze specific anatomical structures, the gain in SNR and CNR obtained with IR enhanced the conspicuity of these structures (Tables 1 and 2, Figs. 4 and 5a and b).

A few other studies have focused on low-dose MDCT of the cervical region (25–28). Mulkens et al. demonstrated that low-dose cervical spine MDCT in patients with blunt trauma allowed to substantially reduce radiation dose (about 65% compared to standard-dose MDCT protocols) with a small increase in image noise but no difference in subjective image quality evaluation (25). Similarly, Gnannt et al. (27) and Hoang et al. (28) recently showed that decreasing tube voltage (from 120 to 70 and 80 kVp, respectively) enabled to significantly reduce the dose without impairing subjective image quality, except for the lower cervical spine (27). However, none of these studies were performed using IR techniques. Furthermore, we did not use a sliding-thin-slab averaging algorithm, which would have improved the detectability of low-contrast lesions and helped to further reduce radiation dose in thin-collimated MDCT imaging (29,30). Therefore, based on our results, additional dose reductions are conceivable depending on the anatomical structure to be analyzed. Further studies are necessary to determine how far we could go in terms of radiation dose reduction in MDCT of the cervical spine.

This study has several limitations. First, the number of patients was relatively small. However, each examination (n ¼ 30) was associated with more than a hundred separate qualitative assessments by each reader. Second, we did not evaluate the impact of increasing the strength level of IR on diagnostic performance, due to the absence of a reliable non-invasive reference standard. Further studies (on cadavers or with surgical correlation) are necessary to assess this parameter. Third, the results obtained with SAFIRE might not be valid with other IR techniques. However, we still believe that general recommendations can be derived from the current study. Finally, we only assessed the optimal strength level of IR for one low-dose protocol. This optimal level might be different for other cervical spine MDCT protocols.

In conclusion, the optimal strength level of IR to be applied in low-dose MDCT of the cervical spine depends on the anatomical structure to be analyzed. Ideally, two strength levels of IR are necessary: a higher for the evaluation of the intervertebral discs and the content of neural foramina, and a lower to assess trabecular bone.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Munera F, Rivas LA, Nunez DB Jr, et al. Imaging evaluation of adult spinal injuries: emphasis on multidetector CT in cervical spine trauma. Radiology 2012;263:645–660.
2. Van Goethem JW, Maes M, Ozsarlak O, et al. Imaging in spinal trauma. Eur Radiol 2005;15:582–590.
3. Douglas-Akinwande AC, Rydberg J, Shah MV, et al. Accuracy of contrast-enhanced MDCT and MRI for identifying the severity and cause of neural foraminal stenosis in cervical radiculopathy: a prospective study. Am J Roentgenol 2010;194:55–61.
4. Freund M, Sartor K. Degenerative spine disorders in the context of clinical findings. Eur J Radiol 2006;58:15–26.
5. Tins B. Technical aspects of CT imaging of the spine. Insights Imaging 2010;1:349–359.
6. Biswas D, Bible JE, Bohan M, et al. Radiation exposure from musculoskeletal computerized tomographic scans. J Bone Joint Surg Am 2009;91:1882–1889.
7. Lee TY, Chhem RK. Impact of new technologies on dose reduction in CT. Eur J Radiol 2010;76:28–35.
8. Fleischmann D, Boas FE. Computed tomography–old ideas and new technology. Eur Radiol 2011;21:510–517.
9. Willemink MJ, de Jong PA, Leiner T, et al. Iterative reconstruction techniques for computed tomography Part 1: Technical principles. Eur Radiol 2013;23:1623–1631.
10. Gervaise A, Osemont B, Lecocq S, et al. CT image quality improvement using Adaptative Iterative Dose Reduction with wide-volume acquisition on 320-detector CT. Eur Radiol 2012;22:295–301.

11. Becce F, Ben Salah Y, Verdun FR, et al. Computed tomography of the cervical spine: comparison of image quality between a standard-dose and a low-dose protocol using filtered back-projection and iterative reconstruction. Skeletal Radiol 2013;42:937–945.
12. Silva AC, Lawder HJ, Hara A, et al. Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm. Am J Roentgenol 2010;194:191–199.
13. Moscariello A, Takx RA, Schoepf UJ, et al. Coronary CT angiography: image quality, diagnostic accuracy, and potential for radiation dose reduction using a novel iterative image reconstruction technique-comparison with traditional filtered back projection. Eur Radiol 2011;21:2130–2138.
14. Winklehner A, Karlo C, Puippe G, et al. Raw data-based iterative reconstruction in body CTA: evaluation of radiation dose saving potential. Eur Radiol 2011;21:2521–2526.
15. Deak PD, Smal Y, Kalender WA. Multisection CT protocols: sex- and age-specific conversion factors used to determine effective dose from dose-length product. Radiology 2010;257:158–166.
16. Willemink MJ, Leiner T, de Jong PA, et al. Iterative reconstruction techniques for computed tomography Part 2: Initial results in dose reduction and image quality. Eur Radiol 2013;23:1632–1642.
17. Bongartz G, Golding SJ, Jurik AG, et al. European guidelines for multislice computed tomography. See http://www.msct.eu/CT_Quality_Criteria.html.
18. American College of Radiology - American Society of Neuroradiology - American Society of Spine Radiology - Society for Pediatric Radiology. Practice guideline for the performance of computed tomography (CT) of the spine. See http://www.acr.org/Quality-Safety/Standards-Guidelines/Practice-Guidelines-by-Modality/CT.
19. Société Française de Radiologie. Guide pratique à l'usage des médecins radiologues pour l'évaluation de leurs pratiques professionnelles. Société Française de Radiologie, 2009: 282–283 and 451–452.
20. Miéville FA, Gudinchet F, Brunelle F, et al. 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.
21. Singh S, Kalra MK, Gilman MD, et al. Adaptive statistical iterative reconstruction technique for radiation dose reduction in chest CT: a pilot study. Radiology 2011;259:565–573.
22. Miéville FA, Gudinchet F, Rizzo E, et al. Paediatric cardiac CT examinations: impact of the iterative reconstruction method ASIR on image quality–preliminary findings. Pediatr Radiol 2011;41:1154–1164.
23. Rampado O, Bossi L, Garabello D, et al. Characterization of a computed tomography iterative reconstruction algorithm by image quality evaluations with an anthropomorphic phantom. Eur J Radiol 2012;81:3172–3177.
24. Rapalino O, Kamalian S, Kamalian S, et al. Cranial CT with adaptive statistical iterative reconstruction: improved image quality with concomitant radiation dose reduction. Am J Neuroradiol 2012;33:609–615.
25. Mulkens TH, Marchal P, Daineffe S, et al. Comparison of low-dose with standard-dose multidetector CT in cervical spine trauma. Am J Neuroradiol 2007;28:1444–1450.
26. Weidemann J, Stamm G, Galanski M, et al. Comparison of the image quality of various fixed and dose modulated protocols for soft tissue neck CT on a GE Lightspeed scanner. Eur J Radiol 2009;69:473–477.
27. Gnannt R, Winklehner A, Goetti R, et al. Low kilovoltage CT of the neck with 70 kVp: comparison with a standard protocol. Am J Neuroradiol 2012;33:1014–1019.
28. Hoang JK, Yoshizumi TT, Nguyen G, et al. Variation in tube voltage for adult neck MDCT: effect on radiation dose and image quality. Am J Roentgenol 2012;198:621–627.
29. von Falck C, Hartung A, Berndzen F, et al. Optimization of low-contrast detectability in thin-collimated modern multidetector CT using an interactive sliding-thin-slab averaging algorithm. Invest Radiol 2008;43:229–235.
30. von Falck C, Galanski M, Shin HO. Informatics in radiology: sliding-thin-slab averaging for improved depiction of low-contrast lesions with radiation dose savings at thin-section CT. Radiographics 2010;30:317–326.

Musculoskeletal Imaging • Original Research

Tobalem et al.
Adaptive Statistical Iterative Reconstruction in Hip MDCT Arthrography

Musculoskeletal Imaging
Original Research

MDCT Arthrography of the Hip: Value of the Adaptive Statistical Iterative Reconstruction Technique and Potential for Radiation Dose Reduction

Frank Tobalem[1]
Eric Dugert[1]
Francis R. Verdun[2]
Vincent Dunet[1]
Julien G. Ott[2]
Hannes A. Rudiger[3]
Stephane Cherix[3]
Reto Meuli[1]
Fabio Becce[1]

Tobalem F, Dugert E, Verdun FR, et al.

OBJECTIVE. The purpose of this article is to assess the effect of the adaptive statistical iterative reconstruction (ASIR) technique on image quality in hip MDCT arthrography and to evaluate its potential for reducing radiation dose.

SUBJECTS AND METHODS. Thirty-seven patients examined with hip MDCT arthrography were prospectively randomized into three different protocols: one with a regular dose (volume CT dose index [CTDI_{vol}], 38.4 mGy) and two with a reduced dose (CTDI_{vol}, 24.6 or 15.4 mGy). Images were reconstructed using filtered back projection (FBP) and four increasing percentages of ASIR (30%, 50%, 70%, and 90%). Image noise and contrast-to-noise ratio (CNR) were measured. Two musculoskeletal radiologists independently evaluated several anatomic structures and image quality parameters using a 4-point scale. They also jointly assessed acetabular labrum tears and articular cartilage lesions.

RESULTS. With decreasing radiation dose level, image noise statistically significantly increased ($p = 0.0009$) and CNR statistically significantly decreased ($p = 0.001$). We also found a statistically significant reduction in noise ($p = 0.0001$) and increase in CNR ($p \leq 0.003$) with increasing percentage of ASIR; in addition, we noted statistically significant increases in image quality scores for the labrum and cartilage, subchondral bone, overall diagnostic quality (up to 50% ASIR), and subjective noise ($p \leq 0.04$), and statistically significant reductions for the trabecular bone and muscles ($p \leq 0.03$). Regardless of the radiation dose level, there were no statistically significant differences in the detection and characterization of labral tears ($n = 24$; $p = 1$) and cartilage lesions ($n = 40$; $p \geq 0.89$) depending on the ASIR percentage.

CONCLUSION. The use of up to 50% ASIR in hip MDCT arthrography helps to reduce radiation dose by approximately 35–60%, while maintaining diagnostic image quality comparable to that of a regular-dose protocol using FBP.

Keywords: CT arthrography, hip, image quality, iterative reconstruction, radiation dose

DOI:10.2214/AJR.14.12821

Received March 11, 2014; accepted after revision June 3, 2014.

F. Tobalem and E. Dugert contributed equally to this work.

[1]Department of Diagnostic and Interventional Radiology, Lausanne University Hospital, Rue du Bugnon 46, Lausanne 1011, Switzerland. Address correspondence to F. Becce (fabio.becce@chuv.ch).

[2]Institute of Radiation Physics, Lausanne University Hospital, Lausanne, Switzerland.

[3]Department of Orthopaedics and Traumatology, Lausanne University Hospital, Lausanne, Switzerland.

WEB
This is a web exclusive article.

AJR 2014; 203:W1–W9

0361–803X/14/2036–W1

© American Roentgen Ray Society

CT arthrography is a mature imaging technique that, over the past 2 decades, has benefited from several advances in CT technology (e.g., the advent of the helical scan method and extension to MDCT systems) to provide submillimeter isotropic imaging of joints [1–3]. MDCT arthrography, with its exquisite spatial resolution and high intraarticular contrast resolution, is a viable alternative (widely available, rapid, robust, and well tolerated) to MRI and MR arthrography in the evaluation of internal derangements of the hip [1, 4, 5]. Specific indications for MDCT arthrography include contraindications to MRI and MR arthrography and failed or inconclusive MRI examinations. It may also be considered when access to MRI is limited, as well as in the preoperative planning of osseous and cartilage abnormalities [1, 2, 4–7]. However, the use of MDCT arthrography is limited by its low soft-tissue contrast resolution and substantial exposure to ionizing radiation [1, 2, 8].

Previous studies on hip MDCT arthrography reported relatively high radiation doses, ranging from approximately 3 to 10 mSv [4, 9–11]. Iterative reconstruction techniques have recently been successfully implemented in MDCT imaging. By reducing image noise, these algorithms have the potential to reduce radiation dose while maintaining an acceptable diagnostic image quality [12, 13]. In fact, their initial application in musculoskeletal imaging has helped to reduce the dose by approximately 40–55% in spine MDCT, compared with regular-dose protocols using filtered back projection (FBP) [14–16]. To our knowledge, the use of iterative reconstruction techniques in MDCT arthrography has not been evaluated so far.

Therefore, the purpose of our study was to assess the impact of the adaptive statistical iterative reconstruction (ASIR; GE Healthcare) technique on the objective and subjective image quality in hip MDCT arthrography and to evaluate its potential for reducing radiation dose.

Subjects and Methods

The institutional ethics committee approved this single-center prospective study (Protocol 61/13). All patients gave written informed consent to be enrolled.

Patients

Over an 18-month period (February 2012 to July 2013), 43 consecutive patients with chronic hip pain were examined with MDCT arthrography in our department. Six patients were excluded, five because of previous total hip replacement surgery and one because he refused to take part in the study. Hence, the final study population consisted of 37 patients (28 women and nine men; mean age, 45 years; range, 22–73 years), whose weight and height were measured before MDCT imaging. Twenty-one patients were suspected of having femoroacetabular impingement, whereas eight had preexisting developmental dysplasia of the hip. All patients were referred by one of three experienced hip surgeons. Indications for hip MDCT arthrography included contraindication to MRI or MR arthrography (*n* = 16), MRI-incompatible implanted medical device (*n* = 7), orthopedic fixation device in close proximity to the hip joint (*n* = 9), failed or inconclusive MR arthrography (*n* = 11; severe claustrophobia [*n* = 5], morbid obesity [*n* = 3], and noncooperative patient generating substantial motion artifacts [*n* = 3]), or preoperative planning of osseous and cartilage abnormalities (*n* = 7). In addition, three patients underwent combined MDCT and MR arthrography. For the study purposes, patients were then randomly assigned to one of three different MDCT protocols: one regular-dose and two reduced-dose (intermediate- or low-dose) protocols.

MDCT Arthrography Protocols

Hip arthrography was performed under aseptic conditions and fluoroscopic guidance. Patients were positioned supine with the leg in slight internal rotation. The superior femoral head-neck junction was targeted through an anterior approach using a 20-gauge needle, and 12–15 mL of an arthrographic blend (15 mL of iodinated contrast medium [Accupaque 300, GE Healthcare], 5 mL of bupivacaine hydrochloride [Bupivacain 0.25%, Sintetica], and two drops of epinephrine hydrochloride [Adrenalin, Sintetica], mixed with 0.1 mL of gadolinium-based contrast agent [Dotarem, Guerbet] when MR arthrography was initially requested) was injected into the joint. Patients were then immediately transferred to the CT (*n* = 26) or MRI (*n* = 11) room, and all MDCT scans were completed within 30 minutes.

All MDCT examinations were performed on a 64-MDCT scanner (LightSpeed VCT, GE Healthcare) using the helical acquisition mode. Patients were positioned supine with the leg in neutral position. Scan length was set to approximately 11 cm, covering from the acetabular roof to the lesser trochanter. Three different fixed tube currents (250, 160, and 100 mA for the regular-, intermediate-, and low-dose protocols, respectively) were used to vary the radiation dose level of hip MDCT arthrograms. Although the ASIR technique is compatible with automatic tube current modulation, the latter was disabled in the current study for the same reasons as described by Singh et al.

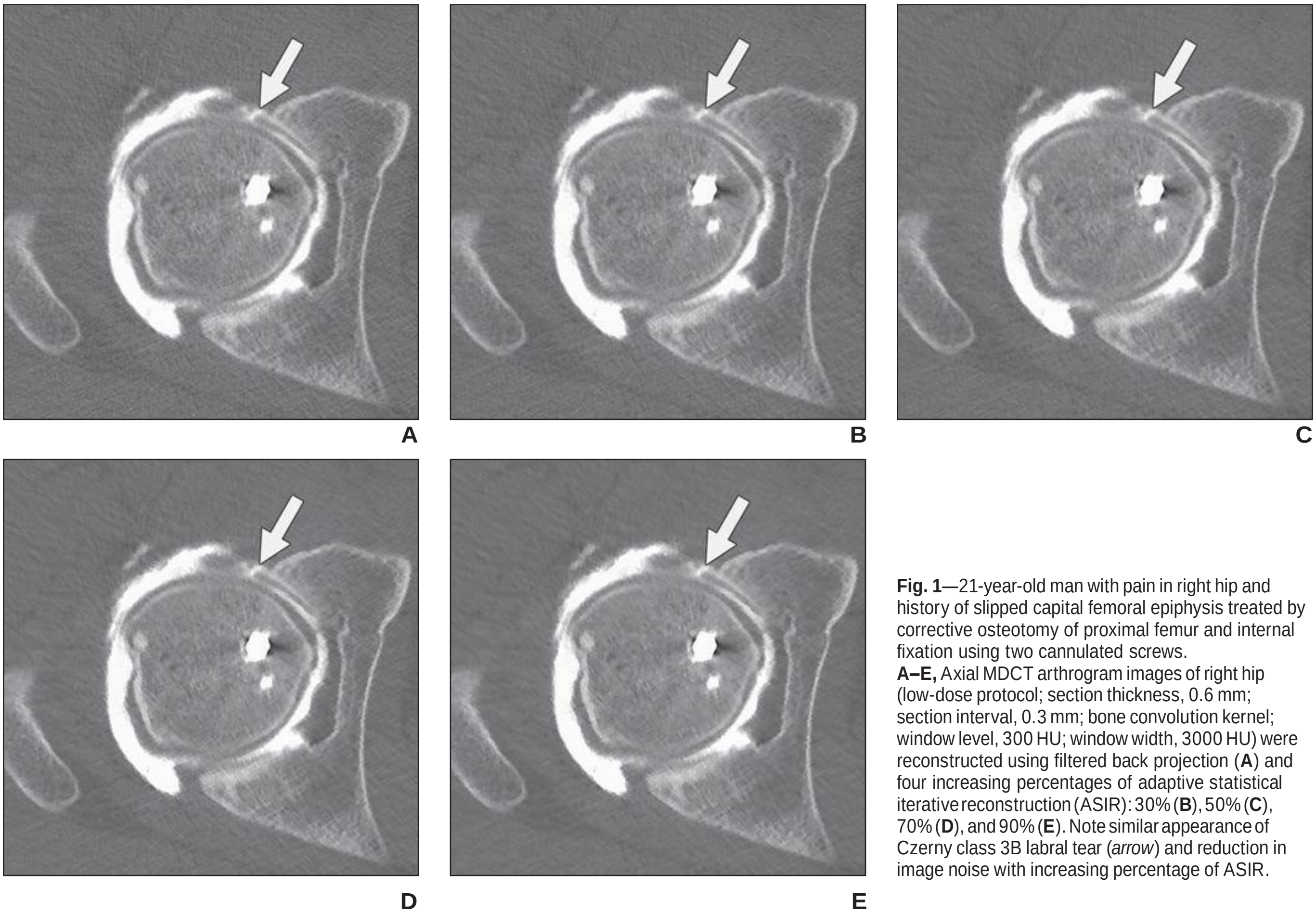


Fig. 1—21-year-old man with pain in right hip and history of slipped capital femoral epiphysis treated by corrective osteotomy of proximal femur and internal fixation using two cannulated screws.
A–E, Axial MDCT arthrogram images of right hip (low-dose protocol; section thickness, 0.6 mm; section interval, 0.3 mm; bone convolution kernel; window level, 300 HU; window width, 3000 HU) were reconstructed using filtered back projection (**A**) and four increasing percentages of adaptive statistical iterative reconstruction (ASIR): 30% (**B**), 50% (**C**), 70% (**D**), and 90% (**E**). Note similar appearance of Czerny class 3B labral tear (*arrow*) and reduction in image noise with increasing percentage of ASIR.

[17]. The tube current for the low-dose protocol was selected according to the lowest-dose protocols previously reported for hip MDCT arthrography [6, 18]. All other scan parameters were held constant as follows: tube voltage, 120 kVp; gantry rotation time, 0.8 second; pitch, 0.516; beam collimation, 40 mm (detector configuration, 64 × 0.625 mm); and scan FOV, 32 × 32 cm.

MDCT raw datasets were reconstructed using both a conventional FBP algorithm and four increasing percentages of the ASIR technique (30%, 50%, 70%, and 90%) (Fig. 1). ASIR is a first-generation dose-reduction technology that extracts noise by photon statistics and object modeling and whose technical principles have been described elsewhere [12, 13, 19]. The forward projection and multiple iterative reconstruction steps help to reduce noise in the image data domain, albeit with a slight increase in reconstruction time. To prevent an overly artificial appearance of MDCT images, the ASIR algorithm further offers the possibility to blend ASIR and FBP images in 10% increments (ASIR 30% thus represents 30% ASIR and 70% FBP in the blended reconstructed image datasets). The other image reconstruction parameters were as follows: display FOV, 19 × 19 cm; section thickness, 0.6 mm; section interval, 0.3 mm; and bone and standard (GE Healthcare) convolution kernels.

Radiation Dose Estimates

The volume CT dose index ($\mathrm{CTDI_{vol}}$), determined for a 32-cm-diameter polymethyl methacrylate cylindric reference phantom, and dose-length product were retrieved from the digital imaging and communications in medicine radiation-dose structured report. The effective dose was then calculated by multiplying the dose-length product by the appropriate conversion factor (0.0129 mSv × $\mathrm{mGy^{-1}} \times \mathrm{cm^{-1}}$ for adult patients, the pelvis region, and 120 kVp) [20]. Furthermore, the size-specific dose estimate (SSDE) was calculated as follows:

$$\mathrm{SSDE} = f_{\mathrm{size}}^{\mathrm{32D}} \times \mathrm{CTDI}_{\mathrm{vol}}^{\mathrm{32D}} \qquad (1),$$

where $f_{\mathrm{size}}^{\mathrm{32D}}$ is the conversion factor as a function of patient's effective diameter [21]. The patient's effective diameter was determined from the maximum anteroposterior and lateral dimensions on the CT localizer radiographs, using the following formula [21]:

$$\text{Effective diameter} = \sqrt{\text{anteroposterior} \times \text{lateral}} \qquad (2).$$

Image Quality Metrics

To investigate the differences in noise texture between FBP and ASIR techniques as a function of both radiation dose level and image reconstruction plane, we scanned a 20-cm-diameter water phantom using exactly the same acquisition parameters as for patients. Images were reconstructed in both axial and coronal planes using FBP and the same four increasing percentages of ASIR (using exactly the same reconstruction parameters as already described in both planes, with the bone convolution kernel only). Furthermore, this step was repeated by reconstructing 0.6-mm images with no overlap. Noise power spectra (NPS) were then calculated using the method described by Miéville et al. [22].

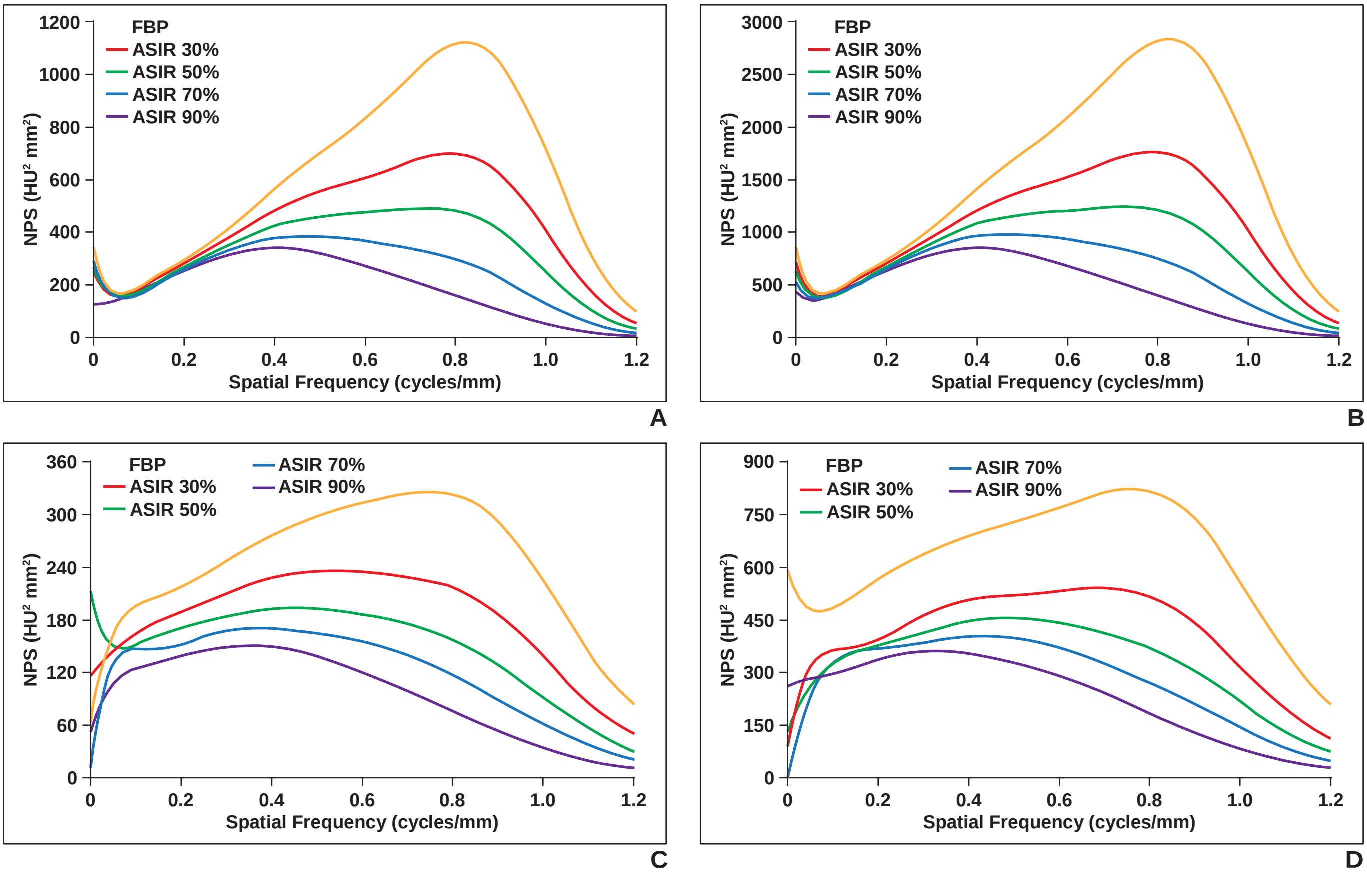


Fig. 2—Noise power spectra (NPS) for 0.6-mm (0.3-mm section interval) images.
A–D, Axial (**A** and **B**) and coronal (**C** and **D**) MDCT images were obtained at two different radiation dose levels, 38 mGy (**A** and **C**) and 15 mGy (**B** and **D**), and were reconstructed using bone convolution kernel with filtered back projection (FBP) and four increasing percentages of adaptive statistical iterative reconstruction (ASIR; 30%, 50%, 70%, and 90%). Curve shape represents noise reduction (*y*-axes) as function of spatial frequency (*x*-axes). Higher curve, higher noise is. Compared with axial images, noise is strongly reduced in coronal plane, and impact of ASIR on noise reduction is less pronounced.

To determine the impact of FBP and ASIR algorithms on spatial resolution depending on the radiation dose level, we subsequently scanned a dedicated phantom (Catphan 600, The Phantom Laboratory) containing a 28-µm-diameter tungsten wire, using the same three MDCT protocols as for patients. Images were reconstructed using FBP and the four increasing percentages of ASIR (same reconstruction parameters as for patients, with the bone convolution kernel only). In-plane modulation transfer functions were then computed as described elsewhere [22].

TABLE 1: Patient Characteristics and Radiation Dose Estimates for the Three Different MDCT Arthrography Protocols

Variable	Regular-Dose Protocol	Intermediate-Dose Protocol	Low-Dose Protocol	p
Patient characteristics				
No. of patients	11	13	13	NA
Sex				0.96
Female	8	10	10	
Male	3	3	3	
Age (y)	42 ± 8	43 ± 13	51 ± 17	0.36
Body mass index (kg/m^2)	24.9 ± 4.1	22.7 ± 2.2	23.3 ± 4.1	0.20
Effective diameter (cm)	30.3 ± 3.7	29.7 ± 1.8	29.0 ± 2.6	0.35
Radiation dose estimates				
Volume CT dose index (mGy), mean	38.4	24.6	15.4	0.0001
Size-specific dose estimate (mGy)	47.1 ± 6.6	30.7 ± 2.1	19.7 ± 1.9	0.0001
Dose-length product (mGy × cm)	588 ± 12	392 ± 42	251 ± 19	0.0001
Effective dose (mSv)	7.6 ± 0.2	5.1 ± 0.5	3.3 ± 0.2	0.0001

Note—Except where noted otherwise, data are mean ± SD. NA = not applicable.

Image Analysis

For the study purposes, only MDCT images reconstructed using the bone convolution kernel were analyzed.

Quantitative Analysis

All measurements were performed by a musculoskeletal fellow on a PACS workstation (Vue, version 11.3, Carestream Health). Three circular ROIs of approximately 50, 25, and 5 mm^2 were drawn into homogeneous areas of the gluteus maximus muscle, intraarticular contrast medium, and acetabular labrum, respectively. Measurements were taken on three consecutive axial sections, and the median values of the CT numbers (mean and SD Hounsfield units) recorded. The selected ROI was then copied and pasted from the FBP image to each of the four corresponding ASIR images (30%, 50%, 70%, and 90% ASIR), to be at exactly the same position. Image noise and contrast-to-noise ratio (CNR) were calculated as follows:

$$\text{Noise} = \text{SD of the mean HU}_{\text{muscle}} \qquad (3),$$

in the ROI drawn into the gluteus maximus muscle; and

$$\text{CNR} = \frac{(\text{mean HU}_{\text{contrast}} - \text{mean HU}_{\text{labrum}})}{\text{SD}_{\text{labrum}}} \qquad (4),$$

in the ROIs drawn into the intraarticular contrast medium and acetabular labrum, respectively [23]. To assess the reproducibility of measurements, they were repeated by the same radiology fellow after a 4-week interval.

Qualitative Analysis

After a consensus reading with three test cases that were not included in the study, two musculoskeletal radiologists (readers 1 and 2, with 5 and 3 years of experience, respectively) independently reviewed all hip MDCT arthrograms on the same PACS workstation as already described. MDCT scans were presented in random order, with both readers blinded to the acquisition parameters (radiation dose level) and image reconstruction technique. Images were displayed with a window level and width set to 300 and 3000 HU, respectively. First, the conspicuity of the following anatomic structures was assessed on axial sections using a 4-point rating scale (4 = exemplary, 3 = diagnostic, 2 = limited, and 1 = nondiagnostic, as described in the Radiologic Society of North America radiology lexicon [24]): acetabular labrum and articular cartilage surfaces, cortical bone, subchondral bone, trabecular bone, and gluteal muscles. Then, the subjective image noise, the severity of beam-hardening (streak) artifacts, and the subjective overall diagnostic image quality were evaluated in the same fashion. Finally, the two readers jointly assessed the acetabular labrum tears and articular cartilage lesions using the multiplanar reformation view mode (with isotropic voxels of 0.6 mm), reaching consensus agreement in cases of discrepancies. Labral tears were categorized using a slightly modified Czerny classification system (class 0–3B lesions, excluding class 1 because purely intrasubstance lesions were not visible with MDCT arthrography) [25] and were localized as being in the anterior, anterosuperior, posterosuperior, or posterior quadrants [26]. Articular cartilage lesions were graded as superficial partial-thickness (< 50% of total cartilage thickness) or deep partial- or full-thickness defects (≥ 50% of thickness). They were localized as being in the anterior, anterosuperior, posterosuperior, or posterior quadrants of the acetabulum, and anteromedial, anterosuperior, posterosuperior, or posteromedial quadrants of the femoral head [27].

Statistical Analysis

Data were analyzed using a statistical software package (Stata, version 12.1, StataCorp). Chi-square and Kruskal-Wallis tests were used to compare patients' characteristics among the three study groups. Quantitative and qualitative data for the three different MDCT protocols and five distinct image reconstruction techniques (ASIR percentages) were compared using the Kruskal-Wallis test and Goldstein equivalence test for means. Intrarater reliability of measurements was evaluated using the Pearson correlation coefficient (ρ). For qualitative variables, interobserver agreement was assessed by calculating the Cohen kappa coefficient. Values for both Pearson correlation and kappa coefficients were interpreted as follows: 0 or less was poor, 0.01–0.20 was slight, 0.21–0.40 was fair, 0.41–0.60 was moderate, 0.61–0.80 was substantial, and 0.81–1 was almost perfect agreement; p values less than 0.05 were considered to be statistically significant.

Results

Patient Characteristics

Patients' characteristics for the three study groups are reported in Table 1. There were no statistically significant differences in sex ($p = 0.96$), age ($p = 0.36$), body mass index ($p = 0.20$), and effective diameter ($p = 0.35$) among patients from the regular-, intermediate-, and low-dose MDCT protocols.

Radiation Dose Estimates

Radiation dose estimates for the three different MDCT protocols are reported in Table 1. The CTDI_{vol}, size-specific dose estimate, dose-length product, and effective dose were all statistically significantly reduced for the low-dose protocol compared with the intermediate- and regular-dose protocols ($p = 0.0001$).

TABLE 2: Image Noise and Contrast-to-Noise Ratio (CNR) for the Three Different MDCT Arthrography Protocols as a Function of the Four Increasing Percentages of Adaptive Statistical Iterative Reconstruction (ASIR)

Variable	Regular-Dose Protocol	Intermediate-Dose Protocol[a]	Low-Dose Protocol[b]	p
Noise				
FBP	**60.4 ± 14.7**	70.7 ± 9.0	81.0 ± 10.6	0.0009
ASIR 30%	NA	**57.8 ± 7.8**	**66.7 ± 8.6**	0.008
ASIR 50%	NA	49.5 ± 6.6	**57.4 ± 7.8**	0.01
ASIR 70%	NA	42.0 ± 6.4	48.8 ± 6.7	0.01
ASIR 90%	NA	35.2 ± 5.9	40.6 ± 6.3	0.02
CNR				
FBP	**36.9 ± 6.4**	25.3 ± 5.1	29.3 ± 7.9	0.001
ASIR 30%	NA	29.2 ± 6.7	**35.3 ± 11.0**	0.12
ASIR 50%	NA	**32.4 ± 8.4**	**40.7 ± 13.9**	0.11
ASIR 70%	NA	**36.4 ± 11.1**	**48.9 ± 21.5**	0.10
ASIR 90%	NA	**41.4 ± 14.8**	56.8 ± 26.1	0.12

Note—Equivalent numbers are shown in **bold**, and those with in bold and italic represent the most equivalent ones. FBP = filtered back projection, NA = not applicable.

[a] $p = 0.0001$ and 0.003 for noise and CNR, respectively.

[b] $p = 0.0001$ and 0.007 for noise and CNR, respectively.

Noise Power Spectrum and Modulation Transfer Function

Variations in NPS as a function of the radiation dose level, image reconstruction algorithm, and plane are illustrated in Figure 2. In the axial plane, the noise level decreased statistically significantly with increasing radiation dose (Figs. 2A and 2B). It further depended heavily on the ASIR percentage. Moreover, the centroids of the frequency ranges were shifted toward low spatial frequencies when switching from FBP to increasing percentages of ASIR. On coronal reformatted images, we noted a strong reduction in image noise compared with the axial plane, due only in part to the slight smoothing performed during the interpolation to form coronal reformats (Figs. 2C and 2D). Increasing the dose also resulted in substantial noise reduction. However, the impact of the ASIR percentage was considerably reduced. Furthermore, the curve shapes were different in that they reached a sort of plateau between 0.1 and 0.8 mm^{-1}. The shape of the NPS was comparable with or without overlap in the reconstructed images, with the exception that all curves were shifted slightly upward when no overlap was used.

Modulation transfer functions for FBP and ASIR algorithms are illustrated in Figure 3 (low-dose protocol only). There were no significant changes in spatial resolution either depending on the radiation dose level or ASIR percentage.

Quantitative Analysis

The results of the quantitative analysis are reported in Table 2. Image noise was statistically significantly higher for the low-dose protocol compared with the intermediate- and regular-dose protocols, both for FBP ($p = 0.0009$) and the four increasing percentages of ASIR ($p \leq 0.02$). For both intermediate- and low-dose protocols, noise was statistically significantly lower with increasing percentage of ASIR ($p = 0.0001$). Image noise was found to be comparable between the regular-dose protocol using FBP, intermediate-dose protocol using ASIR 30%, and low-dose protocol using ASIR 30% and 50%. Intrarater reliability was almost perfect ($\rho = 0.83$).

In contrast, CNR was statistically significantly lower for the low-dose protocol using FBP compared with the intermediate- and regular-dose protocols ($p = 0.001$) (Table 2). However, no statistically significant differences were found between the low- and intermediate-dose protocols using ASIR ($p \geq 0.10$). For both intermediate- and low-dose protocols, CNR was statistically significantly higher with increasing percentage of ASIR ($p \leq 0.003$). CNR was found to be comparable between the regular-dose protocol using FBP, intermediate-dose protocol using ASIR at 50%, 70%, and 90%, and low-dose protocol using ASIR at 30%, 50%, and 70%. Intrarater reliability was moderate ($\rho = 0.45$).

Qualitative Analysis

The major findings of the qualitative analysis are summarized in Table 3 (reader 1 only). Results were comparable for both readers, unless specified otherwise.

With decreasing radiation dose level, we found a statistically significant reduction in image quality scores for the subchondral bone ($p < 0.03$) (Table 3). Statistically nonsignificant trends toward decreased subjective image quality were noted for the labrum and cartilage ($p = 0.15$), trabecular bone (reader 1, $p = 0.10$), as well as the overall diagnostic quality (reader 2, $p = 0.10$). There were no statistically significant differences in scores for the cortical bone ($p = 1$), muscles ($p \geq 0.36$), subjective noise ($p \geq 0.51$), and artifacts ($p = 1$).

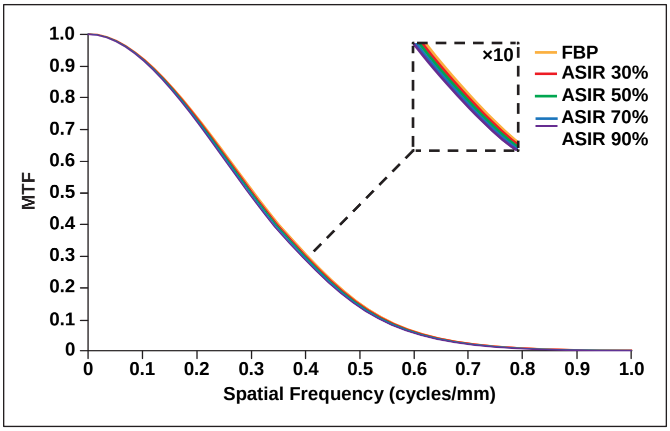


Fig. 3—Modulation transfer functions (MTF) for 0.6-mm MDCT images (low-dose protocol, 15 mGy) reconstructed using bone convolution kernel with filtered back projection (FBP) and four increasing percentages of adaptive statistical iterative reconstruction (ASIR; 30%, 50%, 70%, and 90%). Graph shows no significant differences in spatial resolution between FBP and ASIR images, regardless of ASIR percentage.

With increasing percentage of ASIR, we found a statistically significant increase in image quality scores for subjective noise ($p \leq 0.001$) and statistically significant reductions for the trabecular bone ($p = 0.0001$) and muscles (reader 1, $p \leq 0.03$), both with intermediate- and low-dose protocols (Table 3). In addition, there were statistically significant increases in scores for the labrum and cartilage (reader 2, low-dose protocol, $p = 0.04$), subchondral bone (intermediate-dose protocol, $p \leq 0.03$), as well as the overall diagnostic quality (intermediate-dose protocol, $p \leq 0.04$) up to 50% ASIR, followed by substantial decreases with higher percentages of ASIR. No statistically significant changes were noted for the cortical bone ($p = 1$) and artifacts ($p = 1$).

Interobserver agreement ranged from 0.23 to 1, 0.26 to 1, and 0.41 to 1 for the regular-, intermediate- and low-dose protocols, respectively. It was at least moderate for all items, except for the subjective noise (regular-dose protocol, $\kappa = 0.23$) and subchondral bone (intermediate-dose protocol, $\kappa = 0.26$).

A total of 30 acetabular labrum tears were identified using FBP: 11 partial-thickness (six class 2A and five class 2B) and 19 full-thickness (12 class 3A and seven class 3B) tears. They were located in the anterior ($n = 6$), anterosuperior ($n = 18$), and posterosuperior ($n = 6$) quadrants. Of those 30 labral tears, six lesions (four partial thickness and two full thickness) were found in the regular-dose patient group, and 12 in both intermediate-dose (one partial thickness and 11 full thickness) and low-dose (six partial thickness and six full-thickness) groups. For both intermediate- and low-dose protocols, all labral tears were accurately detected and characterized regardless of the ASIR percentage ($p = 1$) (Figs. 1 and 4).

Overall, the readers identified 59 articular cartilage lesions using FBP: 24 superficial partial-thickness and 35 deep partial-thickness or full-thickness defects. Thirty-one lesions were located on the acetabulum (four in the anterior, eight in the anterosuperior, nine in the posterosuperior, and 10 in the posterior quadrants), and 28 lesions on the femoral head (one in the anteromedial, seven in the anterosuperior, seven in the posterosuperior, and 13 in the posteromedial quadrant). Of those 59 cartilage lesions, 19 were observed in the regular-dose, 24 in the intermediate-dose, and 16 in the low-dose patient groups. The number of cartilage lesions identified as a function of the ASIR percentage is reported in Table 4. For both intermediate- and low-dose protocols, there were no statistically significant differences in detection and grading of lesions depending on the ASIR percentage ($p \geq 0.89$).

TABLE 3: Subjective Image Quality Scores for the Three Different MDCT Arthrography Protocols as a Function of the Four Increasing Percentages of Adaptive Statistical Iterative Reconstruction (ASIR) for Reader 1 Only and Selected Items

Variable	Regular-Dose Protocol	Intermediate-Dose Protocol[a]	Low-Dose Protocol[b]	p
Labrum and cartilage				
FBP	**3.4 ± 0.5**	**3.4 ± 0.5**	2.9 ± 0.5	0.15
ASIR 30%	NA	**3.3 ± 0.5**	**3.1 ± 0.3**	0.32
ASIR 50%	NA	**3.4 ± 0.5**	**3.2 ± 0.6**	0.38
ASIR 70%	NA	**3.2 ± 0.6**	2.8 ± 0.6	0.17
ASIR 90%	NA	**3.1 ± 0.6**	2.8 ± 0.4	0.27
Subchondral bone				
FBP	**3.8 ± 0.4**	3.3 ± 0.5	3.2 ± 0.4	0.03
ASIR 30%	NA	**3.5 ± 0.5**	3.4 ± 0.5	0.74
ASIR 50%	NA	**3.8 ± 0.4**	3.3 ± 0.6	0.08
ASIR 70%	NA	**3.6 ± 0.5**	2.9 ± 0.8	0.03
ASIR 90%	NA	3.0 ± 0.4	2.6 ± 0.7	0.12
Trabecular bone				
FBP	**3.5 ± 0.5**	**3.5 ± 0.5**	3.1 ± 0.3	0.10
ASIR 30%	NA	**3.3 ± 0.5**	3.1 ± 0.3	0.32
ASIR 50%	NA	**3.2 ± 0.4**	2.8 ± 0.4	0.22
ASIR 70%	NA	2.6 ± 0.5	2.2 ± 0.4	0.10
ASIR 90%	NA	2.2 ± 0.4	2.0 ± 0.0	0.32
Overall diagnostic quality				
FBP	**3.4 ± 0.5**	**3.3 ± 0.5**	**3.1 ± 0.3**	0.43
ASIR 30%	NA	**3.4 ± 0.5**	**3.1 ± 0.3**	0.18
ASIR 50%	NA	**3.6 ± 0.5**	**3.2 ± 0.4**	0.10
ASIR 70%	NA	**3.2 ± 0.4**	2.8 ± 0.4	0.13
ASIR 90%	NA	2.8 ± 0.4	2.8 ± 0.4	0.74

Note—Equivalent scores are shown in **bold**, and those in bold and italic represent the most equivalent ones. FBP = filtered back projection, NA = not applicable.

[a] $p = 0.80, 0.03, 0.0001$, and 0.04 for the labrum and cartilage, subchondral bone, trabecular bone, and overall diagnostic quality, respectively.

[b] $p = 0.54, 0.12, 0.0001$, and 0.33 for the labrum and cartilage, subchondral bone, trabecular bone, and overall diagnostic quality, respectively.

TABLE 4: Number and Grade of Articular Cartilage Lesions for the Three Different MDCT Arthrography Protocols as a Function of the Four Increasing Percentages of Adaptive Statistical Iterative Reconstruction (ASIR)

Dose Protocol	FBP	ASIR 30%	ASIR 50%	ASIR 70%	ASIR 90%	p
Regular	19 (4, 15)	NA	NA	NA	NA	NA
Intermediate	24 (11, 13)	25 (12, 13)	25 (11, 14)	29 (11, 18)	29 (10, 19)	0.97
Low	16 (9, 7)	16 (9, 7)	17 (8, 9)	18 (8, 10)	22 (9, 13)	0.89

Note—Data are no. of lesions (superficial partial-thickness, deep partial- or full-thickness). FBP = filtered back projection, NA = not applicable.

Discussion

Owing to several technologic developments, MDCT arthrography has become an established imaging technique for evaluating various internal joint disorders [1–3, 28]. It may further be regarded as the first-line cross-sectional imaging for assessing articular cartilage injuries [5, 10, 29, 30], as well as in the postoperative setting [2, 4, 28]. With the exception of one single study [18], relatively high diagnostic performances have been reported for the hip, with sensitivities of 67–94%, specificities of 82–100%, and accuracies of 75–90% for articular cartilage lesions, and sensitivities of 90–97%, specificities of 80–100%, and accuracies of 88–95% for acetabular labrum tears [5, 9, 11, 29, 31]. However, improvements in diagnostic image quality were achieved at the expense of an increased radiation burden, and MDCT arthrography is still associated with significant radiation dose, particularly in the hip and shoulder [8, 32]. This is one reason why MRI or MR arthrography remains the first-line examination in most cases, the other being its higher soft-tissue contrast resolution.

In the past few years, exposure to ionizing radiation generated by CT has become a growing concern [33]. MDCT arthrography requires thin sections and low pitch values, as well as high tube voltage and loading in the pelvic and scapular girdles, to provide submillimeter isotropic diagnostic-quality images. Widely varying radiation dose levels are used in hip MDCT arthrography, ranging from approximately 3 to 10 mSv ($CTDI_{vol}$, ≈ 10 to 45 mGy), thus emphasizing the need for protocol optimization [4, 9–11, 18, 29–31]. The current study shows that radiation dose in hip MDCT arthrography can be reduced by up to 60% using 30–50% ASIR, while maintaining a similar diagnostic image quality compared with a regular-dose protocol using FBP. This is consistent with the results from previous studies on iterative reconstruction techniques, which have focused primarily on chest and abdominal CT and have shown dose reductions from 23% to 76% compared with locally used default FBP settings [34, 35]. It is worth mentioning, however, that the reduction percentage strongly depends on the initial dose level, and our regular-dose protocol lay in the upper quarter of the previously reported doses for hip MDCT arthrography. Therefore, smaller reductions are to be expected for lower-dose protocols. Moreover, although our low-dose protocol (15.4 mGy) was actually higher than the lowest-dose protocols reported in the literature, it was still significantly lower (−38%) than the diagnostic reference level proposed in the European guidelines for CT of the osseous pelvis (25 mGy) [32, 36]. We selected the two decreasing dose levels (intermediate and low) as an initial reduction step, while trying to preserve an acceptable diagnostic confidence. At the time of writing, Simoni et al. [37] reported intriguing results on the optimization of dose parameters in hip MDCT arthrography. They suggested that the combination of 120 kVp and 50 mA ($CTDI_{vol}$ estimated to be as low as 4 mGy) may be the optimal setting for reducing the dose, without adversely affecting the visibility of the articular cartilage and subchondral bone. However, their preliminary findings were obtained using only one custom-made phantom and one cadaver, without applying iterative reconstruction techniques or assessing the impact of dose parameters on the detection and characterization of joint damage. Accordingly, additional dose reductions are conceiv-

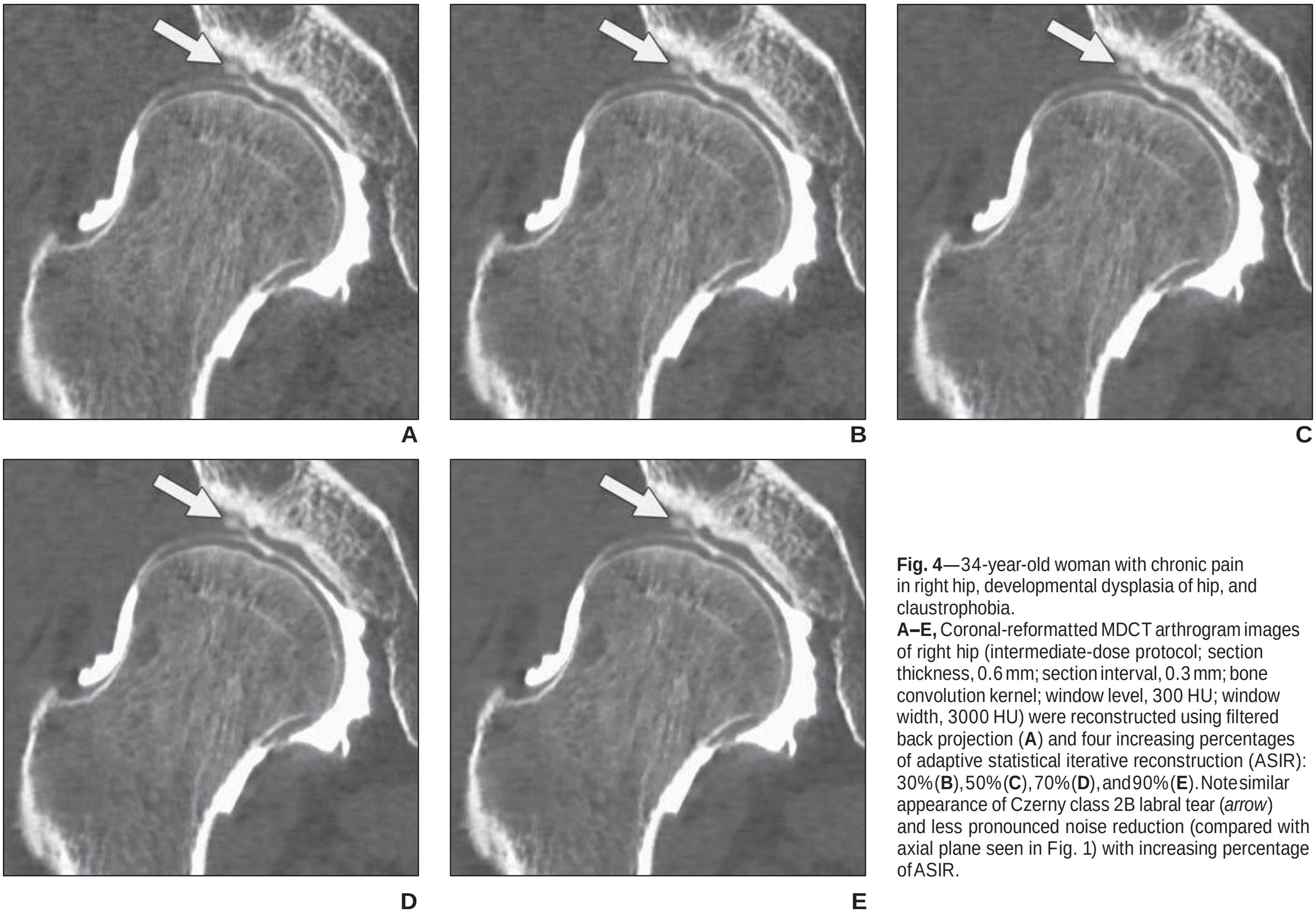


Fig. 4—34-year-old woman with chronic pain in right hip, developmental dysplasia of hip, and claustrophobia.

A–E, Coronal-reformatted MDCT arthrogram images of right hip (intermediate-dose protocol; section thickness, 0.6 mm; section interval, 0.3 mm; bone convolution kernel; window level, 300 HU; window width, 3000 HU) were reconstructed using filtered back projection (**A**) and four increasing percentages of adaptive statistical iterative reconstruction (ASIR): 30% (**B**), 50% (**C**), 70% (**D**), and 90% (**E**). Note similar appearance of Czerny class 2B labral tear (*arrow*) and less pronounced noise reduction (compared with axial plane seen in Fig. 1) with increasing percentage of ASIR.

able, and further research is needed to determine how low we can go in hip MDCT arthrography, particularly with the use of second-generation iterative reconstruction techniques such as model-based iterative reconstruction (MBIR; GE Healthcare).

Iterative reconstruction techniques have recently been introduced in clinical routine as an additional effective tool for reducing radiation dose [12, 32, 34]. In musculoskeletal MDCT imaging, a few studies on the cervical [15, 16] and lumbar [14] spine have shown significant dose reductions, ranging from approximately 40% to 55%, compared with locally used regular-dose FBP protocols. These studies were performed using various iterative reconstruction algorithms (adaptive iterative dose reduction, AIDR; Toshiba [14]; ASIR; GE Healthcare [15]; and sinogram-affirmed iterative reconstruction; Siemens [16]), with either smooth [14, 16] or sharp [15] convolution kernels.

To our knowledge, the present investigation is the first to show the value of iterative reconstruction techniques in MDCT arthrography. As anticipated, reducing the dose resulted in increased image noise and decreased CNR (Table 2). Applying ASIR helped to decrease noise and increase CNR. As a result, image noise and CNR proved to be the most equivalent between the regular-dose FBP, intermediate-dose ASIR 30%, and low-dose ASIR 50% protocols. Interestingly, both readers considered that those three different protocols provided comparable image quality in terms of overall diagnostic confidence, as well as for assessing the acetabular labrum and articular cartilage (Table 3).

Moreover, no statistically significant differences were found regarding the detection and characterization of labral tears and cartilage lesions as a function of the ASIR percentage, for either intermediate- or low-dose protocols (Table 4). However, higher percentages of ASIR (> 50%) were associated with a decrease in image quality scores for the overall diagnostic confidence, the labrum and cartilage, and subchondral bone. This finding might be partly explained by the strong noise-free oversmooth artifactual appearance of MDCT images reconstructed using excessively high percentages of ASIR [12, 13]. This is also consistent with the results of Omoumi et al. [38], who have recently shown that applying exceedingly high strength levels of sinogram-affirmed iterative reconstruction in cervical spine MDCT resulted in a significant decrease in subjective image quality for the trabecular bone. They reported that the optimal strength level of iterative reconstruction depended on the anatomic structure to be analyzed, and thus recommended combining high strength levels of iterative reconstruction when using smooth (soft tissue) convolution kernels with medium levels for sharp (bone) kernels. Similarly, Singh et al. [39] and Mueck et al. [40] both suggested that the optimal percentage of ASIR was between 30% and 50% when using sharp (lung) convolution kernels in chest MDCT.

Regarding image quality metrics, spatial resolution (modulation transfer function) proved not to be affected either by the radiation dose level or ASIR percentage, when using the bone convolution kernel (Fig. 3). This is in contradiction with Singh et al. [17], who reported that ASIR images had higher spatial resolution than FBP images when a smooth (standard) kernel was used. On the other hand, NPS analysis revealed that noise amplitude was statistically significantly lower in the coronal than in the axial plane (Fig. 2). In the axial plane, the ASIR percentage influenced not only the noise amplitude but also the spatial frequency content of MDCT images (Figs. 2A and 2B). However, the standard shape of the curves (ramp filter followed by apodization function) remained similar to the one obtained using FBP. In the coronal plane, we noted a difference in the spatial frequency content of FBP images compared with the axial plane (Figs. 2C and 2D). The standard shape of the reconstruction kernels was less apparent. Applying ASIR strengthened those differences by producing spectra for which a wide range of spatial frequencies were transferred with a similar weight. This certainly had an impact on the appearance of anatomic structures, such as articular cartilage, and needs to be investigated in greater detail in future studies.

This study has several limitations. First, small numbers of patients were included in each of the three study groups, thus limiting the power of the statistical analysis. This is partly because we compared three radiation dose levels instead of two. Second, only two reduced-dose protocols were evaluated, thereby preventing us from establishing the lowest achievable dose level. Additional studies, such as the one recently performed by Simoni et al. [37] on cadavers, are needed to determine whether ultra-low-dose hip MDCT arthrography will be feasible in the near future.

Third, we assessed a single iterative reconstruction technique (ASIR) using only a sharp (bone) convolution kernel. However, few studies have compared all of the different iterative reconstruction algorithms, and only using phantoms [22, 41]. Furthermore, the impact of iterative reconstruction techniques on image quality when using smooth kernels has already been evaluated, including in previous studies on the spine [14, 16, 38].

Fourth, it was not possible to measure the patients' effective diameter on the reconstructed MDCT arthrogram images because a few overweight patients lay partly outside the scan FOV. However, measurements on the localizer are subject to considerable parallax because of the short source-to-detector distance. Finally, we were not able to assess the impact of the ASIR algorithm on the diagnostic performance because surgical correlation was unfortunately not available for all patients (*n* = 11). Nevertheless, no statistically significant differences were found in the detection and characterization of labral tears and cartilage lesions (Table 4).

In conclusion, the use of up to 50% ASIR in hip MDCT arthrography helps to reduce radiation dose, while maintaining diagnostic image quality comparable to that of a regular-dose protocol using FBP. Future studies will determine how low we can go in terms of dose reduction, without impairing diagnostic accuracy.

Acknowledgment

We thank Martine Bernasconi and her team for help with MDCT data acquisition.

References

1. Buckwalter KA. CT arthrography. *Clin Sports Med* 2006; 25:899–915
2. Farber JM. CT arthrography and postoperative musculoskeletal imaging with multichannel computed tomography. *Semin Musculoskelet Radiol* 2004; 8:157–166
3. Buckwalter KA, Rydberg J, Kopecky KK, Crow K, Yang EL. Musculoskeletal imaging with multislice CT. *AJR* 2001; 176:979–986
4. Llopis E, Fernandez E, Cerezal L. MR and CT arthrography of the hip. *Semin Musculoskelet Radiol* 2012; 16:42–56
5. Christie-Large M, Tapp MJ, Theivendran K, James SL. The role of multidetector CT arthrography in the investigation of suspected intraarticular hip pathology. *Br J Radiol* 2010; 83:861–867
6. Allen BC, Peters CL, Brown NA, Anderson AE. Acetabular cartilage thickness: accuracy of three-dimensional reconstructions from multidetector CT arthrograms in a cadaver study. *Radiology* 2010; 255:544–552

7. Dolan MM, Heyworth BE, Bedi A, Duke G, Kelly BT. CT reveals a high incidence of osseous abnormalities in hips with labral tears. *Clin Orthop Relat Res* 2011; 469:831–838
8. Biswas D, Bible JE, Bohan M, Simpson AK, Whang PG, Grauer JN. Radiation exposure from musculoskeletal computerized tomographic scans. *J Bone Joint Surg Am* 2009; 91:1882–1889
9. Ha YC, Choi JA, Lee YK, et al. The diagnostic value of direct CT arthrography using MDCT in the evaluation of acetabular labral tear: with arthroscopic correlation. *Skeletal Radiol* 2013; 42:681–688
10. Tamura S, Nishii T, Shiomi T, et al. Three-dimensional patterns of early acetabular cartilage damage in hip dysplasia: a high-resolutional CT arthrography study. *Osteoarthritis Cartilage* 2012; 20:646–652
11. Yamamoto Y, Tonotsuka H, Ueda T, Hamada Y. Usefulness of radial contrast-enhanced computed tomography for the diagnosis of acetabular labrum injury. *Arthroscopy* 2007; 23:1290–1294
12. 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* 2010; 194:191–199
13. Willemink MJ, de Jong PA, Leiner T, et al. Iterative reconstruction techniques for computed tomography. Part 1. Technical principles. *Eur Radiol* 2013; 23:1623–1631
14. Gervaise A, Osemont B, Lecocq S, et al. CT image quality improvement using Adaptive Iterative Dose Reduction with wide-volume acquisition on 320-detector CT. *Eur Radiol* 2012; 22:295–301
15. Geyer LL, Körner M, Hempel R, et al. Evaluation of a dedicated MDCT protocol using iterative image reconstruction after cervical spine trauma. *Clin Radiol* 2013; 68:e391–e396
16. Becce F, Ben Salah Y, Verdun FR, et al. Computed tomography of the cervical spine: comparison of image quality between a standard-dose and a low-dose protocol using filtered back-projection and iterative reconstruction. *Skeletal Radiol* 2013; 42:937–945
17. Singh S, Kalra MK, Hsieh J, et al. Abdominal CT: comparison of adaptive statistical iterative and filtered back projection reconstruction techniques. *Radiology* 2010; 257:373–383
18. Perdikakis E, Karachalios T, Katonis P, Karantanas A. Comparison of MR-arthrography and MDCT-arthrography for detection of labral and articular cartilage hip pathology. *Skeletal Radiol* 2011; 40:1441–1447
19. Singh S, Khawaja RD, Pourjabbar S, Padole A, Lira D, Kalra MK. Iterative image reconstruction and its role in cardiothoracic computed tomography. *J Thorac Imaging* 2013; 28:355–367
20. Deak PD, Smal Y, Kalender WA. Multisection CT protocols: sex- and age-specific conversion factors used to determine effective dose from dose-length product. *Radiology* 2010; 257:158–166
21. Boone JM, Strauss KJ, Cody DD, et al. Size-specific dose estimates (SSDE) in pediatric and adult body CT examinations: AAPM report no. 204. American Association of Physicists in Medicine website. www.aapm.org/pubs/reports/rpt_204.pdf. Published 2011. Accessed August 3, 2013
22. 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
23. Subhas N, Freire M, Primak AN, et al. CT arthrography: in vitro evaluation of single and dual energy for optimization of technique. *Skeletal Radiol* 2010; 39:1025–1031
24. Radiological Society of North America. RadLex: a lexicon for uniform indexing and retrieval of radiology information resources. RSNA website. www.rsna.org/radlex.aspx. Published 2010. Accessed August 11, 2013
25. Czerny C, Hofmann S, Urban M, et al. MR arthrography of the adult acetabular capsular-labral complex: correlation with surgery and anatomy. *AJR* 1999; 173:345–349
26. Blankenbaker DG, Tuite MJ. Acetabular labrum. *Magn Reson Imaging Clin N Am* 2013; 21:21–33
27. Blankenbaker DG, Ullrick SR, Kijowski R, et al. MR arthrography of the hip: comparison of IDEAL-SPGR volume sequence to standard MR sequences in the detection and grading of cartilage lesions. *Radiology* 2011; 261:863–871
28. Fritz J, Fishman EK, Small KM, et al. MDCT arthrography of the shoulder with datasets of isotropic resolution: indications, technique, and applications. *AJR* 2012; 198:635–646
29. Nishii T, Tanaka H, Nakanishi K, Sugano N, Miki H, Yoshikawa H. Fat-suppressed 3D spoiled gradient-echo MRI and MDCT arthrography of articular cartilage in patients with hip dysplasia. *AJR* 2005; 185:379–385
30. Wyler A, Bousson V, Bergot C, et al. Comparison of MR-arthrography and CT-arthrography in hyaline cartilage-thickness measurement in radiographically normal cadaver hips with anatomy as gold standard. *Osteoarthritis Cartilage* 2009; 17:19–25
31. Nishii T, Tanaka H, Sugano N, Miki H, Takao M, Yoshikawa H. Disorders of acetabular labrum and articular cartilage in hip dysplasia: evaluation using isotropic high-resolutional CT arthrography with sequential radial reformation. *Osteoarthritis Cartilage* 2007; 15:251–257
32. Gervaise A, Teixeira P, Villani N, Lecocq S, Louis M, Blum A. CT dose optimisation and reduction in osteoarticular disease. *Diagn Interv Imaging* 2013; 94:371–388
33. Brenner DJ, Hall EJ. Computed tomography: an increasing source of radiation exposure. *N Engl J Med* 2007; 357:2277–2284
34. Willemink MJ, Leiner T, de Jong PA, et al. Iterative reconstruction techniques for computed tomography. Part 2. Initial results in dose reduction and image quality. *Eur Radiol* 2013; 23:1632–1642
35. 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* 2009; 193:764–771
36. Bongartz G, Golding SJ, Jurik AG, et al. European guidelines on quality criteria for computed tomography. w3.tue.nl/fileadmin/sbd/documenten/leergang/bsm/european_guidelines_quality_criteria_computed_tomography_eur_16252.pdf. Accessed September 28, 2013
37. Simoni P, Leyder PP, Albert A, et al. Optimization of computed tomography (CT) arthrography of hip for the visualization of cartilage: an in vitro study. *Skeletal Radiol* 2014; 43:169–178
38. Omoumi P, Verdun FR, Salah YB, et al. Low-dose multidetector computed tomography of the cervical spine: optimization of iterative reconstruction strength levels. *Acta Radiol* 2014; 55:335–344
39. Singh S, Kalra MK, Gilman MD, et al. Adaptive statistical iterative reconstruction technique for radiation dose reduction in chest CT: a pilot study. *Radiology* 2011; 259:565–573
40. Mueck FG, Michael L, Deak Z, et al. Upgrade to iterative image reconstruction (IR) in MDCT imaging: a clinical study for detailed parameter optimization beyond vendor recommendations using the adaptive statistical iterative reconstruction environment (ASIR). Part 2. The chest. *Rofo* 2013; 185:644–654
41. Löve A, Olsson ML, Siemund R, Stålhammar F, Björkman-Burtscher IM, Söderberg M. Six iterative reconstruction algorithms in brain CT: a phantom study on image quality at different radiation dose levels. *Br J Radiol* 2013; 86:20130388

Optimization of Radiation Dose and Image Quality in Musculoskeletal CT: Emphasis on Iterative Reconstruction Techniques (Part 1)

Patrick Omoumi, MD, MSc, PhD[1] Fabio Becce, MD[1*] Julien G. Ott, MSc[2] Damien Racine, MSc[2]
Francis R. Verdun, PhD[2*]

[1] Department of Diagnostic and Interventional Radiology, Lausanne University Hospital, Lausanne, Switzerland
[2] Institute of Radiation Physics, Lausanne University Hospital, Lausanne, Switzerland

Address for correspondence Patrick Omoumi, MD, MSc, PhD, Department of Diagnostic and Interventional Radiology, Lausanne University Hospital, Rue du Bugnon 46, CH-1011 Lausanne, Switzerland (e-mail: patrick.omoumi@chuv.ch).

Semin Musculoskelet Radiol 2015;19:415–421.

Abstract

Computed tomography (CT) is a modality of choice for the study of the musculoskeletal system for various indications including the study of bone, calcifications, internal derangements of joints (with CT arthrography), as well as periprosthetic complications. However, CT remains intrinsically limited by the fact that it exposes patients to ionizing radiation. Scanning protocols need to be optimized to achieve diagnostic image quality at the lowest radiation dose possible. In this optimization process, the radiologist needs to be familiar with the parameters used to quantify radiation dose and image quality. CT imaging of the musculoskeletal system has certain specificities including the focus on high-contrast objects (i.e., in CT of bone or CT arthrography). These characteristics need to be taken into account when defining a strategy to optimize dose and when choosing the best combination of scanning parameters. In the first part of this review, we present the parameters used for the evaluation and quantification of radiation dose and image quality. In the second part, we discuss different strategies to optimize radiation dose and image quality at CT, with a focus on the musculoskeletal system and the use of novel iterative reconstruction techniques.

Keywords
- computed tomography
- image quality
- radiation dose
- iterative reconstruction techniques
- musculoskeletal imaging

Computed tomography (CT) is a modality of choice for the study of the musculoskeletal system for various indications including the study of bone, calcifications, internal derangements of joints (with CT arthrography), as well as periprosthetic complications.[1–13]

However, CT remains intrinsically limited by the fact that it exposes patients to ionizing radiation. Although the carcinogenic effects of medical imaging have been disputed, there has been an increasing interest in recent years to reduce radiation dose due to CT examinations in accordance with as low as reasonably achievable (ALARA) principles.[14–18] In the process of optimization of scanning protocols to achieve diagnostic image quality at the lowest radiation dose possible, the radiologist needs to be familiar with parameters used to quantify radiation dose and image quality.

In the first part of this review, we present the parameters that are used for the evaluation and quantification of radiation dose and image quality. In the second part, in this issue, we discuss different strategies to optimize the radiation dose and image quality of CT, with a focus on the musculoskeletal system and the use of novel iterative reconstruction (IR) techniques.[19]

Evaluation of Radiation Dose: Definitions

To assess the radiation dose delivered by a CT scan, as well as the radiologic risk associated with it, several indicators exist. The first standardized dose indicator is the volume CT dose

---

[*] F. Becce and F.R. Verdun contributed equally to this work.

Issue Theme Advances in Musculoskeletal Computed Tomography and Tumor Imaging; Guest Editor, Patrick Omoumi, MD, MSc, PhD

Copyright © 2015 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.
Tel: +1(212) 584-4662.

DOI http://dx.doi.org/10.1055/s-0035-1569255.
ISSN 1089-7860.

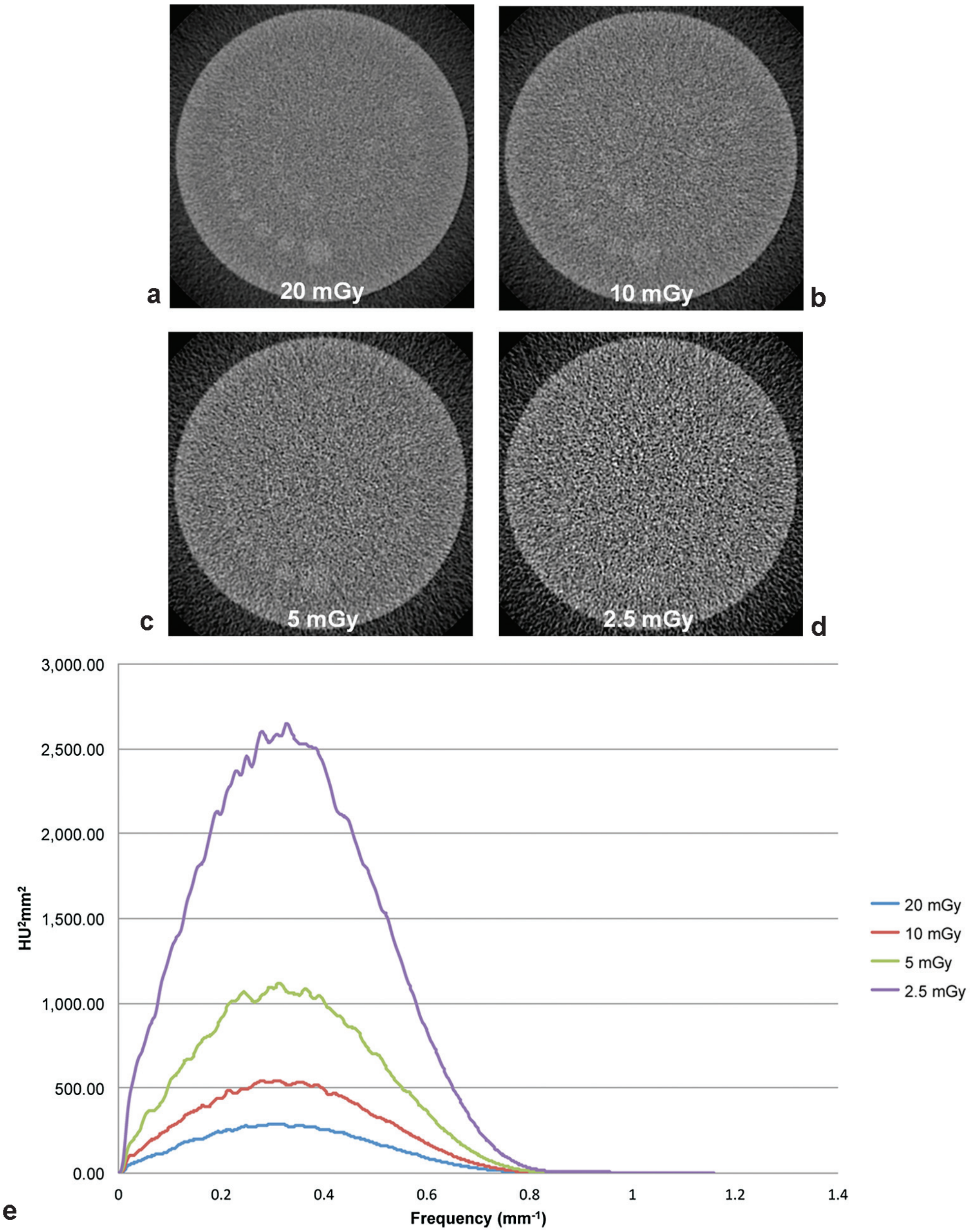


Fig. 1 (**a–d**) Effect of the dose decrease (20, 10, 5, and 2.5 mGy on the low-contrast detectability (slice thickness 1.25 mm; window level/width 180/60 HU = dedicated liver parameters). The increase of the noise level clearly affects the detectability of low contrast structures. (**e**) Corresponding noise power spectra graph.

index ($CTDI_{vol}$), which is related to the noise level in an image and represents the average dose in the scanned region. The $CTDI_{vol}$ is the average dose delivered to a polymethyl-methacrylate phantom (either a head phantom, 16 cm in diameter, or a body phantom, 32 cm in diameter). Therefore, the $CTDI_{vol}$ is the dose to a phantom and reflects the scanner radiation output.[20] Although the $CTDI_{vol}$ can be used to compare different scanners or protocols, or give an idea of the noise level in an image using a certain protocol, it does not represent the dose to a specific patient. The actual dose to a specific patient depends on the size and shape of that patient.[20] To improve the estimation of the average dose delivered in one slice in a certain patient from the $CTDI_{vol}$, the size-specific dose estimator (SSDE) was introduced (with $SSDE = CTDI_{vol} \times f_{size}$, where f_{size} depends on the effective diameter of the patient).[21] To estimate the exposure from the entire CT examination, the dose length product (DLP) (mGy × cm) is used (with $DLP = CTDI_{vol} \times$ scan length). However, $CTDI_{vol}$, SSDE, and DLP are only dose indicators; they cannot be used directly to estimate the stochastic risk associated with an exposure. Stochastic risks should be assessed using the effective dose, E (mSv), which is obtained by multiplying the DLP by a conversion coefficient that depends on the radiosensitivity of each anatomical region. Conversion coefficients for different musculoskeletal anatomical regions have been calculated.[22–25]

Evaluation of Image Quality: Definitions

As already discussed, there has been a renewed interest in the last few years for the optimization of CT radiation dose. To optimize CT protocols, it is necessary to establish a link between image quality and patient radiation dose. Image quality can be assessed both quantitatively and qualitatively. Qualitative image evaluation usually consists of a subjective evaluation of images by observers using semiquantitative scales.[26] Quantitative assessment of image quality, in contrast, corresponds to measuring certain physical parameters of the image. The simplest way to quantify image quality is to consider the output signal to be proportional to the input signal, and spatial resolution and noise properties to be equal in the whole image plane. When using filtered back-projection reconstruction algorithms, these conditions are generally accepted, and the following objective image quality parameters can be used:

- *Image noise*: This parameter assesses the fluctuation of the pixel values in a homogeneous area of the image. It can be estimated by measuring the standard deviation (SD) of the CT numbers in a defined region of interest (ROI) over a homogeneous area of the image. However, SD does not give any information on the texture of noise. To assess image noise texture, it is useful to compare signal and noise at each spatial frequency, by decomposing the noise variance (SD^2) into its spatial frequency component, leading to the noise power spectrum (NPS). Compared with the variance, the NPS is a more complete description of noise because it takes into account texture changes introduced by reconstruction kernels and other image processing parameters.
  Quantifying image noise is of major importance when dealing with the detection of low-contrast structures (i.e., with poor contrast between the structure and surrounding tissues) because noise impairs their detection (►**Figs. 1** and **2**). Low-contrast structures typically are represented by focal liver lesions, and they are relatively rare in the musculoskeletal system. For example, noise can impair the differentiation between a protruded intervertebral disk and nerve roots or dural sac, which all have similar attenuation values (►**Fig. 2**). For the detection of high-contrast lesions (e.g., cortical bone fractures, calcifications, lesions at CT arthrography), higher levels of image noise can be tolerated (►**Figs. 3** and **4**).
- *Signal-to-noise ratio (SNR)*: This parameter represents the average pixel value in a given ROI divided by SD. When SNR increases, image quality increases.
- *Contrast-to-noise ratio (CNR)*: As mentioned previously, image noise impacts the detection of low-contrast structures. One way to estimate the potential detectability of a low-contrast lesion is to calculate the CNR, representing

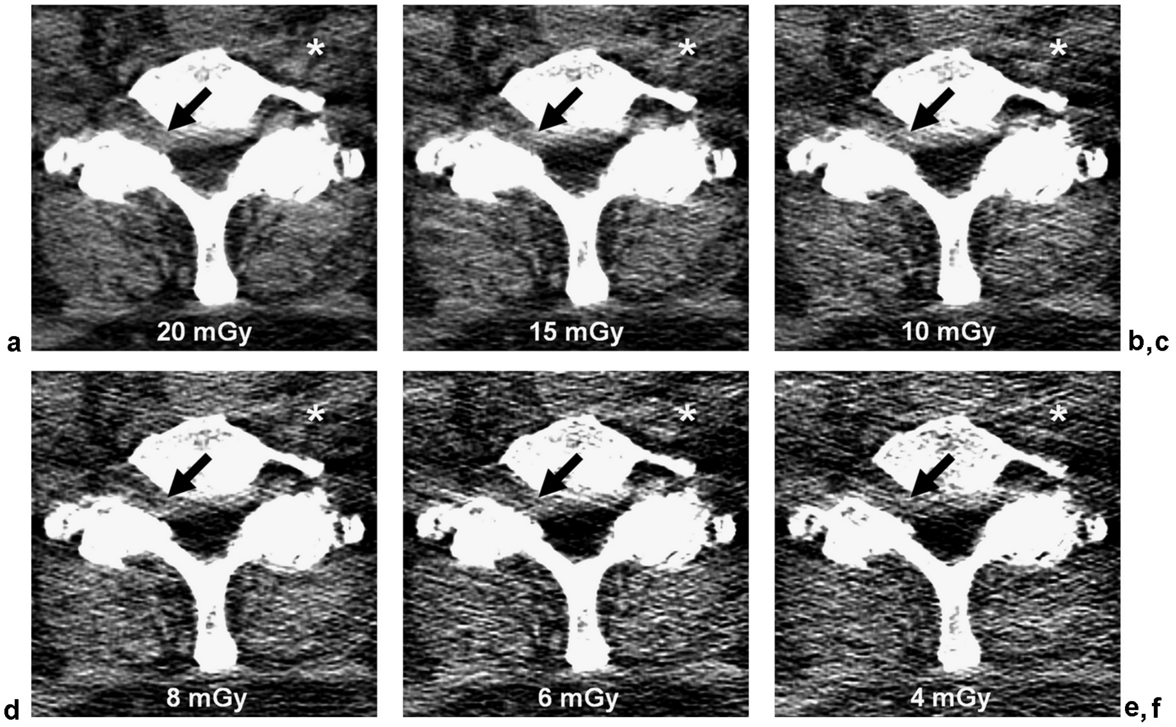


Fig. 2 Axial noncontrast computed tomography images (0.9 mm thickness; smooth kernel; window level/width, 60/300 HU) of the cervical spine at C6–C7 disk level acquired on a fresh frozen cadaver using decreasing radiation doses, from 20 (**a**) to 4 mGy (**f**), respectively. Note the gradual loss of sharp contours of the focal disk protrusion (arrows) and the progressive blurriness of the paraspinal soft tissues (asterisks), with decreasing radiation dose and concomitant increase in image noise.

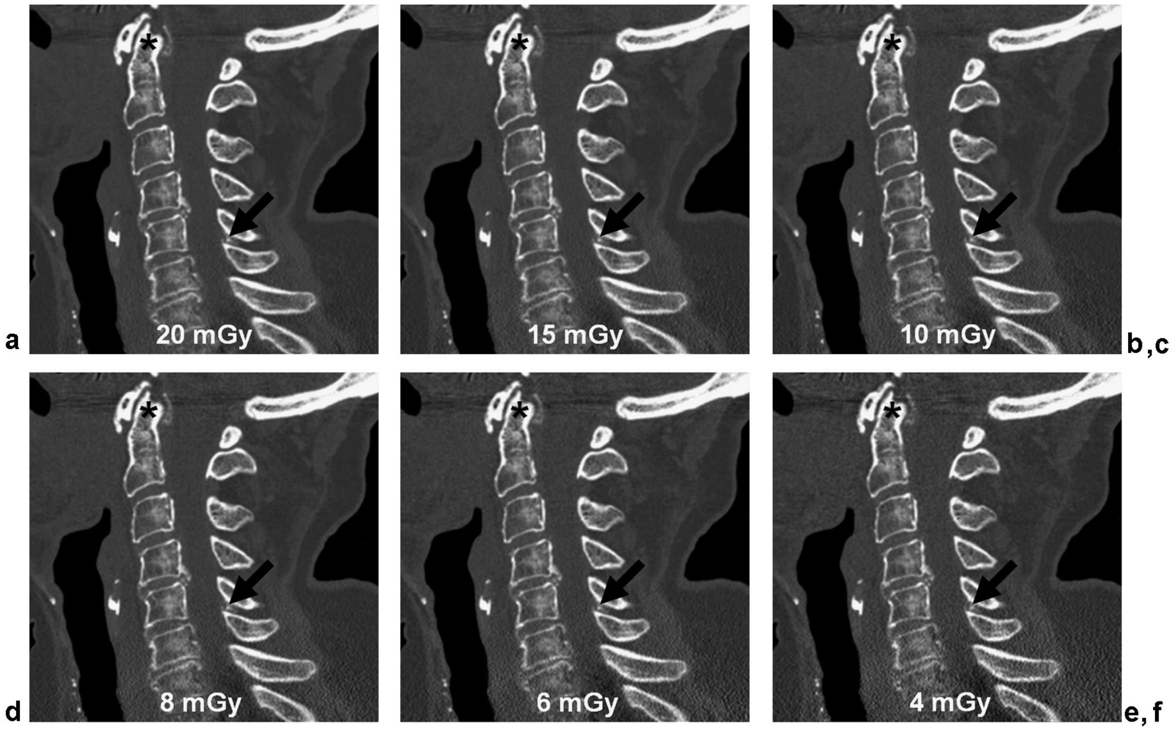


Fig. 3 Sagittal-reformatted noncontrast computed tomography images (0.9 mm thickness; sharp kernel; window level/width, 400/2000 HU) of the cervical spine acquired on a fresh frozen cadaver using decreasing radiation doses, from 20 (**a**) to 4 mGy (**f**), respectively. Note that although image noise significantly increases in the lower cervical spine with decreasing radiation dose, calcium pyrophosphate dihydrate deposits remain visible in ligaments (arrows) and the periodontoid region (asterisks) even at low dose.

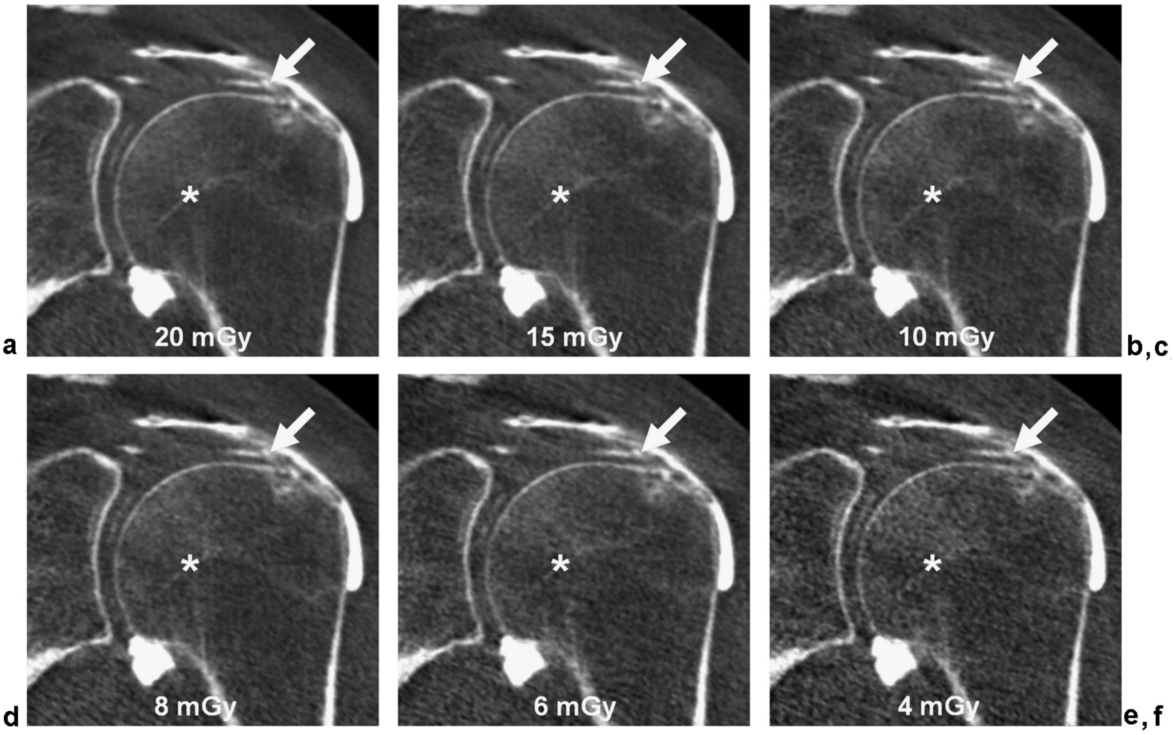


Fig. 4 Coronal-reformatted computed tomography arthrogram images (0.9 mm thickness; sharp kernel; window level/width, 400/2000 HU) of the left shoulder acquired on a fresh frozen cadaver using decreasing radiation doses, from 20 (**a**) to 4 mGy (**f**), respectively. Note that although image noise significantly increases with decreasing radiation dose, the "high-contrast" full-thickness partial supraspinatus tendon tear (arrows) remains clearly visible at low dose, whereas the "lower contrast" physeal line (asterisks) gradually disappears.

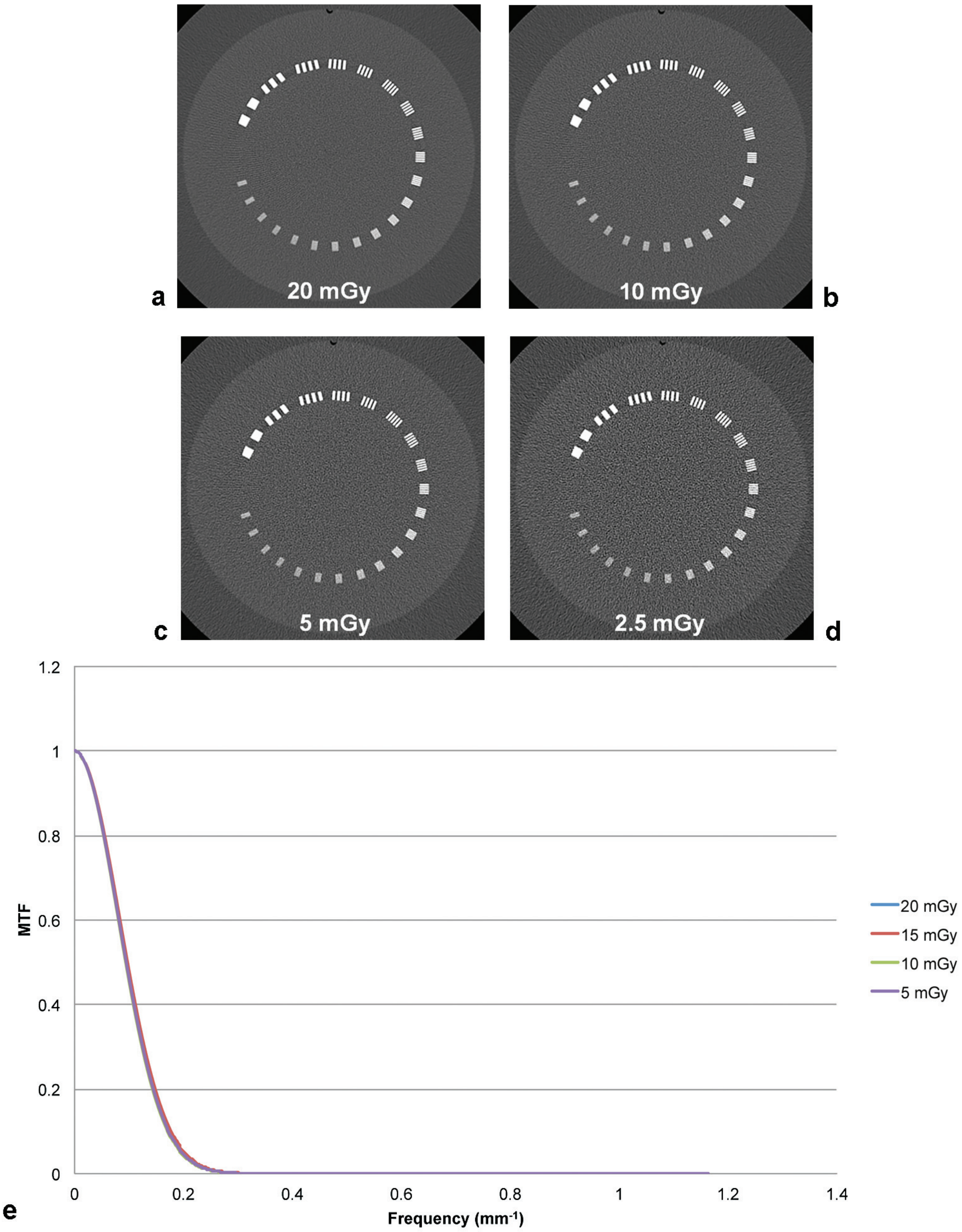


Fig. 5 (**a–d**) Effect of the dose decrease (20, 10, 5, and 2.5 mGy) on image resolution (slice thickness 1.25 mm; window L400 W2000 HU = dedicated bone parameters). The increase of the noise level has no significant effect over image resolution. (**e**) Corresponding modulation transfer function graph.

the contrast between the lesion and its environment (i.e., difference of average pixel values in each of these ROIs divided by the average SD of the two ROIs). Because CNR, contrary to NPS, does not take into account the variations of noise depending on the spatial frequency, a relationship between CNR and low-contrast detectability is only valid for a given reconstruction kernel.

- *Spatial resolution*: Ideally, the image of a punctual structure should be a point. Unfortunately, the image of such a structure will be a spot of a certain width. The poorer the spatial resolution, the wider the spot. One way to assess image quality is to draw a profile across that spot to get a function called the point spread function (PSF). To characterize that function, the full width at half maximum can be measured. In the same way as the noise parameter SD, the PSF can be decomposed into its spatial frequency, leading to the modulation transfer function (MTF). Spatial resolution is of major importance when dealing with very fine and high-contrast structures (when image conspicuity is important) (►**Fig. 5**).

Recent advances in CT technology, including new image reconstruction techniques such as IR algorithms, have led us to reconsider these image quality parameters due to their limitations.[27–29] For low-contrast structures, and to ensure that dose reduction does not impair their detectability, mathematical model observers that reflect the performance of radiologists in their diagnostic tasks have been promoted.[30,31] For high-contrast structures, spatial resolution is key. The standard MTF is not an appropriate parameter to be used with IR algorithms because they are highly nonlinear. Other parameters, such as the task-based transfer function (TTF), have been suggested for this purpose.[32]

Conclusion

Different parameters are useful for the purpose of evaluating the radiation dose in CT imaging. The CTDI_{vol} is the main parameter used to assess the radiation generated by scanning protocols. The effective dose E is the main parameter to assess the stochastic risk associated with a radiation exposure.

Scanning protocols should be optimized according to ALARA principles to deliver the lowest radiation dose that provides diagnostic image quality. Different parameters exist to assess image quality objectively. For the detection of low-contrast structures, noise, SNR, and CNR can be used, together with NPS that is useful to assess the influence on noise texture of image filtering and processing. For the conspicuity of high-contrast structures, the MTF is the main parameter that is currently used.

The implementation of new reconstruction algorithms such as IR techniques has led us to reconsider these traditional image quality parameters. For low-contrast structures, mathematical model observers have been promoted to assess the impact of image processing on lesion detectability. For high-contrast structures, conspicuity can be assessed by nonlinear functions such as TTFs.

References

1 Roth TD, Buckwalter KA, Choplin RH. Musculoskeletal computed tomography: current technology and clinical applications. Semin Roentgenol 2013;48(2):126–139
2 Larbi A, Viala P, Omoumi P, et al. Cartilaginous tumours and calcified lesions of the hand: a pictorial review. Diagn Interv Imaging 2013;94(4):395–409
3 Lepage-Saucier M, Thiéry C, Larbi A, Lecouvet FE, Vande Berg BC, Omoumi P. Femoroacetabular impingement: normal values of the quantitative morphometric parameters in asymptomatic hips. Eur Radiol 2014;24(7):1707–1714
4 Rydberg J, Buckwalter KA, Caldemeyer KS, et al. Multisection CT: scanning techniques and clinical applications. Radiographics 2000;20(6):1787–1806
5 Omoumi P, Bae WC, Du J, et al. Meniscal calcifications: morphologic and quantitative evaluation by using 2D inversion-recovery ultrashort echo time and 3D ultrashort echo time 3.0-T MR imaging techniques—feasibility study. Radiology 2012;264(1):260–268
6 Freire V, Becce F, Feydy A, et al. MDCT imaging of calcinosis in systemic sclerosis. Clin Radiol 2013;68(3):302–309
7 Omoumi P, Vande Berg B, Simoni P, Lecouvet F. Value of CT arthrography in the assessment of cartilage pathology. In: Link TM ed. Cartilage Imaging: Significance, Techniques, and New Developments. New York, NY: Springer; 2011:37–49
8 Omoumi P, Rubini A, Dubuc J-E, Vande Berg BC, Lecouvet FE. Diagnostic performance of CT-arthrography and 1.5T MR-arthrography for the assessment of glenohumeral joint cartilage: a comparative study with arthroscopic correlation. Eur Radiol 2015;25(4):961–969
9 Omoumi P, Mercier GA, Lecouvet F, Simoni P, Vande Berg BC. CT arthrography, MR arthrography, PET, and scintigraphy in osteoarthritis. Radiol Clin North Am 2009;47(4):595–615
10 Omoumi P, Michoux N, Roemer FW, Thienpont E, Vande Berg BC. Cartilage thickness at the posterior medial femoral condyle is increased in femorotibial knee osteoarthritis: a cross-sectional CT arthrography study (Part 2). Osteoarthritis Cartilage 2015;23(2):224–231
11 Omoumi P, Michoux N, Thienpont E, Roemer FW, Vande Berg BC. Anatomical distribution of areas of preserved cartilage in advanced femorotibial osteoarthritis using CT arthrography (Part 1). Osteoarthritis Cartilage 2015;23(1):83–87
12 Omoumi P, Bafort A-C, Dubuc J-E, Malghem J, Vande Berg BC, Lecouvet FE. Evaluation of rotator cuff tendon tears: comparison of multidetector CT arthrography and 1.5-T MR arthrography. Radiology 2012;264(3):812–822
13 Omoumi P, de Gheldere A, Leemrijse T, et al. Value of computed tomography arthrography with delayed acquisitions in the work-up of ganglion cysts of the tarsal tunnel: report of three cases. Skeletal Radiol 2010;39(4):381–386
14 McCollough CH, Primak AN, Braun N, Kofler J, Yu L, Christner J. Strategies for reducing radiation dose in CT. Radiol Clin North Am 2009;47(1):27–40
15 Prabhu V, Rosenkrantz AB. Imbalance of opinions expressed on Twitter relating to CT radiation risk: an opportunity for increased radiologist representation. AJR Am J Roentgenol 2015;204(1):W48–W51
16 Brody AS, Guillerman RP. Don't let radiation scare trump patient care: 10 ways you can harm your patients by fear of radiation-induced cancer from diagnostic imaging. Thorax 2014;69(8):782–784
17 Doss M. COUNTERPOINT: should radiation dose from CT scans be a factor in patient care? No. Chest 2015;147(4):874–877
18 McCunney RJ. POINT: should radiation dose from CT scans be a factor in patient care? Yes. Chest 2015;147(4):872–874
19 Omoumi P, Verdun F, Becce F. Optimization of radiation dose and image quality in musculoskeletal CT: emphasis on iterative reconstruction techniques (part 2). Semin Musculoskelet Radiol 2015;19(5):422–430
20 McCollough CH, Leng S, Yu L, Cody DD, Boone JM, McNitt-Gray MF. CT dose index and patient dose: they are not the same thing. Radiology 2011;259(2):311–316
21 American Association of Physicists in Medicine. Size-specific dose estimates (SSDE) in pediatric and adult body CT examinations: report of AAPM Task Group 204. . Available at: https://www.aapm.org/pubs/reports/RPT_204.pdf
22 Saltybaeva N, Jafari ME, Hupfer M, Kalender WA. Estimates of effective dose for CT scans of the lower extremities. Radiology 2014;273(1):153–159
23 Biswas D, Bible JE, Bohan M, Simpson AK, Whang PG, Grauer JN. Radiation exposure from musculoskeletal computerized tomographic scans. J Bone Joint Surg Am 2009;91(8):1882–1889
24 Gervaise A, Teixeira P, Villani N, Lecocq S, Louis M, Blum A. CT dose optimisation and reduction in osteoarticular disease. Diagn Interv Imaging 2013;94(4):371–388
25 Deak PD, Smal Y, Kalender WA. Multisection CT protocols: sex- and age-specific conversion factors used to determine effective dose from dose-length product. Radiology 2010;257(1):158–166
26 Bongartz G, Golding SJ, Jurik AG, et al. European Guidelines for Multislice Computed Tomography. Luxembourg: European Commission; 2004
27 Omoumi P, Verdun FR, Ben Salah Y, et al. Low-dose multidetector computed tomography of the cervical spine: optimization of iterative reconstruction strength levels. Acta Radiol 2014;55(3):335–344

28 Becce F, Ben Salah Y, Verdun FR, et al. Computed tomography of the cervical spine: comparison of image quality between a standard-dose and a low-dose protocol using filtered back-projection and iterative reconstruction. Skeletal Radiol 2013;42(7):937–945
29 Tobalem F, Dugert E, Verdun FR, et al. MDCT arthrography of the hip: value of the adaptive statistical iterative reconstruction technique and potential for radiation dose reduction. AJR Am J Roentgenol 2014;203(6):W665–W73
30 Christianson O, Chen JJS, Yang Z, et al. An improved index of image quality for task-based performance of CT iterative reconstruction across three commercial implementations. Radiology 2015; 275(3):725–734
31 Schindera ST, Odedra D, Raza SA, et al. Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? Radiology 2013;269(2):511–518
32 Ott JG, Becce F, Monnin P, Schmidt S, Bochud FO, Verdun FR. Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms. Phys Med Biol 2014;59(15): 4047–4064

3.2 Physical approach

- [**Ott, 2014**] J. G. Ott, F. Becce, P. Monnin, S. Schmidt, F.O. Bochud, F. R. Verdun, Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms. 2014 Phys Med Biol **59(15)** pp. 4047-4064.

- [**Ott, 2015 a**] J. G. Ott, A. Ba, D. Racine, A. Viry, F. O. Bochud and F. R. Verdun, Assessment of low contrast detection in CT using model observers: Developing a clinically-relevant tool for characterizing adaptive statistical and model-based iterative reconstruction. Submitted to Zeitschrift für Medizinische Physik in 2015.

- [**Ott, 2015 b**] J. G. Ott, A. Ba, D. Racine, N. Ryckx, F. O. Bochud, H. Alkadhi and F. R. Verdun, Patient exposure optimisation through task-based assessment of a new model-based iterative reconstruction technique. 2015 Accepted for publication in Radiation Protection Dosimetry.

- [**Ott, 2015 c**] J. G. Ott, P. Omoumi, E. Dugert, F. O. Bochud, F. R. Verdun and F. Becce, Task-based assessment of an MBIR algorithm: A study in all reconstruction planes. Submitted to investigative radiology.

- [**Verdun & Racine, 2015**] F.R. Verdun & D. Racine, J. G. Ott, M. J. Tapiovaara, P. Toroi, F. O. Bochud, W. J. H. Veldkamp, A. Schegerer, R. W. Bouwman, I. Hernandez Giron, N. Marshall, S. Edyvean 2015 Characterising clinically relevant image quality in CT: From physical measurements to model observers. 2015 Phys. Med. **31(8)**:823-43.

iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2014 Phys. Med. Biol. 59 4047

(http://iopscience.iop.org/0031-9155/59/15/4047)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
This content was downloaded by: jyvesray
IP Address: 153.109.40.250
This content was downloaded on 01/10/2014 at 06:25

Please note that terms and conditions apply.

IOP Publishing | Institute of Physics and Engineering in Medicine Physics in Medicine & Biology
Phys. Med. Biol. **59** (2014) 5203–5203 doi:10.1088/0031-9155/59/17/5203

Corrigendum:Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms (2014 *Phys. Med. Biol.* 59 4047–64)

Julien G Ott[1], **Fabio Becce**[2], **Pascal Monnin**[1], **Sabine Schmidt**[2], **François O Bochud**[1] and **Francis R Verdun**[1]

[1] Institute of Radiation Physics – Lausanne University Hospital – Rue du Grand-Pré, 1 – 1007 Lausanne – Switzerland
[2] Department of Diagnostic and Interventional Radiology – Lausanne University Hospital – Rue du Bugnon, 46 – 1011 Lausanne – Switzerland

E-mail: julien.ott@chuv.ch

Received 11 July 2014
Accepted for publication 23 July 2014
Published 19 August 2014

The wrong figure was used for figure 10. This is the correct one:

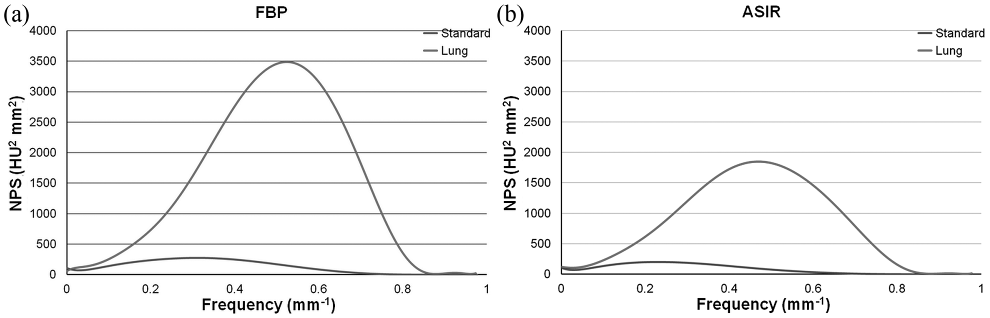


Figure 10. Comparison between NPS of standard and lung kernels for (*a*) FBP and (*b*) ASIR algorithms at 35 mAs.

IOP Publishing | Institute of Physics and Engineering in Medicine Physics in Medicine & Biology

Phys. Med. Biol. **59** (2014) 4047–4064 doi:10.1088/0031-9155/59/4/4047

Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms

Julien G Ott[1], Fabio Becce[2], Pascal Monnin[1], Sabine Schmidt[2], François O Bochud[1] and Francis R Verdun[1]

[1] Institute of Radiation Physics, Lausanne University Hospital, Rue du Grand-Pré, 1-1007 Lausanne, Switzerland
[2] Department of Diagnostic and Interventional Radiology, Lausanne University Hospital, Rue du Bugnon, 46-1011 Lausanne, Switzerland

E-mail: julien.ott@chuv.ch

Received 25 January 2014, revised 19 May 2014
Accepted for publication 10 June 2014
Published 3 July 2014

Abstract

The state of the art to describe image quality in medical imaging is to assess the performance of an observer conducting a task of clinical interest. This can be done by using a model observer leading to a figure of merit such as the signal-to-noise ratio (SNR). Using the non-prewhitening (NPW) model observer, we objectively characterised the evolution of its figure of merit in various acquisition conditions. The NPW model observer usually requires the use of the modulation transfer function (MTF) as well as noise power spectra. However, although the computation of the MTF poses no problem when dealing with the traditional filtered back-projection (FBP) algorithm, this is not the case when using iterative reconstruction (IR) algorithms, such as adaptive statistical iterative reconstruction (ASIR) or model-based iterative reconstruction (MBIR). Given that the target transfer function (TTF) had already shown it could accurately express the system resolution even with non-linear algorithms, we decided to tune the NPW model observer, replacing the standard MTF by the TTF. It was estimated using a custom-made phantom containing cylindrical inserts surrounded by water. The contrast differences between the inserts and water were plotted for each acquisition condition. Then, mathematical transformations were performed leading to the TTF. As expected, the first results showed a dependency of the image contrast and

© 2014 Institute of Physics and Engineering in Medicine

noise levels on the TTF for both ASIR and MBIR. Moreover, FBP also proved to be dependent of the contrast and noise when using the lung kernel. Those results were then introduced in the NPW model observer. We observed an enhancement of SNR every time we switched from FBP to ASIR to MBIR. IR algorithms greatly improve image quality, especially in low-dose conditions. Based on our results, the use of MBIR could lead to further dose reduction in several clinical applications.

Keywords: computed tomography, image quality, iterative reconstruction, modulation transfer function, noise power spectrum, target transfer function

(Some figures may appear in colour only in the online journal)

1. Introduction

Since the beginning of the 70s, the use of computed tomography (CT) has been steadily on the rise, and nowadays it has become an indispensable tool in the clinical environment. Patients certainly benefit from CT examinations; however, in the last decade its impact on the collective radiation dose has drastically increased. In order to ensure that the benefit of CT procedures continues to significantly outweigh the potential risk from radiation exposure, particular attention needs to be given to optimise the relationship between clinical effectiveness and radiation risks. Furthermore, the assessment of clinical effectiveness remains difficult to quantify, and image quality parameters, such as those described in ICRU 87, are generally used as a surrogate thereof.

An objective assessment of image quality relies in general on the application of the signal detection theory with which image quality can be quantified using the way a signal (i.e. a given structure), as a function of its spatial frequency, is transferred by the system via the so-called modulation transfer function (MTF). Image noise properties can be characterised by the noise power spectrum (NPS). To better relate image quality performance to specific tasks, the use of model observers developed for simple detection tasks is an interesting approach, since it incorporates standard image quality criteria to provide an objective score or signal-to-noise ratio (SNR) (Hernandez-Giron et al 2011).

A relatively simple model observer such as the non-prewhitening (NPW) model observer can be expressed in the Fourier domain with the MTF and NPS by assuming that the image system is linear (Monnin et al 2011). While this is practically the case for the standard filtered back-projection (FBP) reconstruction algorithm, this is no longer valid with the various iterative reconstruction (IR) algorithms currently available. IR algorithms are known to be highly non-linear (Hsieh et al 2013, Thibault et al 2007) and to introduce a dependency of the image contrast and noise on the spatial resolution properties of the reconstructed images (Richard et al 2012). In order to overcome this problem, we used a metric defined as the target transfer function (TTF) (Richard et al 2012, Brunner and Kyprianou 2013), which allows characterisation of the resolution even if it depends on noise and contrast. Those two metrics are similar but differ from one another in the sense that MTF only applies to a single given contrast level. TTF was already used by Brunner and Richard, but in this work we adapted it for IR algorithms of first and second generation combined with standard and sharp kernels. Eventually, we wanted to compute a detectability index based on this metric.

The aim of the present study was first to further investigate the behaviour of spatial resolution as a function of the image contrast, image noise and radiation dose for various IR algorithms. The TTF was plotted using three relatively high contrast levels in the presence of scatter

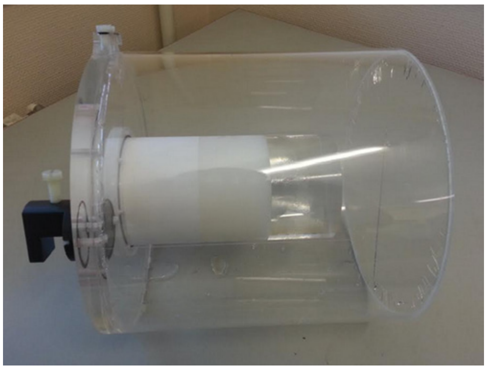


Figure 1. Custom-made TTF phantom: a 10 cm-diameter cylinder made of Teflon, low-density Polyethylene and Plexiglas.

to better represent a range of clinically relevant situations. The second goal of our investigation was to study the evolution of the SNR estimated with the NPW model observer, replacing the standard MTF by the TTF. It is noteworthy that this study aimed to characterise the image quality properties of a CT system in situations where image noise is not a limiting factor.

2. Materials and methods

Images from our custom-made phantom with its three cylindrical inserts were used to compute the TTFs. We built our TTF phantom with a 10 cm diameter cylinder made of Teflon (polytetrafluoroethylene), low-density Polyethylene and Plexiglas (polymethyl methacrylate) centred in water (figure 1). The materials of the cylinder mimic cortical bone, fatty tissue and cartilaginous tissue, respectively. The phantom also contained a homogeneous region used to calculate NPS. Those were then used to estimate the SNR (or the detectability index) according to the NPW model. The images were acquired at the Lausanne University Hospital (CHUV) with a Discovery HD 750 CT scanner (GE Healthcare, USA). We used the GE software installed on the machine for our acquisitions. Next, we acquired several image sets of the phantom; each set was composed of 93 image slices and corresponded to a particular scanning condition. We performed all the acquisitions at 120 kVp, with a 260 mm field-of-view (DFOV), a 512 × 512 matrix size, a slice thickness of 2.5 mm, a small focal spot, and using the helical mode with a pitch close to one. Four different tube loads were used: 5 mAs (10 mA and 0.5 s rotation time), 35 mAs (35 mA and 1 s rotation time), 170 mAs (170 mA and 1 s rotation time) and finally 450 mAs (450 mA and 1 s rotation time). These tube outputs led to 0.4, 3, 15 and 40 mGy of $\mathrm{CTDI}_{\mathrm{vol}}$, respectively. Those values were calculated as described in the IEC 60601-2-44. We used the classic FBP reconstruction algorithm as well as the adaptive statistical iterative reconstruction (ASIR) at a percentage of 40%, with both the GE smooth 'standard' and sharp 'lung' kernels. The last-generation model-based iterative reconstruction (MBIR) developed by GE, the 'VEO' algorithm, was also used, but it was only compatible with the standard kernel. We made a comparison between those different categories of algorithms as already suggested in the literature (Kaza et al 2014). A total of 24 different acquisition conditions were obtained.

The image noise was investigated within a 6 cm-long uniform region of our TTF water phantom. NPS were calculated with a home-made Matlab® routine (The MathWorks, Natick, MA, USA) based on 25 image slices of the homogeneous water region containing regions of interests (ROIs) of 128 × 128 pixels. A radial-averaged NPS was obtained based on the guidelines described in the ICRU report 54. The methodology we used was similar to the one used by Miéville (Miéville et al 2011a, b). Resolution was investigated through an object-specific MTF, which we referred to as the TTF. It allowed us to take into account noise and contrast effect when characterising image resolution in different acquisition conditions. Common methods to compute MTF (and therefore also TTF) consist of using an edge spread function (ESF) (Judy 1976), a line spread function (LSF) (Boone 2001), a point spread function (PSF) (Nickoloff and Riley 1985) or methods like the ones proposed by Droege (Droege and Morin 1982), Friedman (Friedman et al 2013) and Nakaya (Nakaya et al 2012). Extensive work has already been done to compare and choose the metric (either ESF or LSF or even PSF) the most adapted to our experimental conditions when computing the MTF (Miéville et al 2010, Samei et al 2006). Based on those elements and on the geometry of our phantom we computed the TTF starting from the ESF. The main mathematical steps of this process are described below, but extensive details can be found in the ICRU report 41.

We fitted the ESF and used the curve equation in order to compute the TTF, performing only analytical calculation. This method allowed us to diminish noise in the TTF curves as already investigated by Boone and Maidment (Boone and Seibert 1994, Maidment and Albert 2002). Several fitting models have already been tested in the literature (Schneiders and Bushong 1978, Schneiders and Bushong 1980, Yin *et al* 1990) but they all appear to fail when dealing with non-monotonic ESFs. To overcome this problem we used a fitting model developed by Brunner which is based on a combination of Gaussian and Boltzmann functions (Brunner and Kyprianou 2013). All the fitting techniques and image processing steps were performed using Matlab®. We began by generating a ROI to select only the cylinder and the water background. Then, two ROIs of 40 × 40 pixels, one in the background made of water, the other in the cylinder (either Teflon, or low-density Polyethylene or Plexiglas) were generated and used to produce a binary image. The mean value of those two ROIs was calculated. The pixels were set to white if they were closer to the mean value of the cylinder and to black if they were closer to the mean background. This transformation allowed us to determine the geometrical centre of the cylinder by computing the centre of gravity of the white pixels. This point was then used as the origin of coordinates on the acquired image (figure 2). The raw ESF was obtained by plotting the grayscale value of each pixel (in Hounsfield Units (HU)) as a function of its distance from the coordinate centre. Then, we performed rebinning on the data set points obtained by averaging the HU value of all the pixels in 0.01 mm bins. This value was chosen in order to find a good compromise between a short time for the convergence of the fit while maintaining a good precision. In the meantime, processing was performed on the raw ESF, i.e. background subtraction and normalisation. Moreover, Brunner *et al* (Brunner and Kyprianou 2013) already underlined that the ESF obtained this way is asymmetric because the scatter from the background is different from the scatter from the cylinder. Since the TTF characterises the diffusion in the direction from the cylinder to the background, only the background part of the processed ESF is relevant. Thus, we performed a central symmetry at the centre of the slope of our normalised and resampled ESF and kept only the background part.

The obtained symmetric ESF was then fitted by the following function:

$$\mathrm{ESF}(r) = \frac{1}{1+\mathrm{e}^{-r/m}} + C\left(\mathrm{e}^{-s(r+h)^2} - \mathrm{e}^{-s(r-h)^2}\right) \tag{1}$$

The first term of the sum corresponds to a normalised sigmoid and the second term adds two Gaussians in order to take into account the effects of the edge enhancement filter (EEF) when using sharp kernels (such as a lung kernel, for example). This EEF results in bumps on

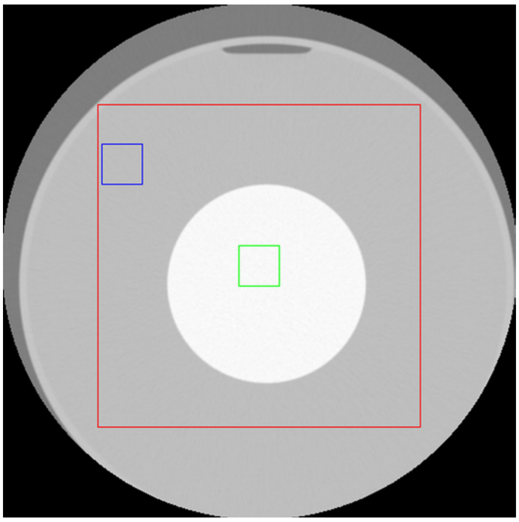


Figure 2. Regions of interest in the phantom: red, analysed ROI; blue, background ROI; green, ROI in the analysed material.

each side of the ESF slope (figure 3(*b*)). Before performing the fit, the resampled ESF was displayed on the screen and the user had to visually check if an edge enhancement was performed by checking for bumps on each side of the sigmoid (figure 3). If there were none, s was set equal to zero in equation (1), otherwise no modification was done. In order to produce the LSF, equation (1) is analytically differentiated.

$$\mathrm{LSF}(r)=\begin{cases}\dfrac{1}{4m\left(\cosh\,\frac{x}{2m}\right)^{2}}-C\ \mathrm{e}^{s(-r^{2}-2hr-h^{2})}\,2\\ \quad s\left(\mathrm{e}^{4hsr}(r-h)+r+h\right)\quad(\text{with EEF})\\ \quad\dfrac{1}{4m\left(\cosh\,\frac{r}{2m}\right)^{2}}\quad(\text{without EEF})\end{cases}\tag{2}$$

Finally, performing the Fourier transform (FT) of the LSF we can obtain the TTF normalised at one for $f=0$:

$$\mathrm{TTF}(f)=\begin{cases}\left|\dfrac{2\pi^{2}fm}{\sinh\,(2\pi^{2}fm)}-C\dfrac{\sqrt{8\pi^{3}}\ \mathrm{if}\ \sqrt{2s}}{2s}(\mathrm{e}^{4\mathrm{i}\pi fh}-1)\,\mathrm{e}^{-\frac{2\mathrm{i}\pi fsh+(\pi f)^{2}}{s}}\right|\quad(\text{with EEF})\\ \qquad\left|\dfrac{2\pi^{2}fm}{\sinh\,(2\pi^{2}fm)}\right|\quad(\text{without EEF})\end{cases}\tag{3}$$

To produce this formula, the calculation was performed analytically. Since the calculation of the FT of (2) was not trivial, a formal calculus software was used (Mathematica, Wolfram research, Champaign, IL, USA) to yield (3) (figure 4).

The goodness of our fitting model for the ESF was assessed using the sum square method. The sum square error (SSE) is defined by $\mathrm{SSE}=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$, where y_i is the result of

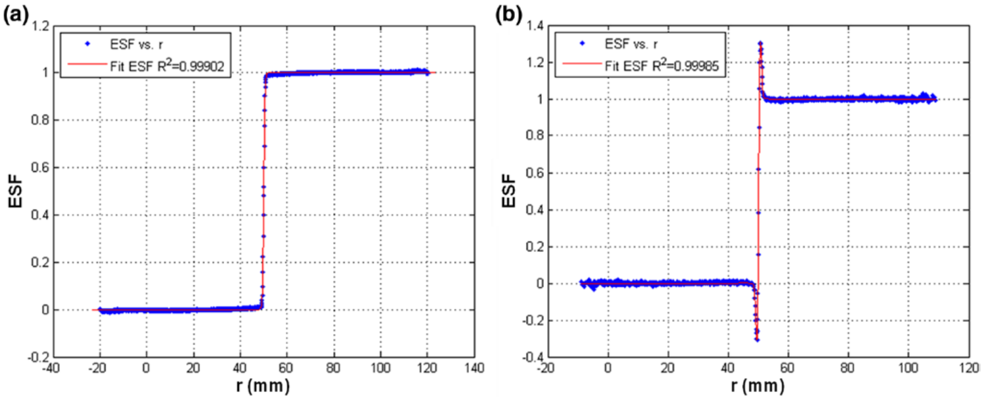


Figure 3. Two ESFs, (*a*) without and (*b*) with EEF.

the measurement of the *i*th variable and $\hat{y}_i$ is the value predicted by the fit. In the meantime the sum square residual (SSR) was calculated using $\mathrm{SSR} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$, where $\bar{y}$ is the mean of the measured values. Thus we can define the total sum of square (SST): $\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR} = \sum_{i=1}^{n} (y_i - \bar{y})^2$. In the end, the coefficient R^2 was used to assess the goodness of the fit: $R^2 = \mathrm{SSR}/\mathrm{SST} = 1 - \mathrm{SSE}/\mathrm{SST}$. A R^2 equal to one shows a perfect fit whereas a result close to zero indicates a poorly fitted curve.

The error on the TTF was evaluated by computing the TTF curve ten times using different images under the same conditions. The extreme values provided the amplitudes of the error bars.

The SNR was evaluated based on the work of Monnin *et al* (Monnin et al 2011) and adapting it to CT:

$$d' = \mathrm{SNR} = \frac{\sqrt{2\pi}\ \Delta\mathrm{HU} \int_0^{f_{Ny}} S^2(f)\,\mathrm{TTF}^2(f)\, f\ \mathrm{d}f}{\sqrt{\int_0^{f_{Ny}} S^2(f)\,\mathrm{TTF}^2(f)\,\mathrm{NPS}(f) f\ \mathrm{d}f}} \tag{4}$$

where SNR is the signal-to-noise ratio, f_{Ny} is the Nyquist frequency and $S(f)$ is the FT of the input signal. For a disc of radius R we have:

$$S(f) = \frac{R}{f}\, J_1(2\pi R f) \tag{5}$$

J_1 being the first-order Bessel function of the first kind.

3. Results

The ESF was computed for each acquisition protocol and each material of the cylinder. Figures 3 and 5 show examples of some ESFs as well as the corresponding LSFs. The regression coefficient (R^2) accounting for the goodness of the fit was checked for all the curves we obtained. Its value indicated that some acquisition protocols presented a low R^2 (inferior to 0.90000). We chose to exclude them from the study, making the mean value of R^2 equal to

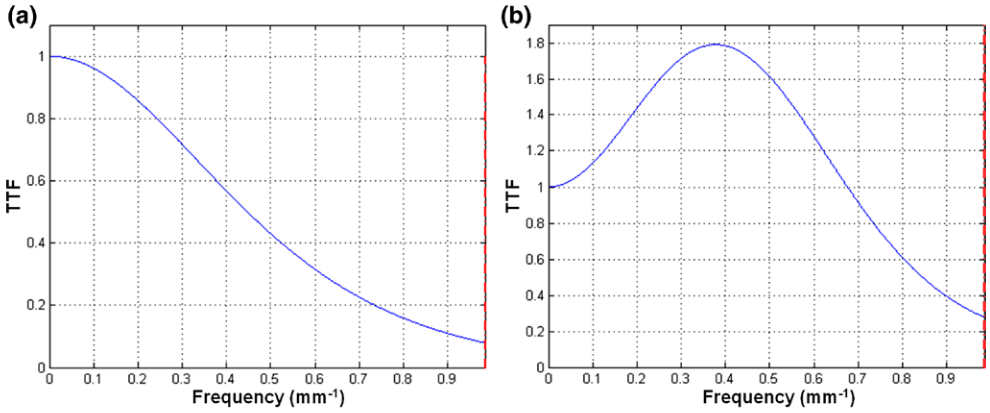


Figure 4. Two TTFs, resulting from ESFs (*a*) without and (*b*) with EEF.

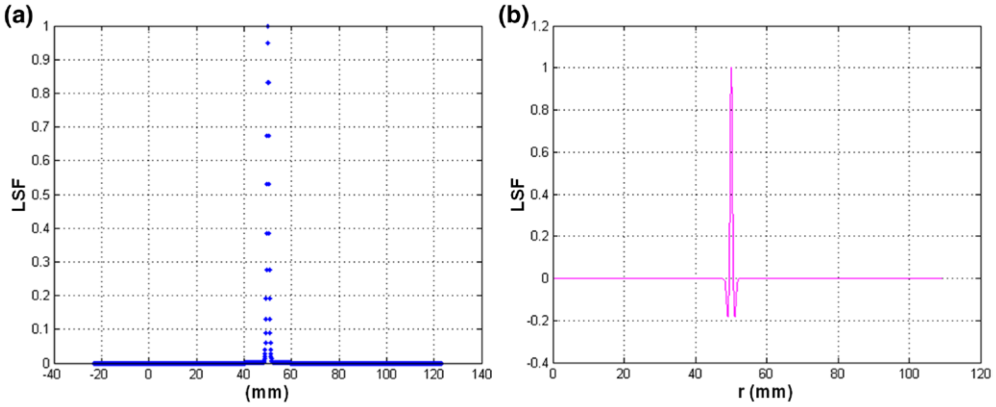


Figure 5. Line spread functions for (*a*) standard kernel, ASIR, Teflon and 450 mAs; and (*b*) lung kernel, ASIR, Teflon and 450 mAs.

0.99875 for the remaining cases. A low R^2 was observed for most acquisitions at 5 mAs as well as 35 mAs in the Polyethylene and Plexiglas with the lung kernel and the FBP algorithm. High noise levels in the images acquired under those protocols are responsible for these low values of R^2.

Figures 6(*a*)–(*c*) summarise the TTFs acquired at the highest dose level with the standard kernel for (*a*) FBP, (*b*) ASIR and (*c*) MBIR in the three materials. For the FBP algorithm, the results in Polyethylene were slightly inferior to those obtained with Teflon and Plexiglas, but this can be explained by the low contrast difference between Polyethylene and water (±80 HU) resulting in a less precise fit than with Teflon or Plexiglas (±1000 HU and ±125 HU). In the end, the error bars allow the curves to overlay each other, meaning that when using the FBP algorithm, the TTFs for the different materials can be considered similar. This result is not surprising since spatial resolution is supposed to be independent of contrast. When focusing on ASIR results, we observed a drop in the spatial resolution of low-contrast materials (Polyethylene and Plexiglas) and an enhancement in the spatial resolution for high-contrast material (Teflon) compared to FBP. The error bars do not allow TTFs to overlay one

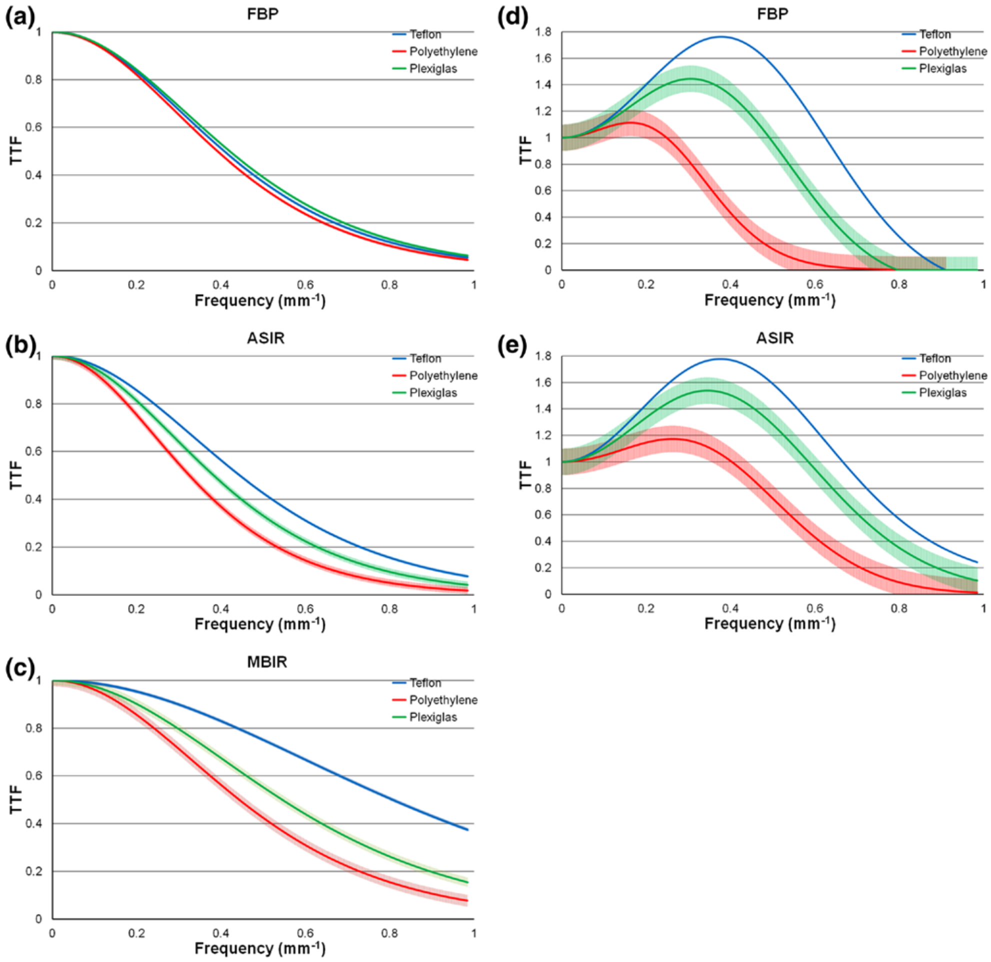


Figure 6. Comparison of TTFs at 450 mAs in different materials with the standard kernel using (*a*) FBP, (*b*) ASIR and (*c*) MBIR algorithms; and with the lung kernel using (*d*) FBP and (*e*) ASIR.

another, indicating that spatial resolution depends on contrast for the ASIR algorithm. This trend is even more pronounced for the MBIR algorithm, with large differences between the three materials. The TTF of Teflon is strongly increased, whereas the TTF of Polyethylene is comparable to the results of the FBP algorithm, and Plexiglas lies in between. When switching from FBP to IR algorithms such as ASIR or MBIR, we observe that signal transfer is dependent on the contrast level.

Figures 6(*d*) and (*e*) show the TTFs under the same conditions as in (*a*–*c*) but using the lung kernel. MBIR is not represented here since it does not provide a lung mode. With the lung kernel and the FBP algorithm, we observe that the TTF depends on the material. This trend was not observed with the standard kernel. The lung kernel is non-linear as it emphasises high spatial frequencies, which accounts for our observations on the TTF, even though the FBP algorithm was thought to be linear. ASIR follows the same trend as FBP even though, in this case, the result was more predictable given its behaviour with the standard kernel (figure 6).

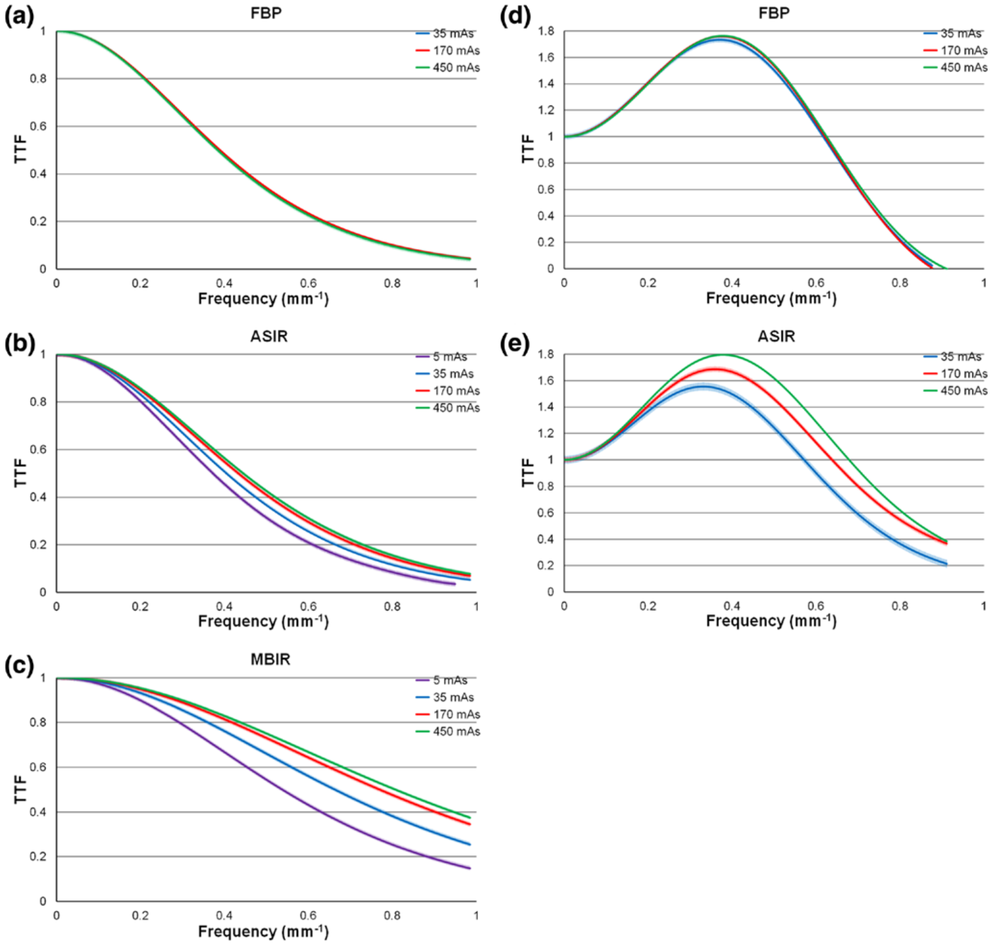


Figure 7. Comparison of TTFs at different tube loads in Teflon, using the standard kernel with (*a*) FBP, (*b*) ASIR and (*c*) MBIR algorithms; and the lung kernel with (*d*) FBP and (*e*) ASIR algorithms.

Figure 7 shows the TTFs with the standard kernel in Teflon for different tube loads and reconstruction with (*a*) FBP, (*b*) ASIR and (*c*) MBIR. For the FBP reconstruction, there is no dependence on tube load. Indeed, when using a classical linear reconstruction algorithm such as FBP with the standard kernel, tube load only modifies the amount of noise in the image. Therefore, when dropping the tube load the quality is lowered because of the increase of noise in the image. However, provided the amount of noise does not prevent the signal transfer, the resolution still remains the same when increasing noise. Therefore, no change is observed in TTF. Using the ASIR reconstruction, a slight tube load dependence is visible (the TTF tends to increase with the tube load). Furthermore, when switching from FBP to ASIR and MBIR the greater noise reduction allowed for the computation of the TTF at 5 mAs, illustrating one of the benefits of IR algorithms. Indeed, when trying to calculate the TTF at 5 mAs with the FBP algorithm, the high noise level did not allow us to produce a reliable fit for the ESF. Finally, when reconstructing the images using the MBIR algorithm, we observed that the TTF depends strongly on tube load. It is interesting to note that images acquired at 5 mAs and reconstructed

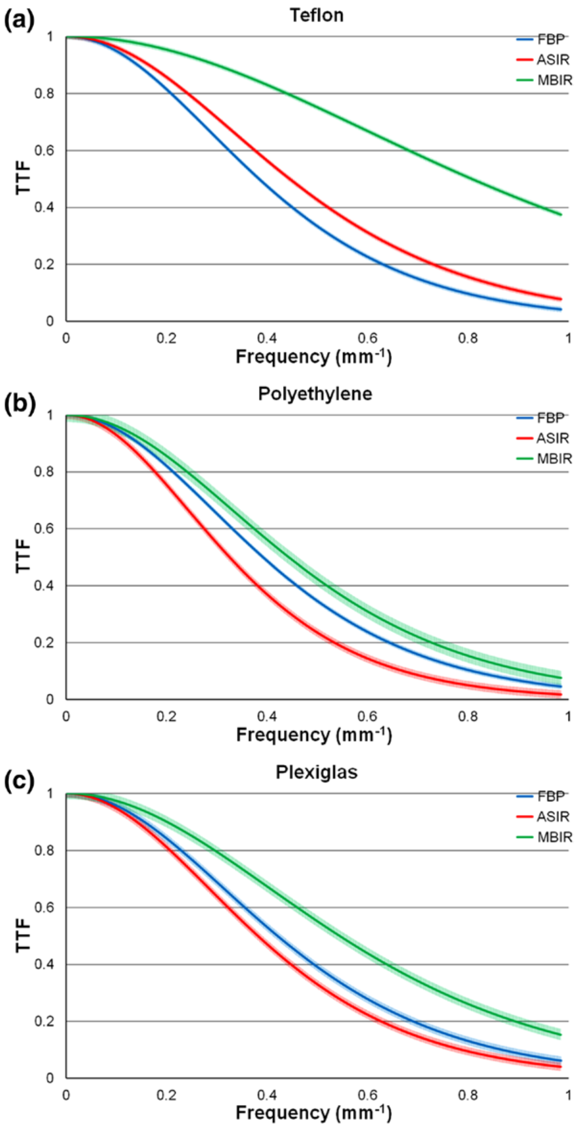


Figure 8. Comparison of the TTFs acquired with the standard kernel at 450 mAs in the different materials and with the three different algorithms.

with MBIR yielded a higher TTF than those acquired at 450 mAs and reconstructed with ASIR. Moreover, we observed that at one point a strong increase in mAs produces only a small increase in TTF. For high mAs values, a high-contrast object reconstructed with MBIR yields a much better spatial resolution than expected. Figures 7(*d*) and (*e*) show the TTFs with the lung kernel for Teflon at three different tube loads. For the FBP algorithm (*d*), the curves acquired with 170 and 450 mAs tube loads are identical. The TTF at 35 mAs is slightly lower but the error bars show that the curves overlap. The additional noise due to the low flux explains the difference between the TTF measured at 35 mAs and the other curves. When switching to ASIR (*e*), an increase in the tube load produces a noticeable increase in the TTFs.

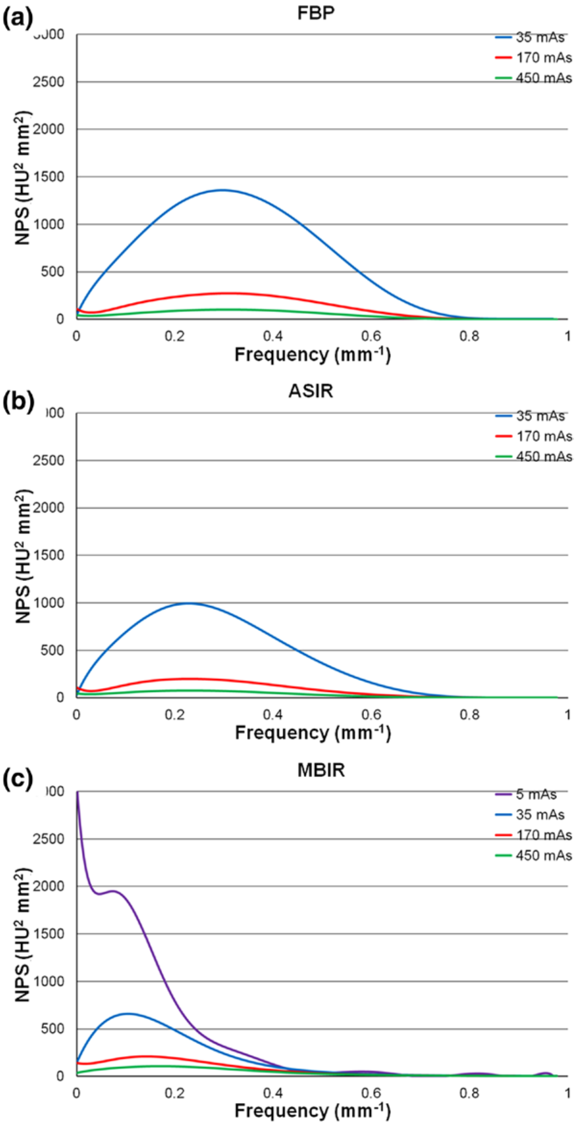


Figure 9. Noise power spectra at different tube loads and with the standard kernel for (*a*) FBP, (*b*) ASIR and (*c*) MBIR algorithms.

Figure 8 plots the TTFs acquired with the standard kernel at 450 mAs for (*a*) Teflon, (*b*) Polyethylene and (*c*) Plexiglas for the three different reconstruction algorithms. The TTFs are higher for the high-contrast object (Teflon) and lower for the low-contrast object (Polyethylene). We also observed that MBIR produces the highest TTF enhancement in every material: the higher the contrast of the material, the stronger the TTF enhancement with MBIR. In low-contrast material (Plexiglas and Polyethylene), the TTF was higher for FBP than for ASIR, indicating that a reduction in contrast is likely to produce a decrease in spatial resolution in ASIR-reconstructed images.

Figure 9 shows the NPS at different tube loads for (*a*) FBP, (*b*) ASIR and (*c*) MBIR algorithms using the standard kernel. As expected, the noise decreases with increasing tube load.

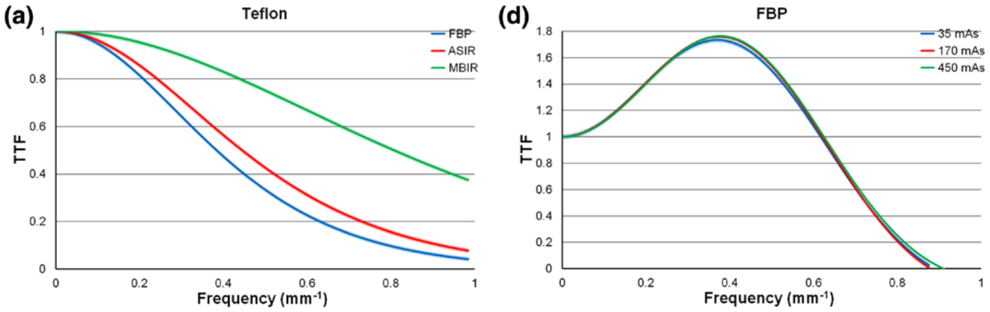


Figure 10. Comparison between NPS of standard and lung kernels for (*a*) FBP and (*b*) ASIR algorithms at 35 mAs.

Increasing the dose also leads to a shift in the mean frequency toward higher frequencies. When using the lung kernel (figure 10), we observe an increase in the noise amplitude as well as a shift of the frequencies toward higher frequencies compared to the standard kernel for both (*a*) FBP and (*b*) ASIR algorithms. Indeed, lung kernels are designed to reinforce high frequencies. Finally, at a fixed dose level, when comparing FBP, ASIR and MBIR (figure 9), we observed a strong reduction of the noise amplitude as well as a shift toward low frequencies. Indeed, switching from FBP to ASIR to MBIR induces a variation of the mean frequency from 0.334 mm^{-1} to 0.292 mm^{-1} to 0.193 mm^{-1}. This highlights the modification of image texture that occurs when using IR algorithms.

Finally, the SNR was calculated accordingly (4) for a 1 mm-radius disc in order to simulate a clinical lesion. The results are summarised in Figure 11. The SNR is always much higher in Teflon than in the other two materials. It is also slightly higher in Plexiglas than in Polyethylene. This is due to the contrast values, which are 925, −80 and 125 HU for Teflon, Polyethylene and Plexiglas, respectively. Moreover, an SNR increase is observed with increasing tube load. For all materials and acquisition protocols, the SNR also tends to reach a plateau for high tube loads. We also noticed that the lung kernel always yields a worse SNR than the standard kernel.

4. Discussion

The goal of our study was to investigate a CT task-based image quality metric, applicable when switching from FBP to different generations of non-linear IR algorithms. This was accomplished by investigating the use of TTF for a specific object instead of the classical MTF system for computing the SNR in the NPW model observer. It was observed that the classical MTF system was well adapted for linear reconstruction algorithms such as FBP with the standard kernel, with our results indicating that spatial resolution is independent of the contrast level. However, the results also indicated that spatial resolution was dependent on the contrast level for IR algorithms, and for FBP as well when using the lung kernel. This clearly shows the need to adapt the guidelines for spatial resolution measurement: the noise and contrast levels at which the resolution is measured have to be mentioned. This also has to be taken into account when comparing two CT systems and protocols. Moreover, the method we proposed also presents some advantages compared to the classic point source method usually used to compute the MTF. Indeed, when using the point source method, a small DFOV has to be used in order to obtain a reliable PSF. We could eliminate this problem by using a

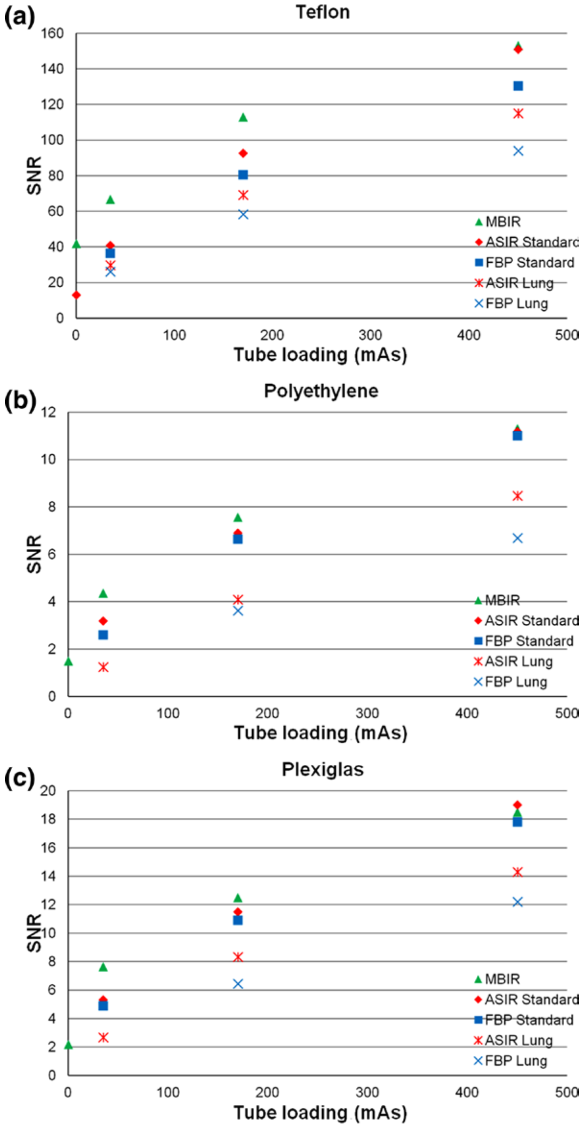


Figure 11. SNR for all acquisition conditions and in all materials.

10 cm-diameter cylinder to compute the ESF. In addition, using a circular phantom leads to natural oversampling.

The SNRs were calculated for a 1 mm-radius disc made of the three materials of our cylinder and under all acquisition protocols. Coherent results were obtained since they showed an increase of the detectability with the dose as well as the contrast level. With a mean R^2 value of 0.99875, the fit could be considered reliable and even very strong for high dose and high contrast. The R^2 values indicated that fits for low contrast, low dose, sharp filter and FBP algorithms were less reliable. This was due to the high image noise generated under those conditions. Consequently, the error on TTFs was very small for low noise data and relatively higher for noisy data. In the end, we were able to estimate a reliable SNR for almost every acquisition protocol. Those which generated too much noise and that we were unable to fit were not investigated for the calculation of SNR.

When analysing the results, the lung filter was found to lead to a lower SNR than the standard filter for all materials and tube loads. This is due to the fact that the lung kernel is designed to enhance high frequencies, and that noise appears stronger than the TTF, thus leading to a reduction in SNR. Regarding the contribution of the IR algorithms, we observed that in the three materials the same trends seemed to occur, showing that the SNR increases when switching from the FBP lung kernel to ASIR lung kernel, FBP standard kernel, ASIR standard kernel and MBIR. However, an exception to this trend can be seen for Plexiglas and Polyethylene at 450 mAs, where ASIR with the standard kernel yields a slightly better SNR than MBIR. This is due to the fact that at these contrast and dose levels, both algorithms provide a very high and similar signal value. However, at this dose level, ASIR reduces noise slightly more than MBIR, which explains the value of the SNR under those conditions. For the other cases, results showed that ASIR led to a significant noise reduction while maintaining the TTF at the same level as FBP. All this explains the better results of ASIR compared to FBP in terms of detectability. Finally, the improvement enabled by MBIR algorithms is clearly visible at low doses. Indeed, the more the dose decreases, the larger the gap between the SNR of MBIR and that of the other algorithms. This is due to the fact that MBIR not only leads to a drastic noise reduction, but also to an enhancement of the TTF. MBIR also happens to be the only algorithm for which we were able to estimate the SNR at 5 mAs in all materials.

The results of this study may also have clinical implications. Indeed, the three materials used in the phantom mimic the contrast levels of cortical bone, fatty tissues and cartilaginous structures. Moreover, the $\mathrm{CTDI}_{\mathrm{vol}}$ range (3–40 mGy) corresponds to the range of $\mathrm{CTDI}_{\mathrm{vol}}$ values for the majority of clinical CT examinations.

We conclude that the more the material is contrasted, the stronger the TTF enhancement produced by MBIR. These results are especially promising for CT urography, at present performed daily for the detection of urinary calculi. Indeed, highly contrasted structures, in particular urinary stones, despite their small size, may benefit the most from the application of MBIR, thus allowing an important reduction in radiation dose to the patient. Since the patients undergoing these kinds of examinations are often young, they may particularly benefit from a radiation dose reduction. Moreover, our study revealed that for small lesions (1 mm-radius disc in our case), MBIR outperformed ASIR (and FBP) in detecting and characterising structures mimicking cortical bone (Teflon), fat (Polyethylene), and cartilage (Plexiglas), respectively. This could be particularly helpful when performing low-dose CT imaging of the thorax and abdomen (figure 12) or the whole spine (e.g. evaluation of scoliosis in children and young adults; Kalra et al 2013). It could also be useful to accurately assess fat-containing lesions such as adrenal adenomas, soft-tissue lipomas, vertebral haemangiomas or even fatty infiltration of the muscles (e.g. in rotator cuff tears). At an ultra-low dose, MBIR was found to be the only technique able to identify all three structures. This could be interesting for young cancer patients undergoing regular follow-ups with CT. It may also be useful to further develop dynamic/functional CT imaging of the joints, which is currently limited by substantial radiation dose (Kalia et al 2009, Leng et al 2011).

However, the long reconstruction times of MBIR still limit its widespread use in emergency CT imaging. While the first-generation IR algorithms have been shown not to substantially delay image reconstruction in an emergency setting (Willemink et al 2013), this is currently not the case for the second-generation IR algorithms such as MBIR. The application of ASIR in cervical spine CT has recently allowed reducing the radiation dose to a level comparable to plain radiography with no loss of subjective image quality (Geyer et al 2013, Becce et al 2013). Nevertheless, in the future, additional dose reductions will certainly be achievable in

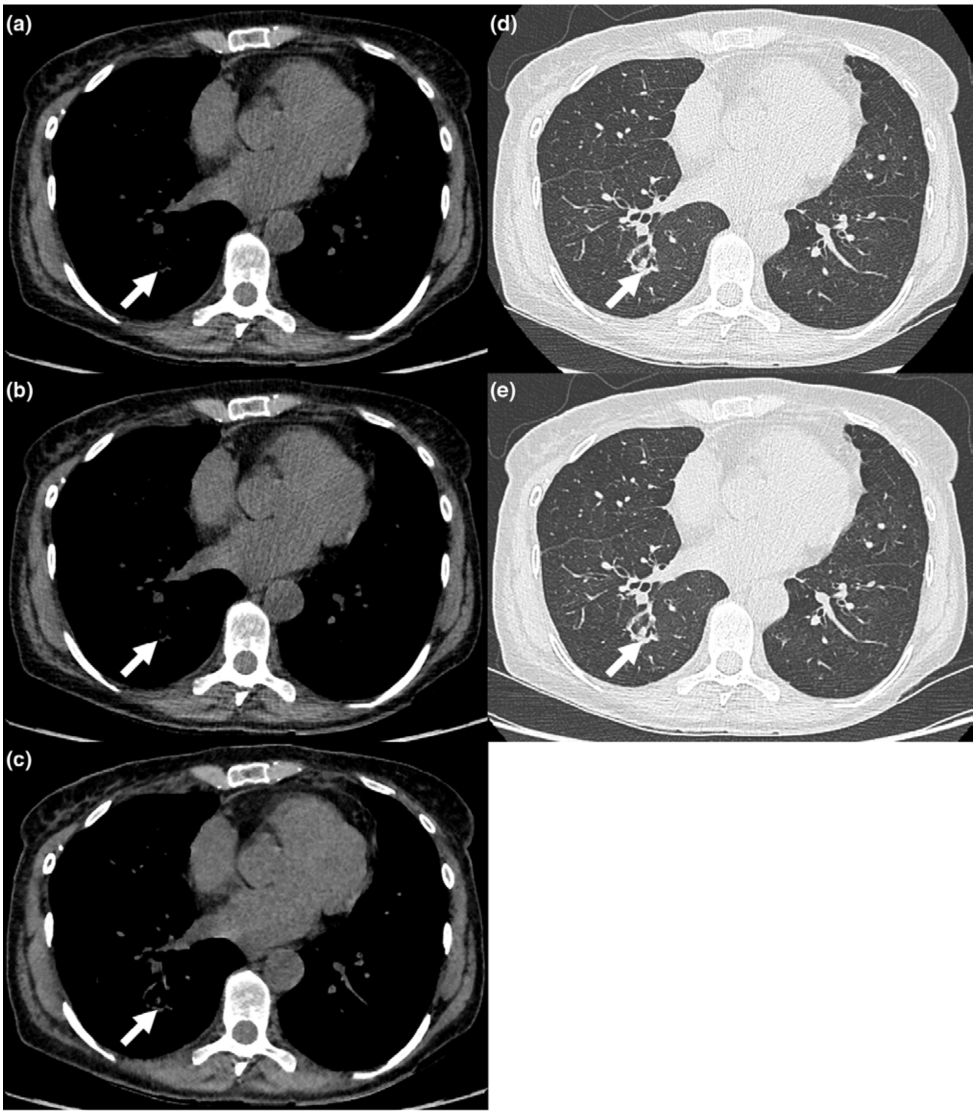


Figure 12. Follow-up axial CT images of the thorax in a 59 year-old woman with pneumonia of the right lower lobe acquired with a $\mathrm{CTDI}_{\mathrm{vol}}$ of 3 mGy, and reconstructed using the (*a*–*c*) standard, and (*d* and *e*) lung kernels. Note the changes in appearance of the right lower lobe opacity (arrows) when the image is reconstructed using the (*a* and *d*) FBP, (*b* and *e*) ASIR 40%, and (*c*) MBIR algorithms.

emergency CT imaging of the spine. Indeed, reconstruction delay will certainly be shortened thanks to the evolution of hardware and computational techniques, and it is probably only a matter of time until the requirements for use in an emergency setting are fulfilled.

Even though our study allowed us to fully answer our research questions, some limitations need to be mentioned. First, the results of our NPW model have not been compared to human observer models. Our study focused more on image quality comparisons based on physical metrics (such as objectively measured noise and output signal) allowing the benchmarking of CT units, for example. The measured SNR reached very high values, particularly for high-contrast materials (Teflon). Such values will almost never be reached by a human observer, even though human SNR will still be very high for those kinds of objects. In order to reach results more comparable to human observers, the introduction of the NPW with the eye model (NPWE) would be necessary (Hernandez *et al* 2011). Similar work has already been done by Monnin for mammography (Monnin et al 2011) and Richard in CT (Richard et al 2011), but more work needs to be done on this subject for the IR algorithms in CT. However, it was already demonstrated that the NPW model observer in the spatial-frequency domain correctly predicts the detection by human observers of well-defined signals in homogeneous backgrounds, and this allowed us to estimate the detectability behaviour with it. The second limitation we were confronted to was that, even if SNR is a key parameter to detection, the clinical results we extrapolated based on our physics metrics cannot be used directly. Indeed, the behaviour of the SNR shows that using IR would allow maintaining the detectability while reducing the dose, but more work has to be done on the subject in order to quantify the dose reduction we could achieve in real practice. In order to estimate the exact dose that could be spared, this work needs to be backed up with studies made on anatomical images and interpreted by radiologists. The third limitation is the fact that we did not investigate the impact of the object's form and size on its detectability. A simple modification of the object's transfer function would provide the results. We can observe that similar investigations were already done on the subject by Brunner (Brunner et al 2012) and seem to indicate an influence on the signal form on the SNR values. However, this was beyond the scope of this work. Another limitation is that we assumed noise to be stationary when computing our NPS. The ESF was also rotationally averaged assuming rotational symmetry of the acquired images. Finally, the study was performed on a single CT system onto which we could only investigate a single version of a first-generation IR algorithm (ASIR) and a second-generation IR algorithm (MBIR). More work needs to be done to include different CT vendors and the latest generations of reconstruction algorithms in future studies.

5. Conclusion

In this work, we built a robust and reliable tool to investigate the behaviour of the TTF metric for different reconstruction algorithms and kernels. We used a custom-made CT phantom with a three-material cylinder in water to generate an ESF using a circular edge technique. The ESF was fitted and analytically modified to yield the TTF, which provides spatial resolution performance for objects of different contrasts. Using the TTF and the NPS, we were able to compute the SNR under all acquisition conditions and thus compare image quality performance. We found that, when using IR algorithms, the resolution varied with contrast and noise, which highlights the importance of using a measurement technique such as the TTF. The TTF was used to compute the SNR, which was our final indicator of image quality. In the end, the calculation of the SNR through the NPW model was made based on physics metrics (TTF and NPS), making it well-suited to comparing CT systems based on purely objective image quality. The results provided by those SNR calculations make it clear that IR significantly improves image quality, especially for low doses. In terms of detectability, ASIR always outperforms FBP, and MBIR almost always outperforms ASIR. This is particularly true for low doses where the noise is high and the signal is weak. In conclusion, the calculation of the SNR through the NPW model observer using the TTF could be a useful tool to assess image quality and benchmark CT systems.

Acknowledgement

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

References

Becce F, Ben Salah Y, Verdun F R, Vande Berg B C, Lecouvet F E, Meuli R and Omoumi P 2013 Computed tomography of the cervical spine: comparison of image quality between a standard-dose and a low-dose protocol using filtered back-projection and iterative reconstruction *Skeletal Radiol.* **42** 937–45

Boone J M 2001 Determination of the presampled MTF in computed tomography *Med. Phys.* **28** 356–60

Boone J M and Seibert J A 1994 An analytical edge spread function model for computer fitting and subsequent calculation of the LSF and MTF *Med. Phys.* **21** 1541–5

Brunner C C, Abboud S F, Hoeschen C and Kyprianou I S 2012 Signal detection and location-dependent noise in cone-beam computed tomography using the spatial definition of the Hotelling SNR *Med. Phys.* **39** 3214–28

Brunner C C and Kyprianou I S 2013 Material-specific transfer function model and SNR in CT *Phys. Med. Biol.* **58** 7447–61

Droege R T and Morin R L 1982 A practical method to measure the MTF of CT scanners *Med. Phys.* **9** 758–60

Friedman S N, Fung G K, Siewerdsen J H, and Tsui B W 2013 A standardized approach to measure computed tomography (CT) modulation transfer function (MTF) and noise power spectrum (NPS) using the American College of Radiology (ACR) accreditation phantom *Med. Phys.* **40** 051907

Kaza R K, Platt J F, Goodsitt M M, Al-Hawary M M, Maturen K E, Wasnik A P and Pandya A 2014 Emerging techniques for dose optimization in abdominal CT *RadioGraphics* **34** 4–17

Geyer L L, Körner M, Hempel R, Deak Z, Mueck F G, Linsenmaier U, Reiser M F and Wirth S 2013 Evaluation of a dedicated MDCT protocol using iterative image reconstruction after cervical spine trauma *Clin. Radiol.* **68** 391–6

Hernandez-Giron I, Geleijns J, Calzado A and Veldkamp W J 2011 Automated assessment of low contrast sensitivity for CT systems using a model observer *Med. Phys.* **38** S25

Hsieh J, Nett B, Yu Z, Sauer K, Thibault J-B and Bouman C A 2013 Recent advances in CT image reconstruction *Curr. Radiol. Rep.* **1** 39–51

Judy P F 1976 The line spread function and modulation transfer function of a computed tomographic scanner *Med. Phys.* **3** 233–6

Kalia V, Obray R W, Filice R, Fayad L M, Murphy K and Carrino J A 2009 Functional joint imaging using 256-MDCT: technical feasibility *Am. J. Roentgenol.* **192** 295–9

Kalra M K, Quick P, Singh S, Sandborg M and Persson A 2013 Whole spine CT for evaluation of scoliosis in children: feasibility of sub-milliSievert scanning protocol *Acta Radiol.* **54** 226–30

Leng S, Zhao K, Qu M, An K-N, Berger R and McCollough C H 2011 Dynamic CT technique for assessment of wrist joint instabilities *Med. Phys.* **38** S50

Maidment A D A and Albert M 2002 Conditioning data for calculation of the modulation transfer function *Med. Phys.* **30** 248–53

Miéville F, Beaumont S, Torfeh T, Gudinchet F and Verdun F R 2010 Computed tomography commissioning programmes: how to obtain a reliable MTF with an automatic approach? *Radiat. Prot. Dosim.* **139** 443–8

Miéville F A, Bolard G, Benkreira M, Ayestaran P, Gudinchet F, Bochud F O and Verdun F R 2011a 3D noise power spectrum applied on clinical MDCT scanners: effects of reconstruction algorithms and reconstruction filters *Proc. SPIE* **7961** 7961 0E

Miéville F A, Bolard G, Bulling S, Gudinchet F, Bochud F O and Verdun F R 2011b Effects of computing parameters and measurement locations on the estimation of 3D NPS in non-stationary MDCT images *Phys. Med.* **29** 684–94

Monnin P, Marshall N W, Bosmans H, Bochud F O and Verdun F R 2011 Image quality assessment in digital mammography: part II. NPWE as a validated alternative for contrast detail analysis *Phys. Med. Biol.* **56** 4221–38

Nakaya Y, Kawata Y, Niki N, Umetatni K, Ohmatsu H and Moriyama N 2012 A method for determining the modulation transfer function from thick microwire profiles measured with x-ray microcomputed tomography *Med. Phys.* **39** 4347–64

Nickoloff E L and Riley R 1985 A simplified approach for modulation transfer function determinations in computed tomography *Med. Phys.* **12** 437–42

Richard S, Husarik D B, Yadava G, Murphy S N and Samei E 2012 Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms *Med. Phys.* **39** 4115–22

Richard S, Li X, Yadava G and Samei E 2011 Predictive models for observer performance in CT: applications in protocol optimization *Proc. SPIE* **7961** 79610H

Samei E, Ranger N T, Dobbins J T III and Chen T 2006 Intercomparison of methods for image quality characterization. I. Modulation transfer function *Med. Phys.* **33** 1454–65

Schneiders N J and Bushong S C 1978 Single-step calculation of the MTF from the ERF *Med. Phys.* **5** 31–3

Schneiders N J and Bushong S C 1980 Computer assisted MTF determination in CT *Med. Phys.* **7** 76–8

Thibault J-B, Sauer K D, Bouman C A and Hsieh J 2007 A three-dimensional statistical approach to improved image quality for multislice helical CT *Med. Phys.* **34** 4526–44

Willemink M J, Schilham A M R, Leiner T, Mali W P Th M, De Jong P A and Budde R P J 2013 Iterative reconstruction does not substantially delay CT imaging in an emergency setting *Insights Imag.* **4** 391–7

Yin F F, Giger M L and Doi K 1990 Measurement of the presampling modulation transfer function of film digitizers using a curve fitting technique *Med. Phys.* **17** 962–6

International Commission on Radiation Units and Measurements 1986 *Modulation Transfer Function of Screen-Film Systems. Report* 41 (Bethesda, MD: The Commission)

International Commission on Radiation Units and Measurements 1996 Medical imaging–the assessment of image quality. Report 54 *Radiography* **3** 243–4

International Commission on Radiation Units and Measurements 2012 Radiation dose and image—quality assessment in computed tomography. Report 87 *J. ICRU* **12** 9–149

IEC 2009 *Medical Electrical Equipment. Part 2-44: Particular Requirements for the Basic Safety and Essential Performance of x-ray Equipment for Computed Tomography* (Geneva: International Electrotechnical Commission)

Assessment of low contrast detection in CT using model observers: Developing a clinically-relevant tool for characterizing adaptive statistical and model-based iterative reconstruction.

Julien G. Ott[1], Alexandre Ba[1], Damien Racine[1], Anais Viry[2], François O. Bochud[1] and Francis R. Verdun[1]

[1]Institute of Radiation Physics –Lausanne University Hospital, Rue du Grand Pré, 1 – 1007 Lausanne – Switzerland
[2]National Institute for Nuclear Science and Technology – CEA Saclay – 91191 Gif-Sur-Yvette – France

Abstract:

Purpose: This study aims to assess CT image quality in a way that would meet specific requirements of clinical practice. Classic physics metrics like Fourier metrics were traditionally employed for that. However, assessment methods through a detection task have also developed quite extensively lately, and we chose here to rely on this modality for image quality assessment. Our goal was to develop a tool adapted for a fast and reliable CT image quality assessment in order to pave the way for new CT benchmarking techniques in a clinical context. Additionally, we also used this method to estimate the benefits brought by some IR algorithms.

Material and Methods: A modified QRM chest phantom containing spheres of 5 and 8 mm at contrast levels of 10 and 20 HU at 120 kVp was used. Images of the phantom were acquired at $\mathrm{CTDI}_{\mathrm{vol}}$ of 0.8, 3.6, 8.2 and 14.5 mGy, before being reconstructed using FBP, ASIR 40 and MBIR. They were then assessed by eight human observers undergoing a 4-AFC test. After that, these data were compared with the results obtained from two different model observers (NPWE and CHO with DDoG channels). The study investigated impacts of the acquisition conditions as well as reconstruction methods.

Results: NPWE and CHO models both gave coherent results and were also in accordance with human observer. Moreover, the reconstruction technique used to retrieve the images had a clear impact on the PC values. Both models suggest that switching from FBP to ASIR 40 and particularly to MBIR produces an increase of the low contrast detection, provided a minimum level of exposure is reached.

Conclusion: Our work showed that both CHO with DDoG channels and NPWE models provided responses similar to humans when performing a detection task. Both models also suggest that the use of IR goes along with an increase of the PCs, indicating that further dose reduction is still possible when using those techniques. Eventually, the CHO model associated to the protocol we described in this study happened to be an efficient way to assess CT images in a clinical environment. In the future, this method could represent a sound basis to benchmark clinical practice and CT devices.

Keywords: Computed Tomography (CT), model observer, image quality, dose reduction, iterative reconstruction (IR)

1. Introduction

Since the beginning of the 1980’s, the radiation dose delivered to patients via diagnostic X-ray imaging has continuously increased and now contributes to 25% of the total (man-made plus natural) dose. Among that 25%, the amount for which Computed Tomography (CT) is responsible raises a particular concern. Indeed, this modality’s impact on the population’s collective dose increased so much (NCRP report 160, 2009) that it has now become the largest single source of X-rays, representing in Switzerland for example, 8% of examinations yet 68% of the collective dose (Samara *et al* 2012). Patients certainly benefit from this kind of exam, but many efforts still have to be made in order to ensure that the benefits-risk ratio remains acceptable. Thus, the ongoing challenge in radiology is to ensure diagnostic image quality while minimizing delivered dose. An efficient way to perform this optimisation scheme is to use an objective, task-related image quality assessment method.

Characterisation of image quality in CT has been done successfully by using Fourier metrics (Samei *et al* 2006 and Torfeh *et al* 2007). Among those metrics, noise power spectrum (NPS) or modulation transfer function (MTF) have been used extensively (details on their computation can be found in ICRU report 87). However, it is necessary to be careful when applying those metrics on non-stationary images or on images produced by nonlinear IR algorithms or by FBP when nonlinear filters are used (Vaishnav *et al* 2014). In order to overcome those limitations, several authors have successfully adapted the above-mentioned Fourier metrics. Mieville for example averaged several NPS calculated at different radial positions on the slices to deal with the non stationarity problem (Mieville et al 2010). Richard and Brunner (Richard et al 2012, Brunner and Kyprianou 2013) developed an alternative to the MTF called target transfer function that uses an object at different contrasts in order to take the non linearity into account when estimating the resolution. Concomitantly, quantitative measurements methods based on statistical decision theory and designed to assess diagnostic image quality in non linear and non stationary domains have emerged (ICRU report 54). Extensive details on those task-based methods can also be found in Barrett’s work (Barrett H H and Myers K J 2004). Several groups decided to use or develop those methods in order to investigate image quality in general radiography (Beutel, Kundel and Van Metter 2000, He and Park 2013 and Tseng et *al* 2014), mammography (Hill *et al* 2010), nuclear medicine (He X *et al* 2010, Brankov J G 2013 and Gifford H C *et al* 2013) and also for CT (Abbey *et al* 2002 and Wunderlich *et al* 2010). These methods are used to assess how well the desired information can be extracted from the image, using four elements: the task, the signal and the background properties, the observer and the figure of merit (FOM). In other words, the principle consists of measuring the performance of an observer conducting a task of clinical interest. The task may be the detection of a signal into a noisy background, the observer is the person or algorithm performing the task and the FOM measures how well the observer performed the task. Other studies characterising IR algorithms and relying on the radiologist's perception can already be found in the literature (Schindera *et al* 2013, Goenka *et al* 2014). They are however time-consuming. In our study, we used model observer to assess image quality of FBP and IR images. This approach has already been successfully used by several groups for CT modality (Chen et al 2014, Tseng et al 2014 and Samei et al 2015).

Our first goal in this work was to present a robust method to investigate the image quality yielded by CT devices used in clinical routine. These elements introduced several constraints for our tool and for the acquisition conditions to be tested, namely the implementation of clinical settings in the experiment. This approach is also part of the current guidelines of the HERCA committee (Heads of European Radiological protection Competent Authorities) whose first objective was to find a unified criterion for the benchmarking of image quality in CT (http://www.herca.org/WGs.asp?WGnr=3 visited on March 3rd 2015). The second objective of our work was to conduct investigations using this tool, in order to estimate the progress driven by some first and second generation of IR (Tseng *et al* 2014). In our work, performances of GE (GE Healthcare, USA) filtered back projection (FBP) algorithm were compared to those of adaptive statistical iterative reconstruction (ASIR), and model based iterative reconstruction (MBIR). Thus, the framework of this experiment would enable us to determine up to what point does the dose reduction remain acceptable in terms of low contrast detectability.

For this, a chest phantom containing spheres of different sizes and contrast levels was scanned at several doses and reconstructed using three algorithms: FBP, ASIR and MBIR. Next, a group of human observers assessed the images obtained. A four-alternative forced-choice (4-AFC) experiment together with a Kappa statistical analysis were performed in order to define the low contrast resolution level of the images. After that, two model observers, the Non-Prewhitening with Eye-filter (NPWE) model (Burgess *et al* 1994) and the Channelized Hotelling Observers (CHO) model (Barrett *et al* 1993), were applied to the images. We used these models on 4-alternative-forced-choice (4-AFC) tests and computed a percentage of correct responses (PC) in order to quantify their performance. In the end, the results from the human observers and the model observers were compared in order to predict the detection of low contrast lesions (Barrett *et al* 1993, Leng *et al* 2013 and Yu *et al* 2013).

2. Materials and Methods

2.1 Phantom and Data acquisition

We used a commercially available chest phantom (QRM, Moehrendorf, Germany) in the size of an adult's chest (30 cm wide and 20 thick) that mimicked the attenuation produced by a patient (CF Figure 1).

It included a removable module in its middle. For this study, two different modules were used. The first one was a uniform module, whereas the other contained 48 low contrast spherical targets. Targets of 3, 4, 5, 6 and 8 mm in diameter at two contrast levels (10 and 20 HU at 120 kVp) could be found inside the low contrast module. Using spheres implied a careful selection of the images in order to ensure that the central part of the sphere was used and processed by the model observer. Despite this difficulty, we think that a sphere is closer to a clinical lesion and allow us to take the partial volume effect into account.

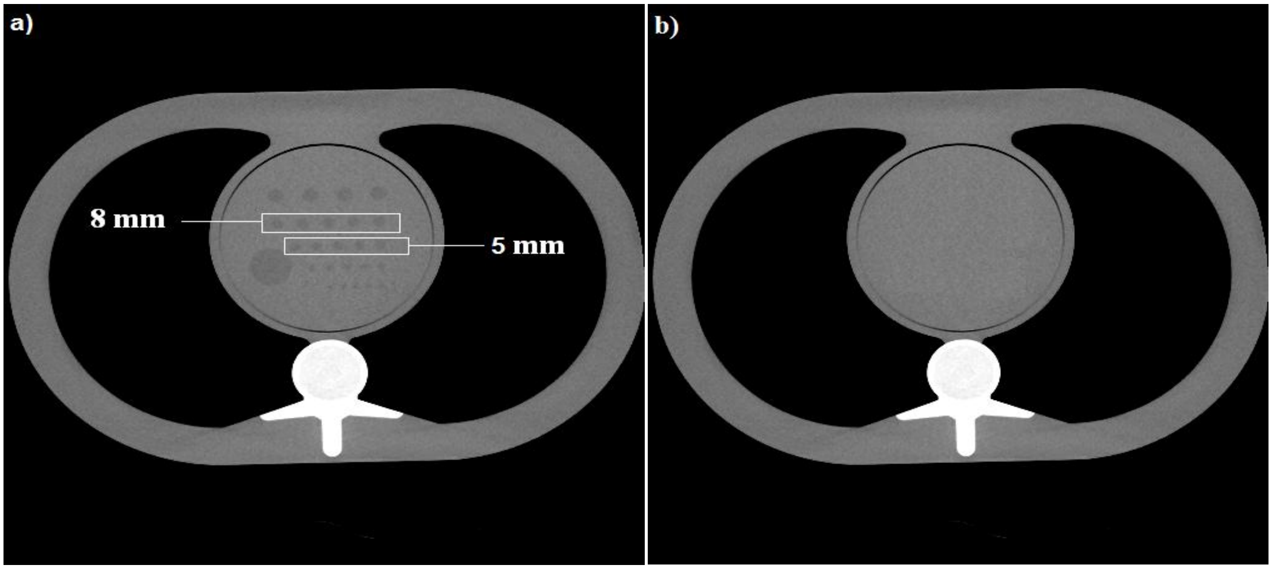


Figure 1: The QRM phantom with the module containing the signals (a) and the uniform module (b)

Data acquisition was performed at the Lausanne University Hospital (CHUV) on a Discovery HD 750 CT (GE Healthcare, USA). We performed all acquisitions at a tube tension of 120 kVp, using a 500 mm scan field of view (SFOV), a 300 mm display field-of-view (DFOV), a 512×512 matrix size, a small focal spot and a reconstructed slice thickness of 2.5 mm associated to an interval of 1.25 mm (As Veo only support 0.625 mm reconstruction thickness, a post reconstruction reformatting was performed on the MBIR data sets to obtain the values previously mentioned). We used the helical mode with a pitch of 0.9 and a 40 mm X-ray beam collimation. $CTDI_{vol}$ values were calculated as described in the IEC 60601-2-44. Reconstructions were performed using the classic FBP algorithm, as well as ASIR at a percentage of 40% and a commercial implementation of the MBIR (Veo, GE Healthcare), all together with the smooth "standard" kernel. In the end, a total of 48 different categories were obtained: four dose levels, three reconstruction techniques, two target sizes and two contrasts (in this work we only investigated the 5 mm and 8 mm targets at 10 and 20 HU). From these sets of acquisitions, regions of interest (ROIs) of 41x41 pixels containing the centred targets were extracted. For each dose level (0.8, 3.6, 8.2 and 14.5 mGy expressed in a $\mathrm{CTDI_{vol}}$ phantom of 32 cm in diameter), the phantom was scanned 20 times. Thus, we obtained 100 images containing a signal (20 scans x 5 targets with identical size and contrast in the phantom (cf. figure 1)) and 1000 images with only noise for each category. The noise-only ROIs were extracted from the uniform module, whereas the signals came from the low contrast module. This way, the same position on the (x, y) plane was ensured for both signal and noise ROIs, thus solving the noise stationarity problem. Eventually, this allowed us to estimate image quality and also to make a comparison between the performances of different reconstruction algorithms as already suggested in the literature (Kaza *et al* 2014).

2.2 Tasks and human observers

In our study, eight human observers -namely eight medical physicists- with five years of experience regarding image quality assessment using M-AFC tests were blinded to the CT acquisitions and reconstruction conditions submitted to a 4-AFC experiment. After having undergone a training session, they had to select the signal-present image in a batch of three signal-absent images and one signal-present image, which were presented together in a randomized order. In each image of the 4-AFC test, the signal was centred in the “location known exactly” paradigm. We displayed the images on a DICOM calibrated screen (EIZO RadiForce MX210, Eizo, Hakusan, Japan) using cues to indicate the centre of the images in order to minimize location uncertainty (Figure 2). The tests were performed with an ambient light level of about 50 lux in a radiology reading room. Observers were asked to maintain a fixed distance of 50 cm between the monitor and their eyes and make a decision for all 48 different categories we acquired. Thus, for each category, signal-present images as well as signal-absent images were chosen at random among our 100 signal-present images and 1000 signal-absent images. The correct answer for each trial was stored in a database in order to be compared to the observers’ answer and thus yield a PC for each category.

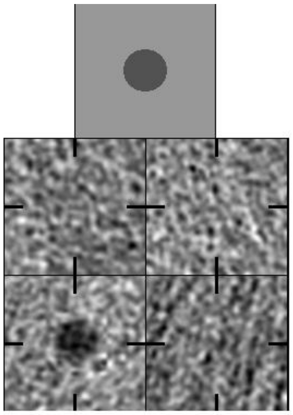


Figure 2: Example of a 4-AFC test

In order to characterize the repeatability, each observer took the test three times. Then, a Kappa statistical analysis was used to compare two tests from the same observer and to rule on their correlation. This statistical test relies on the number of times the observer agreed on having seen a signal twice (denoted by a1), disagreed on having seen a signal twice (denoted by b1) and when the observer once agreed and then disagreed on having seen a signal and inversely (respectively denoted by a2 and b2) (Ker 1991).

The observed agreement (OA) and the chance agreement (CA) were calculated as follow:

$$OA = \frac{\mathrm{a1} + \mathrm{b1}}{\mathrm{a1} + \mathrm{b1} + \mathrm{a2} + \mathrm{b2}}$$

$$\mathrm{CA} = \frac{\mathrm{AF} + \mathrm{DF}}{\mathrm{a1} + \mathrm{a2} + \mathrm{b1} + \mathrm{b2}}$$

Where AF is the agreement fraction and DF the disagreement fraction.

$$AF = \frac{(a1 + a2)(a1 + b2)}{a1 + a2 + b1 + b2}$$

$$DF = \frac{(b1 + b2)(b1 + a2)}{a1 + a2 + b1 + b2}$$

Finally, the kappa factors κ were calculated using the formula:

$$\kappa = \frac{OA - CA}{1 - CA} \qquad (1)$$

2.3 Model observers

Two model observers were tested: the NPWE model and the CHO model, both of which make it possible to predict the detection of low contrast signals.

For each category, extracted ROIs of 41x41 pixels were used to calculate a scalar response. The application of a linear model observer enabled the computation of this decision variable denoted by λ_i, which is given by the dot product between a template of the model $\boldsymbol{w}$ and the image of the analysed ROI $\boldsymbol{g_i}$ (i = 0 or i=1 respectively represents signal-absent or signal-present hypothesis):

$$\lambda_i = \boldsymbol{w}^T . \boldsymbol{g_i} \qquad (2)$$

The NPWE, which is an anthropomorphic model observer, is also one of the simplest models. In order to take limitations of the visual properties of a human observer into account, the original template is filtered in the spatial frequencies domain by the eye contrast sensitivity function, which represents the sensitivity of the human eye at different contrast level and spatial frequencies (Barten 1992, Beutel, Kundel and Van Metter 2000 and Monnin *et al* 2010). Its functional form in the frequency domain at a viewing distance of 50 cm can be modelled by the Eq. (3) (Castella *et al* 2007):

$$\boldsymbol{E}(\rho) = \rho^{n} e^{-c\rho^{2}} \qquad (3)$$

With ρ the frequency in cycles/deg, n=1.3 and c=0.041 (values from Burgess 1994).

Then, the calculation of the template is achieved using the following equation:

$$\boldsymbol{w}_{NPWE} = \boldsymbol{E}^{T}[\overline{\boldsymbol{g_s}} - \overline{\boldsymbol{g_n}}]\boldsymbol{E} \qquad (4)$$

Where $\overline{\boldsymbol{g_s}}$ and $\overline{\boldsymbol{g_n}}$ are vectors corresponding to the data of the means of the signal-present and noise-only images, and $\boldsymbol{E}$ is a matrix containing the eye filter, which is based on Eq. (3).

The template $\boldsymbol{w}_{NPWE}$ can then be used in (2), in order to estimate the decision variable λ.

The CHO also belongs to the family of anthropomorphic models, and includes pre-processing of the image by a set of channels that are tuned to respond to specific spatial frequencies (Myers and Barrett 1987 and Gallas and Barrett 2003). The template $\boldsymbol{w}_{CHO}$ of this model observer is obtained by performing the following calculations (Extensive details can be found in Beutel, Kundel and Van Metter 2000, Castella *et al* 2007, Yu *et al* 2013 and Tseng *et al* 2014):

$$\boldsymbol{w}_{CHO} = \left[\frac{1}{2}(\boldsymbol{K_{cs}} + \boldsymbol{K_{cn}})\right]^{-1} (\overline{\boldsymbol{g_{cs}}} - \overline{\boldsymbol{g_{cn}}}) \qquad (5)$$

Where:
$$\boldsymbol{K_{cn}} = \boldsymbol{U}^{T}\boldsymbol{K_n}\boldsymbol{U} \;\; and \;\; \boldsymbol{K_{cs}} = \boldsymbol{U}^{T}\boldsymbol{K_s}\boldsymbol{U} \qquad (6)$$

$\boldsymbol{K_n}$ and $\boldsymbol{K_s}$ are the covariance matrix calculated respectively from the signal-absent and signal-present data and $\boldsymbol{U}$ is the matrix representation of the channel filters, which are described more extensively below.

In (5), $\overline{\boldsymbol{g_{cs}}}$ and $\overline{\boldsymbol{g_{cn}}}$ are the means of the channel outputs under a signal-present and signal-absent hypothesis and can be estimated according Eq. (7):

$$\overline{\boldsymbol{g_{cs}}} = \boldsymbol{U}^{T}\overline{\boldsymbol{g_s}} \;\; and \;\; \overline{\boldsymbol{g_{cn}}} = \boldsymbol{U}^{T}\overline{\boldsymbol{g_n}} \qquad (7)$$

The set of channels used for this study includes ten channels and is represented in Figure 3. This model was described by Abbey and Barrett in *2001* (Abbey and Barrett 2001) and is called dense difference of Gaussian (DDoG) channels. The radial frequency profile of the jth channel is given by:

$$\boldsymbol{U}_j(\rho) = exp\left[-\frac{1}{2}\left(\frac{\rho}{W\sigma_j}\right)^2\right] - exp\left[-\frac{1}{2}\left(\frac{\rho}{\sigma_j}\right)^2\right] \qquad (8)$$

Where ρ is the spatial-frequency, W=1.67 defines the bandwidth of the channel and σ_j is the standard deviation of the jth channel. Each σ_j value is defined starting from $\sigma_0 = 0.005$ by $\sigma_j = \sigma_0\alpha^j$ with $\alpha = 1.4$ (Values from Abbey and Barrett 2001).

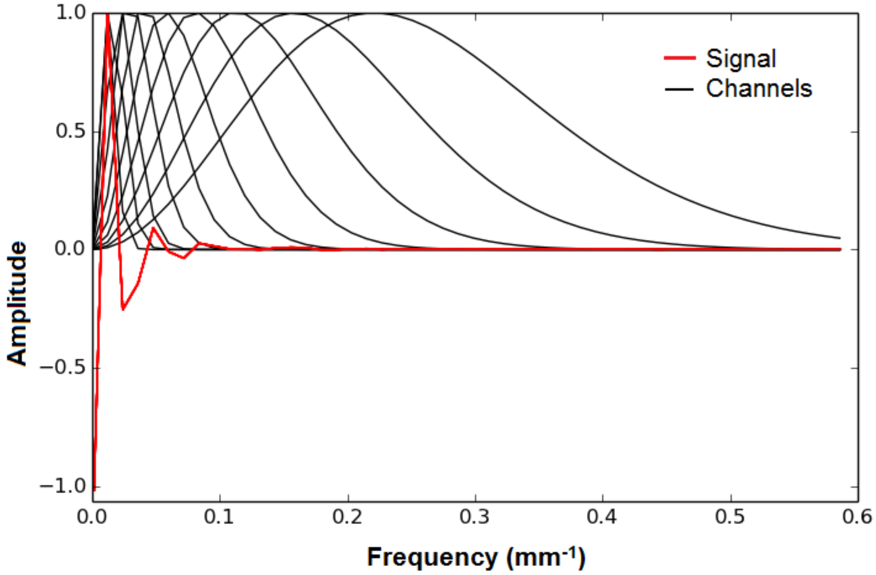


Figure 3: Profile of the signal and DDoG channels

Eventually, the decision variable can be calculated by injecting (5) into (2). Note, however, that care must be taken because in this case the decision variable is calculated using the channel output of the ROI (denoted by $\boldsymbol{g}_{ci}$) instead of simply the analysed ROI $\boldsymbol{g}_{\boldsymbol{i}}$. Eventually we obtain:

$$\lambda_i = \boldsymbol{w}_{CHO}^T . \boldsymbol{g}_{ci} \qquad (9)$$

2.4 Figure of merit

In order to obtain a PC from the model observer, we performed 4-AFC tests on the acquired images with our two model observers. The value of the decision variable λ was used to determine which of the four images of the test was recognised as the signal-present by the model observers (the highest value of λ is supposed to be the signal-present image). Then, the correct answers which were stored in a database were compared to those results, enabling the computation of PCs for the model observers.

We chose the PC rather than the detectability as FOM in this study, because no assumptions on the statistical properties of the model responses are necessary. In other studies, detectability is frequently computed in order to compare human and model results. In an earlier work, Bochud described in detail how to proceed (Bochud *et al* 2000). The method consists of converting the PC values obtained by humans into detectability values. However, this requires the distribution of the λ values to be Gaussian in order to function. If this requirement is satisfied when dealing with large amount of images, it may not be the case when the number of images is limited.

In our case, we dealt with 100 signal-present images per category. This number was chosen because it facilitated a good compromise between the amount of images acquired and the time the device is used (which has to be reasonable when using a clinical device). However, even if 100 images remains similar to what was done in previous work (Yi *et al* 2014), this amount may still be seen as insufficient to ensure the Gaussian distribution of λ values. As a consequence, we chose to rely on PC (which requires no particular distribution of the responses) rather than detectability for the calculation of our FOM.

2.5 Uncertainty estimation and adequacy of the models with humans

The uncertainties on the PC values produced by the models were estimated by performing bootstrap (Efron and Tbishirani 1993). In our experiment, the 4-AFC tests were made using 100 trials for each category. Thus, in order to estimate the error on the PCs for one particular category, we performed 100 draws with replacement on the signal-present images of this chosen category, as well as 100 draws with replacement of three signal-absent images. In the end, a PC value was calculated based on the results of this 4-AFC test. This process was repeated 150 times, leading to 150 values of PC. Computing the mean and standard deviation of those 150 values enabled us to determine the final PC value as well as its standard deviation for each acquisition condition. We used this to estimate a 95% confidence interval.

The PCs and uncertainties of the human observers were calculated using the different results of the eight observers who scored the images. For each category, the mean PC value of the eight observers and its standard error were calculated. The error bars correspond to a 95% confidence interval.

Finally, the adequacy of the models with humans was estimated using a Pearson correlation coefficient:

$$\rho_{XY} = \frac{COV(X,Y)}{\sigma_X \sigma_Y} \qquad (10)$$

Where $COV(X,Y)$ is the covariance between two vector data sets X and Y, and σ_X and σ_Y are their standard deviations.

3. Results

3.1 Human observer results

Table 1 presents the Kappa test results of one human observer for all investigated conditions as well as for each individual dose (mixing twelve parameters: two sizes, two contrast levels and three reconstruction algorithms).

Table 1: Agreement between three repetitions of the same experiment for one observer using a Kappa analysis

	1st & 2nd test	1st & 3rd test
0.8 mGy	0.27	0.24
3.6 mGy	0.35	0.29
8.2 mGy	0.53	0.50
14.5 mGy	0.62	0.60
Entire test	0.61	0.58

On the Kappa scale, a value between 0.2 and 0.4 indicates a weak agreement. A value between 0.4 and 0.6 indicates a moderate agreement, and a value between 0.6 and 0.8 indicates a strong agreement.

3.2 Impact of IR reconstruction on human and model observer performance

In this section, the qualities of the images obtained from FBP, ASIR 40 and MBIR algorithms are compared using the performances of model observers (NPWE in Figure 4 and CHO in Figure 5) as well as human observers (in Figure 6). Note that the results for the 8 mm 20 HU signals are not presented because they were close to 100% for all observers and doses.

Taking the uncertainties into account, we see that both observers exhibited the same tendency with the lowest performance for the FBP algorithm, the best for MBIR and in-between for ASIR 40. The performance of the NPWE observer did not differ between FBP and ASIR 40 with the smallest signal, the lowest contrast, and the lowest dose (0.8 mGy). In this case, given the acquisition conditions, the detection task was too complicated and remained barely better than random (PC equal to 35%) with both FBP and ASIR 40.

Results of the CHO model with the DDoG channels also suggest an improvement of the PCs when switching from FBP to ASIR 40, but it is much less pronounced than with the NPWE model observer since an increase of only a few percent was observed on the PCs depending on the signal and dose range investigated. Oppositely, switching from FBP to MBIR did provide a more visible improvement on the

PCs. This is particularly true for the 5 mm 10 HU signal under all CTDIs and also for the 8 mm 10 HU and 5 mm 20 HU signals at 0.8 and 3.6 mGy. For the remaining conditions, the detection task yielded high and identical PC values for all algorithms.

Human observer results suggest that switching from FBP to ASIR 40 leads to an improvement in the PC for every condition except those were the signal was already highly visible (that is to say in this case 5 mm 20 HU at 8.2 mGy or more and 8 mm 10 HU at 14.5 mGy). In those cases, the PC remained equivalent regardless whether FBP or ASIR 40 was used. When using the MBIR algorithm, strong PC enhancements were generally observed for the 10 HU signals. Nevertheless, there were also several cases where MBIR was outperformed by ASIR 40, especially at the lowest dose level for the 5 mm signals. The 5 mm 10 HU signal at 0.8 mGy with ASIR 40 yielded a PC of 42% whereas only 34% were reached with MBIR. Similarly, the 5 mm 20 HU signal yielded 64% correct responses with ASIR 40 and 47% with MBIR. These results were surprising since a first generation algorithm seemed to outperform a second generation in this particular case.

We observed that there was no improvement in the detection when switching from FBP to ASIR 40 or MBIR when the signal is acquired under conditions making it difficult to detect. This is because the signal was too noisy to be properly detected whatever the type of reconstruction we used. However, once a sufficient dose amount was reached (often around 3 mGy according our measurements), the use of IR is a real help to detect the signals, and it brings a significant improvement of PCs compared to what was observed when using FBP. This behaviour was observed for both human and model observers. We also witnessed that the PC increases with the given dose until it reaches an asymptote and that this asymptote is reached faster when using IR. All those results are true for both IR types, but a stronger increase of the PCs was almost always observed with MBIR than with ASIR 40.

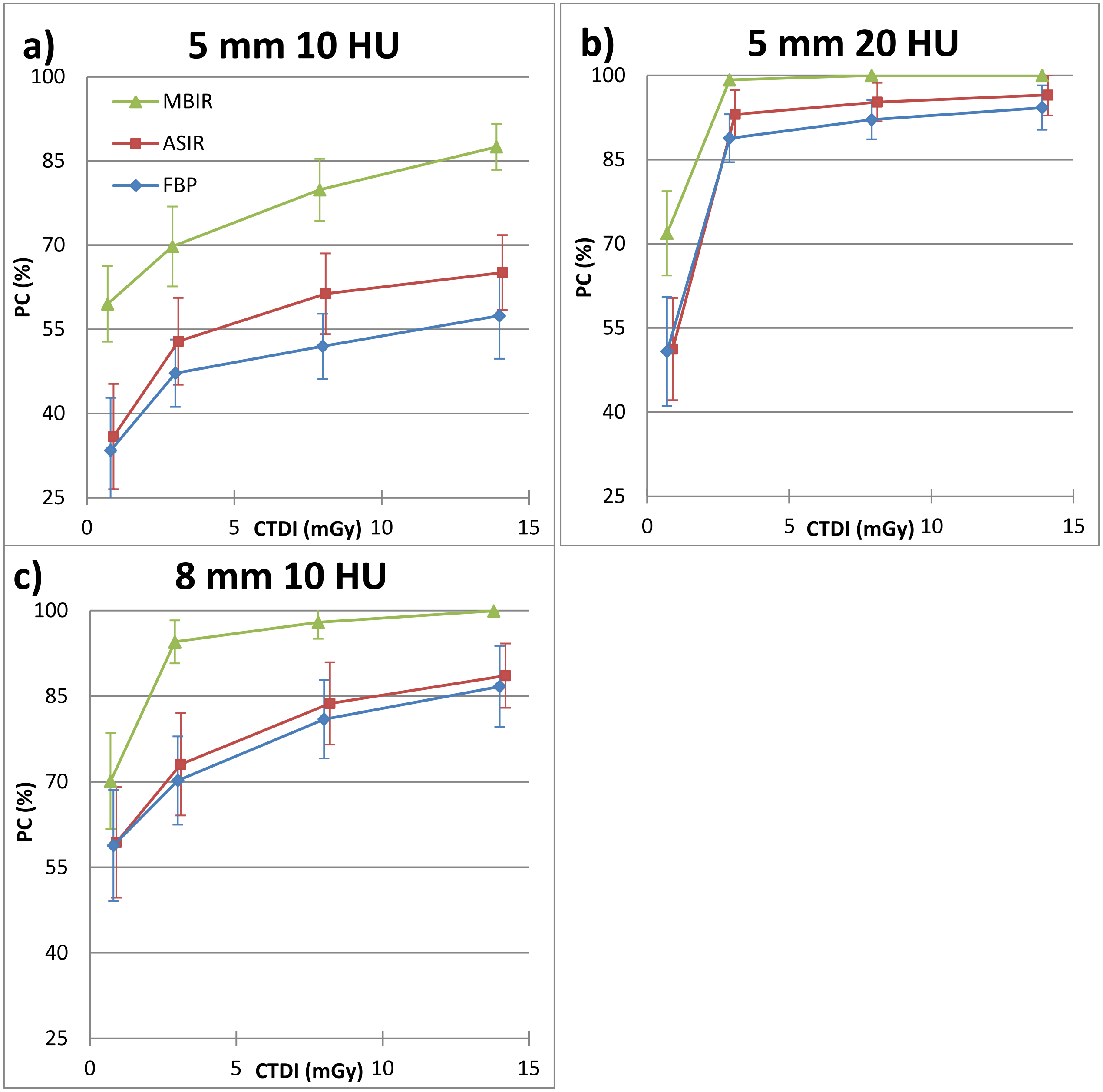


Figure 4: PC results obtained by performing 4-AFC tests on the NPWE model with the three reconstruction techniques

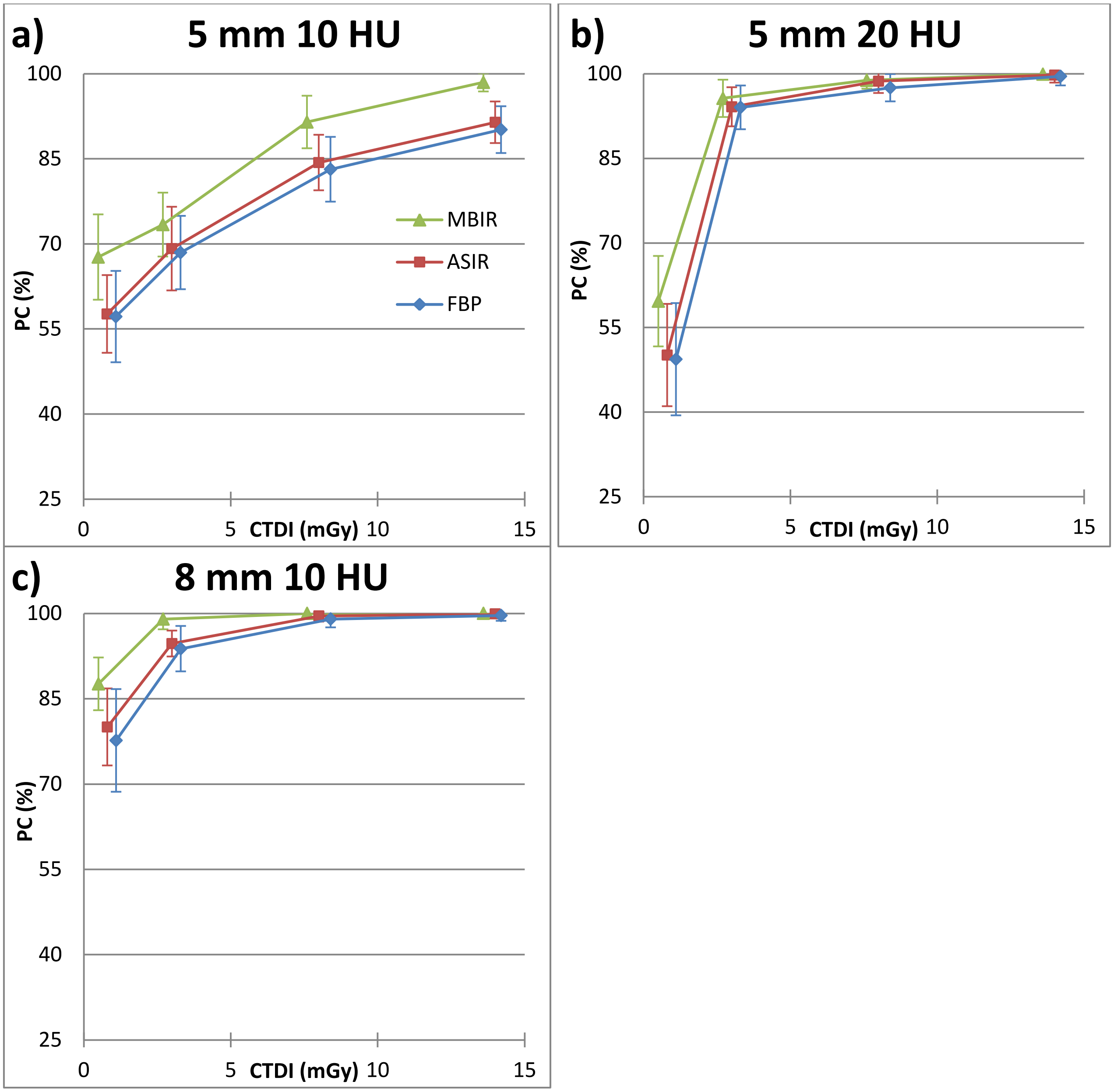


Figure 5: PC results obtained by performing 4-AFC tests on the CHO model with the three reconstruction techniques

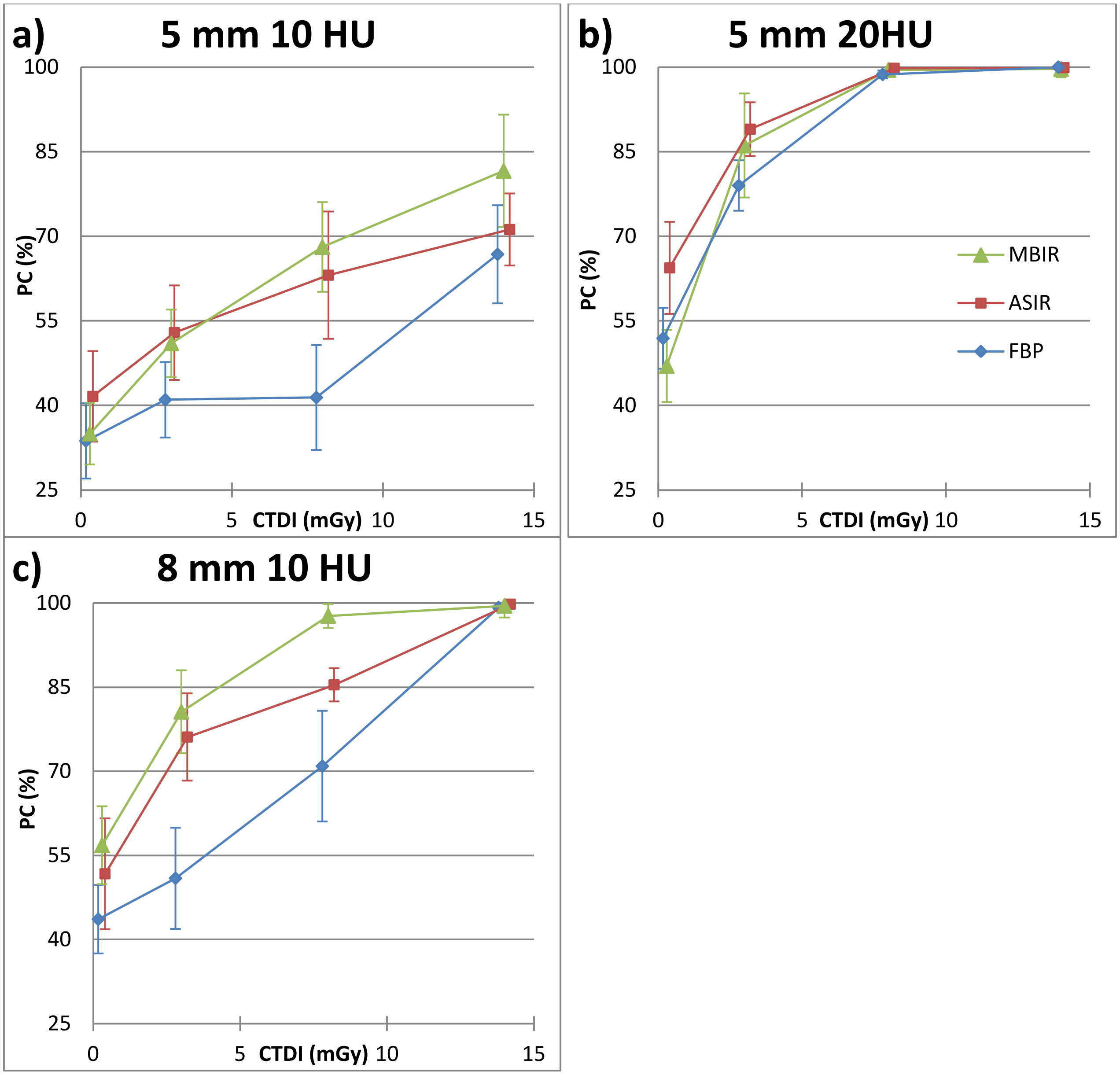


Figure 6: PC results obtained by performing 4-AFC tests on the human observers with the three reconstruction techniques

3.3 Correlation between human and model observer

The CHO and NPWE models were compared to human observer responses in order to investigate if they were reliable for predicting low contrast detection. Both models and humans reported an increase of target visibility with the delivered dose up to a plateau, as well as an improvement with the size and the contrasts of the targets. The NPWE induced a bigger increase in the PC than the CHO when switching from

FBP to IR, but they both displayed a similar trend when looking at the behaviour of the PC. These results express the coherence of the models we tested.

The Pearson correlation coefficient was calculated for all reconstruction algorithms and all signals in order to estimate the correlation between humans and each model observer. Results are presented in Table 2. They show an excellent agreement between humans and the two models, since they are all far higher than 0.9 except the 5 mm 10 HU reconstructed with FBP. Moreover, note that the correlation between humans and models increases when switching from FBP to first and to second generation of IR.

Table 2: Pearson coefficient between humans and the two model observers for all signals and reconstruction algorithms

NPWE-Humans	FBP	ASIR 40%	MBIR
5 mm 10 HU	0.845	0.985	0.992
5 mm 20 HU	0.951	0.994	0.985
8 mm 10 HU	0.904	0.991	0.977

CHO-Humans	FBP	ASIR 40%	MBIR
5 mm 10 HU	0.792	0.951	0.995
5 mm 20 HU	0.942	0.996	0.992
8 mm 10 HU	0.782	0.918	0.936

4. Discussion

Traditionally, frequency metrics like MTF or NPS are used to determine if the devices are in compliance with the acceptable tolerances. This approach based on Fourier metrics for image quality assessment has been done successfully in previous studies (Mieville *et al* 2012 and Friedman *et al* 2013). Yet, other groups have used other approaches based on task based assessment (Ott *et al* 2014, Solomon *et al* 2015). Moreover, as radiation protection has become a major concern, there has been a parallel tendency to lower doses. This results in a loss of the low contrast resolution in the clinical images. This information loss could affect the radiologist's diagnosis and means that it is essential now to find ways to characterise these new low dose images. Thus, our first goal was to provide a reliable and easy-to-implement method to evaluate the whole image chain with one figure of merit. Choices had to be made in the implementation so that this technique would be adapted to a clinical practice.

The kappa test performed on the observers indicated that our experiment was reliable and that the signal visibility was clearly linked to the amount of dose given. Indeed, its results showed a good agreement between the three intra observer tests, meaning that our experiment was reproducible (Table 1). Analysing the four categories, the agreement between the observers was better for the higher dose level (the quasi-totality of the targets is detected by each observer). When the dose decreased, the agreement between the tests became a random occurrence.

When analysing the outcomes of the models, we found out that the NPWE model gave good results since it reproduced the behaviour of humans under a wide range of conditions and signal types. This is particularly true for low contrast and small size signals. Analysis of the Pearson coefficients also suggests that the model fits a human's response for every algorithm and all signal types. These results are in accordance with previous studies on low contrast detection in CT with the NPWE model (Hernandez-Giron *et al* 2011).

Similarly, the results showed that the CHO model with DDoG channels works well for a large dose range (from 0.8 to 14.5 mGy) and for all reconstruction algorithms. It seems to be very efficient in low contrast detection even if it sometimes tends to overestimate human results in those particular conditions. This trend is also revealed when looking at the Pearson coefficients which indicate a drop for 10 HU signal reconstructed with FBP. Otherwise, very good accordance with humans was observed. These results are coherent with what can be found in the literature since recent studies from Leng and Yu (Leng *et al* 2012 and Yu *et al* 2013) showed the efficiency of the CHO model for the low contrast detection.

The second goal of our investigation was to determine if the use of MBIR and ASIR instead of FBP could lead to an improvement of the task-based detection performance and therefore also to a further dose reduction.

Our results indicate that using ASIR at 40% systematically led to PC improvement for NPWE and CHO models. Humans witnessed the same trend for every parameter.

Considering our results, we conclude that the second generation of IR offers a superior image quality at an equivalent dose level compared to FBP. Indeed, both models and humans reported a slightly to significant (depending on the parameters we were investigating) improvement in the low contrast detection when switching from FBP to MBIR. Taking FBP as a reference point, these results also suggest that PC improvement is better when using MBIR compared to ASIR 40. This observation is always verified for model observers and remains true for a lot of conditions tested by humans. Some exceptions can be found, for example when considering the 5 mm targets at 0.8 mGy. In these cases the humans' PC is superior with ASIR 40 than with MBIR. This is because at such low doses and low size signals the image contains a lot of noise, making the detection by the observer more or less arbitrary. This observation is confirmed by the low values of the PCs as well as the consequent size of the error bars. In the end, the use of MBIR makes it possible to diminish the dose (even more than with ASIR at 40%) without losing the information which was in the FBP image.

However, no significant image quality improvement could be obtained for either the first or second generation of IR algorithms at very low dose (0.8 mGy in our work). This could be due to the high noise level encountered at those $\mathrm{CTDI_{vol}}$ levels, making the signal very hard to see no matter which algorithm we used. We also noticed that the PCs for the higher $\mathrm{CTDI_{vol}}$ (14.5 mGy) tends to be close to perfect with all three algorithms, leading to no benefits when using IR instead of FBP in this dose range. In the end, these elements explain that the benefits involved when using IR are greater for the two intermediate doses we investigated (that is to say 3.6 and 8.2 mGy).

The first limitation of our study concerns the number of image acquisitions we made. Indeed, a larger number of images would allow us to separate the data into two sets. The first would be used for the computation of the covariance matrix (and therefore the template of the CHO model) whereas the second would be the one we use to compute the detectabilities. In our case we did not dispose of enough images to make two different sets, forcing us to compute both the template and the detectabilities with the same images. It is however interesting to note that the number of images acquired is not sufficient to separate the data in the above-mentioned two sets, but this allows us to work under a reasonable acquisition time, which is adapted to a clinical environment.

The second limitation of our study is that we had to focus on a very simple detection task (when the signal, location and background are exactly known), a situation that is far from true clinical conditions. A study done in more complex conditions, like with an anatomical background and a location unknown paradigm would be more realistic, but this was outside the scope of this work.

The last limitation of this work is that no internal noise was added to the model observers we tested. Indeed, the models sometimes overestimate humans' results, requiring the use of this feature in order to improve the coherence with humans. Elaborated work has been done on both NPWE and CHO models in order to include internal noise (Zhang *et al* 2007, Richard and Siewerdsen 2008 and Brankov 2013) and therefore managed to obtain a response even more similar to humans. However, we voluntarily chose not to include this feature. Indeed, our main goal was to create a reliable tool dedicated for use in clinical conditions which would remain simple and the easiest possible to apply. It is also possible to change the set of channel used with the CHO model in order to produce responses more close to humans (Park *et al* 2007), but we only tested one set of channels because they were well adapted to our selected detection-task, and trying to test other sets would have been outside the scope of this work.

5. Conclusion

The assessment of image quality can be done according to different methods. Among them, Fourier metrics have been widely used and been adapted in order to be compatible with images produced by IR algorithms. In this study, we used a task-based approach with model observers in order to build a tool usable in clinical routine. The model observers matched human detection results' in our experiment for both FBP and IR algorithms, confirming that this approach is certainly well suited to perform an efficient assessment of CT image quality.

The method we have described here should be an efficient way to benchmark clinical CT devices and therefore improve our optimisation procedure between image quality and patient exposure. Eventually, both model observers we tested produced excellent results in terms of mimicking human behaviour for low contrast detection.

Results also showed that using IR, especially MBIR, clearly had benefits compared to FBP in terms of image quality in the 3 to 10 mGy $CTDI_{vol}$ range. Therefore, using MBIR at the same time as keeping the image quality previously obtained with FBP would enable a significant dose reduction. Our study also indicated that the use of ASIR at a percentage of 40% would enable a dose reduction too, but not to the same extent as MBIR.

Finally, this work made it possible for us to create the basis for a tool we can then use to perform an automatic benchmark of CT devices and clinical protocols. It has also demonstrated that the tool is adapted to image quality assessment with IR.

Acknowledgements

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

The author would also like to thank Christel Elandoy for her active participation in acquisitions, as well as Georg Kropat for his statistics advice and all the readers for their help in detection studies.

References

Abbey C K and Barrett H H 2001 Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability *J. Opt. Soc. Am.* **18(3):**473-88

Abbey C K, Eckstein M P, Shimozaki S S, Baydush A H, Catarious D M and Floyd C E 2002 Human-observer templates for detection of a simulated lesion in mammographic images *Proc. SPIE* **4686**

Barrett H H, Yao J, Rolland J P and Myers K J 1993 Model observer for assessment of image quality *Proc. Natl. Acad. Sci.* **90** 9758-65

Barrett H H and Myers K J 2004 Foundations of Image Science (Hoboken, NJ: Wiley)

Barten P G 1992 Physical model for the contrast sensitivity of the human eye *Proc. SPIE* **1666** *Human Vision Visual Processing, and Digital Display III*

Beutel J, Kundel H L and Van Metter R L 2000 Handbook of medical imaging Vol 1. Physics and psychophysics *SPIE press*

Bochud F O Abbey C K Eckstein M P 2000 Visual signal detection in structured backgrounds. III. Calculation of figures of merit for model observers in statistically non stationary backgrounds *JOSA A* **17(2)**:193-205

Boone J M 2001 Determination of the presampled MTF in computed tomography *Med. Phys.* **28** 356-60

Brankov J G 2013 Evaluation of the channelized Hotelling observer with an internal-noise model in a train-test paradigm for cardiac SPECT defect detection *Phys. Med. Biol.* **58(20)**:7159-82

Brunner C C and Kyprianou I S 2013 Material-specific transfer function model and SNR in CT *Phys. Med. Biol.* **58** 7447-61

Burgess A E 1994 Statistically defined backgrounds: performance of a modified non prewhitening observer model *JOSA A* **11(4):**1237-42

Burgess A E 1999 The Rose model, revisited *J. Opt. Soc. Am.* **16** 633-646

Castella C, Abbey C K, Eckstein M P, Verdun F R, Kinkel K and Bochud F O 2007 Human linear template with mammographic backgrounds estimated with a genetic algorithm *J. Opt. Soc. Am.* **24(12)**:B1-12

Chen B, Ramirez Giraldo J C, Solomon J and Samei E 2014 Evaluating iterative reconstruction performance in computed tomography *Med. Phys.* **41(12)**:121913

Efron B and Tbishirani R J 1994 An introduction to the bootstrap: Monograph on statistics and applied probability *CRC Press*

Friedman S N, Fung G K, Siewerdsen J H, and Tsui B W 2013 A standardized approach to measure computed tomography (CT) modulation transfer function (MTF) and noise power spectrum (NPS) using the American College of Radiology (ACR) accreditation phantom *Med. Phys.* **40**(5): 051907-1-9

Gallas B D and Barrett H H 2003 Validating the use of channels to estimate the linear model observer *J. Opt. Soc. Am.* **20(9)**:*1725*-38

Hernandez-Giron I, Geleijns J, Calzado A and Veldkamp W J 2011 Automated assessment of low contrast sensitivity for CT systems using a model observer *Med. Phys.* **38** (7)

Hill M L, Mainprize J G, YaffeAn M J 2010 Observer Model for Lesion Detectability in Contrast-Enhanced Digital Mammography Lecture Notes in Computer Science **6136**: 720-7

Gifford H C, King M A, de Vries D J and Soares E J 2000 Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT imaging *J. Nucl. Med.* **41(3)**:514-21

Goenka A H, Herts B R, Obuchowski N A, Primak A N, Dong F, Karim W and Baker M E 2014 Effect of reduced radiation exposure and IR on detection of low contrast low attenuation lesions in an anthropomorphic liver phantom *Radiology* **272(1)** 154-63

He X, Links JM, Frey E C 2010 An investigation of the trade-off between the count level and image quality in myocardial perfusion SPECT using simulated images: the effects of static noise and object variability on defect detectability *Phys. Med. Biol.* **55** 4949-61

He X and Park S 2013 Model observer in medical imaging research *Theranostics* **3(10)** 774–86

Hernandez-Giron I, Geleijns J, Calzado A, Veldkamp W J 2011 Automated assessment of low contrast sensitivity for CT systems using a model observer *Med. Phys.* **38** Suppl 1:S25

IEC 2009 '60601-2-44:2009 – Part 2-44: Particular requirements for the basic safety and essential performance of X-ray equipment for computed tomography

International Commission on Radiation Units and Measurements (2008) Radiation dose and image-quality assessment in computed tomography. Report 79

Kaza R K, Platt J F, Goodsitt M M, Al-Hawary M M, Maturen K E, Wasnik A P and Pandya A 2014 Emerging Techniques for Dose Optimization in Abdominal CT *RadioGraphics* **34**

Ker M 1991 Issues in the use of kappa *Invest. Radiol.* **26(1)**:78-83

Leng S, Yu L, Chen L, Ramirez Giraldo J C and McCollough C H 2013 Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain *Med. Phys.* **40(8)**:081908

Miéville F A, Ayestaran P, Argaud C, Rizzo E, Ou P, Brunelle F, Gudinchet F, Bochud F and Verdun F R 2010 Potential benefit of the CT adaptive statistical iterative reconstruction method for pediatric cardiac diagnosis *Proc. SPIE* 7622, Medical Imaging 2010: Physics of Medical Imaging

Miéville F A, Gudinchet F, Brunelle F, Bochud F O and Verdun F R 2013 Iterative reconstruction methods in two different MDCT scanners: Physical metrics and 4-alternative forced-choice detectability experiments– A phantom approach *Phys. Med.* **29** 99–110

Monnin P, Marshall N W, Bosmans H, Bochud F O and Verdun F R 2011 Image quality assessment in digital mammography: part II. NPWE as a validated alternative for contrast detail analysis *Phys. Med. Biol.* **56** 4221-38

Myers K J and Barrett H H 1987 Addition of a channel mechanism to the ideal-observer model *J. Opt. Soc. Am.* **4(12)**:2447-57

National Council on Radiation Protection and Measurements 2009 Ionizing radiation exposure of the population of the United States. Report 160

Ott J G, Becce F, Monnin P, Schmidt S, Bochud F O and Verdun F R 2014 An update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms *Phys. Med. Biol.* **59** 4047-64

Park S, Barrett H H, Clarkson E, Kupinski M A and Myers K J 2007 Channelized-ideal observer using Laguerre–Gauss channels in detection tasks involving non-Gaussian distributed lumpy backgrounds and a Gaussian signal *JOSA A* **24(12)**:136-50

Richard S and Siewerdsen J H 2008 Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images *Med. Phys.* **35(11)**:5043–5053

Richard S, Husarik D B, Yadava G, Murphy S N and Samei E 2012 Towards task-based assessment of CT performance: System and object MTF across different reconstruction algorithms *Med. Phys.* **39** 4115-22

Solomon J, Mileto A, Ramirez-Giraldo JC and Samei E 2015 Diagnostic Performance of an Advanced Modeled Iterative Reconstruction Algorithm for Low-Contrast Detectability with a Third-Generation Dual-Source Multidetector CT Scanner: Potential for Radiation Dose Reduction in a Multireader Study Radiology **275(3)**:735-45

Samara E T, Aroua A, Bochud F O, Ott B, Theiler T, Treier R, Trueb P R, Vader J P and Verdun F R 2012 Exposure of the Swiss population by medical x-rays: 2008 review *Health Phys.* **102(3)** 263-70

Samei E, Ranger N T, Dobbins J T III and Chen T 2006 Intercomparison of methods for image quality characterization. I. Modulation transfer function *Med. Phys.* **33** 1454-65

Samei E and Samuel R 2015 Assessment of the dose reduction potential of a model-based iterative reconstruction algorithm using a task-based performance metrology *Med. Phys.* **42(1)**:314-23

Schindera S T, Odedra D, Raza S A, Kim T K, Jang H J, Szucs-Farkas Z and Rogalla P 2013 Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? *Radiology* 269(2) 511-8

Torfeh T, Beaumont S, Guédron J P and Denis E 2007 Evaluation of two Software Tools Dedicated for an Automatic Analysis of the CT Scanner Image Spatial Resolution IEEE Eng. Biol. Soc. conference 2007:3910-3

Tseng H W, Fan J, Kupinski M A, Sainath P and Hsieh J 2014 Assessing image quality and dose reduction of a new x-ray computed tomography iterative reconstruction algorithm using model observers *Med. Phys.* **41(7)**:071910

Vaishnav J Y, Jung W C, Popescu L M, Zeng R, Myers K J 2014 Objective assessment of image quality and dose reduction in CT iterative reconstruction *Med. Phys.* **41(7)**:071904

Wunderlich A and Noo F 2010 Image Covariance and Lesion Detectability in Direct Fan-Beam X-Ray Computed Tomography *Phys. Med. Biol.* **53(10)**:2471–93

Yu L, Leng S, Chen L, Kofler M, Carter R E and McCollough C H 2013 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.* **40(4**)

Zhang Y, Pham B T and Eckstein M P 2007 Evaluation of internal noise methods for Hotelling observer models *Med. Phys.* **34**:3312

Zhang Y, Leng S, Yu L, Carter RE and McCollough CH 2014 Correlation between human and model observer performance for discrimination task in CT *Phys. Med. Biol.* **59**:3389

International Commission on Radiation Units and Measurements (1996) Medical Imaging–The assessment of image quality. Report 54

International Commission on Radiation Units and Measurements (2012) Radiation dose and image-quality assessment in computed tomography. Report 87

PATIENT EXPOSURE OPTIMISATION THROUGH TASK-BASED ASSESSMENT OF A NEW MODEL-BASED ITERATIVE RECONSTRUCTION TECHNIQUE

Julien G Ott[1,*], Alexandre Ba[1], Damien Racine[1], Nick Ryckx[1], François O Bochud[1], Hatem Alkadhi[2] and Francis R Verdun[1]
[1]Institute of Radiation Physics, CHUV, Lausanne, Switzerland
[2]Institute of diagnostic and interventional radiology, USZ, Zürich, Switzerland

Received month date year, amended month date year, accepted month date year

Our goal is to report and investigate the performances of a new iterative reconstruction algorithm, using a model observer. For that, a dedicated low contrast phantom containing different targets was scanned at four $CTDI_{vol}$ levels on a Siemens SOMATOM Force CT. The acquired images were reconstructed using the ADMIRE algorithm and were then assessed by three human observers who performed alternative forced choice experiments. Next, a channelized hotelling observer model was applied on the same set of images. The comparison between the two was performed using the percentage correct as a figure of merit. Our results indicated a strong agreement between human and model observer as well as an improvement in the low contrast detection when switching from an ADMIRE strength of 1 to 3. Good results were also observed even in situations where the target was hard to detect, suggesting that patient dose could be further reduced and optimised.

INTRODUCTION

Over the last decade, the radiation dose delivered to patients via diagnostic X-ray imaging has continuously increased until today, where it reaches 25 of the accumulated man-made and natural radiation. Among that 25%, computed tomography (CT) raises a particular concern, since this imaging modality represents in Switzerland for example 68% of the collective dose, yet 8% of the examinations only [1]. In this context, CT manufacturers have developed new strategies like iterative reconstruction (IR) algorithms in order to ensure that the benefits-risk ratio remains in favour of the patient. This new technology certainly improved the clinical practice [2], but it has also led to drastic changes in image perception. Thus, ensuring an adequate level of image quality while keeping patient's exposure as low as reasonably achievable constitutes a new challenge to be addressed. The use of task-based image quality assessment method could represent an efficient way to perform this optimisation scheme. Therefore, our goal in this study is to report and investigate the performances of a new IR technique using a model observer that mimics human detection of low contrast targets: the channelized hotelling observer (CHO) model.

MATERIAL AND METHODS

Data acquisition

We used a dedicated low contrast phantom (QRM, Moehrendorf, Germany) mimicking the attenuation produced by a patient's chest. The phantom could embed two different custom-made modules in its middle: a homogeneous modulus and another containing low contrast spherical targets of 6 and 8 mm in diameter with contrast levels of 10 and 20 HU at 120 kVp.

Data acquisition was performed at the University Hospital Zurich on a third-generation dual-source 192-slice CT scanner (SOMATOM Force, Siemens Healthcare, Erlangen, Germany). We used a tube voltage of 120 kVp, a 300 mm field-of-view (DFOV), a 512×512 matrix size and 2.0 mm thick slices, which were reconstructed every 1.0 mm. Acquisitions were performed in the helical mode with a pitch of 0.98. We investigated four doses levels (1.0, 3.5, 8.0 and 15.0 mGy expressed in a $CTDI_{vol}$ phantom of 32 cm in diameter), using the procedure described in the IEC 60601-2-44 to measure the $CTDI_{vol}$. The phantom was scanned 20 times for each condition. Reconstructions were performed using the Siemens advanced model iterative reconstruction (ADMIRE) with the strength levels 1 and 3. On the machine, users can choose ADMIRE levels ranging from 1 to 5, with level 1 being closest to the image impression of traditional, filtered back projection, and level 5 showing the strongest noise reduction. In the end, 32 different categories were obtained (four dose levels, two ADMIRE levels, two contrast levels and two target sizes). From these sets of acquisitions, regions of interest (ROIs) of 41x41 pixels (0.59 mm pixel size) containing the centred targets were extracted. For each category we extracted 100 images containing a signal (20 scans x 5 targets with identical size and contrast in the phantom) and 1000 images with only noise. The ROIs which contain noise only were extracted from the homogeneous modulus, whereas the ROIs containing the signals came from the low contrast modulus. This methodology enabled us to hold the same position on the (x,y) plane for both signal and noise ROIs, enabling us to solve the noise stationarity problem.

Human observer

In our experiment, three human observers (medical physicists) took part in four alternative forced choice (4- AFC) experiments in order to yield a percentage of correct responses (PC) indicating how well they managed to detect the signals. This 4-AFC experiment consisted in selecting the signal-present image in a batch of three signal-absent images and one signal-present image, which were presented together in a randomized order. All observers were blinded to the CT acquisition and reconstruction conditions and began their test with a training session which was made of images acquired at high dose level. They were then asked to make decisions for all 32 categories we acquired. Our previously made acquisitions provided us with 100 signal-present ROIs and 1000 signal-absent ROIs for each category. ROIs among those data were selected randomly and used for the 4-AFC tests. For each observer and category, every answer to the 100 trials was stored and compared to the correct response, allowing the computation of the PC.

CHO model observer

A model observer enables to predict the detection of low contrast signals by calculating a scalar response called the decision variable and denoted by λ_i. This parameter is given by:

$$\lambda_i = \boldsymbol{w}^T . \boldsymbol{g_i} \qquad (1)$$

With $\boldsymbol{w}$ the template of the model observer and $\boldsymbol{g_i}$ the analysed ROI ($i=n$ or $i=s$ represents signal-absent, respectively signal-present hypothesis).
The CHO model used in this study is an anthropomorphic model observer also including pre-processing of the image by a set of channels that enhance some spatial frequencies [3] [4]. The template $\boldsymbol{w}_{CHO}$ of this model is obtained as explained in the following part (Extensive details can be found in [5] [6] [7] [8]):

$$\boldsymbol{w}_{CHO} = \left[\frac{1}{2}\left(\boldsymbol{K}_{cs} + \boldsymbol{K}_{cn}\right)\right]^{-1}\left(\overline{\boldsymbol{g}_{cs}} - \overline{\boldsymbol{g}_{cn}}\right) \qquad (2)$$

Where: $$\boldsymbol{K}_{cn} = \boldsymbol{U}^T \boldsymbol{K}_n \boldsymbol{U} \ \ and \ \ \boldsymbol{K}_{cs} = \boldsymbol{U}^T \boldsymbol{K}_s \boldsymbol{U} \qquad (3)$$

$\boldsymbol{K_n}$ and $\boldsymbol{K_s}$ are the covariance matrix calculated respectively from the signal-absent and signal-present data and $\boldsymbol{U}$ is the matrix representation of the channel filters described more extensively below.

In (2), $\overline{\boldsymbol{g}_{cs}}$ and $\overline{\boldsymbol{g}_{cn}}$ are the means of the channel outputs under a signal-present and signal-absent hypothesis and can be estimated according Eq. (4):

$$\overline{\boldsymbol{g}_{cs}} = \boldsymbol{U}^T \overline{\boldsymbol{g}_s} \ \ and \ \ \overline{\boldsymbol{g}_{cn}} = \boldsymbol{U}^T \overline{\boldsymbol{g}_n} \qquad (4)$$

The employed set of channels is called dense difference of Gaussian (DDoG) channels and was described by Abbey and Barrett in *2001* [9]. It includes ten channels for which the radial frequency profile of the jth channel is given by:

$$\boldsymbol{U}_j(\rho) = exp\left[-\frac{1}{2}\left(\frac{\rho}{W\sigma_j}\right)^2\right] - exp\left[-\frac{1}{2}\left(\frac{\rho}{\sigma_j}\right)^2\right] \qquad (5)$$

Where ρ is the spatial-frequency, W=1.67 defines the bandwidth of the channel and σ_j is the standard deviation of the jth channel. Each σ_j value is defined by the equation $\sigma_j = \sigma_o \alpha^j$ with $\sigma_o = 0.005$ and $\alpha = 1.4$ (Cf. [9]).

The decision variable can then be calculated by injecting the channel output of the ROI (denoted by $\boldsymbol{g_{ci}}$) and the template $\boldsymbol{w}_{CHO}$ in Eq. (1):

$$\lambda_i = \boldsymbol{w}_{CHO}^T . \boldsymbol{g_{ci}} \qquad (6)$$

The PC was obtained by using the CHO model to perform 4-AFC tests on the acquired images. The value of the decision variable was used to determine which of the four images contained the signal (the highest value of λ is supposed to be the signal-present image). Then, the results of the model observers, were compared to the truth in order to enabling the computation of PCs.

Uncertainty estimation

The uncertainties on the models' results were estimated by performing bootstrap [10]. We performed the 4-AFC test 150 times for each category, leading to 150 values of PC. Then, we computed the mean and standard deviation of the 150 values we obtained in order to determine the final mean PC value as well as its standard deviation for each acquisition condition. This allowed us to estimate a 95% confidence interval. The uncertainties for the human observers were calculated using the results of the three different observers. For each category, the mean PC value of the observers and its standard error were calculated in order to display a 95% confidence interval.

---

*Corresponding author: Julien.Ott@chuv.ch

RESULTS

In this section, the qualities of the images obtained using the ADMIRE algorithm with the strength levels 1 and 3 are assessed through the performances of model (Figure 1) and human observers (Figure 2).

Results of the CHO model with the DDoG channels suggest that both strength levels exhibited PCs in the same range with an increase of few percentages in the results when switching from a strength level 1 to 3. This trend was however only observed for certain signals and dose range, namely the lowest contrast (10 HU) associated to low dose levels (1.0 and 3.5 mGy).

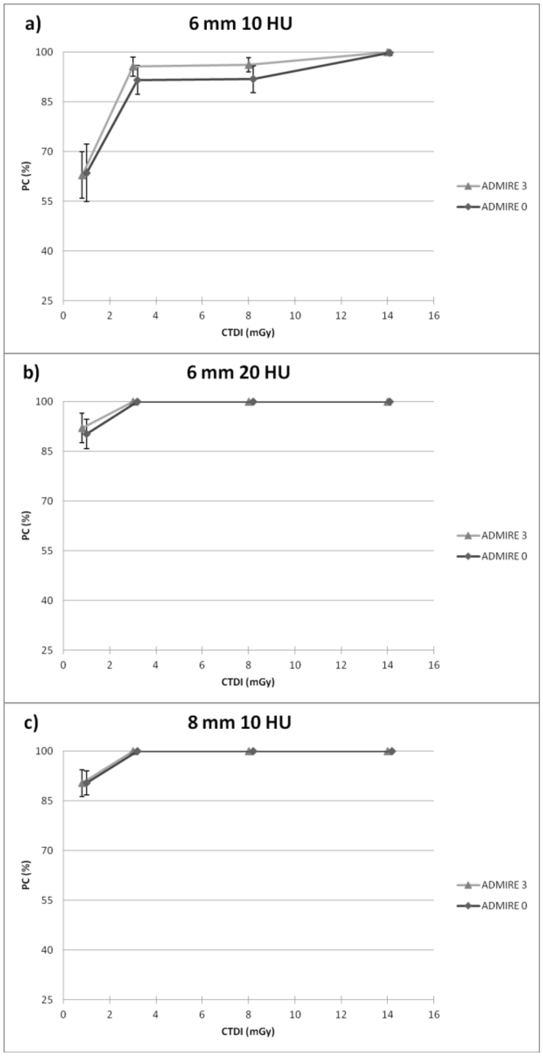


Figure 1: Results of the CHO model observer (in PC, obtained by performing 4-AFC tests)

Human observers exhibit results very similar to the ones obtained with the CHO model in terms of PC values. The results also suggest that the use of higher ADMIRE strength level is useful to improve the detection of small size and low contrast signal under low $CTDI_{vol}$. On top of that, we observed that the largest signal size (8 mm) with the highest contrast (20 HU) corresponded to a trivially easy task with PC always equal to 100% no matter which dose and ADMIRE level was employed. Those results remained true for both human and model observers and are therefore not represented on figure 1 and 2.

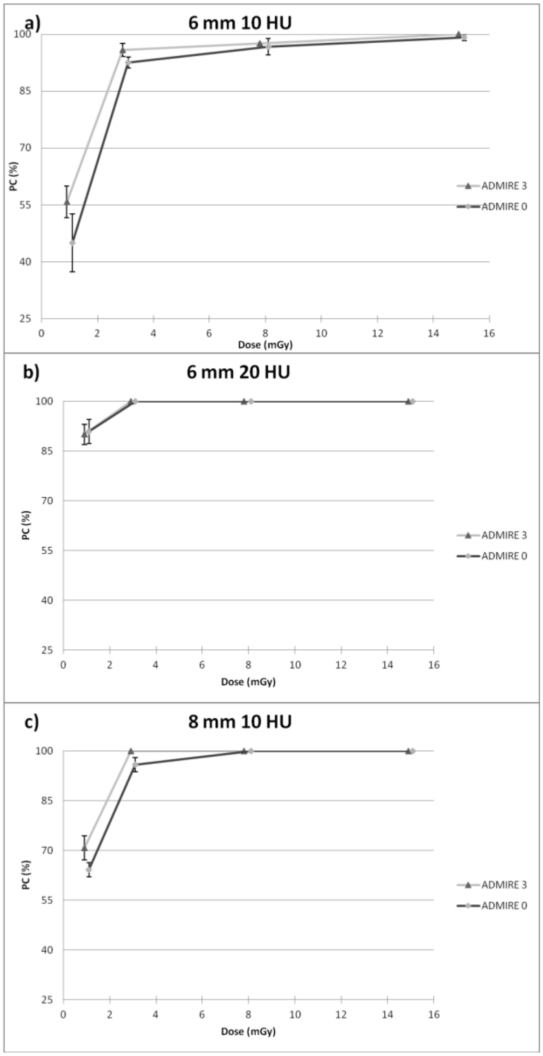


Figure 2: Results of the three human observers (in PC, obtained by performing 4-AFC tests)

Model and human observers exhibit a great adequacy in their results, both of them indicating that the use of a higher ADMIRE level enhances the detection when working under conditions where the signal is hard to detect. We also witnessed that the PC increases with the given dose until it reaches an asymptote and that this asymptote is reached faster when using higher ADMIRE strength level. Furthermore, the results enlighten that no more dose increase is needed once an amount of approximately 4 mGy is reached because the performance of the human observer is already almost maximal at this point.

DISCUSSION

The goal of our investigation was to determine if the use of the ADMIRE algorithm at different strength could lead to a high detection performance, therefore allowing a further dose reduction in the clinical practice.

The results showed that the CHO model with DDoG channels gave coherent results since it reproduced the behaviour of humans very well and under a wide range of conditions and signal contrasts. The model seems to be very efficient in low contrast detection and even sometimes overestimates human results for low contrast and low doses. These results are coherent with recent studies from Leng and Yu [11] which showed the efficiency of the CHO model for the low contrast detection. Also, both model and human observers reported a visible improvement in the low contrast detection when increasing the ADMIRE strength. This trend was observed when working at low dose levels (less than 4 mGy) for all signal types. Indeed, when working at higher dose levels, the PCs always reached 100%, letting no room for improvement. In the end, the use of ADMIRE makes it possible to diminish the dose without losing information in the image. Indeed, the PC results obtained in the study reach very high or perfect values for almost every acquisition condition and signal, indicating that a dose reduction without impacting the detection performance would be possible.

Some limitations of our study have however to be underlined. First among them, the number of images we acquired may be considered as low since we did not dispose of enough ROIs to separate them in two exploitable data sets. Usually, a first set is used for the computation of the covariance matrix in the determination of the template of the CHO model, while the second data set is used for the computation of the PCs. However, it is worth to underline that Barrett and Myers [12] who studied this problem concluded that using one single set of data remained a reliable way to proceed. Moreover, performing a limited number of acquisitions (20 scans for each acquisition condition in our case) allowed to reduce the operating time of the device, which is appreciable when working in a clinical environment.

The second limitation we faced is that the paradigm we worked with (when the signal, location and background are exactly known) was simplified and therefore different from real anatomical conditions. Our results could nevertheless be used in order to assess the performances of the tested IR algorithms, but it is worth to mention that there is room for a more complex study on the subject.

CONCLUSION

Nowadays, assessing CT image quality cannot be done with image space metrics anymore. Moreover, evidence indicates that frequency metrics should not be used either when working with IR. However, the task-based tool used in this investigation (CHO model observer associated to the DDoG channels) successfully demonstrated its ability to reproduce the human's response in a low contrast detection task, thus establishing its reliability for image quality assessment. Our results obtained with this tool revealed that the ADMIRE algorithm led to high PCs even in situations where the target was harder to detect (i.e. low $\mathrm{CTDI}_{\mathrm{vol}}$ and contrast level). Also, using higher ADMIRE strength, led to PC improvement, particularly in the low $\mathrm{CTDI}_{\mathrm{vol}}$ range. Therefore, using those benefits to keep the image quality equivalent to what was previously obtained would enable to spare some delivered dose.

All those elements suggest that the patient dose could be further optimised and reduced thanks to the use of the ADMIRE algorithm and this new CT unit.

ACKNOWLEDGEMENTS

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

The author would also like to thank the personal of the USZ for their active participation in the acquisitions, as well as all the readers who took part in the detection studies.

REFERENCES

1. Samara E T, Aroua A, Bochud F O, Ott B, Theiler T, Treier R, Trueb P R, Vader J P and Verdun F R 2012 Exposure of the Swiss population by medical x-rays: 2008 review Health Phys. 102(3) 263-70
2. Schindera S T, Odedra D, Raza S A, Kim T K, Jang H J, Szucs-Farkas Z and Rogalla P 2013 Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? Radiology 269(2):511-8.
3. Myers K J and Barrett H H 1987 Addition of a channel mechanism to the ideal-observer model J. Opt. Soc. Am. 4(12):2447-57
4. Gallas B D and Barrett H H 2003 Validating the use of channels to estimate the linear model observer J. Opt. Soc. Am. 20(9):1725-38

5. Beutel J, Kundel H L and Van Metter R L 2000 Handbook of medical imaging Vol 1. Physics and psychophysics SPIE press
6. Castella C, Abbey C K, Eckstein M P, Verdun F R, Kinkel K and Bochud F O 2007 Human linear template with mammographic backgrounds estimated with a genetic algorithm J. Opt. Soc. Am. 24(12):B1-12
7. Yu L, Leng S, Chen L, Kofler M, Carter R E and McCollough C H 2013 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. 40(4)
8. Tseng H W, Fan J, Kupinski M A, Sainath P and Hsieh J 2014 Assessing image quality and dose reduction of a new x-ray computed tomography iterative reconstruction algorithm using model observers Med. Phys. 41(7):071910
9. Abbey C K and Barrett H H 2001 Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability J. Opt. Soc. Am. 18(3):473-88
10. Efron B and Tbishirani R J 1994 An introduction to the bootstrap: Monograph on statistics and applied probability CRC Press
11. Leng S, Yu L, Chen L, Ramirez Giraldo J C and McCollough C H 2013 Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain Med. Phys. 40(8):081908
12. H.H. Barett and K.J. Myers, Foundation of image Science (John Wiley's & Son, Hoboken, NJ, 2004)

Task-based assessment of an MBIR algorithm: A study in all reconstruction planes

Julien G. Ott[1], Francis R. Verdun[1], Patrick Omoumi[2], Eric Dugert[2], François O. Bochud[1] and Fabio Becce[2]

[1] Institute of Radiation Physics, CHUV, Lausanne, Switzerland
[2] Department of Diagnostic and Interventional Radiology, CHUV, Lausanne, Switzerland

Introduction

Computed tomography (CT) images are traditionally acquired, reconstructed and analyzed in the axial plane. However, in several clinical situations, CT images need to be analyzed in the coronal and/or sagittal planes, particularly in cardiovascular, thoracic and musculoskeletal imaging [Fang et al 2015]. With the arrival of CT 3D imaging it has become possible today to acquire and reconstruct CT images in all reconstruction planes [Dalrymple et al 2007]. Several authors [Long 2010 *et al;* Rydberg *et al* 2007; Von Falck *et al* 2011] already investigated the impact of sagittal and coronal reformatting for various clinical CT acquisition conditions. Most of the work conducted so far was however only done for classical filtered-back projection (FBP) algorithms and using subjective image quality assessment methods like visual grading analysis made by human observers.

The recent implementation of iterative reconstruction (IR) techniques in clinical CT protocols has helped to significantly reduce radiation dose, but with a potential change in image quality. The impact of the various manufacturer-specific IR techniques on CT image quality has already been extensively studied in the axial plane using both visual grading analysis and physics methods fully adapted to IR [Schindera *et al* 2011, Mieville *et al* 2012; Ott *et al* 2014]. However, IR impact on the coronal and sagittal reconstruction planes has not yet been done using objective and fully IR adapted methods.

Our work will focus on comparing image quality in all three reconstruction planes using objective assessment methods adapted to IR. We focus on a great variety of algorithms designed by a single manufacturer.

Material and Methods

Phantom

In order to evaluate key CT image quality parameters like image noise or spatial resolution, we built our own phantom, containing a 10-cm-diameter cylinder made of Teflon® (polytetrafluoroethylene), low-density Polyethylene and Plexiglas® (polymethyl methacrylate) and centered in the middle of a solution made of diluted (dilution at 25%) iodinated contrast material and saline (figure 1). Among those materials, Plexiglas® revealed itself to be particularly useful since its contrast with background was close to the one of articular cartilage (120 HU). The solution was diluted at 25% rather than 50% (usual clinical value) because no resorption by the synovial tissue occurred in the phantom contrary to clinical cases. This experimental paradigm enabled us to produce phantom images similar to clinical CT arthrogram images (figure 1 b). Analyzing the Hounsfield units (HU) transition between the contrast medium and the Plexiglas® insert allowed us to obtain a resolution estimate based on the computation of a metric called target transfer function (TTF) [Richard *et al* 2012]. The phantom also contained a homogeneous region where only diluted contrast material was present. This region was used to calculate the noise power spectrum (NPS) which is a good estimate of the noise texture within an image. Those noise and spatial resolution assessment tools were employed in order to estimate the quality of the images regarding the modality which was used to perform their reconstruction and the plane in which they were acquired and displayed.

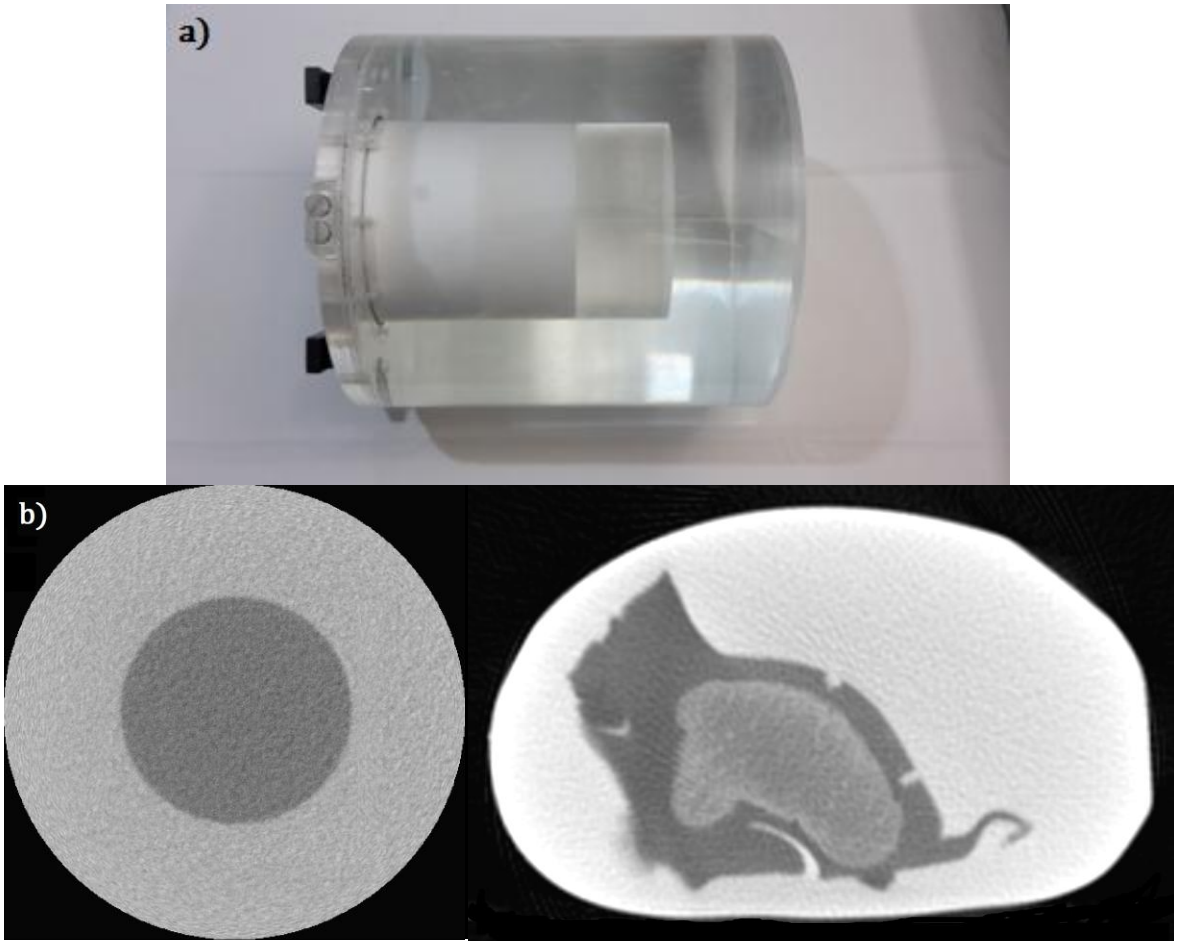


Figure 1: a) Custom-made CT image quality phantom containing a 10-cm-diameter cylinder made of Teflon®, low-density Polyethylene and Plexiglas® (from left to right). **b)** Comparison between a slice of our custom made phantom and a clinical CT arthrogram image (Axial plane, FBP algorithm,).

Protocol

Our custom-made image quality phantom was scanned in all three reconstruction planes on a GE HD 750 CT scanner (GE Healthcare, USA). The acquisitions were performed with a protocol similar to the one used in clinical routine for patients undergoing CT arthrography, that is to say with a $\mathrm{CTDI}_{\mathrm{vol}}$ of 7.3 mGy, a tube voltage of 120 kVp, a pitch equal to 0.97 and a field of view of 200 mm with a matrix size of 512x512 pixels. The acquired data sets were then reconstructed in the axial, sagittal and coronal planes, using a nominal slice thickness of 0.625 mm, and three different manufacturer-specific reconstruction algorithms: the classical filtered-back projection (FBP), the adaptive statistical iterative reconstruction (ASIR) at a percentage of 40 and 80% (both with the GE bone kernel) as well as two versions of GE model-based iterative reconstruction (MBIR), the “VEO” algorithm. The oldest version of VEO used in our study (VEO 2.0) was only compatible with the standard kernel. However, new presets are provided in the VEO 3.0 version, including different noise preference (NR**) and resolution preference (RP**). The preset index exactly describes the expected noise reduction or resolution improvement as measured on a GE performance phantom. For example, NR05 implies 5% noise reduction over standard preset. Similarly RP05 implies 5% higher resolution compared to standard preset. Also, the image model and noise model of the algorithm were improved to make the noise covariance more isotropic in all three dimensions.

First, we used those acquisitions to compare the quality of the images in a given reconstruction plane depending on the reconstruction algorithm which was used. Then, we compared the differences in term of image quality in the different reconstruction planes. Comparisons were based on objective physical metrics which will be detailed in the next part. Our approach consisted in integrating those metrics in a customized model observer to estimate the signal-to-noise ratio (SNR) of simulated articular cartilage lesions of several sizes surrounded by contrast material.

Physics Metrics

The image noise was investigated within a 6-cm-long uniform region of our phantom. NPS were calculated with a home-made Matlab® routine (The MathWorks, Natick, MA, USA) based on 25 image slices of the homogenous contrast-media region containing regions of interests (ROIs) of 128×128 pixels. A radial-averaged NPS was obtained based on the guidelines described in the ICRU report 54 [ICRU 54]. Extensive details on the NPS metric as well as elements to perform its computation can be found in Miéville’s work [Miéville *et al* 2011].

Iterative reconstruction (IR) algorithms are known to be highly non-linear and therefore to introduce a dependency of the image contrast and noise over the spatial resolution of the image [Thibault *et al* 2007; Richard *et al* 2012]. In order to overcome those problems, spatial resolution was investigated through an object-specific modulation transfer function (MTF), which we referred to as TTF. It allowed us to take into account noise and contrast effect when characterizing image resolution in different acquisition conditions. A common method to compute this TTF consist in using an edge spread function (ESF) [Judy 1976], before deriving the result and switching to the Fourier space. In our case, this object specific metric was obtained using the contrast variation between the circular edge of the Plexiglas® cylinder and the contrast product around it. Eventually, computations of TTF for resolution estimation allowed us to overcome IR specific problems. Main mathematical steps, as well as extensive details and explanations on the methodology can be found in the paper of Ott et al. [Ott *et al*, 2014].

Eventually, we integrated those two metrics into a model observer in order to yield an objective assessment of the image quality in the three reconstruction planes when using different reconstruction algorithms. Model observers rely on the concept of task-based assessment in order to assess how well the desired information can be extracted from the image [Barrett and Meyers 2004]. The principle of such models consists of measuring the performance of an observer conducting a task of clinical interest. The task may be the detection of a signal into a noisy background, the observer is the person or algorithm performing the task and the figure of merit (FOM) measures how well the observer performed the task. In our case, we used an updated non-prewhitening with eye (NPWE) model observer to compute the SNR of a simulated lesion of the articular cartilage. The NPWE is a mathematical model observer developed by Burgess [Burgess 1994] and which can be used to perform task-based assessment. Some authors [Brunner and Kyprianou 2011; Ott *et al* 2014] successfully integrated some modifications to this model, like the use of the TTF instead of the MTF in order to meet the requirements due to the use of IR algorithms. Based on those works we were able to compute SNR values for our simulated cartilaginous lesion using Equation 1:

$$SNR = \frac{\sqrt{2\pi}\,\Delta HU\ \int_0^{f_{Ny}} S^2(f) TTF^2(f)\, VTF^2(f) f\, df}{\sqrt{\int_0^{f_{Ny}} S^2(f) TTF^2(f)\, NPS(f) VTF^4(f) f\, df}} \qquad (1)$$

f_{Ny} is the Nyquist frequency, ΔHU is the contrast difference, $VTF(f)$ is the visual transfer function of the human eye [Burgess 1994] and $S(f)$ is the Fourier transform of the input signal (in our case, $S(f) = \frac{R}{f} J_1(2\pi Rf)$, J_1 being a Bessel function of the first kind).

Results

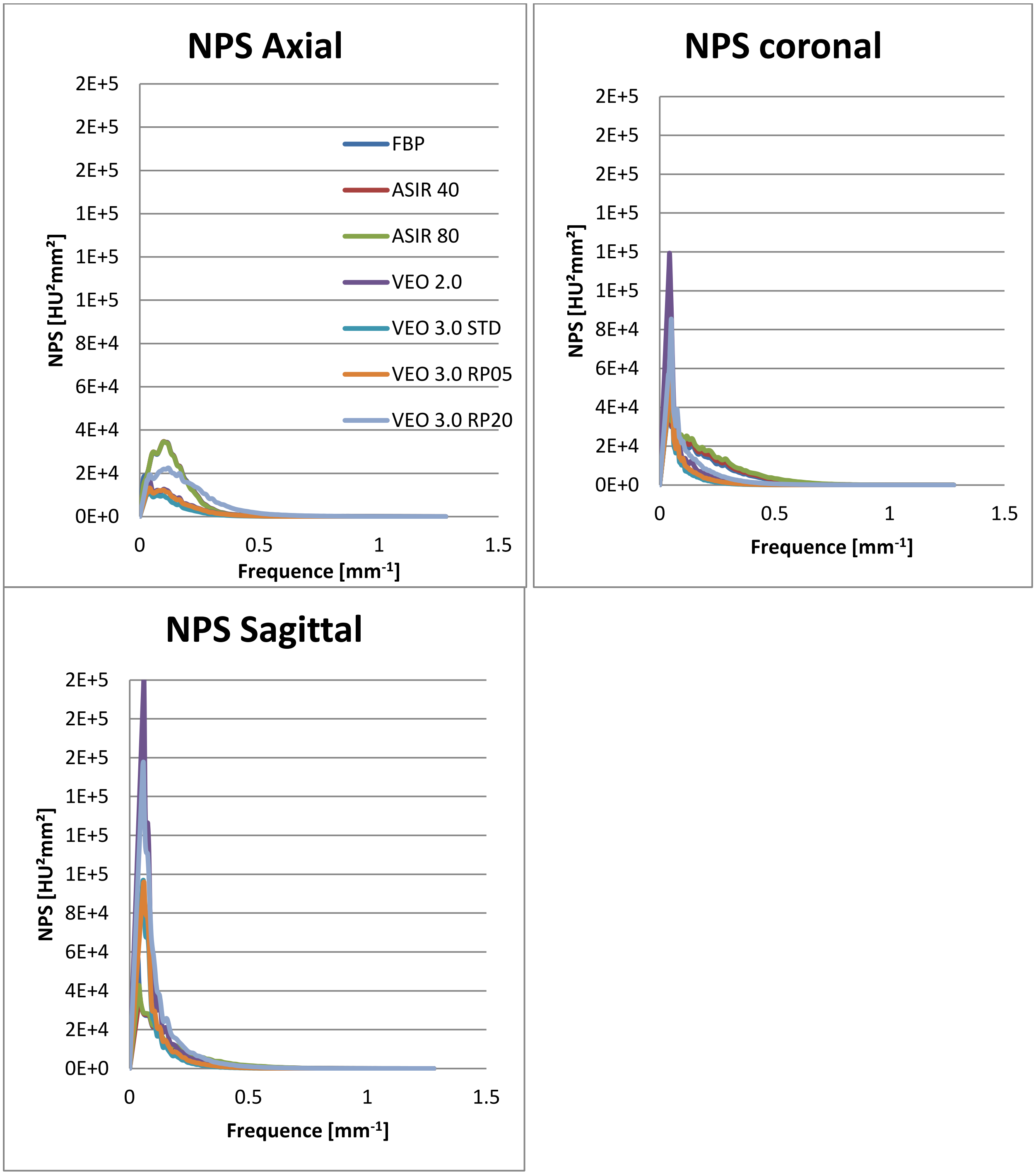


Figure 2: NPS curves computed for FBP, ASIR, VEO 2.0 and VEO 3.0 reconstruction algorithms in the three reconstruction planes.

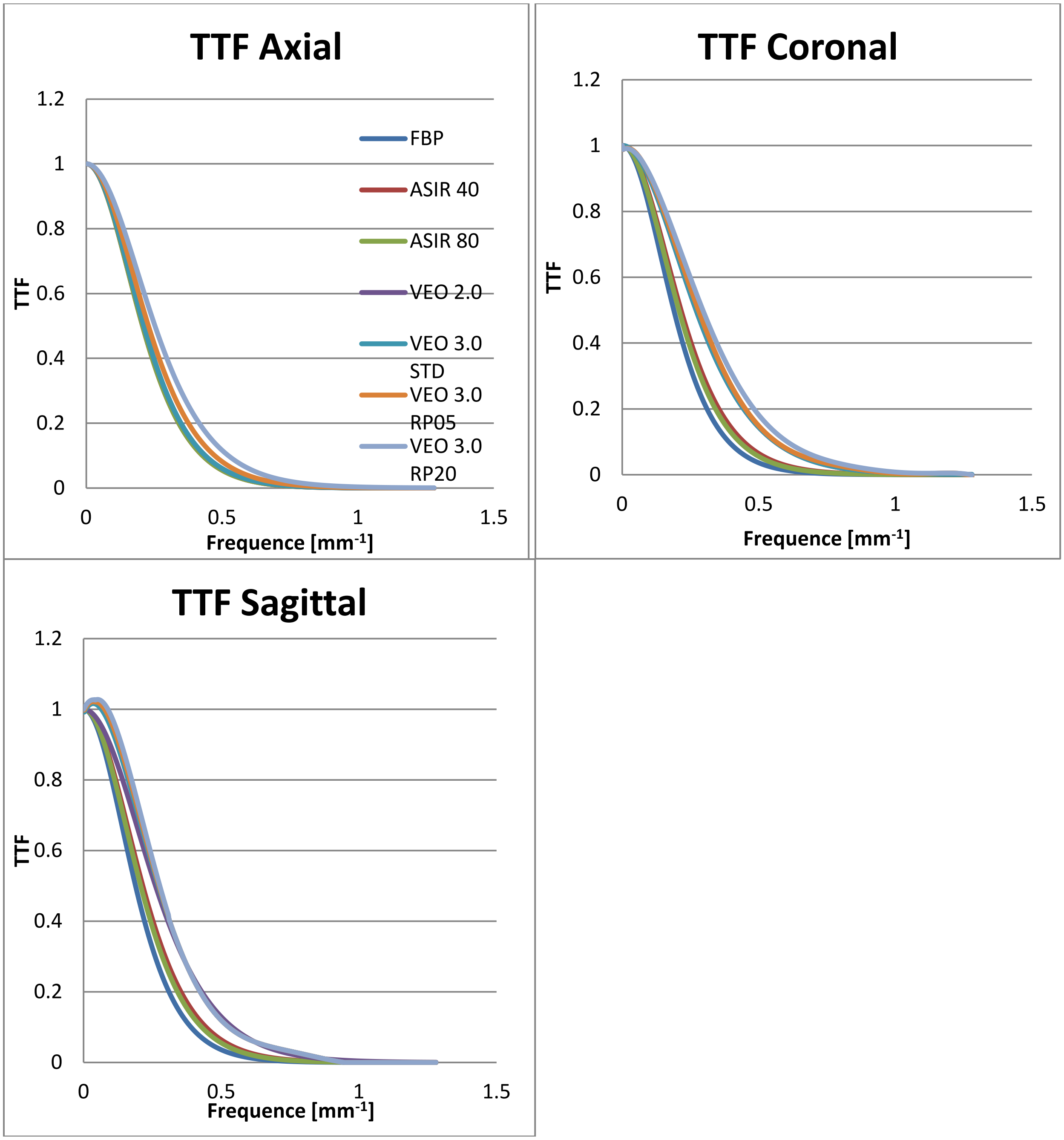


Figure 3: TTF curves computed for FBP, ASIR, VEO 2.0 and VEO 3.0 reconstruction algorithms in the three reconstruction planes.

FBP, ASIR 40 & 80, VEO 2.0

Variations in NPS as a function of the image reconstruction algorithm and reconstruction plane were observed (Figure 2). Those changes in the shape of the NPS curves indicate a modification of the overall image texture. Coronal and sagittal reformatted images thus display a higher proportion of low range frequencies, accounting for their smoother aspect compared to axial images.

In the axial plane, and compared to FBP, no changes in the noise level were observed when using ASIR 40 or 80. Switching from FBP to VEO 2.0 did however yield to a consequent noise decrease, especially in the low frequencies range. On coronal and sagittal-reformatted images, we noted an increase in image noise compared with the axial plane. This trend was even slightly more pronounced in the sagittal plane. Moreover, the impact of both ASIR and the VEO 2.0 algorithm on noise level was considerably different in those two planes. Indeed, no difference between NPS curves of ASIR 40, ASIR 80 and FBP images could be observed, while VEO 2.0 happened to lead to higher noise levels than FBP and both ASIR, especially in the low frequency range (inferior to 0.2 mm^{-1}).

TTF curves (figure 3) in the axial plane indicated that no improvement of the spatial resolution was perceived when using ASIR at both 40% and 80% instead of FBP. On the contrary, the use of VEO 2.0 produced an increase of the resolution for the whole frequency range of the image. A similar behavior was observed in the coronal and sagittal planes for all algorithms. Eventually, the TTFs curves suggest that, for a given algorithm, an increase of the spatial resolution happens when switching from axial to coronal or sagittal reconstruction planes. This resolution increase happened to be of the same magnitude in both sagittal and coronal planes, leading to very similar TTF curves in those two planes.

Our results suggest no major changes in the detectability as estimated by the NPWE model observer in the sagittal and particularly the coronal plane compared with the axial plane (figure 4). Those observations are in accordance with the noise and resolution changes which were reported above using TTF and NPS. However, we observed a noticeable enhancement of detectability in all reconstruction planes when using VEO 2.0 instead of FBP or ASIR, demonstrating that the use of this MBIR algorithm could help to increase SNR even when switching reconstruction planes.

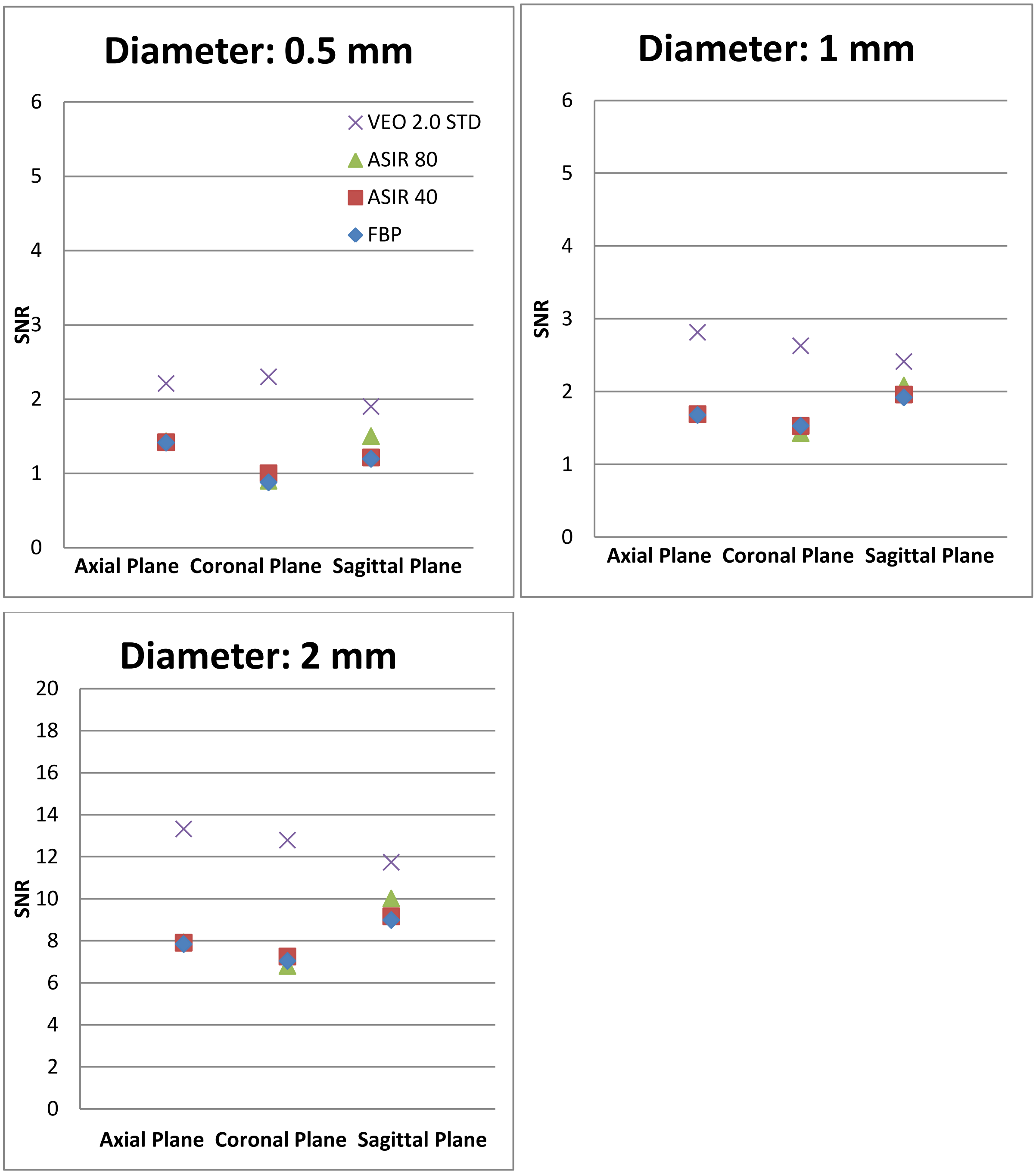


Figure 4: SNR computed for FBP, ASIR and VEO 2.0 reconstruction algorithms in the three reconstruction planes for simulated cartilaginous lesion of 0.5 mm 1.0 mm and 2.0 mm.

VEO 3.0

NPS obtained using VEO 3.0 algorithm to reconstruct the images revealed to contain very few noise in the high frequency domain (almost no noise is detected above frequencies of 0.2 mm^{-1}) and to be influenced by the type of reconstruction kernel being used. Indeed, when looking at NPS of images reconstructed in the axial plane, we observed that the noise level slightly increases when switching from the standard kernel, to the RP05 kernel to the RP20 kernel. Those results also remain true in the coronal and sagittal plane, in addition of changes observed in the magnitude of the NPS curves. Indeed, we observe that the amplitude of the noise is gradually increased when switching from axial to coronal to sagittal plane.

TTF curves in the axial plane allow us to witness an increase of the resolution when switching the kernels from standard to RP05 to RP20. In our case we measured an 8% and 19% increase of the frequencies at 50% of the TTF height. The same kind of behavior was also observed for TTF curves in the coronal and sagittal planes. In addition, our results also suggest that the resolution as estimated by the TTF increases when switching from axial to coronal to sagittal planes.

Results regarding the detectability of simulated cartilaginous lesions using the VEO 3.0 algorithm are summarized in figure 5. They globally suggest that using a standard kernel yields a higher SNR than the RP05, which itself produces higher SNR than the RP20 kernel. This trend is due to a higher noise level when using sharper resolution kernels and which is not fully compensated by the signal increase those kernels also generate. Also, the resolution increase we measured when switching from axial to sagittal and coronal planes allows to compensate the noise increase that occur in the same time, leading to an stable SNR in the three planes for a given reconstruction kernel.

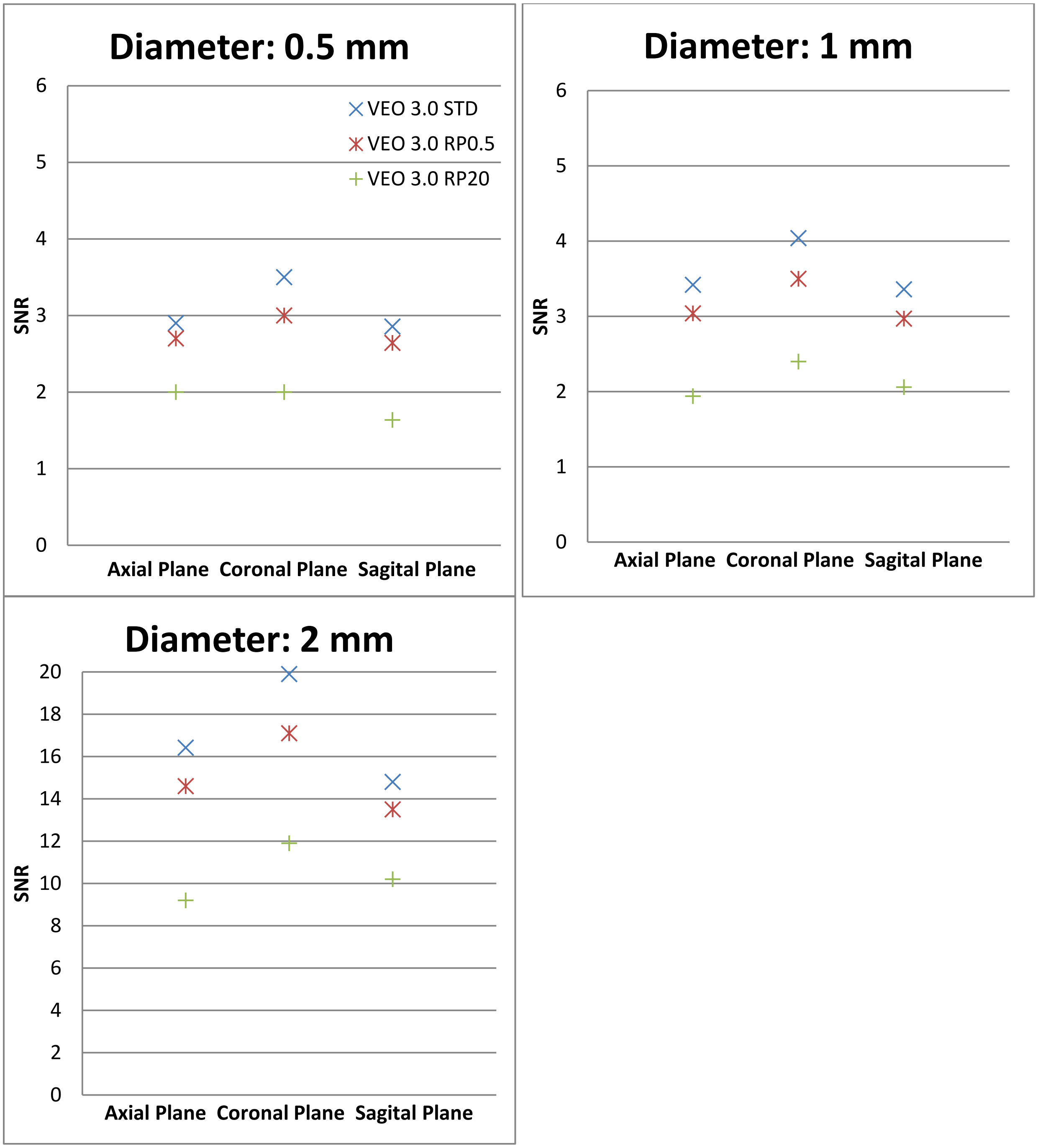


Figure 5: SNR computed for VEO 3.0 in the three reconstruction planes for simulated cartilaginous lesion of 0.5 mm 1.0 mm and 2.0 mm.

Discussion

Our results indicate that in our paradigm, using ASIR algorithms instead of FBP does not bring any major improvement regarding the SNR. This result was already observed in other studies focusing on the axial plane exclusively [Racine et al 2015] and remained true in the three different reconstruction planes. Yet, the use of MBIR like VEO 2.0 happened to lead to more conclusive effects on the SNR values. In the axial plane, a consequent noise reduction was indeed observed with VEO 2.0 compared to FBP, whereas an increase of the noise was observed in the coronal and sagittal planes when switching from FBP to VEO 2.0. In the mean time, resolution found out to be improved with VEO 2.0 when switching from the axial to the other planes. Eventually, using VEO 2.0 instead of FBP led to a significant SNR increase in all reconstruction planes. The previously stated results regarding noise and resolution behavior remained true and even reinforced when switching from FBP to VEO 3.0. Results suggest that the 3.0 version leads to a higher SNR in all three reconstruction planes than the 2.0. In the end, switching reconstruction planes and algorithms has repercussions on image quality, so that lesion detection and characterization by the radiologist are therefore modified.

Image quality depending on the reconstruction plane was already investigated by several authors [Rydberg *et al* 2007; Singla Long *et al* 2010] who successfully demonstrated that multi-planar reconstructions could improve diagnostic accuracy and interpretation speed. Those studies were however conducted relying on subjective image quality assessment methods, i.e. scoring of the images by human observers in this case. If this kind of methods has already demonstrated its great potential to assess the quality of CT images, there is also a need for objective methods to perform this task. Thus, other authors chose to adopt those objective methods in order to study the impact of the reconstruction plane over the image quality [Von Falck *et al* 2011] and to conclude on the usefulness of the multi-planar reconstructions. Most of the objective methods used to characterize image quality however often consist in the computation of pixels' standard deviation, or of contrast-to-noise ratio. The study therefore provided useful and novel results regarding the quality of images reconstructed in different planes, but the outcome of these simple metrics is systematically improved when working with IR, because those algorithms integrate the knowledge of the noise statistics to penalise pixel variations and reduce noise.

This therefore enlightens the need to develop more elaborated tools, which are in compliance with the requirements of IR techniques, and we believe that methods such as task-based assessment could represent an efficient way out of the problem. Some authors like Guggenberger [Guggenberger *et al* 2013] already tackled the problematic of the multi-planar reconstruction using this kind of method. The study in question was however limited to FBP algorithms and we believe that IR algorithms, now that they are widely used in clinical routine should also be assessed this way. Our study was therefore conducted in order to objectively assess several types of IR, including some recently released MBIR algorithm, in different planes and using a task-based assessment method.

There are however some limitations to our work that have to be mentioned, the first one being that no image quality assessment was performed in our experimental paradigm. Indeed, if it is true that this kind of assessment is subject to some drawback like the subjectivity of the observer, they still represent the gold standard and comparing our objective results to human observers could be interesting. The second limitation comes from the fact that image quality assessment was performed but the diagnostic accuracy was not evaluated. To date, most of the published literature on IR techniques has reported reduction in radiation dose while maintaining diagnostic image quality, but very few have evaluated the impact of IR techniques on the diagnostic performance.

The results we obtained show that MBIR are indubitably the reconstruction algorithms leading to the highest SNR level in every reconstruction planes. The 3.0 version even surpasses the 2.0 version thanks to its possibility to change the reconstruction kernel and its results leading to higher SNR values.

Acknowledgment

The authors would like to thank Virginie Berclaz for her active participations in the acquisitions, and Debashish Pal for providing us the CT images reconstructed with the VEO 3.0 algorithm. This work was partly supported by a grant from the Swiss National Science Foundation (SNSF) [No. 320030-140995].

References

[Barrett and Meyers 2004] H. H. Barrett and K. J. Myers, "*Foundations of Image Science*." Wiley, New York, 2004.

[Brunner and Kyprianou 2011] C. C. Brunner and I. S. Kyprianou, *"Material-specific transfer function model and SNR in CT"* 2013 PhysMed Biol **58** 7447-61.

[Burgess 1994] A. E.Burgess, *"Statistically defined backgrounds: performance of a modified non prewhitening observer model"* 1994 JOSA A **11(4)**:1237-42.

[Dalrymple *et al* 2007] N. C. Dalrymple, S. R. Prasad, F. M. El-Merhi, K. N. Chintapalli, "Price of isotropy in multidetector CT." 2007 Radiographics. **27**(1):49-62.

[Fang *et al* 2015] Q. Fang , F. Chen, A. Jiang, Y. Huang and X. Deng, *"Computed tomographic angiography of the superficial cerebral venous anastomosis based on volume rendering, multi-planar reconstruction, and integral imaging display."* 2015 Australas Phys Eng Sci Med.

[Guggenberger *et al* 2013] R. Guggenberger, S. Winklhofer, J. V. Spiczak, G. Andreisek and H. Alkadhi, *" In vitro high-resolution flat-panel computed tomographic arthrography for artificial cartilage defect detection: comparison with multidetector computed tomography."* 2013 Invest Radiol. **48(8)**:614-21.

[ICRU 54] International Commission on Radiation Units and Measurements, "*Medical Imaging: The Assessment of Image Quality*." 1996 Report 54 ICRU Publications, Bethesda, MD.

[Judy 1976] P. F. Judy, *"The line spread function and modulation transfer function of a computed tomographic scanner."* 1976 Med Phys **3**:233–6.

[Miéville *et al* 2011] F. A. Miéville, F. Gudinchet, E. Rizzo, P. Ou, F. Brunelle, F. Bochud, and F R. Verdun, *"Pediatric cardiac CT examinations: impact of the iterative reconstruction method (ASIR) on image quality – a preliminary study"* 2011 Pediatr Radiol **41**, 1154–64.

[Mieville *et al* 2012] F. A. Miéville, F. Gudinchet, F. Brunelle, F. O. Bochud and F. R. Verdun, *"Iterative reconstruction methods in two different MDCT scanners: physical metrics and 4-alternative forced-choice detectability experiments--a phantom approach."* 2012 Phys Med. **29(1)**:99-110.

[Ott *et al*, 2014] [Ott, 2014] J. G. Ott, F. Becce, P. Monnin, S. Schmidt, F.O. Bochud and F. R. Verdun, *"Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms."* 2014 Phys Med Biol **59(15)** 4047-64.

[Racine *et al* 2015] D. Racine, A. H. Ba, J. G. Ott, F. O. Bochud and F. R. Verdun, "Objective assessment of low contrast detectability in computed tomography with Channelized Hotelling Observer." 2015 Phys Med. S1120-1797(15)00997-7

[Richard *et al* 2012] S. Richard, D. B. Husarik, G. Yadava, S. N. Murphy and E. Samei, *"Towards task-based assessment of CT performance: System and object MTF across different reconstruction algorithms"* 2012 Med Phys **39** 4115-22.

[Rydberg *et al* 2007] J Rydberg, K. Sandrasegaran, R. D. Tarver, M. S. Frank, D. J. Conces, R. H. Choplin, *"Routine Isotropic Computed Tomography Scanning of Chest Value of Coronal and Sagittal Reformations."* 2007 Invest Radiol 42(1):23-8.

[Schindera *et al* 2011] S. T. Schindera, L. Diedrichsen, H. C. Müller, O. Rusch, D. Marin, B. Schmidt, R. Raupach, P. Vock and Z. Szucs-Farkas, *"Iterative reconstruction algorithm for abdominal multidetector CT at different tube voltages: assessment of diagnostic accuracy, image quality, and radiation dose in a phantom study."* 2011 Radiology **260(2)**:454-62.

[Singla Long *et al* 2010] S. S. Long, P. T. Johnson, Karen M. Horton, and Elliot K. Fishman, *"Are Multiplanar Reconstructions Necessary in Routine Body Computed Tomography Practice?: What Is the Published Evidence?"* 2010 J Comput Assist Tomogr **34**: 689-98.

[Thibault *et al* 2007] [Thibault, 2007] J. B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, *"A three-dimensional statistical approach to improved image quality for multislice helical CT"* 2007 Med Phys **34**, 4526–4544.

[Von Falck *et al* 2011] C. von Falck, P. Hollmann, T. Rodt, S. Waldeck, B. Meyer, F. Wacker and H. O. Shin, *"Influence of multiplanar reformations on low-contrast performance in thin-collimated multidetector computed tomography."* 2011 Invest Radiol. **46(10)**:632-8

Physica Medica 31 (2015) 823–843

Contents lists available at ScienceDirect

Physica Medica

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

Review Paper

Image quality in CT: From physical measurements to model observers

F.R. Verdun [a,*,1], D. Racine [a,1], J.G. Ott [a], M.J. Tapiovaara [b], P. Toroi [b], F.O. Bochud [a], W.J.H. Veldkamp [c,d], A. Schegerer [e], R.W. Bouwman [d], I. Hernandez Giron [c], N.W. Marshall [f], S. Edyvean [g]

[a] Institute of Radiation Physics, Lausanne University Hospital, 1 Rue du Grand-Pré, 1007 Lausanne, Switzerland
[b] STUK-Radiation and Nuclear Safety Authority, PO Box 14, FIN-00881 Helsinki, Finland
[c] Department of Radiology, Leiden University Medical Center, C2-S, PO Box 9600, 2300RC Leiden, The Netherlands
[d] Dutch reference Centre for Screening (LRCB), Radboud University Medical Centre, PO Box 6873, 6503 GJ Nijmegen, The Netherlands
[e] Department for Radiation Protection and Health External and Internal Dosimetry, Biokinetics, Ingolstädter Landstr, 1, 85764 Neuherberg, Germany
[f] Katholieke Universiteit Leuven, Oude Markt 13, 3000 Leuven, Belgium
[g] Medical Dosimetry Group, Centre for Radiation Chemicals and Environmental Hazards, Public Health England, Didcot, UK

ARTICLE INFO

Article history:
Received 8 May 2015
Received in revised form 4 August 2015
Accepted 23 August 2015
Available online 12 October 2015

Keywords:
Computed tomography
Image quality
Patient dose optimisation
Model observer

ABSTRACT

Evaluation of image quality (IQ) in Computed Tomography (CT) is important to ensure that diagnostic questions are correctly answered, whilst keeping radiation dose to the patient as low as is reasonably possible. The assessment of individual aspects of IQ is already a key component of routine quality control of medical x-ray devices. These values together with standard dose indicators can be used to give rise to 'figures of merit' (FOM) to characterise the dose efficiency of the CT scanners operating in certain modes. The demand for clinically relevant IQ characterisation has naturally increased with the development of CT technology (detectors efficiency, image reconstruction and processing), resulting in the adaptation and evolution of assessment methods. The purpose of this review is to present the spectrum of various methods that have been used to characterise image quality in CT: from objective measurements of physical parameters to clinically task-based approaches (i.e. model observer (MO) approach) including pure human observer approach. When combined together with a dose indicator, a generalised dose efficiency index can be explored in a framework of system and patient dose optimisation. We will focus on the IQ methodologies that are required for dealing with standard reconstruction, but also for iterative reconstruction algorithms. With this concept the previously used FOM will be presented with a proposal to update them in order to make them relevant and up to date with technological progress. The MO that objectively assesses IQ for clinically relevant tasks represents the most promising method in terms of radiologist sensitivity performance and therefore of most relevance in the clinical environment.

© 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

Diagnostic x-rays contribute to nearly 50% of the total annual collective effective dose of radiations from man-made and natural sources to the general population in western countries; computed tomography (CT) is the largest single source of this medical exposure.

The contribution of CT to collective dose has significantly increased in recent years and a considerable effort is required to control this trend and ensure that the benefits from the use of this technology outweigh the risks [1]. For example, in 2007–2008 the average dose per inhabitant, due to CT, was about 0.8 mSv in France and Switzerland, and about 0.7 mSv in Germany (as part of an average for all x-ray imaging of about 1.2 mSv and 1.7 mSv, respectively) [2–4]. An update of the French and German data showed that in 2012 the contribution of CT exposure had increased to approximately 1.15 mSv, with a similar increase shown in the last Swiss survey performed for 2013 [5].

In this context the radiation protection requirements in diagnostic radiology (justification of the examination and optimisation of the imaging protocol) need to be re-enforced. Justifying a CT scan is a clinical consideration and therefore will not be addressed in this work. However, the optimisation of a CT examination is achieved when image quality enables the clinical question to be answered whilst keeping patient radiation dose as low as reasonably possible. For this purpose the clinical question needs to be formulated as concretely as possible to enable a clear description of the image quality level required. To achieve this, appropriate and clinically

---

[*] Corresponding author. Institute of Radiation Physics, Lausanne University Hospital, 1 Rue du Grand-Pré, 1007 Lausanne, Switzerland. Tel.: +41 21 314 82 50; fax: +41 21 314 8299.

E-mail address: Francis.Verdun@chuv.ch (F.R. Verdun).

[1] Both authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ejmp.2015.08.007

1120-1797/© 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/).

relevant image quality parameters and radiation dose indices must be defined, described, and used. This paper concentrates on image quality parameters.

The first step of the optimisation process should ensure that x-ray conversion into image information is performed as efficiently as possible. In projection radiology such as radiology or mammography one can use the DQE (Detective Quantum Efficiency as described in IEC 62220-1/2) as a global figure of merit. Unfortunately, due to the geometry and data processing required for CT, the use of such a quantity is not feasible. In general, one will assess the amount of radiation required to achieve a certain level of image quality. As a surrogate of the radiation received by the detector one uses the standardised CT dose index ($\mathrm{CTDI_{vol}}$). This quantity represents the average dose delivered in PMMA phantoms of 16 and 32 cm in diameter and is related to the amount of noise present in an image. According to its definition $\mathrm{CTDI_{vol}}$ is different from the actual average dose delivered in a slice of a patient, and the latter should be estimated using the Size Specific Dose Estimator (SSDE) proposed by the AAPM (American Association of Physics in Medicine) [6]. For a given $\mathrm{CTDI_{vol}}$ level, image quality parameters are generally assessed using the signal detection theory that considers the imaging system linear and shift invariant.

The next step of the optimisation process should be done with the clinical applications in mind. Direct determination of clinical performance is, however, difficult, expensive, and time-consuming. Furthermore, the results in these studies can be strongly dependent on the patient sample and on the radiologists involved. As an alternative, one can assess image quality using task-oriented image quality criteria. They will necessarily be simplistic in comparison to the clinical situations but make it possible to predict the perception of simple structures within an image. The phantoms available for this type of study remain quite simple whilst trying to mimic important disease-related structures in actual patients. It is likely that 3D printing techniques will improve phantom and task realism in the future [7–9]. To seek optimisation, task-oriented image quality metrics could be studied as a function of $\mathrm{CTDI_{vol}}$ or SSDE. Figure 1 summarises this optimisation process.

Part 1 of this review focuses on signal detection theory and summarises the methods used to assess image quality in an objective way. When CT images are reconstructed using the standard filtered back-projection (FBP), these methods are commonly used to characterise a CT unit. The objective image quality metrics assess separate aspects of the features of the image, and therefore need to be combined to give an overall representation of the image quality.

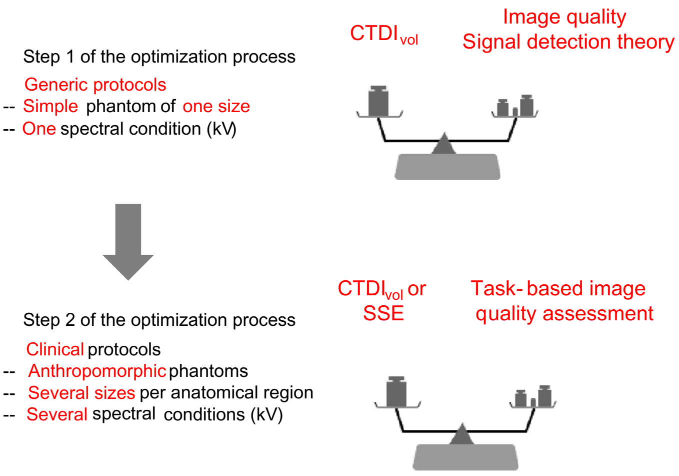

Step 1 of the optimization process
Generic protocols
-- Simple phantom of one size
-- One spectral condition (kV)
CTDI vol
Image quality
Signal detection theory
Step 2 of the optimization process
Clinical protocols
-- Anthropomorphic phantoms
-- Several sizes per anatomical region
-- Several spectral conditions (kV)
CTDIvol or SSE
Task- based image quality assessment

Figure 1. CT optimisation process in two steps: generic acquisition optimisation and clinical protocol optimisation.

To synthesise the information, and balance image quality with radiation doses, several figures of merit have been developed by combining image quality parameters such as the standard deviation in a region of interest (ROI) and the modulation transfer function (MTF). They were applied for specific clinical protocols to enable appropriate comparison of systems. This approach was quite useful during the development of CT technology, where performances between different units could vary drastically. These figures of merit can be based on simplified assumptions requiring caution in their interpretation. However it appears that the sensitivity of such methods is quite limited for newer systems, and, in addition, the effect of iterative reconstruction on the standard image quality parameters would mean that this approach would be difficult to implement.

Both clinical and phantom images can be assessed using the ROC paradigm or one of its derivatives (Localisation ROC, Free-response ROC). These methods give an accurate estimate of clinical image quality but, although carefully controlled measurements, they are still subjective because human observers are involved. These methods are time consuming and require large samples to obtain precise results. In spite of these limitations these methods can be used either by radiologists (when dealing with clinical images) or naïve observers when dealing with phantom images. To avoid the burden associated with ROC methods more simplified methods have been developed; for example, VGA (Visual Grading Analysis) in which image quality criteria can be used to give a relatively quick image quality assessment, without the explicit need for pathology or a task. Alternatively, phantom images can be assessed using the 2-AFC (two-alternative forced-choice) or M-AFC (multiple-alternative forced-choice) methods. Part 2 of this review discusses these methodologies, and these methods are used to validate the results produced by model observers presented in Part 3.

The introduction of iterative reconstruction in CT poses a new challenge in image quality assessment since most of the standard metrics presented in Part 1 cannot be used directly. In order to establish a bridge between radiologists and medical physicists, and therefore between clinical and physical image qualities, task related metrics can be used (even if the tasks are simplified versions of actual clinical tasks). Mathematical model observers are particularly suited to the routine image quality measurement of clinical protocols, with the results indicated to the user together with the standard dose report. Part 3 summarises the concepts behind these model observers, focusing on the anthropomorphic model observers that mimic human detection of simple targets in images, since the aim is to present tools for practical applications. The theory and description of the ideal observer can be found in the literature and a brief introduction to this model is done at the beginning of Part 3. Note that model observers can also be used when images are reconstructed with FBP. The inconvenience associated with the use of model observers is that they all lead to an overall outcome without the separation of the image quality parameters as with signal detection theory.

This paper is structured into three separate sections that provide an overview of the most common approaches taken when dealing with image quality in CT imaging. This structure is described in Fig. 2.

Traditional objective metrics

CT is a 3D imaging technique in which image quality assessment must be approached with some caution. Objective assessment of parameters that influence image quality is often made using physical metrics specified in either the spatial or spatial frequency domain. This duality is due to the fact that some features will produce overall responses which are independent of the location in the image, whereas other features will produce responses that are spatially correlated.

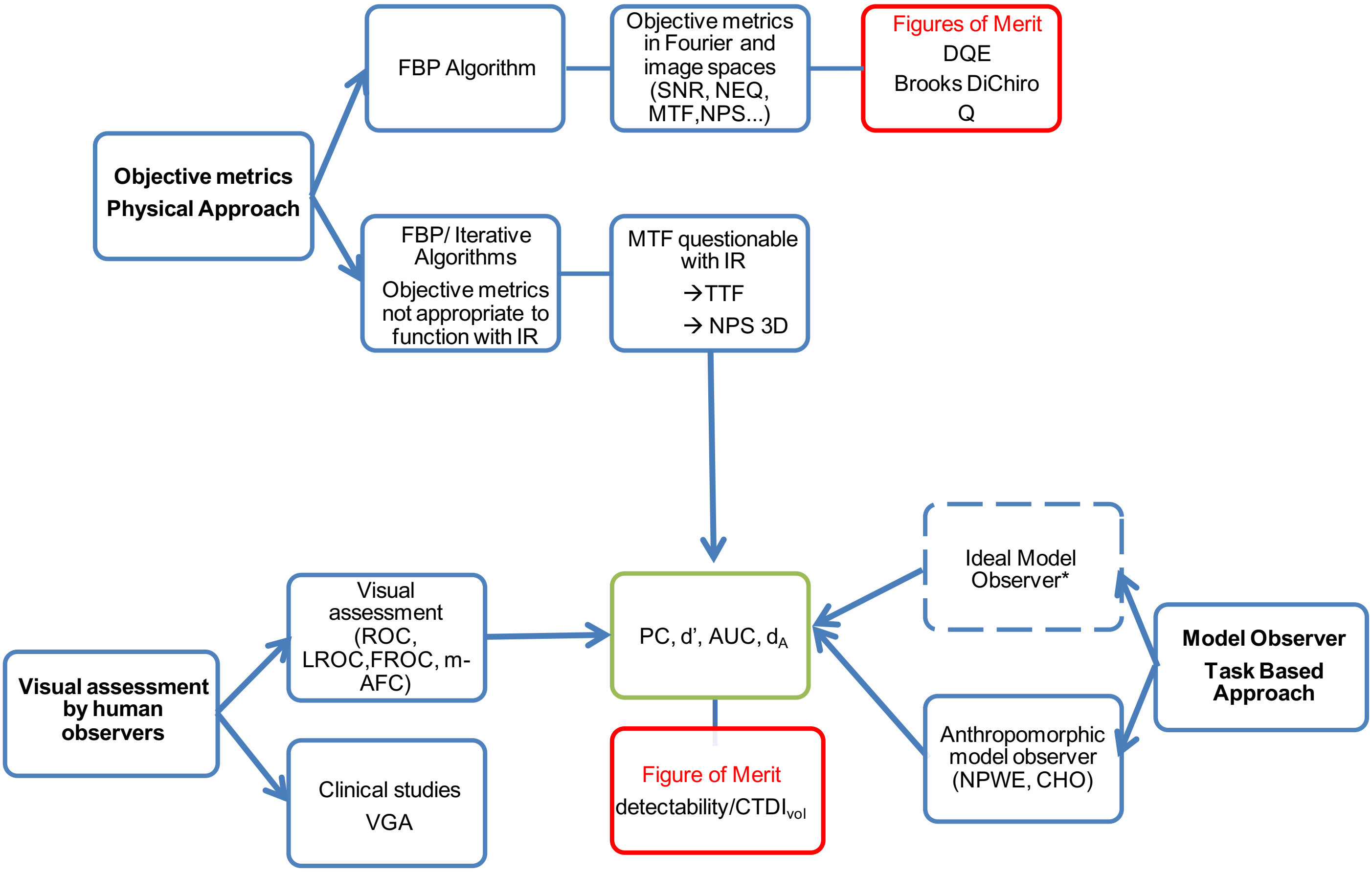

Objective metrics
Physical Approach
FBP Algorithm
Objective metrics in Fourier and image spaces (SNR, NEQ, MTF,NPS…)
Figures of Merit
DQE
Brooks DiChiro
Q
FBP/ Iterative Algorithms
Objective metrics not appropriate to function with IR
MTF questionable with IR
→TTF
→ NPS 3D
Ideal Model Observer*
Model Observer
Task Based Approach
Visual assessment by human observers
Visual assessment (ROC, LROC,FROC, m-AFC)
PC, d', AUC, d_A
Anthropomorphic model observer (NPWE, CHO)
Clinical studies
VGA
Figure of Merit
detectability/CTDI_vol

Figure 2. Summary of the content of the review (*this part will not be presented).

Objective metrics in spatial domain

Image signal and image noise are key parameters in image quality assessment. In the ideal and linear case, image signal (S) is directly linked to the detected number of photons N, whilst the noise (σ) may be seen as the pixel's stochastic fluctuation around their mean value. The photons are distributed according to Poisson's law, meaning that the quantity σ is equal to $\sqrt{N}$. The ratio of these two quantities yields the signal-to-noise ratio (SNR), expressed as:

$$SNR \propto \frac{S}{\sigma} = \frac{N}{\sqrt{N}} = \sqrt{N} \tag{1}$$

In an ideal device, each quantum could be counted by the detector and contributes towards the image. We could thus transpose Eq. (1) as:

$$SNR_{Ideal} \propto \frac{N_{Ideal}}{\sqrt{N_{Ideal}}} = \sqrt{N_{Ideal}} \tag{2}$$

However, due to the properties of the detector and its limited efficiency, a real measurement of the SNR would give the following result:

$$SNR_{Real} = \frac{N_{Real}}{\sqrt{N_{Real}}} = \sqrt{N_{Real}} < \sqrt{N_{Ideal}} \tag{3}$$

In Eq. (3), N_{Real} gives the number of quanta that contribute to the image for the real device and is also called noise-equivalent quanta (NEQ). Thus:

$$SNR^2{}_{Real} = N_{Real} = NEQ \tag{4}$$

Based on those parameters, we can eventually estimate the efficiency of a device by making the ratio between the number of photons actually used for the imaging and the incoming number of photons to the detector. This quantity is called detective quantum efficiency (DQE) and is defined as:

$$DQE = \frac{SNR^2_{Real}}{SNR^2_{Ideal}} = \frac{NEQ}{N_{Ideal}} \tag{5}$$

In Eq. (5), the NEQ can be measured in a straightforward manner, but some care must be taken when estimating quantity SNR^2_{Ideal}. Indeed, when considering a monochromatic beam, SNR^2_{Ideal} is simply the number of photons produced. However, for a polychromatic beam, SNR^2_{Ideal} should be the summed variance of the number of photons in each energy bin. In fact, some authors prefer to use an energy weighted variance because most detectors integrate energy [10] to form an image.

Another commonly used global image quality index is the signal difference-to-noise ratio ($SDNR$), defined for an object as the intensity difference from the background divided by the standard deviation:

$$SDNR = \frac{I_{object} - I_{Background}}{\sigma} \tag{6}$$

These metrics are extended to the spatial frequency domain in the following section.

Objective metrics in Fourier domain

Spatial resolution can be defined as the ability to distinguish two separate objects and is directly linked to the pixel size, the reconstruction kernel as well as the hardware properties of the imaging device. In order to derive an expression for image resolution, it is necessary to describe the imaging process generating a CT slice. Our analysis will be restricted to the axial plane. $I(x,y)$, which is the image slice of an input object denoted by $f(x,y)$, can be mathematically expressed as:

$$I(x,y)=\iint f(x-x',y-y')PSF(x',y')dx'dy' \tag{7}$$

with $PSF(x,y)$ being the point spread function in the axial plane and describing resolution properties of the device. It corresponds to the impulse response of a system, the response of the system to a Dirac input ($\delta(x,y)$).

Resolution can also be estimated through the line spread function (*LSF*), which is the response of the system to a straight line. Thus, the relationship between the *LSF* and the *PSF* can be derived from Eq. (7) in which the input function is replaced by the equation of a straight line in the axial plane (that is to say replacing $f(x,y)$ by $\delta(x)$ in Eq. (7)), yielding:

$$LSF(x)=\iint \delta(x-x')PSF(x',y')dx'dy'$$

leading to:

$$LSF(x)=\int_{-\infty}^{+\infty} PSF(x,y)dy \tag{8}$$

The point spread function needs to be similar at each location in the image (shift invariance) in order to ensure that the *LSF* will remain the same at every localisation. However, isotropy of the axial plane is a hypothesis which is not always true, especially when dealing with CT. In this case, the *LSF* will depend on the direction of the straight line in the axial plane. Assuming the straight line is positioned tilted with an angle θ the expression of the *LSF* will become:

$$LSF_\theta(x,y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} PSF(x',y')\delta((x-x')\cos\theta+(y-y')\sin\theta)dx'dy' \tag{9}$$

Besides those two metrics, it is also possible to estimate the resolution using the edge spread function (*ESF*), that is to say the response of the device to an edge. An edge can be mathematically approached by the Heaviside function $H(x,y)=\begin{cases}1 & if\ x>0\\ 1/2 & if\ x=0\\ 0 & if\ x<0\end{cases}$. This function has the property: $\frac{dH(x)}{dx}=\delta(x)$.

Using this property, injecting $f(x,y)=H(x)$ in Eq. (7) and using Eq. (8) we obtain:

$$LSF(x)=\frac{\partial ESF(x)}{\partial x} \tag{10}$$

Hence, *PSF*, *LSF* and *ESF* are all related to each other and it is possible to use their representation in the frequency space thanks to the Fourier transform.

The Fourier representation of the *PSF* is the optical transfer function (*OTF*), which is defined as following:

$$OTF(u,v)\overset{\text{def}}{=}FT\{PSF(x,y)\} \tag{11}$$

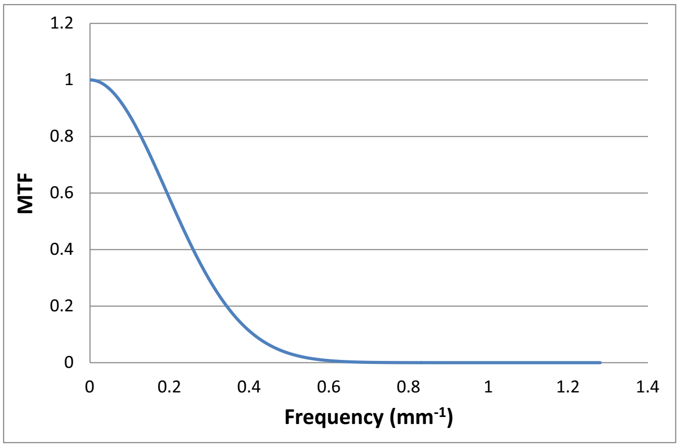


Figure 3. Example of a 1 dimension MTF curve of a GE VCT system with a 0.40 mm pixel size.

What is commonly used in order to estimate the resolution is the modulation transfer function (*MTF*), defined as the modulus of the *OTF normalised by its zero-frequency value*:

$$MTF(u,v)\overset{\text{def}}{=}\frac{|OTF(u,v)|}{|OTF(0,0)|} \tag{12}$$

Using Eqs. (8), (11) and (12) together with the Fourier slice theorem and assuming shift-invariance in the axial plane, we can state that a normalised radial *MTF* of the system is given by:

$$MTF_{1D}(f)=\left|\frac{FT\{LSF(x)\}}{\int_{-\infty}^{+\infty}LSF(x)dx}\right| \tag{13}$$

This metric describes how well frequencies are transferred through the system and is therefore used to make objective resolution estimation (Fig. 3).

Practically, the MTF can be computed from the image of a point (~PSF), a line (~LSF) or an edge (~ESF) [11–13]. In calculating MTF from the image of a point source (effectively from the PSF), a metal bead or taut wire fixed within a dedicated phantom is used to generate the signal [14]. Boone [12] used a tilted aluminium foil of thickness 50 µm to generate an oversampled *LSF*; the MTF is then computed using Eq. (13). Judy [13] was the first to describe calculation of MTF from an edge method in which the ESF was differentiated to give the LSF. This method has been developed over the years by various authors to include the use of spheres from which the oversampled ESF is built [15–17]. An older method was proposed by Droege and Morin, in which MTF is estimated from line pair test object images using the Coltman formula. Extensive details on the practical implementation of these techniques are given in ICRU Report 87 [18]. Several of these methods have been investigated by Miéville et al. in order to compare and contrast the advantages and drawbacks [19].

As with resolution, and of equal importance for SNR transfer, image noise can also be estimated in the frequency space. There are different sources of noise within the CT system, such as the electronic noise caused by the detector readout circuits (amplifiers) and the primary quantum noise which is inherent to the statistics of the limited quanta building the image. In a stationary system, the Wiener spectrum or noise power spectrum (NPS) gives a complete description of the noise by providing its amplitude over the entire frequency range of the image [20]. If the image noise is not stationary, the Wiener spectrum is not a complete description and the whole covariance matrix would be needed for complete description. However,

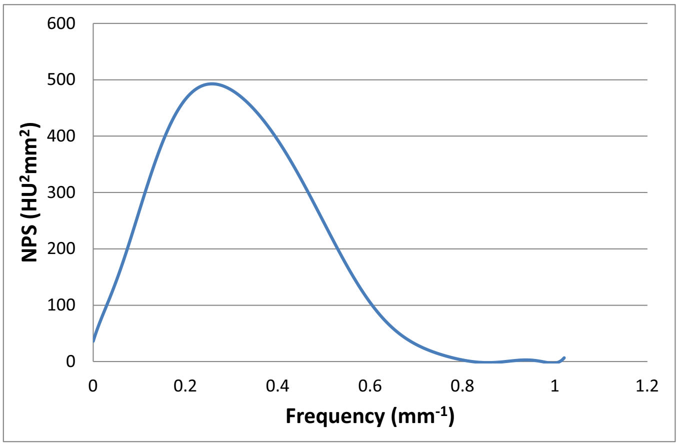


Figure 4. Example of a radially averaged NPS obtained on a GE revolution system with a standard convolution kernel.

if applied with care – for example working with small ROIs, extracted from a restricted region of the image – the NPS can be applied to both conventionally (i.e. FBP based) and iteratively reconstructed images. For NPS calculation, the assumption of 'small signal linearity' has to be made in order to apply Fourier analysis, which requires system linearity in order to be valid. This is the case for the logarithmic step applied to all reconstruction processes and also to the explicitly non-linear iterative methods.

In order to compute the *NPS* of an image, it is necessary to acquire homogeneous CT images and select region of interests (ROI) in this stack. The 2D *NPS* can then be computed as:

$$NPS_{2D}(f_x, f_y) = \frac{\Delta_x \Delta_y}{L_x L_y} \frac{1}{N_{ROI}} \sum_{i=1}^{N_{ROI}} \left| FT_{2D}\left\{ ROI_i(x, y) - \overline{ROI_i} \right\} \right|^2 \tag{14}$$

where Δ_x, Δ_y are the pixel sizes in the x and y dimension, $L_x L_y$ are the ROI's lengths (in pixel) for both dimensions, N_{ROI} is the number of ROIs used in the average operation and $\overline{ROI_i}$ is the mean pixel value of the ith ROI.

In practice, the *NPS* is largely affected by the detector dose, the hardware properties and the reconstruction kernel and algorithm. From each image of the stack a ROI is extracted and a custom computer program is generally used to compute the *NPS* according to Eq. (14). It is of common use to average the 2D *NPS* along a 1D radial frequency using the equation $f_r = \sqrt{f_x^2 + f_y^2}$ (Fig. 4). More details on the *NPS* computing can be found in ICRU Report 87 [18]. In the end, the NPS characterises the noise texture, thus giving a better and more complete description of noise than the simple pixel's standard deviation. Moreover, information about the pixel's standard deviation can still be retrieved with knowledge of the Wiener spectrum. Indeed, the Parseval theorem ensures that the total energy is obtained by summing the contribution of the different harmonics and that its value does not depend on the chosen space (image or frequency space). Since the *NPS* is a spectral decomposition of noise over frequencies, we have:

$$\sigma^2 = \iint NPS_{2D}(f_x, f_y)\, df_x df_y \tag{15}$$

As explained before, *MTF* shows how well the signal frequencies are transferred through an imaging system, that is to say it exhibits the signal response of a system at a given spatial frequency. As for the spatial domain, the ratio of signal (i.e. MTF) and noise (i.e. NPS) yields the output signal to noise ratio (the NEQ) and therefore the frequency dependent *NEQ* can be calculated as:

$$NEQ(f) = SNR^2_{Real}(f) = \frac{a^2 MTF^2_{1D}(f)}{NPS_{1D}(f)} \tag{16}$$

where a^2 is the mean pixel value squared.

The DQE in the frequency space can therefore be estimated by:

$$DQE(f) = \frac{SNR^2_{Real}(f)}{SNR^2_{Ideal}(f)} = \frac{a^2 MTF^2_{1D}(f)}{N_{Ideal} NPS_{1D}(f)} \tag{17}$$

Limitations of conventional and Fourier-based image quality metrics for the assessment of IR images

In order to compute an MTF that represents the spatial resolution of the entire image, the assumption of shift-invariance has to be made. That is to say that the device's response has to remain the same, whether measured at the image centre or periphery. If this assumption is not fulfilled it is necessary to make the measurements at the same location in different images to obtain an *MTF* that can be used to compare the resolution of different devices. Furthermore, the linearity hypothesis also needs to be fulfilled for the *MTF* to be reliable. That is to say, the output signal has to remain within the optimal range of response of the imaging system in terms of Hounsfield units (HU), usually in the range from −200 to +200 HU for clinical CT scanners [18]. Consequently, estimating the *MTF* with a high Z material can give a signal outside this range, yielding an incorrect assessment of resolution. In practice, estimating the *MTF* with high Z materials generally leads to a resolution overestimation because of the high SNR they generate [18].

Those two assumptions are approximately satisfied for CT images reconstructed with filtered back projection (FBP) algorithms and a standard reconstruction kernel, but the introduction of iterative reconstruction (IR) has changed the game [21]. Indeed, IR images exhibit stronger non-linear and non-stationary properties that force a change in the MTF measurement paradigm. Several authors have already highlighted the non-linearity problem of these algorithms, which manifests itself as contrast dependency of the resolution [21–23]. Also, investigations on how Fourier-based metrics are influenced by the characteristics of IR images have been described [24,25]. They showed, for example, that the shape of the NPS for some IR algorithms also depends on the dose level and that the resolution not only depends on the contrast but also on the radiation dose levels. These elements have highlighted the need to adapt the existing metrics to IR algorithms.

Adaption of Fourier metrics

These difficulties in estimating resolution can be overcome by using an adapted metric, such as the target transfer function (*TTF*), which makes it possible to characterise the resolution even in the presence of noise and contrast dependency [24,26]. *MTF* and *TTF* are similar but differ from one another in the sense that *MTF* only applies to a single given contrast level, whereas a *TTF* will exhibit three different curves at three different contrasts (corresponding to three different materials) for one measurement (Fig. 5). This enables a characterisation of the resolution when dealing with non linear algorithms for which contrast influences the resolution. As already demonstrated by several authors this will make full characterisation of the resolution possible when dealing with IR [24,27].

The technological evolution of CTs has also led to changes in the way *NPS* must be computed. The 2D axial *NPS* was well suited for the first generations of devices where only one CT image per axial scan could be acquired without noise correlation between slices. Now that the acquisitions are also made in helical mode and that the number of detectors along the z-axis is higher, a 3D NPS is required to fully characterise the noise (Fig. 6) [12,28]. 3D NPS can

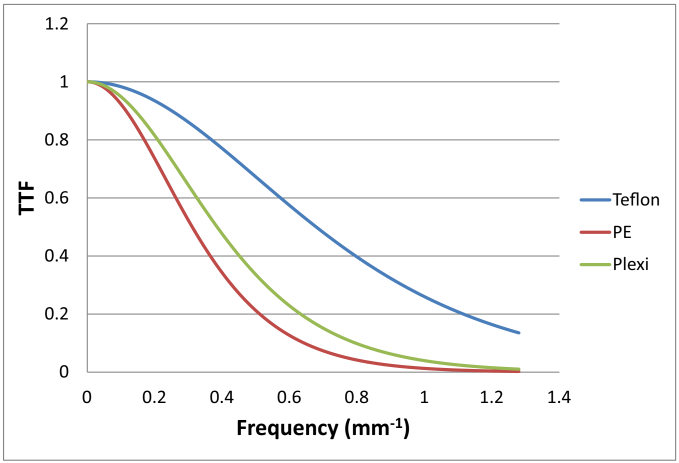


Figure 5. Resolution estimation through the *TTF* on a GE HD 750 system with a 0.4 mm pixel size and three different materials (Teflon polyethylene and plexiglass). Differences are observed on the resolution depending on the material. Such changes could not be observed when using the *MTF*.

be measured in a similar manner to the 2D NPS, but working with volumes of interests (VOI) instead of ROIs:

$$NPS_{3D}(f_x, f_y, f_z)=\frac{\Delta_x\Delta_y\Delta_z}{L_xL_yL_z}\frac{1}{N_{VOI}}\sum_{i=1}^{N_{VOI}}\left|FT_{3D}\left\{VOI_i(x,y,z)-\overline{VOI_i}\right\}\right|^2 \quad (18)$$

For this case, the units of NPS are HU^{2}mm^3.

In this particular paradigm, Eq. (15) becomes:

$$\sigma^2=\iiint NPS_{3D}(f_x, f_y, f_z)\,df_x df_y df_z \quad (19)$$

How to synthesise the information towards a figure of merit

Combining image quality and dose

In the clinical setting the focus for optimisation is balancing image quality and radiation dose in the context of the clinical question. Statistical noise, spatial resolution and imaged slice width are the fundamental parameters which describe the amount of object information retrievable from an image, and give rise to the perceived image quality. X-ray dose can be regarded as the cost of this information. It is meaningless to quote any of these image quality quantities without reference to the others, or to the radiation cost. The ‘holy grail’ is to try to find a way to combine the relevant parameters objectively and appropriately in a dose efficiency factor.

A dose efficiency factor, or figure of merit, can take a number of forms depending on how the various parameters are measured and quoted. Correctly developed and applied it can be used as a tool to compare scanner models, or simply different scan settings to optimise the balance of image quality and radiation dose.

How these parameters, resolution and noise in particular, are balanced is dependent on the clinical question and examination type. An important aspect that must be addressed is the influence of scan and protocol parameters that can be adjusted by the operator and how they affect image quality and radiation dose performance.

Clinical scanner settings – scan and protocol parameters

Any consideration of a theoretical approach to investigate a dose efficiency value needs to be in the scenario of the clinical question and the parameters used to create the image (Fig. 1).

Image quality and dose can be affected by the scanner design and also by the scan settings in the selected protocol (Table 1). The effect of the scan parameters, which form the examination protocol for the clinical question, can be seen in Table 2.

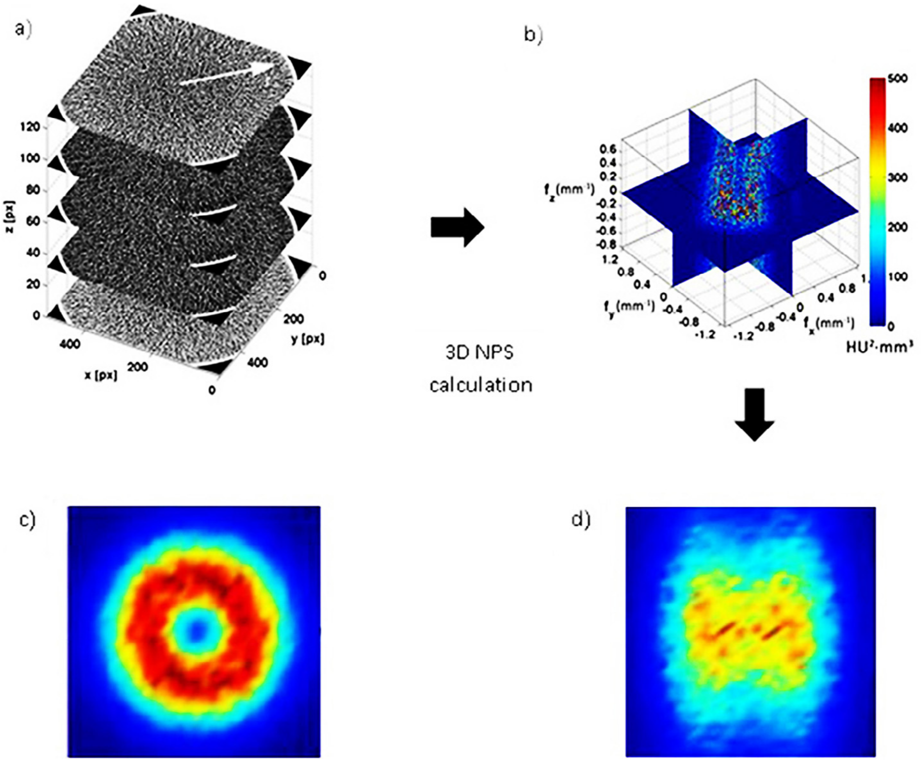


Figure 6. (a) The 3D homogeneous volume from which the 3D NPS is extracted. (b) The 3D NPS and the NPS sectioned in the (c) x–y (axial) and (d) the x–z (sagittal) planes. Figures extracted from Reference 22.

Table 1
Scanner design and scanner settings which can affect image quality and dose on scanner settings (courtesy of ImPACT [29]).

Scanner design factors	Scan protocol factors
Detectors material	Clinical application
Detector configuration	Tube current, tube voltage, focal spot size
Numbers of detectors, rows	Image reconstruction algorithms
Data acquisition rates	X ray Collimation width, detector acquisition width
Software corrections	
Filtration	Reconstructed image slice thickness
Focal spot size	Helical pitch
Geometry (i.e. focus–axis, focus–detector distances)	Interpolation algorithms

Combining image quality and dose metrics – theoretical background

The basic starting premise for a figure of merit for a dose efficiency parameter is that a dose efficient scanner will produce good resolution at minimum dose and noise.

There are a number of mathematical relationships that can be found in the scientific literature, both in terms of general imaging theory and for CT in particular [30–32]. The two of interest for CT are Brooks and Di Chiro [33] and Riederer et al. [34]. These were used in the development of the ImPACT Q value which became a useful, and relatively widely known, approach for comparing CT systems in the 1990s [35,36]. It was also explored by Fuchs and Kalender [37], more recently Kalender devoted a section to this subject in his book Computed Tomography: Fundamentals, System Technology, Image Quality, Applications [37,38]. However the fundamental relationship can also be found in standard textbooks on imaging with radiation [39,40]. The core of all these approaches is that the noise squared is inversely proportional to dose, and also inversely (in real or image space) proportional to the spatial resolution to the power 4. This encompasses spatial resolution in the x,y (to power 3) plane and also the z plane and quoted either as a size or frequency. In some equations the resolution is separated out into frequency for the x and y plane resolution, and the image thickness for the z-axis (z,x and z,y planes).

The relationship can be explored in more detail using the Brooks and Di Chiro equation [33]:

$$\sigma^2(\mu) = \frac{\pi^2 \beta \gamma(E) e^{\alpha} \mu_{en} E}{1200 \omega^3 h D} \tag{20}$$

Here σ^2 is the statistical error in the reconstructed image (i.e. the image noise); β is a beam spreading factor (non-parallel rays), $\gamma(E)$ is the average depth dose factor for photon energy (E), e^{α} is the logarithmic attenuation, μ_{en} is the energy absorption coefficient, E is the photon energy, ω is the detector aperture, h is the slice width, and D is the radiation dose.

For the purposes in this chapter, this can be simplified to:

$$\sigma^2 \propto \frac{1}{\omega^3 h D} \tag{21}$$

Similarly the Reiderer, Pelc and Chesler relationship is given as [33]:

$$\sigma^2 = \frac{\pi}{m N_p} \int_0^{2\pi} d\phi \int_0^{\infty} k dk \frac{|G(k)|^2}{k} \tag{22}$$

where m is the number of projections, N_p is the number of photons per projection, and $G(k)$ is the convolution function with frequency. The product mN_p could be regarded as a measure of radiation dose.

This essentially becomes $\sigma^2 \propto k_c{}^3/mN_p$ (where k_c is the cut-off frequency, i.e. the limiting resolution). Or, indeed as the paper states; 'for all valid correction filters … σ^2 varies with the cube of the resolution'.

This is, in effect, the relationship of:

$$\sigma^2 \propto \frac{1}{N} \quad \text{i.e.} \quad \sigma^2 \propto \frac{1}{D} \tag{23}$$

where N is the number of photons and D is a measure of radiation dose for a fixed value of tube voltage. This can also be seen as a direct result from Eq. (1), assuming Poisson noise and without additive electronics noise.

Combining image quality and dose metrics – a practical approach

The discussion that follows is the approach taken by the UK CT scanner evaluation facility ImPACT [36]. It is a pragmatic solution to a complicated scenario of practical and computational effects on resultant image quality and dose for the operational CT scanner. This approach was reasonably successful for a number of years, and many scanner comparison reports were produced using this factor [32]. There is no other known work in this area covering a number of decades of scanner development. All measurements were undertaken according to a procedure with strict criteria, and in consultation with manufacturers as to the nature of their scan protocols, scanner features and reconstruction parameters. Measurements and analysis were carried out using typical clinical protocols, using the same image quality and dose assessment and calculation methods, and the same team of people. As scanners developed it became harder to apply such strict criteria, and with the development of adaptive filtration, and iterative reconstruction methods, it became very difficult to minimise the effects of other variables on ascertaining a dose efficiency value for a typical scan protocol.

$$Q \alpha \sqrt{\frac{f^3}{\sigma^2 z D}} \tag{24}$$

Table 2
Dependence of image quality and dose parameters on scanner settings (courtesy of ImPACT, adapted from Reference 29).

	Noise	Slice width	Scan plane resolution	Dose
kV	■			■
Effective mA (mA/pitch)	■			■
Focal spot selection			■	
Pitch	░	■		░
X-ray beam collimation	■			■
Detector configuration (e.g. 16 × 1.25 versus 32 × 0.62)	■	■		■
Scan time (for a given mAs)			■	
Interpolation algorithm	■		░	░
Convolution kernel	■		■	■
Reconstructed slice thickness	■	■		■
Use of iterative reconstruction	■	░	■	■

The dark blue represents a major dependence of image quality and dose on scanner settings and the soft blue represents a minor dependence.

where σ is the image noise, f is a measure of the in-plane spatial resolution (in frequency space), z is a measure of the spatial resolution along the z-axis (in image space, and a measure of the z-sensitivity), and D, as indicated above, a is measure of the radiation dose. This is the approach used by the ImPACT CT scanner evaluation facility [32,36] and first proposed in 1978 by Atkinson [35]. Initially one form of the generic equation was used, and then altered some of the definitions of the parameters involved, to create what became known as Q2 [31,41] as shown in Eq. (25).

The Q-factor (Q_2 factor) is in part empirical, it was used with caution and with strict adherence to the calculation procedure, which included standardising certain scan and protocol variables. Since it is not an absolute figure, it cannot be applied to the overall scanner, only to the examination protocol. Each set of image quality and dose parameters was therefore focussed on a typical clinical type of examination; for example a standard brain or standard abdomen.

The first step in the process was to ascertain this scan protocol in conjunction with the manufacturer. Consideration of the effects of the scanner settings, as shown in Table 1, required some adjustment of the protocol. This was in order to minimise the effects of scan parameters whose effects confounded the aim of comparison of image quality and dose, in the context of dose efficiency of the system. The associated challenge was to maintain the integrity of the suggested protocol for that type of examination. The second step was to undertake the various image quality and dose measurements and calculations, and then finally to apply the Q2 relationship.

$$Q_2 = \sqrt{\frac{f_{av}^3}{\sigma^2 z_1 CTDI_{vol}}} \tag{25}$$

The specific parameters used in calculating this value were measured using standard techniques and quoted parameters, such as would be used for quality control or acceptance testing:

σ = the image noise, the standard deviation from the CT numbers of a specified sized region of interest (5 cm^2), expressed as a percentage (for water, standard deviation in HU divided by 10), measured at the centre of the field of view in a standard water phantom.
f_{av} = spatial resolution, given as $(MTF_{50} + MTF_{10})/2$, where MTF_{50} and MTF_{10} are the spatial frequencies corresponding to the 50% and 10% modulation transfer function values respectively (in line pairs per cm).
z_1 = the full width at half maximum (FWHM), (mm), of the imaged slice profile (z-sensitivity). This is measured using the inclined high contrast plates method (mm).
$CTDI_{vol}$ = volume weighted CT dose index (mGy).

To understand the dose efficiency relationship further in a practical manner, it can be helpful to consider the basic equation (Eq. 24) to be formed of three components:

$$\sigma^2 \propto \frac{1}{D}, \quad \sigma^2 \propto \frac{1}{z} \quad \text{and} \quad \sigma^2 \propto f^3 \tag{26}$$

which, in the Q_2 relationship, translate to:

$$\sigma^2 \propto \frac{1}{CTDI_{vol}}, \quad \sigma^2 \propto \frac{1}{z_1} \quad \text{and} \quad \sigma^2 \propto f_{av}{}^3 \tag{27}$$

Each of these relationships will be addressed more fully in the following sub-sections.

Dose value. The dose value in an earlier formulation of Q was the surface dose to a phantom, measured using thermoluminescent dosimeters. This was changed for Q_2 with the introduction of the standardised $CTDI_{vol}$ parameter. The cross-sectional averaging that contributes to the creation of the $CTDI_{vol}$ is more representative of the overall dose to the phantom and therefore a more appropriate value to be used.

The inverse relationship of dose with σ^2, $\left(\sigma^2 \propto \frac{1}{CTDI_{vol}}\right)$ has to be carefully considered with multi-slice CT beams. In CT it is generally acknowledged that the $CTDI_{vol}$ is a suitable dosimetry parameter; however the proportionality breaks down in MSCT since the penumbra contribution to the beam width is a constant value, and as such is a factor that affects the relative dose, and is not accounted for in the relationship. Therefore to accommodate this, the beam width needs to be kept as a constant when comparing one scanner to another, or to take it into account separately with a beam width correction factor.

Image slice width (z_1) – z-axis resolution. The effect on noise from the thickness of the slice (z_1) is from the imaged, as opposed to the nominal, slice width, with a dependence on the inverse proportionality of photons contributing to the image. For testing purposes the full width at half maximum (FWHM) of the imaged slice profile is a suitable parameter to use. However this does not fully describe the imaged slice profile, in terms of the photon distribution contributing to the reconstructed image. For ease of application the FWHM is used, even though a fuller description of this sensitivity profile would be better.

Spatial resolution (f_{av}). A similar approach is taken with the spatial resolution parameter. Rather than using a single value from the modulation transfer curve, a more complete description of the resolution takes into account the full function over all frequencies, and a resolution value based on an average of the 50% and the 10% values of the modulation transfer function is therefore used. These values, averaged, do not completely describe the spatial resolution function, however they are common values automatically extracted from MTF curves as part of a standard testing process, and together were deemed to provide a better indication of the compromise between high and low spatial resolutions, compared to only one of the parameters alone.

The derivation of the cubed relationship of noise with spatial resolution ($\sigma^2 \propto f_{av}{}^3$) relies on assumptions of the shape of the convolution filter used (for example in Brooks and Di Chiro [33], the convolution filter is a ramp filter). In this way comparisons between scanners are likely to be more reliable when comparing images reconstructed with similar convolution filters and, in particular, algorithms that best represent ramp filters. These are in general, the filtered back projection filters named for 'standard' applications, providing reasonably low spatial resolution in order to preserve the contrast detectability in an image. Filters that are slightly smoothed or slightly enhanced would be considered as close; however those with strong smoothing or strong edge enhancing would not be suitable. Reconstruction filters with 'standard' spatial resolution values were therefore chosen to minimise the dependency of Q_2 upon non-ramp like reconstruction filters. Fortunately, or appropriately, these were also the algorithms usually used in the standard clinical protocols under investigation. This aspect of the Q_2 equation is a pragmatic solution for the complexity of modern reconstruction algorithms. The reconstruction filter with MTF_{50} and MTF_{10} values as close as possible to 3.4 lp/cm and 6.0 lp/cm was used.

When investigating the empirical relationship with actual reconstruction filters, which range from ramp-like standard filters with conventional apodisation functions, to edge-enhancing high spatial resolution filters, it was found that the relationship was closer to a power of 4 or 5 [29,42].

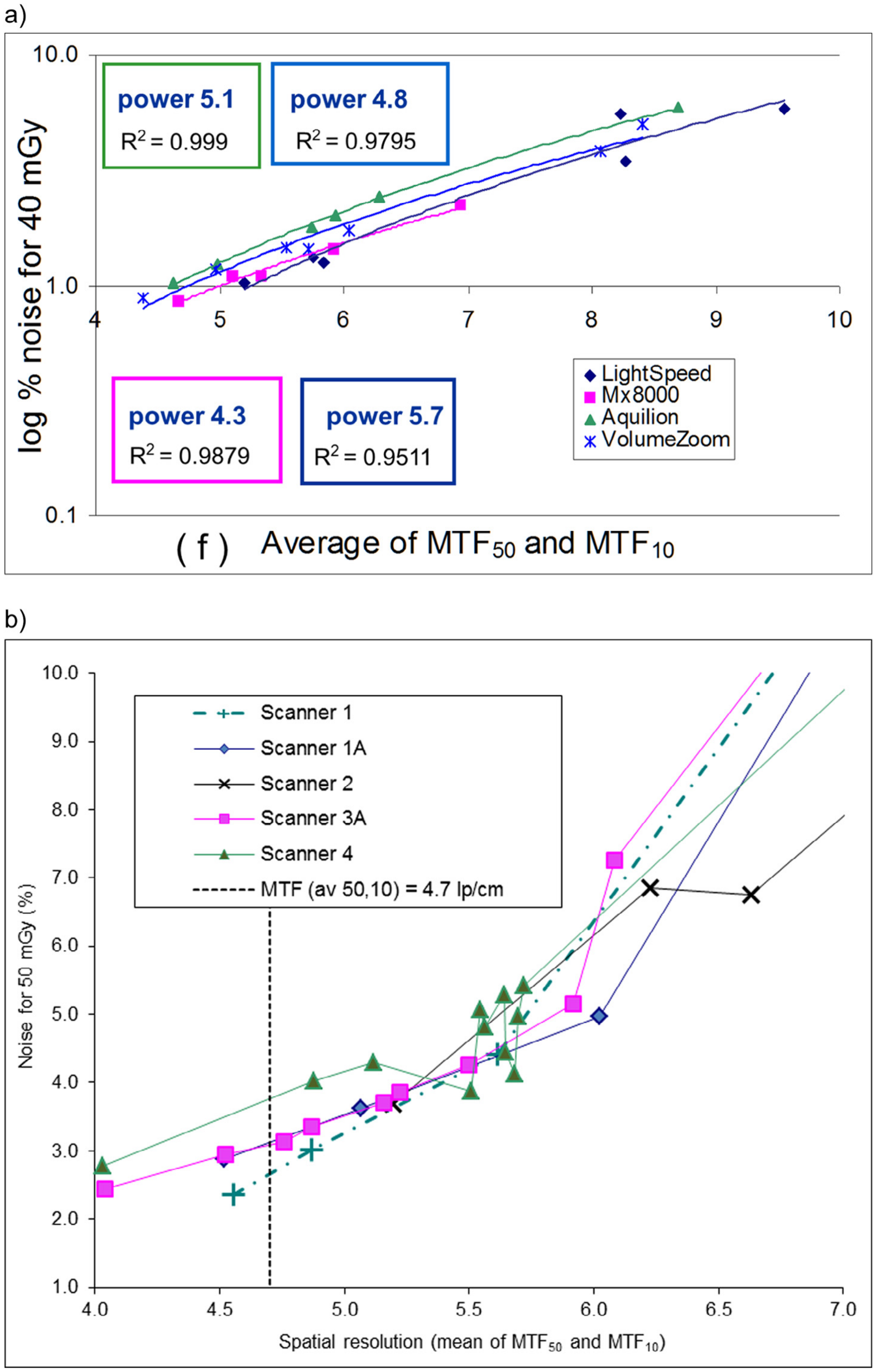


Figure 7. (a) Example for body algorithms, of logarithmic image noise against spatial resolution, with normalised dose (CTDI), demonstrating the deviation from the expected relationship. (The 'power' is the power to which f_{av} is raised against σ^2) (courtesy of ImPACT). (b) Head algorithms showing associated image noise against spatial resolution, with normalised dose (CTDI), demonstrating, particularly for scanner4, how small changes in spatial resolution give rise to large changes in measured noise [from data in Reference 41].

$$\sigma^2 \alpha \frac{f^{\approx 4-5}}{zD} \qquad (28)$$

This is illustrated in the following graph (Fig. 7a), for the body scans. The different points on the graph relate to different reconstruction algorithms. This reinforces the need to compare the 'Q' for scanners with image quality parameters measured using standard algorithms only, as the cubed power relationship is not valid across the whole range of spatial resolutions.

However, with modern scanners and reconstruction algorithms, even with a 'standard' algorithm there can be anomalies in the expected relationships. With adaptive filtration and special reconstruction techniques, even selecting the lower spatial resolution algorithms, inconsistencies in the 'straight line' relationship can

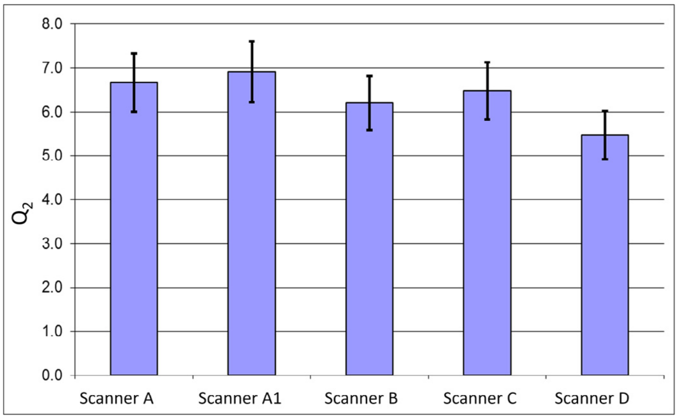


Figure 8. Q2 values for several 16-slice scanners for standard head scans (courtesy of ImPACT).

appear, where a small increase in spatial resolution may not bring the expected associated increase in image noise, as shown in Fig. 7b [41,43].

The uncertainty in the Q value was estimated to be about 15%, and therefore, even once the confounding variables are standardised, it cannot be used to look for fine differences in the image quality and dose relationship [36,41,43], as shown for a set of 16-slice scanners in Fig. 8 [41].

However, it can demonstrate larger differences – such as with the difference between the dose efficiency of xenon gas and solid state detectors. Figure 9 shows data from the original 'Q' value, where surface phantom dose measurements were used (giving the surface multiple scan average dose (MSAD)). By normalising for the spatial resolution both in the z-axis (the image slice thickness) and the scan plane, this can be shown graphically as a relative dose.

Alternative method for combining parameters. Another approach to define CT dose efficiency was suggested by Nagel [44]. This approach for image quality determination is based on a statistical method of determining low contrast detectability (LCD) as previously suggested by Chao et al. [45]. In this method, a uniform phantom is scanned with specified dose and parameter settings. An array of square regions of interest (ROIs) is defined on the uniform image that is covering approximately a third of the central image area. By measuring the distribution of mean CT numbers of the ROIs and assuming a normal distribution, a prediction can be made of the CT number threshold of a low contrast detail having the same size as the ROIs in order to detect it at a 95% confidence level. This threshold contrast C is 3.29 times the standard deviation σ. This parameter is obtained by measuring the mean CT numbers of the ROIs before calculating their standard deviations. There is a 95% probability that a low contrast object of the same size as the ROIs is missed if the contrast is within the normal variation in the ROI means, i.e. if $C < 3.29\ \sigma$. Similarly, with a probability of 5%, a randomly high fluctuation of some ROI numbers could be mistaken for an actual low contrast object if the contrast of interest is sufficiently small. According to the Nyquist theorem, the ROI size limits the noise power spectrum (NPS) at a relatively low spatial frequency (here, approximately 1 lp/cm). Therefore, a measure of the detectability of low contrast objects having the same size as the ROIs suppresses spectral noise components at high spatial frequencies that are strongly affected by the detector and reconstruction algorithm.

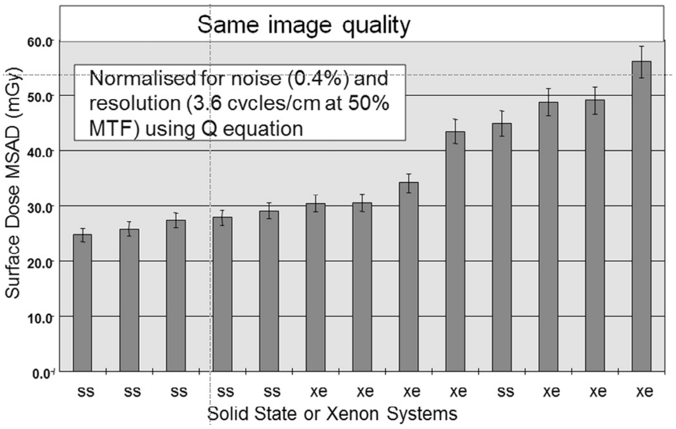


Figure 9. The use of a pervious version of Q to illustrate the relative dose, normalised by the other factors (courtesy of ImPACT) [slide 36 from Reference 29].

The CT dose efficiency value (CTDEV) puts all parameters that are relevant for the specification of LCD into a single number that is based on the fundamental theory of Rose [46]:

$$CTDEV = 10^5 \frac{e^{0.207(D_{\mathrm{eq}}-16)}}{d^2C^2h_{\mathrm{rec}}CTDI_{\mathrm{Vol,H}}} \tag{29}$$

with the diameter d of the low contrast detail (here, $d = 5$ mm), the slice thickness h_{rec} (in mm), the volume CT dose index $CTDI_{\mathrm{vol,H}}$ for the 16 cm head phantom, the PMMA-equivalent phantom diameter (in cm) D_{eq}, and the detail contrast (in %, with 1% = 10 HU) $C = 3.29\ \sigma$.

The method of Chao et al. can be easily implemented by applying customary CT phantoms and reduces the variability in LCD visually specified by human observers in conventional image quality assessments [45]. Chao's method has been applied by two CT manufacturers for the assessment of low contrast specifications [47]. The result of the method, however, depends on the size of the predefined ROI, the location of the CT image slice within the cone beam, and the filter used for image reconstruction [48]. As with other figures of merit, such as the Q_2 value, to apply the CTDEV for CT benchmarking, certain features must be standardised in detail. These are the protocol parameter set, reconstruction filter, phantom and method used.

Measures of diagnostic performance

Visual grading analysis (VGA)

Complementing the physical measurements of image quality, the assessment by observers is a subjective way to evaluate the image quality. Several general principles apply to all subjective observer studies: patients should be selected to have a wide range of body habitus, they should involve as many observers as possible, and they should cover the range of expected competencies in the field [49]. When these assumptions are verified, the visual grading analysis (VGA) based on observer scorings can be used to assess image quality.

VGA provides two types of information [50]:

Firstly, this subjective analysis provides information on the acceptability of the appearance (i.e. image noise level) of the clinical images and how the anatomical structures are visualised. For example the VGA grades the visibility of important structures for different noise levels, because the detectability of low contrast structures is affected by noise, decreasing as the noise level increases.

Secondly, the subjective evaluation provides a context to interpret the physical metrics (i.e. *MTF*, *NPS*). Human observer evaluation is subject to change depending on context (i.e. brightness, tiredness), so the variability is not negligible and it is important to have a sufficient number of observers. For instance if a CT has 40% better *MTF* at high frequencies than another, but both CTs are rated by a single observer the difference between both systems will not become significant.

The VGA paradigm is split into two categories: relative grading and absolute grading.

Relative grading: The observer grades the image quality compared to a reference image or to the other images. The images should be displayed in random order to avoid any bias (i.e. first image read bias) and the viewing conditions should reproduce the darkened environment of the reading diagnosis room [51]. The parameter studied should be as specific as possible, but it is possible to ask more than one question in order to evaluate several specifications. The rating scale used in relative grading can have 3, 5, or more steps/ranks. The scale with 3 steps is not ideal because it is impossible to differentiate sufficiently. But when the degree of difference is small, a two step scale can be a possibility. The quality of the test is dependent on the reference image.

For instance, a scale with 5 steps can be represented by:

–2: A is much better than B
–1: A is slightly better than B
0: A and B are equal
+1: B is slightly better than A
+2: B is much better than A

Absolute grading: The observers do not have any references and the images are displayed one by one. The evaluation is performed for one image at a time unlike the relative grading. To avoid bias from observer learning, the reading sessions must be separated in time. The grading scale should be numerical (i.e. from 1 to 10) or adjectival. With the adjectival scale, the descriptor should be expressive in order to create a difference between the worst and best cases. For instance, the Likert scale is a non-comparative ordinal scale used especially in psychometric studies where the participants express their level of agreement with a given statement. Note that reproducibility is low with this type of study [52–54].

The results of a VGA study can be summarised with the VGA Score (VGAS):

$$VGAS = \frac{\sum_{o,i} S_c}{N_i N_o} \tag{30}$$

where Sc = the given individual scores for observer (o) and image (i), Ni = total number of images, and No = total number of observers. In a VGA study to analyse the statistical difference, the analysis of variance (ANOVA) is calculated, associated with procedures for multiple comparisons.

For VGA, clinical images are required, which increases the implementation difficulties and also forces the avoidance of naïve observers. Indeed, to assess image quality in the VGA paradigm, the observer experience is very important if we want the obtained results to be as little distorted as possible. Nonetheless, VGA results are subjective and the analysis may be influenced by the experience of the radiologist, for instance in visualising different noise textures.

Decision theory: the statistical approach

It is common practice to specify the performance of diagnostic systems in physical terms as described in Part 1. However, it is complicated to translate these results to clinical performance. For instance, in detection tasks, certainty is rarely present. When an observer is asked to detect a signal on a medical image **g**, the result is a degree of belief that the signal is present. This degree of belief is commonly called the response $\lambda(\mathbf{g})$ of the observer: a low value denotes a confidence that the signal is absent, whereas a high value corresponds to the conviction that the signal is present. As shown in Fig. 10, the probability of obtaining a response can be plotted over all possible responses for two categories of images: those that do not contain a signal (top) and those that do contain a signal (bottom). These two curves are called probability density functions (pdf): respectively $P(\lambda|H_0)$ and $P(\lambda|H_1)$, where H_0 is the null-hypothesis corresponding to signal absent and H_1 is the alternative hypothesis corresponding to signal present. In radiology, the observer is forced to make a decision. In the present framework, this means that the observer chooses a threshold λ_c above which a positive decision is made. Below λ_c the observer makes a negative decision.

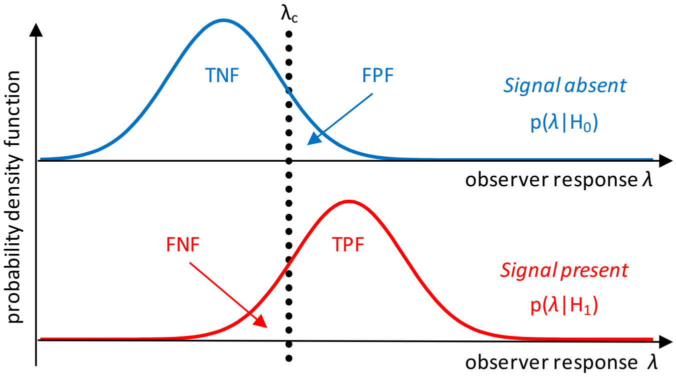


Figure 10. Probability density function of the observer response λ when presented with signal-absent images (top) or signal-present images (bottom). The vertical line λ_c indicates the threshold response above which the observer gives a positive response. TNF: true negative fraction; FPF: false positive fraction; FNF: false negative fraction; TPF: true positive fraction.

The integral of the distribution $P(\lambda|H_0)$ that is below the threshold is called the true negative fraction (TNF), or specificity. On the other hand, the integral of the distribution $P(\lambda|H_1)$ that is above the threshold is called the true positive fraction (TPF), or sensitivity. If the detection strategy is good, one expects both specificity and sensitivity to be as high as possible. However, Fig. 10 shows that changing the threshold changes the balance between specificity and sensitivity: increasing one parameter leads to a decrease of the other.

There are mainly two ways to quantify the effectiveness of the strategy. The first is the signal to noise ratio defined as follows:

$$SNR_\lambda = \frac{|\mu_1 - \mu_0|}{\sqrt{\frac{1}{2}\left(\sigma_0^2 + \sigma_1^2\right)}} \tag{31}$$

where μ_0 and μ_1 are the means of $P(\lambda|H_0)$ and $P(\lambda|H_1)$, respectively, and σ_0 and σ_1 are the corresponding standard deviations. SNR_λ is a global figure of merit that broadly describes how two distributions are separated. This equation is similar to Eq. (6) about SDNR and its purpose is to compare two situations (with and without noise). However, Eq. (31) characterises the response of an observer and not a signal or a noise directly measurable on an image.

$SNR_\lambda = 0$ corresponds to the situation where the two pdfs have the same mean. If their shapes are the same, the decision based on such a strategy will be just guessing, and therefore the image does not transfer any information about the presence of the signal. A large SNR_t corresponds to well-separated pdfs. If the threshold is chosen between the distributions, then a large number of correct responses are expected.

A second way to quantify the effectiveness of the strategy is the receiver operating characteristics (ROC) curve, which displays all the possible combinations of sensitivity and specificity obtainable whilst

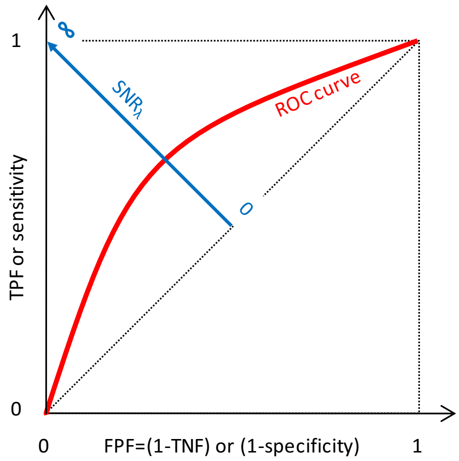


Figure 11. The ROC curve displays the true positive fraction versus the false positive fraction. If both response distributions are Gaussian with the same variances, then the intercept between the ROC curve and the second diagonal corresponds to SNR_λ.

we vary the threshold from the lowest to the highest possible values [55].

For historical reasons, the ROC curve displays the TPF versus the FPF, which is the sensitivity versus the (1 – specificity). If the pdfs are superimposed, the ROC curve is the straight line TPF = FPF. If the pdfs are well separated then the ROC curve has a square shape that passes close to the perfect point defined by sensitivity = 1 and specificity = 1. If pdfs are Gaussian with equal variances (this is often assumed in practice), the ROC curve is symmetrical and its intercept with the secondary diagonal corresponds to SNR_λ (Fig. 11). The value computed from the intercept between the ROC curve and SNR_λ is called the detectability index and usually represented with the symbol *d′*.

In practice, the observer (e.g. the radiologist) chooses a given threshold that corresponds to an operating point on the ROC curve. An objective way to define an optimal combination of sensitivity and specificity consists of computing the mean cost associated with all possible combinations of decision (negative or positive) and reality (signal absent or present):

$$\begin{aligned}\bar{C} = {} & C_{00}P(D_0|H_0)P(H_0) + C_{01}P(D_0|H_1)P(H_1)\\ & + C_{10}P(D_1|H_0)P(H_0) + C_{11}P(D_1|H_1)P(H_1)\end{aligned} \tag{32}$$

where C_{ij} is the cost associated with decision D_i and reality H_j, $P(D_i|H_j)$ is the pdf to make a decision D_i when the reality is H_j, and $P(H_1)$ is the probability to have a signal present. The latter is called prevalence in the case of the disease present in a population. By taking into account the basic properties of probabilities (e.g. $P(H_1) = 1 - P(H_0)$), Eq. (31) can be easily rewritten in terms of the four costs, sensitivity, specificity and prevalence.

All measures of clinical image quality using the decision theory are based on the truth. This truth can either be the ground truth (the truth is known exactly) or a gold standard (based on for instance the pathology outcome or experts opinion). Human observer studies are valuable as they are able to directly measure clinical image quality. Unfortunately, these methods are time consuming, expensive, and the inter- and intra-observer variability is often large. As a result assessment of clinical image quality is only applied incidentally. These limitations, together with the growing awareness of the importance of the evaluation of clinical image quality, make it more relevant to investigate whether model observers can be used as an objective alternative to human observers. This section is however limited to the discussion of rating scale experiments and m-AFC experiments using human observers. Part 3 provides an in-depth discussion about the use of model observers for this purpose. To gain insight into the decision making process rating scale experiments where observers are asked about their decision confidence can be performed. By varying variation in the decision threshold ROC curves can be drawn. The section “Rating Scale Experiments” provides more in-depth background of rating scale experiments. Another way to deal with observer decision criteria is by using multiple-alternative forced choice (m-AFC) experiments. In m-AFC experiments multiple alternatives are shown to the observer who is asked (forced) to choose the m-alternative which is most likely to contain the signal. This type of experiment will be discussed in detail in the section “Alternative Forced Choice Experiments.”

Rating scale experiments

ROC analysis is a quantitative method applicable to a binary decision task. The method results in a graphical plot, the so-called ROC curve (Fig. 11), that illustrates the performance of observers (either humans or computer models) in the detection or classification tasks [50,56–58]. In this chapter we focus on the use of ROC analysis with respect to diagnostic imaging. In diagnostic imaging ROC studies, observers are asked to evaluate different cases and give a confidence about the presence or absence of an abnormality in each case. The TPF and the FPF depend on the choice of the confidence level which results in a positive decision (threshold). Generally, the ROC curve will be determined from the continuous confidence scale by varying the discrimination threshold. However, discrete binary confidence intervals can also be used in ROC analysis. An example of a continuous data experiment could be the assessment of the average CT number of pulmonary nodules from CT images to classify benign from malignant nodules (nodules with higher CT numbers are more likely to be calcified which is a sign of benignity; the average CT number will generate the continuous data). Discrete data could be obtained, for example, in a study with radiologists providing a five-point discrete confidence rating of abnormality concerning a set of normal and abnormal diagnostic images. For examples of ROC analysis used in computed tomography see References 59–61.

Theoretically, ROC curves are continuous and smooth. Unfortunately, the empirically derived ROC curves are most often jagged. Fitting algorithms can aim to create the smoothest curve according to the available data points. A wide range of algorithms is available for this purpose [56]. Often the area under the ROC curve (AUC or Az) is determined as figure of merit for ROC studies. This AUC provides a summary measure of the accuracy of the diagnostic test that is independent of class prevalence (in contrast to accuracy measures mentioned earlier). The AUC would be 1.0 for a perfectly performed test. A test performance that is equal to chance results in an AUC value of 0.5. Sometimes it can be more useful to look at a specific region of the ROC curve rather than at the whole curve. In these scenarios, it is possible to compute partial AUC. For example, one could focus on the region of the curve with a low false positive rate, which could be relevant for population screening tests [56]. The detectability, d_A, related to a rating scale experiment can be derived from the AUC:

$$d_A = \sqrt{2}\Phi^{-1}(AUC) \tag{33}$$

where, $\Phi = \int_{-\infty}^{x} \phi(y)dy$ is the cumulative Gaussian function and $\phi = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ is a Gaussian function.

If the decision variable distribution is Gaussian under both hypotheses (signal present and signal absent), and their variances are equal, then d_A is equivalent to d′.

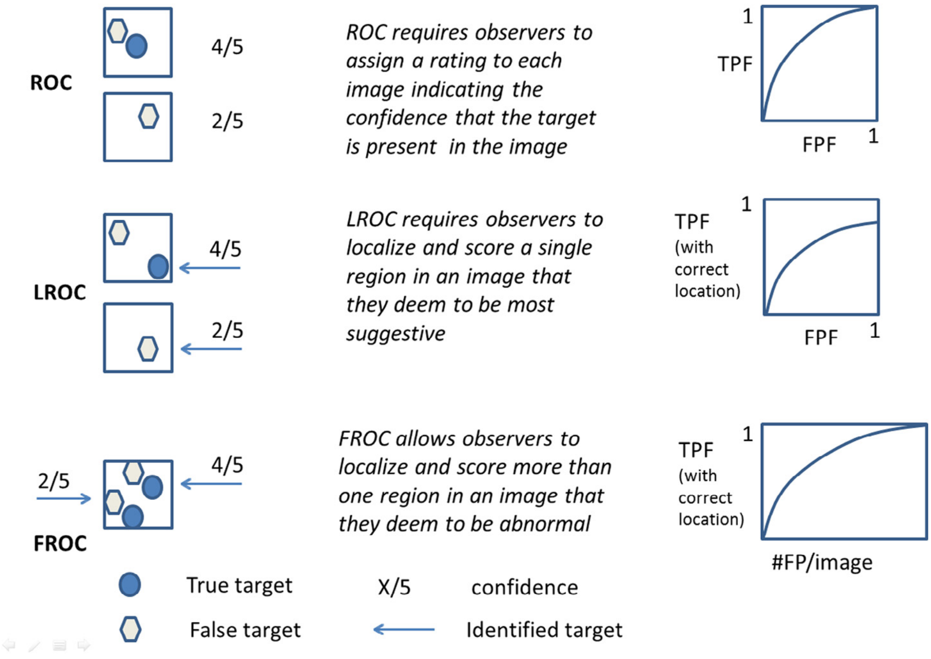

ROC
4/5
2/5
ROC requires observers to assign a rating to each image indicating the confidence that the target is present in the image
1
TPF
FPF
1
LROC
4/5
2/5
LROC requires observers to localize and score a single region in an image that they deem to be most suggestive
1
TPF (with correct location)
FPF
1
2/5
4/5
FROC
FROC allows observers to localize and score more than one region in an image that they deem to be abnormal
1
TPF (with correct location)
#FP/image
True target
False target
X/5
confidence
Identified target

Figure 12. Related methodologies: ROC, LROC, FROC. The task in each of the methods is to give a confidence level concerning the presence of a true target (ROC) eventually in combination with the perceived location (LROC/FROC). In these examples the confidence level scale runs from 1 to 5. A rating of 4 on this scale is given as 4/5 (4 out of 5). Arrows indicate the perceived location.

Several advantages of ROC analysis can be considered. Among these is for instance the fact that the ROC approach provides a simple graphical plot that facilitates visual interpretation of data. Furthermore, depending on the implications of false positive and false negative results, and the prevalence of the condition, one can choose the optimal cut-off for a test from this graph, as the method provides a description of diagnostic accuracy for the full range of sensitivity and specificity. Moreover, two or more tests (for instance radiologists and a Computer Aided Diagnosis (CAD) system) can be compared, for example, analysing the area under each curve (where the better test has the largest AUC) [62]. Shortcomings of ROC analysis are related to its need for specialised computer software (regarding the curve fitting, AUC value calculation and confidence analysis on the ROC curve). Also, large sample sizes may be needed to generate reliable ROC curves. Finally, the ROC methodology does not optimally take the localisation task or the option of multiple abnormalities into account. For this purpose the so-called localisation ROC (LROC) and free response ROC (FROC) have been introduced. Figure 12 gives a graphical impression of the different methods and their concepts. Figure 13 gives a decision tree that illustrates the application of the different methods.

In LROC studies the observers' task is to mark a single location of a suspicious region in each case with a confidence level regarding the observed suspiciousness [56,57,63]. If the marked region is "close enough" to the true abnormal location, the observers' mark is considered a correct localisation. The definition of closeness is not uniformly defined and changes from study to study. Images with no targets (controls, benign, or negative cases) are also scored by marking a "most suspicious" area in the image and by giving this suspicious area a rating (forced localisation choice). To create an LROC curve, the TPF of decisions with correct localisation versus the FPF are plotted. It should be noted that the LROC curve does not necessarily pass the point (1, 1). Unlike the ROC methodology, in LROC the TPF of decisions with correct localisation may well be less than 1.0 at FPF = 1.0 because of incorrect localisations. Similar to the ROC methodology, the area under the LROC curve is considered to be a figure of merit for task performance.

To account for both the localisation and detection of abnormalities in images containing an arbitrary number of them, the free-response ROC (FROC) methodology can be used [56,57,63]. If the localisation mark is within a tolerance range around the true location and the rating of this mark is above a threshold, then a TP is realised. Otherwise a FP decision occurs. The free-response ROC curves are plotted by plotting the TPF (y-axis) versus the number of FP detections per image (x-axis) [64,65].

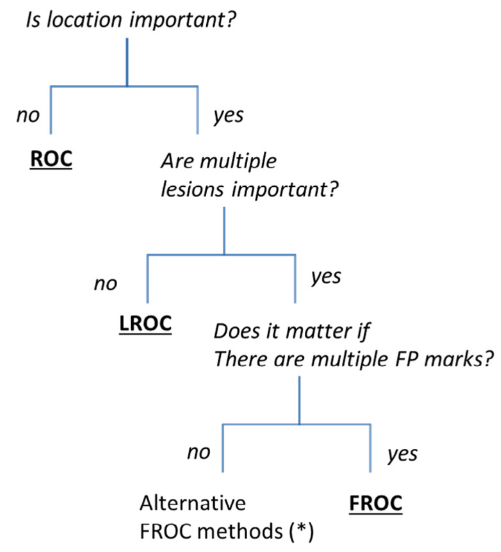

Is location important?
no
yes
ROC
Are multiple lesions important?
no
yes
LROC
Does it matter if There are multiple FP marks?
no
yes
Alternative FROC methods (*)
FROC

Figure 13. Decision tree illustrating the application of the different methods. The figure is a simplification of a figure provided by Wunderlich and Abbey [63]. Alternative methods (*) concern so-called Alternative FROC (AFROC) methods [54].

Alternative forced choice experiments

In forced choice experiments the observer has to make the decision ‘signal present’ between alternatives which are offered, even if this means that he has to guess. Compared to ROC studies, m-AFC experiments are faster and easier to perform [66]. However, m-AFC experiments do not provide insight into the underlying distribution functions and the trade-off between sensitivity and specificity [56]. Therefore, m-AFC is sometimes referred to as a poor measure of sensitivity [67].

The natural outcome of m-AFC experiments is a proportion of correct (PC) response. In m-AFC experiments and under assumption of Gaussian distribution of the decision variables (λ), d′ and PC_m of a m-AFC task are related by:

$$PC_m = \int_{-\infty}^{\infty} \Phi^{m-1}(d')\phi(d') \tag{34}$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ and $\Phi(x) = \int_{-\infty}^{x} \phi(y)dy$ are respectively Gaussian and cumulative Gaussian functions [68].

This formula can be solved using tabulated values or numerical analysis (standard root finding methods) [69–72]. In the 2AFC experiment, this can be rewritten to:

$$d' = \sqrt{2}\Phi^{-1}(PC_2) \tag{35}$$

For 2-AFC experiments, the PC is equal to the AUC but with human observers, the detectability obtained with the alternative forced-choice paradigm is larger than the detectability obtained with the ROC paradigm [50].

An example of setting for 2-AFC Signal Known Exactly/Background Known Exactly detection experiments is depicted in Fig. 14, where samples with signal present or absent are displayed together with a template of the target.

A detailed comparison and discussion about the use of ROC and AFC experiments as well as the optimum selection of m has been presented by Burgess [66]. This paper concludes that depending on the research question, a deliberate choice between ROC – m-AFC experiments and the value of m is possible. In general m-AFC experiments are chosen if the study goal is to determine how well a certain task can be performed and when there is full control over both the ground truth and the SNR associated with the task. Most commonly m has a value of 2 or 4 but any scalar number larger than two is possible [73]. Burgess has demonstrated that a higher value of m will result in a smaller coefficient of variance. Besides this, he has shown that if d′, for experiments with different values of m, is plotted against the signal to noise ratio (SNR) of the task they will fall on a the same line, independent of m [74]. From this it can be concluded that the choice of m depends essentially on the SNR range for the experiments and the accuracy needed. The SNR range which can be used for an experiment is dictated by the SNR related to the lower threshold (halfway between chance and 1) and 0.95–0.98. This upper level is advised to avoid issues due to observer inattention and their impact on d′ [66,75]. This means that in a 2-AFC experiment, the SNR range should be chosen to result in d′ values between 0.95 and 2.33, whilst this should be between 1.23 and 2.92 for 4-AFC (Fig. 15).

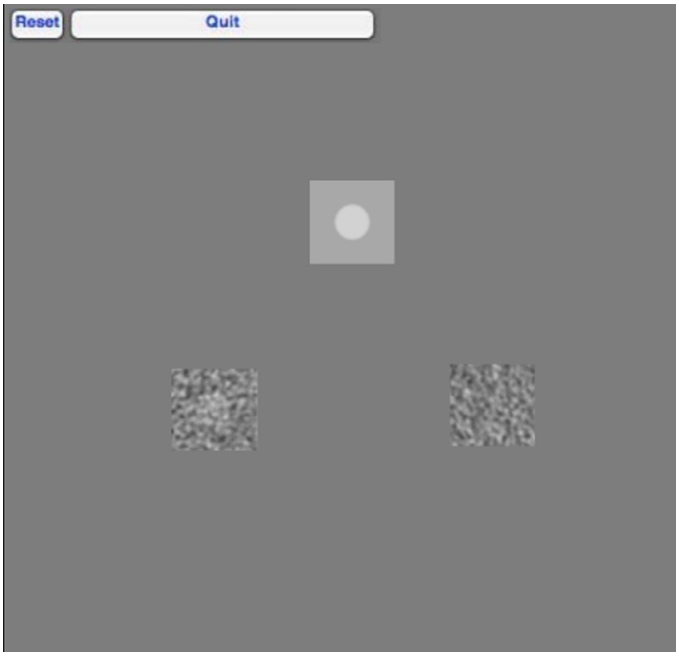


Figure 14. Interface of a 2-AFC human observer SKE/BKE detection experiment.

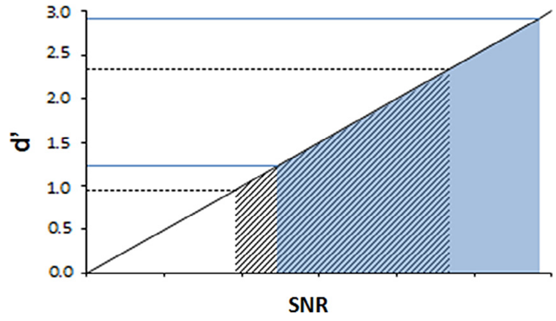


Figure 15. Selecting SNR range for a 2-AFC experiment (dotted black line) and a 4-AFC experiment (solid grey line).

m-AFC experiments can be designed with m independent image combinations or single images which are divided into m areas in which the task can either be signal detection (present-absent) or classification (benign–malignant) [76,77]. Sample sizes for m-AFC experiments are based on the comparison of the expected difference between the PCs of the settings under evaluation for which standard statistical approaches can be followed. m-AFC experiments are based on the signal-known-exactly (SKE) paradigm, which implies that clues should be provided regarding the signal and its position. Therefore, a template of the signal should be visualised together with the m alternatives and an indication of the possible position of the lesion should be indicated. Failure to provide clues on the signal position will result in non-linearity between SNR and observer d′ [66]. Finally, when designing m-AFC experiments care should be taken to avoid bias. For this purpose, the signal should be randomly assigned to one of the m alternatives and the observer PCs should be investigated for the tendencies to favour certain alternatives (e.g. the observer tends to choose left when he is unsure) [66].

Simulated and phantom images are generally well suited to conduct m-AFC experiments because of the full control of ground truth and SNR related to the task [66]. Phantom studies with the m-AFC paradigm are used to evaluate image quality of CT with both human and model observers [77–79]. But also for other modalities m-AFC methodologies are adapted into phantoms for quality control procedures like the CDMAM test object in mammography [80] or the CDRAD for general radiology [81,82].

Yes–no detectability experiments

In yes–no experiments observers only need to decide about the presence of an abnormality. Since yes–no experiments do not provide insight into the decision-making process of the observer they are not often used for measuring clinical performance very often. In the yes–no experiment the observer inspects one displayed image at a time and must indicate if the signal is present or absent. For a model observer, the yes–no performance is computed by comparing the decision variable to a threshold [50]. If the decision variable is higher than a threshold, the decision is: the signal is present. If the decision variable is less than the same threshold, the decision is: the signal is absent. In this test we assume that the case where the decision variable is equal to a threshold is negligible. With this performance it is possible to obtain four outcomes: true positive (the signal is present and the observer outcome is present), false positive (the signal is absent and the observer outcome is present), true negative (the signal is absent and the observer outcome is absent) and finally false negative (the signal is present and the observer outcome is absent). In the yes–no experiment the detectability index is given by:

$$d_{YN} = \Phi^{-1}(TPF) - \Phi^{-1}(FPF) \tag{36}$$

The TPF represents the True Positive Fraction, and it means the probability given that the signal is truly present in the image. The False Positive Fraction represents the probability that when the signal is absent the observer indicates that the signal is present.

Model observers

ICRU Report 54 suggests that methodologies based on statistical decision theory should be used in medical imaging [58]. Under this framework it is understood that the imaging performance depends on various factors: (1) measures describing the image contrast, image sharpness and the quantity and character of noise; (2) the detailed nature of the diagnostic task, including the clinically important details and the figure of the patient, and the complexities arising from variability between patients; and (3) the degree to which information provided in the image is perceived by the clinician. Points (1) and (2) above are related to the information that is being recorded in the image data, but the ability of the human observer to extract the image information (Point 3 above) may also be an important or even the single limiting factor affecting diagnostic outcome.

Related to this, to simplify image quality assessment, the imaging process is often divided into two separate stages: the first stage consists of the image data acquisition and image formation stage; the second stage consists of the further processing of these data and their actual display to the human observer [58,83]. The first stage can be analysed rigorously by using the concept of the ideal observer, at least in principle and also in practice in simple cases. The ideal observer uses all available information in an optimal way for its decision; the performance of the ideal observer in a given imaging task can then be taken as a measure of the image information related to this task. The ability of the human observer to extract this image information can be measured separately; if the human observer is not able to use the recorded image information this implies leeway –and a need– to improve the image processing or display stage to be better suited to the human observer. This chapter will mainly concentrate on the imaging stage and leave the display stage largely outside the scope; the main aim of this paper is to review methods for evaluating CT scanners and their performance and not the quality of display equipment and display conditions. However, some methods which try to include features of human observers are shortly presented.

The performance of the ideal observer can usually be evaluated only for simplified classification tasks, such as the signal-known-exactly/background-known-exactly case, denoted as SKE/BKE. In this case the ideal observer has all a-priori information of the task, and its performance for classifying images to signal-present and signal-absent cases depends only on the amount of information in the image [58]. The performance of the ideal observer can therefore be taken as a measure of the task-related image information. Other tasks, involving uncertainty of the signal and the background, would be better related to clinical image quality assessment than the SKE/BKE. In such tasks the performance of the observer is not just dependent on the information in the image. The amount of a-priori information about the task that the observer has needs to be taken into account and will affect the performance. It may then sometimes be difficult to quantify the actual effect that this a-priori image information has in the task performance.

Relying on stylised imaging tasks based on the SKE/BKE paradigm may not always be reasonable; see, e.g., Myers et al., where the problem of aperture-size optimisation in emission imaging was considered and it was shown that the optimal aperture would be highly different for the detection of a simple signal in a known background and in a lumpy background [84]. Often, however, it may be considered plausible that the performance of an imaging system in tasks involving incomplete a-priori information could be monotonically related to the outcome in similar detection tasks in the case of full a-priori information (SKE/BKE) [85–88]. This appears to be the case in the paper of Brown et al., where the ideal observer's performance was studied for the signal position unknown case [89]. However, we are still far from completely understanding how a-priori information and the actual image information interact in medical imaging.

In phantom measurements the variability and non-uniformity of real patient images are not usually present. In the SKE/BKE paradigm any background structure is treated as being a deterministic known structure, which does not impair detail detectability. This may not always be realistic for a human observer, whose detection performance may in some cases be more impaired because of background variability than because of actual stochastic noise [90–94], but is certainly applicable to the ideal observer. Human observers seem to operate somewhere between two interpretations: background variability appears to function as a mixture of noise and deterministic masking components. For a more detailed discussion on this matter, see, e.g. Burgess and the references therein [91].

For a thorough presentation of modern image science, see the book by Barrett and Myers [57] and by Samei and Krupinski [56]. Another useful handbook on imaging systems, image quality and measurements has been published by the International Society for Optical Engineering [50]. Also, a discussion and review of task-based methods for assessing the quality of iteratively reconstructed CT images have been published recently [25]. They conclude that Fourier-based metrics of image quality are convenient and useful in many contexts, e.g., in quality assurance, but the assessment of iteratively reconstructed CT images requires more sophisticated methods which do not rely on assumptions of system linearity and noise stationarity; these assumptions are prerequisites in the Fourier-based methods [95–97].

Linear observers

Mathematical theory

A linear observer can be described with a decision statistic $\lambda(g)$ which is a linear function of the image data, instead of being a more general function. In the vector notation of images this can be written as an inner product of a template **w** and the image **g**

$$\lambda(\mathbf{g}) = \mathbf{w}^T\mathbf{g} \tag{37}$$

The non-zero elements of the template correspond to image locations where the pixel value needs to be taken into account, and by what weight. The weight can be either positive or negative. Pixels with the value zero in the template do not influence the decision statistic at all, and the observer considers the data in those pixels to be irrelevant for the decision.

The importance and frequent use of linear observers stems mainly from their manageability and ease of use. Further, as was seen in the preceding chapter, the ideal observer of many cases may be obtained in a linear form. This is not the case for all detectability tasks, however. For example, the ideal detection in the case involving uncertainty of the signal position will result in a non-linear test statistic (see, e.g., Brown et al. [89]). A linear observer for this task would consist just of a template which is obtained as the convolution of the pdf of the signal position and shape. Therefore, essentially, this observer would measure only the mean brightness of the image and it seems clear that it would be much less efficient than a human observer, for example.

In order to compute the SNR of a linear observer, we first need to express the mean response under hypothesis H_j as well as its associated variance:

$$\begin{aligned}\bar{\lambda}_j &= \langle \lambda(\mathbf{g})|H_j\rangle = \mathbf{w}^T\langle \mathbf{g}|H_j\rangle \\ \sigma_j^2 &= \left\langle \left(\mathbf{w}^T\mathbf{g} - \langle \mathbf{w}^T\mathbf{g}|\mathbf{H_j}\rangle\right)^2 \middle| H_j\right\rangle = \mathbf{w}^T\mathbf{K}_j\mathbf{w}\end{aligned} \tag{38}$$

This allows us to easily express the signal to noise ratio of a linear observer by injecting Eq. (38) into Eq. (31):

$$\mathrm{SNR}_\lambda^2 = \frac{\left(\mathbf{w}^T(\langle \mathbf{g}|H_1\rangle - \langle \mathbf{g}|H_0\rangle)\right)^2}{\mathbf{w}^T\,\frac{1}{2}(\mathbf{K}_0 + \mathbf{K}_1)\mathbf{w}} \tag{39}$$

Here, it is important to recall the assumptions required for Eq. (39) to be meaningful. First, this requires that the conditional distributions of λ are normal. This is the case at least when the noise in the images is multivariate normal. Secondly, if the covariance matrices for the signal and background cases are different, the SNR does not define the entire ROC curve, but the area under the ROC curve and the percentage of correct answers in a two-alternative forced-choice test using the same images are still specified by the SNR. An inequality of covariance matrices $\mathbf{K}_0$ and $\mathbf{K}_1$ would also infer that a linear observer is not ideal, and may fall far beyond the true ideal observer [98]; however, if measured covariance data are used, it is useful to improve the precision of the **K**-estimate by including both measured covariance, $\mathbf{K_0}$ and $\mathbf{K_1}$.

By inserting the w-templates of the PWMF and the NPWMF to Eq. (39) we obtain the well-known expressions for their SNR

$$\mathrm{SNR}^2{}_{\mathrm{PWMF}} = \mathbf{s}^T\mathbf{K}^{-1}\mathbf{s} = \mathbf{S}^T\mathbf{W}^{-1}\mathbf{S} \tag{40}$$

and

$$\mathrm{SNR}^2{}_{\mathrm{NPWMF}} = (\mathbf{s}^T\mathbf{s})^2/\mathbf{s}^T\mathbf{K}\mathbf{s} == (\mathbf{S}^T\mathbf{S})^2/\mathbf{S}^T\mathbf{W}\mathbf{S} \tag{41}$$

where we have denoted the Fourier transform of **s** by **S** and that of matrix **K** by **W**. If the noise is stationary, **W** is a diagonal matrix and its diagonal values represent the NPS. Then, decomposing the SNR^2 to components: each frequency k contributes by amount

$$\mathrm{SNR}^2{}_{\mathrm{PWMF},k} = \mathrm{I}|\mathrm{S}_k|^2/\mathrm{W}_k \tag{42}$$

to the total $\mathrm{SNR}^2{}_{\mathrm{PWMF}}$. This simplicity is lost if W is not diagonal.

The best possible linear observer is called the Hotelling observer. The Hotelling observer is equal to the PWMF in the case of signal-independent (additive), normally distributed noise and both of these reduce to the NPWMF, when the noise is white. As discussed above, the Hotelling observer may also fall far below ideal performance, for example, in the signal position unknown case, where the ideal decision statistic is not a linear function of image data [89].

The strategy of the ideal observer may be complicated by **K** not being diagonal. However, in the case of uncorrelated image noise the strategy is self-evident: the ideal observer then just looks more keenly to image pixels where the presence of the signal is known to have a strong effect and where the uncertainty of the measurement (noise) is small. Image areas that are not affected by signal presence need not be observed at all. This same interpretation applies to the case of coloured, stationary noise as well; then the Fourier transformed data will have a diagonal covariance matrix, where the diagonal elements constitute the noise power spectrum. In this case the ideal observer puts more emphasis on spatial frequencies where the signal presence makes a large contribution and less emphasis on frequencies which contain more noise.

If the image noise is not white, the NPWMF observer is suboptimal because it does not take into account the noise correlations between pixels, or equivalently, the different noise power at various spatial frequencies. Therefore, in this case, the observer is not tuned against the noise similarly as the ideal observer and it shows a penalty of this in its performance. However, if the frequency spectrum of the signal is concentrated on a relatively narrow band of frequencies where the frequency dependence of the NPS is modest, one can expect the NPWMF observer to perform nearly as well as the ideal PWMF does. This may happen, for example, when the signal to be detected does not have sharp details and is of a relatively large size.

However, note that by definition, the NPWMF believes that the background level is equal in all images and therefore needs not be observed. The NPWMF measures the image intensity only in the pixels that belong to the expected signal position and discards the data in all other pixels. For a disk signal this would be equivalent to observing just the total image intensity of the signal disk area and masking away all other image areas: no reference to the contrast between the signal and the background is made. If in fact, there is any – even small – variation in the background level from image to image, or if there is any low-frequency background variability (e.g., variable lumpy background structures) which in reality can have an effect on the image intensity in the signal detail area, the NPWMF can be considered as being a misled observer, which will perform extremely poorly and often performs worse than human observers. This was the case, for example, in a paper that considered signal detection in added low-pass correlated noise and found that the NPWMF observer was very inefficient and even humans significantly outperformed it [99]. This and other similar results greatly diminished the interest in the NPWMF observer.

To improve this situation, Tapiovaara and Wagner [98] introduced the DC-suppressing observer, which leaves the average brightness of the image (or the zero-frequency channel) outside of the decision.[1] This observer is achieved by subtracting the mean pixel value of the NPWMF-template from every pixel of the template

$$\lambda_{\mathrm{DCS}} = \left[\mathbf{s} - \left(\mathrm{N}^{-1}\Sigma \mathrm{s}_k\right)\mathbf{1}\right]^T\mathbf{g} \tag{43}$$

Here, N is the number of pixels in the analysed image area and **1** denotes a vector with all elements equal to unity. In the Fourier domain this observer is:

$$\lambda_{\mathrm{DCS}} = \left[\mathbf{S} - \mathrm{S}_0\mathbf{e}^0\right]^T\mathbf{G} = \sum\nolimits_{k=1}^{N-1} S_k^* G_k \tag{44}$$

---

[1] In practical imaging measurements one often does not analyse the whole image area, but considers only a relatively small sub-area containing the signal and a reasonable surround of it. Then the image vector **g** corresponds to this sub-area, and the zero-frequency of this image data includes contributions from very low-frequencies in addition to the strict zero-frequency of the whole image data.

This modification of the NPWMF-observer turned out to be crucial for the performance of the observer in measurements of fluoroscopic imaging, where excess noise in the mean image brightness strongly and variably impaired the performance of the NPWMF [100]. This zero-frequency variability can be assumed to be common in other fields of radiology as well: the exact mean image brightness is not probably an important diagnostic feature in any imaging modality, and, on the other hand, if there is excess variability in the brightness, including it – as the NPWMF does – will result in a notable performance penalty. Such a variability in average brightness can be seen as a delta spike at the origin of the NPS and can be properly weighted by the PWMF, of course. However, in many recipes for measuring the NPS, the DC-component is normalised out and therefore equals zero in the NPS results (e.g. Boedeker et al. [101]). Whilst noiseless data in real systems are not realistic, it is then important not to include the zero frequency signal component in the SNR calculation either.

Non-prewhitening with eye filter

Another modification of the NPWMF includes filtering of the image with an eye-filter, intended to obtain a better agreement of the performance of this model observer and human observers. The observer is often denoted as NPWE [102] (a similar observer model has been presented earlier by Loo et al. [103]). This observer is usually expressed in the spatial frequency domain and the eye filter **E** mimics the visual spatial frequency response function (or the contrast sensitivity function) of the human eye. The application of **E** requires knowing the dimension of the image and the viewing distance. The decision function of this observer is then:

$$\lambda_{\mathrm{NPWE}} = [\mathbf{ES}]^{\mathrm{T}}\,\mathbf{EG} = \mathbf{S}^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}\mathbf{EG} \tag{45}$$

It is noted here that the eye filter also suppresses the zero-frequency, like the DCS-observer above, but the NPWE observer also takes very low frequencies into account with a low weighting. This is the main factor for the NPWE observer performing much better than the NPWMF in studies involving excess noise in very low frequencies [25,102]. This means that the usefulness of this observer model may actually be more related to its suppressing of low-frequency noise than in its attempt to mimic human vision.

As an example of NPWE performance, Fig. 16 shows the detectability index (d′) or SNR as a function of object diameter for the 0.5% contrast group of the Catphan and three mAs levels acquired in a Toshiba Aquilion ONE 320 detector-row CT scanner. The NPWE detectability improved with increasing mAs, as the noise level of the images decreased, for all the objects [50].

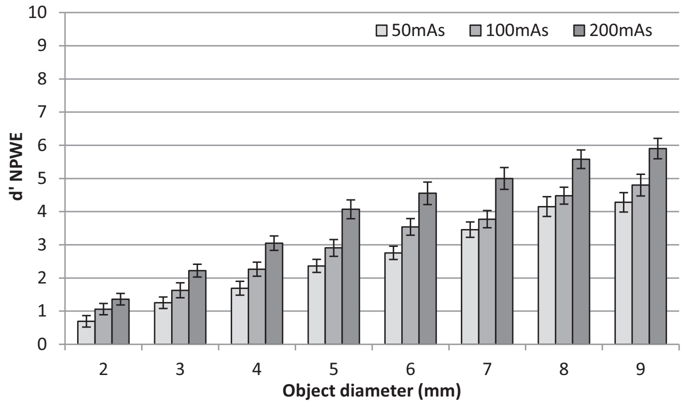


Figure 16. Detectability index (d′) as a function of object diameter for the different levels of mAs for the 0.5% contrast group (2-9 mm) in the Catphan 600 Phantom (Phantom Laboratories, New York).

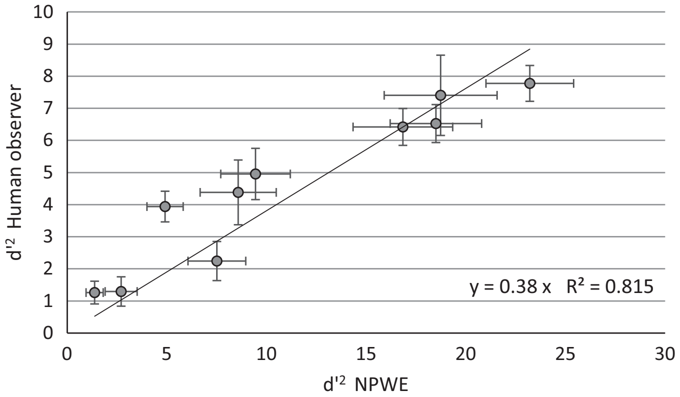


Figure 17. Detectability index (d′) of the average human observer as a function of the NPWE model observer d′, both squared, for 1% contrast objects and all dose values. The efficiency, η, tallies the slope of the linear fit.

The detectability index is given when two assumptions are verified [73]. Firstly the template responses must be Gaussian and secondly the template responses are statistically independent [90]. This performance is given in terms of distance in standard deviation units between the signal distribution and the noise distribution.

$$d' = \frac{\langle\lambda_s\rangle - \langle\lambda_n\rangle}{\sigma_\lambda} \tag{46}$$

where λ_s is the mean model response to the signal, and λ_n is the mean model response to the background. σ_λ is the standard deviation of the model response.

The advantage of this metric is that it computes directly from the image statistic.

Model observers can also be otherwise modified in order to mimic human performance better, for example, by including internal noise [104,105]. Internal noise degrades the model's performance, and takes into account the fact that human observers have "noise" by not giving necessarily the same answer when a certain image is presented twice or more to be scored [106]. Many approaches can be used to decrease the model's performance, and each has pros and cons [105,107]. Such models are of interest in efforts to reproduce the efficiency of the visual detection performance of humans, but are not explained in this review. In Fig. 17 the PC values were translated into d′ and an efficiency (η) was calculated to normalise the model observer results, fitting d′ human as a function of d′NPWE, both squared. As the curve representing d′ as a function of PC saturates above 3 (PC ≈ 0.98) for 2-AFC experiments, only the values below this threshold were taken into account [108].

Channelised Hotelling Observer

Another type of linear observer models is the Channelised Hotelling Observer [109] (CHO) either with or without internal noise; only the latter model is considered here. A thorough treatment of both can be found in Abbey and Barrett [110]. The motivation for this observer results both from its effect in reducing the image data from a large number of pixels to a much lower number of scalars, called channel responses and by the ability of such models to mimic the detectability results of human observers. If the channels in the model are selected such that they help in the tuning against the noisy frequencies without losing too much of the signal energy they may also provide an improvement over the non-prewhitening observer types and a useful approximation for the ideal observer. The reduction of dimensionality especially simplifies computing and inversing the covariance matrix.

The CHO does not have access directly to the pixel values (or the Fourier transform) of the image. Instead, first the image data (**g**) are linearly combined to a small number of channelised data (**u**) by multiplication with a matrix **T**:

$$\mathbf{u} = \mathbf{T}^{\mathrm{T}} g \tag{47}$$

Here the column vectors of **T** represent the spatial profiles of the channels. These channelised data are then combined with a weighting template **v** to a linear decision function:

$$\lambda = \mathbf{v}^{\mathrm{T}} \mathbf{u} \tag{48}$$

If the noise in the image data **g** is Gaussian, it is also Gaussian in the channel **u**, and we already know that the ideal observer (which, however, has access only to the channelised data) is obtained with weighting $\mathbf{v}^{\mathrm{T}} = (\mathbf{u}_1 - \mathbf{u}_0)^{\mathrm{T}} \mathbf{K}_{\mathbf{u}}^{-1}$, and the decision function of this observer is:

$$\lambda_{\mathrm{CHO,T}} = (\mathbf{u}_1 - \mathbf{u}_0)^{\mathrm{T}} \mathbf{K}_{\mathrm{u}}^{-1} \mathbf{u} = (\mathbf{u}_1 - \mathbf{u}_0)^{\mathrm{T}} \mathbf{K}_{\mathrm{u}}^{-1} \mathbf{T}^{\mathrm{T}} \mathbf{g} \tag{49}$$

Above, the channels were presented in the image domain. Usually, however, the channels are specified in the frequency domain, and may be either non-overlapping frequency intervals or overlapping functions of various forms, such as sparse or dense difference-of Gaussians, Laguerre–Gauss polynomials or other functions [109,111].

Note that in the case of stationary image noise the non-overlapping channel models result also to a diagonal covariance matrix, because the frequency channels remain independent, whereas the overlapping channels cause correlations in the noise. If one prefers working in the image domain, one can obtain the spatial representations of the frequency selective channels by taking the inverse Fourier transforms of the latter.

In image quality assessment when using these channelised models it is important to note that the channelised Hotelling observer can adapt to the signal and the image covariance only after they have passed through **T**. Then, for example, the observer is sensitive only to signals that cause a change in the channelised signal $\mathbf{T}^{\mathrm{T}}\mathbf{s}$ (or, equivalently, in the frequency domain representation). For sparse channel models with just a few channels, a significant loss of information may occur in the formation of the channel responses [110].

Also, these observers are typically zero-frequency suppressing, although, being tuned against the noise in the different channels, they could also otherwise handle variability in the average image brightness better than the NPWMF. This would require, however, that if zero-frequency is included in the lowest frequency channel, not much of the important signal energy shall be included in this channel.

Usually, in applications related to medical imaging, the channels are defined to be cylindrically symmetric and are specified in terms of the radial frequency. The use of such models is usually restricted to image signals that are also cylinder-symmetric. Channelised Hotelling observers have been used with good success to predict the performance of human observers in detection tests.

As an example, Fig. 18 shows the CHO performance (detectability index (d′)) with dense of difference of Gaussian for an 8 mm sphere at 20 HU of the QRM 401 phantom and three $\mathrm{CTDI_{vol}}$ levels acquired in GE HD 750 CT scanner.

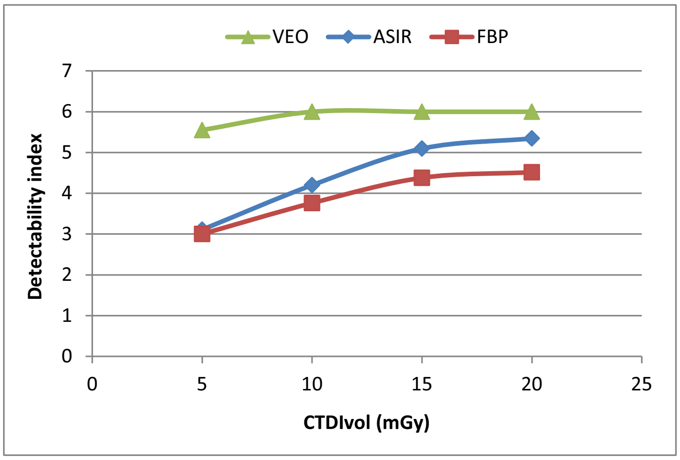


Figure 18. Detectability index (d_A) as a function of $\mathrm{CTDI_{vol}}$ for the different algorithms for a sphere of Ø 8 mm and a contrast with background of 20 HU in the QRM 401 Abdomen Phantom (QRM, Moehrendorf, Germany).

Agreement between observers

The first step to compare model observers and model/human observers is to have the same metrics to measure their performance. For a specific task, background, signal and model the investigator must choose between the area under the curve (AUC), sensitivity/specificity pairs, the percent correct (PC), the signal to noise ratio (SNR) or the detectability index (d′), then a comparison is possible.

Kappa test

To measure the agreement between observers it is common to use the Kappa coefficient. When observers are two or more the inter-observer variation can be computed. The Kappa test is based on the difference between the observer agreement (percentage where observers agree among themselves) and the expected agreement (agreement obtain just by chance). The formula for the Kappa test is then as follows:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{50}$$

where p_0 is the relative observed agreement among reviewers, and p_e is the probability of chance agreement.

The Kappa scale ranges from –1 to 1. 1 represents a perfect agreement; 0, the agreement is obtained just by chance; and –1 represents a systematic disagreement. A generic scale proposed by Landis and Koch is used to help the investigator to interpret the Kappa coefficient (Table 3) [112].

The Kappa coefficient estimated itself could be obtained just by chance, so a P value can be calculated to interpret the result of the Kappa test. The P value is sensitive to sample size, so another Kappa test can be used to interpret the result, the weighted Kappa assigns weighting more or less important to different categories, to focus on categories where the difference is significant. But the weighting is defined by the investigator, and the expert can disagree on the tuning of the weighted Kappa. The Kappa test is used to interpret the agreement, but this test is affected by the prevalence of the disease [50] (Fig. 19); in rare cases a low Kappa test does not reflect a low agreement. Moreover, the Kappa test can give strange results when the observers have a high degree of agreement and when they are close to PC = 1.

Table 3
Genetic scale investigator to interpret the Kappa coefficient.

0.01–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–0.99	Almost perfect agreement

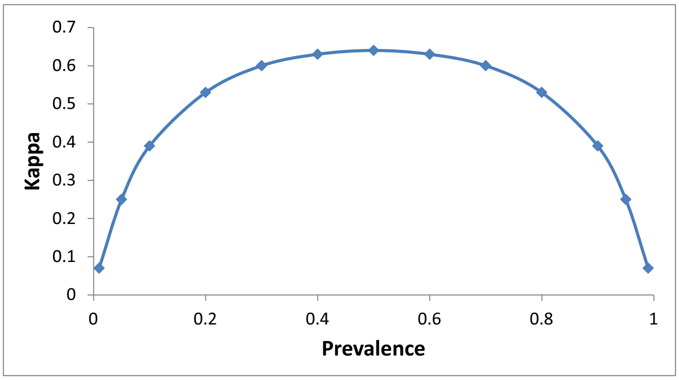


Figure 19. Kappa coefficient in terms of prevalence.

Bland–Altman test

A Bland–Altman plot is often used to compare results between model observers and human observers [113]. When both observers measure the same parameter (i.e. d′ or PC) with the same images, most of the time the correlation is good [57,108]. A good correlation for two observers that measure the same parameter does not imply a good agreement between the two observers.

A Bland–Altman plot shows the mean of the two observers in the abscissa, and the difference between the two observers in the ordinate. The limits of agreement are defined by the mean of the difference and the standard deviation of the difference. If a method is the gold standard then d represents the bias, whereas if any methods are standard, d represents only systematic differences. Figure 20 shows an example comparing the performance of the NPWE model and human observers for a given detection task.

Conclusion and perspectives

Since the introduction of CT many efforts have been made to balance image quality with patient exposure. Image quality was first assessed using signal detection theory, and basic parameters such as image noise and spatial resolution, which made it possible to evaluate the strengths and weaknesses of acquisition protocols. With the technological developments of CT it became necessary to assess units in order to objectively enhance the benefit of new technological solutions. Global figures of merit of image quality were derived, still using signal theory functions, normalising the result by a standardised dose indicator: the $\mathrm{CTDI}_{\mathrm{vol}}$. If this approach seems enticing one has to remember that the use of one number to judge image quality is a simplified solution that can lead to false conclusions. Moreover, image quality assessment methods based on signal theory only do not include a clinically relevant task. With this kind of approach one could optimise aiming at getting the best theoretical image quality, rather than ensuring that images convey the relevant clinical information to make a correct diagnosis. In such a context, image quality assessment in the field of medical imaging should be task oriented and clinically relevant.

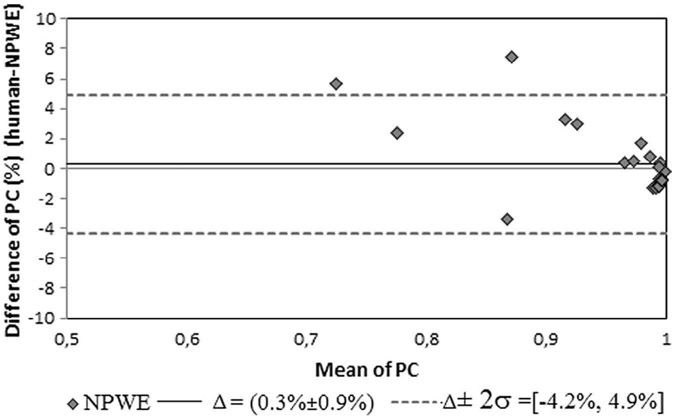


Figure 20. Bland–Altman plot of proportion correct (PC) difference between human and NPWE for 1% contrast and all mAs. The straight line represents the average difference (Δ) and the dash lines, the range of the differences [Δ ± 2σ], where σ is the standard deviation of the differences. The NPWE model was corrected by an efficiency of 0.38.

The use of mathematical model observers may be an appropriate solution, opening a way forward, even if the tasks investigated remain very simple and far from clinical reality. As shown in the review, there are several types of model observers, and the choice of a single solution might not be optimal. The disadvantage of model observers is that they are defined for simple situations, like the detection of a representative signal in a given phantom, and surely do not cover the whole range of characteristics that define image quality at the clinical level. This drawback can nonetheless become an advantage because their calculation can be kept relatively simple; they are objective and compatible with new image reconstruction techniques such as iterative reconstruction. They also lead to reproducible results which can be representative of human perception whilst avoiding the burden of actual studies with human observers. They could be used to compare clinical protocols in terms of image quality and dose levels to initiate an optimisation process. Nevertheless, more studies should be performed in the future on correlations between model observer outcomes and human diagnostic accuracy.

Acknowledgements

This work was supported by the German Radiation Protection Agency (BfS – UFO-Plan Vorhabens 3613S20007) and the Swiss National Science Foundation for research (FNS 320030_140995/1). The authors are thankful to R. van Engen from LRCB, Nijmegen and Pr H. Bosmans from Katholieke Universiteit, Leuven, and P. Monnin from CHUV, Lausanne, for their help for improving the structure of the manuscript.

References

[1] Thurston J. NCRP report no. 160: ionizing radiation exposure of the population of the United States. Phys Med Biol 2010;55:6327. doi:10.1088/0031-9155/55/20/6327.
[2] Aubert B, Sinno-Tellier S, Etard C. Exposition de la population française aux rayonnements ionisants liée aux actes de diagnostic médical en 2007. Institut de Radioprotection et de Sûreté Nucléaire et the Institut de Veille Sanitaire: 2010.
[3] 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.
[4] Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety. Bernhard-Ströl C, Hachenberger C, Trugenberger-Schnabel A, Loebke-Reinl A, Peter J, editors. Umweltradioaktivitaet und strahlenbelastung. Annual report. 2012.
[5] IRSN. Exposition de la population française aux rayonnements ionisants liés aux actes de diagnostic médical en 2012. Rapport PRP-HOM N°2014-6. 2012. <http://www.irsn.fr/FR/Actualites_presse/Communiques_et_dossiers_de_presse/Pages/20141013_Rapport-Expri-Exposition-rayonnements-diagnostic-medical.aspx#.VOHGuuK1Z8E> [accessed 16.02.15].
[6] Boone J, Strauss K, Cody D, McCollough C, McNitt-Gray M, Toth T. Size-specific dose estimates (SSDE) in pediatric and adult body CT exams. Report of AAPM Task Group 204. 2011.
[7] Nguyen TTA, Le HND, Vo M, Wang Z, Luu L, Ramella-Roman JC. Three-dimensional phantoms for curvature correction in spatial frequency domain imaging. Biomed Opt Express 2012;3:1200–14. doi:10.1364/BOE.3.001200.
[8] Mironov V, Boland T, Trusk T, Forgacs G, Markwald RR. Organ printing: computer-aided jet-based 3D tissue engineering. Trends Biotechnol 2003;21:157–61. doi:10.1016/S0167-7799(03)00033-7.

[9] Solomon J, Samei E. Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE. Med Phys 2014;41:091908. doi:10.1118/1.4893497.

[10] Tapiovaara MJ, Wagner R. SNR and DQE analysis of broad spectrum X-ray imaging. Phys Med Biol 1985;30:519. doi:10.1088/0031-9155/30/6/002.

[11] Nickoloff EL, Riley R. A simplified approach for modulation transfer function determinations in computed tomography. Med Phys 1985;12:437–42.

[12] Boone JM. Determination of the presampled MTF in computed tomography. Med Phys 2001;28:356–60.

[13] Judy PF. The line spread function and modulation transfer function of a computed tomographic scanner. Med Phys 1976;3:233–6.

[14] Nakaya Y, Kawata Y, Niki N, Umetatni K, Ohmatsu H, Moriyama N. A method for determining the modulation transfer function from thick microwire profiles measured with x-ray microcomputed tomography. Med Phys 2012;39:4347–64. doi:10.1118/1.4729711.

[15] Thornton MM, Flynn MJ. Measurement of the spatial resolution of a clinical volumetric computed tomography scanner using a sphere phantom, vol. 6142. 2006. p. 61421Z doi:10.1117/12.654969.

[16] Grimmer R, Krause J, Karolczak M, Lapp R, Kachelriess M. Assessment of spatial resolution in CT. IEEE Nucl Sci Symp Conf Rec 2008;5562–6. doi:10.1109/NSSMIC.2008.4774508. 2008 NSS 08.

[17] Friedman SN, Fung GSK, Siewerdsen JH, Tsui BMW. A simple approach to measure computed tomography (CT) modulation transfer function (MTF) and noise-power spectrum (NPS) using the American College of Radiology (ACR) accreditation phantom. Med Phys 2013;40:051907. doi:10.1118/1.4800795.

[18] International Commission on Radiation Units and Measurements. ICRU Report No. 87: radiation dose and image-quality assessment in computed tomography. J ICRU 2012;12:1–149. doi:10.1093/jicru/ndt007.

[19] Miéville F, Beaumont S, Torfeh T, Gudinchet F, Verdun FR. Computed tomography commissioning programmes: how to obtain a reliable Mtf with an automatic approach? Radiat Prot Dosimetry 2010;ncq050. doi:10.1093/rpd/ncq050.

[20] Dainty JC, Shaw R. Image science: principles, analysis and evaluation of photographic-type imaging processes. Academic Press; 1974.

[21] 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.

[22] Hsieh J, Nett B, Yu Z, Sauer K, Thibault J-B, Bouman CA. Recent advances in CT image reconstruction. Curr Radiol Rep 2013;1:39–51. doi:10.1007/s40134-012-0003-7.

[23] Richard S, Li X, Yadava G, Samei E. Predictive models for observer performance in CT: applications in protocol optimization, vol. 7961. 2011. p. 79610H doi:10.1117/12.877069.

[24] Ott JG, Becce F, Monnin P, Schmidt S, Bochud FO, Verdun FR. Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms. Phys Med Biol 2014;59:4047–64. doi:10.1088/0031-9155/59/4/4047.

[25] 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.

[26] Richard S, Husarik DB, Yadava G, Murphy SN, Samei E. Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms. Med Phys 2012;39:4115–22. doi:10.1118/1.4725171.

[27] Brunner CC, Abboud SF, Hoeschen C, Kyprianou IS. Signal detection and location-dependent noise in cone-beam computed tomography using the spatial definition of the Hotelling SNR. Med Phys 2012;39:3214–28. doi:10.1118/1.4718572.

[28] Miéville FA, Bolard G, Bulling S, Gudinchet F, Bochud FO, Verdun FR. Effects of computing parameters and measurement locations on the estimation of 3D NPS in non-stationary MDCT images. Phys Med 2013;29:684–94. doi:10.1016/j.ejmp.2012.07.001.

[29] Edyvean S. Understanding image quality and dose. ImPACT Feb. 2007. <http://www.impactscan.org/slides/course07/lect9/frame.htm> [accessed 29.04.15].

[30] Seeram E. Computed tomography: physical principles, clinical applications, and quality control, 3e. 2nd ed. Saunders; 2000.

[31] Edyvean S. The relationship between image noise and spatial resolution of CT scanners, <http://www.ctug.org.uk/meet02/noiseandspatialresct.pdf>; 2002 [accessed 11.03.15].

[32] Edyvean S, Keat N. Comparison of CT scanner image noise, image width, dose and spatial resolution using standard test methods. <http://www.aapm.org/meetings/04AM/pdf/14-2350-75226.pdf>; 2004 [accessed 11.03.15].

[33] Brooks RA, Di Chiro GD. Statistical limitations in x-ray reconstructive tomography. Med Phys 1976;3:237–40. doi:10.1118/1.594240.

[34] Riederer SJ, Pelc NJ, Chesler DA. The noise power spectrum in computed X-ray tomography. Phys Med Biol 1978;23:446–54.

[35] Atkinson JK. The quantitative assessment of CT scanners. University of London; 1980.

[36] Edyvean S. ImPACT MDA Report type testing of CT scanners – methods and MEthodology, <http://www.impactscan.org/reports/MDA9825.htm>; 1998 [accessed 11.03.15].

[37] Fuchs T, Kalender WA. On the correlation of pixel noise, spatial resolution and dose in computed tomography: theoretical prediction and verification by simulation and measurement. Phys Med 2003;XIX(2):153–64.

[38] Kalender WA. Computed tomography: fundamentals, system technology, image quality, applications. 3rd ed. Erlangen: Publicis; 2011.

[39] Bassano DA. Specification and quality assurance for CT scanners. AAPM Summer Sch 1980.

[40] Allisy-Roberts P, Williams JR. Farr's physics for medical imaging. Elsevier Health Sciences; 2007.

[41] Edyvean S. ImPACT NHS PASA 16 slice CT scanner comparison report version 14. <http://www.impactscan.org/reports/Report06012.htm>; 2006 [accessed 11.03.15].

[42] Edyvean MS. A methodical approach for comparison of CT scanner image quality relative to dose. Radiological Society of North America 2003 Scientific Assembly and Annual Meeting, November 30–December 5, 2003, Chicago, IL. <http://archive.rsna.org/2003/3107396.html> [accessed 29.04.15].

[43] Platten D, Keat N, Lewis M, Barret J, Edyvean S. ImPACT MHRA 04045 Toshiba 16 Report, page 26, <http://www.impactscan.org/reports/MHRA04045.htm>; 2003 [accessed 30.03.15].

[44] Nagel H-D. CT dose efficiency parameters. European medical ALARA network. WG 1: optimisation of patient exposure in CT procedures, synthesis document 2012. <https://www.yumpu.com/en/document/view/42382097/wg1-synthesis-report-pdf-a-11-mb-european-medical-alara-/5>; 2004 [accessed 17.02.15].

[45] Chao EH, Toth TL, Williams EC, Fox SH, Carleton CA, Bromberg NB. A statistical method of defining low contrast detectability, poster presented at RSNA Meeting; 2000.

[46] Rose A. Vision: human and electronic. New York: Plenum Press; 1973.

[47] Torgersen GR, Hol C, Møystad A, Hellén-Halme K, Nilsson M. A phantom for simplified image quality control of dental cone beam computed tomography units. Oral Surg Oral Med Oral Pathol Oral Radiol 2014;118:603–11. doi:10.1016/j.oooo.2014.08.003.

[48] Nagel H-D. Methoden zur bestimmung der dosiseffizienz von CT-scannern, presented at APT meeting; 2008.

[49] Månsson LG. Methods for the evaluation of image quality: a review. Radiat Prot Dosimetry 2000;90:89–99.

[50] Beutel J, Kundel HL, Van Metter RL. Handbook of medical imaging: physics and psychophysics. SPIE Press; 2000.

[51] 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.

[52] Likert R. A technique for the measurement of attitudes. Arch Psychol 1932;22(140):55.

[53] Jamieson S. Likert scales: how to (ab)use them. Med Educ 2004;38:1217–18. doi:10.1111/j.1365-2929.2004.02012.x.

[54] Norman G. Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ 2010;15:625–32. doi:10.1007/s10459-010-9222-y.

[55] International Commission on Radiation Units and Measurements. Receiver operating characteristic analysis in medical imaging, vol. 79. ICRU Rep. N°79. Bethesda (MD): International Commission on Radiation Units and Measurements; 2008.

[56] Samei E, Krupinski E. The handbook of medical image perception and techniques. Cambridge UK: Cambridge University Press; 2014.

[57] Barrett HH, Myers KJ. Foundations of image science. Hoboken (NJ): Wiley-Interscience; 2004.

[58] Vennart W. ICRU Report 54: medical imaging – the assessment of image quality. Radiography 1996;3:243–4. doi:10.1016/S1078-8174(97)90038-9.

[59] Way T, Chan H-P, Hadjiiski L, Sahiner B, Chughtai A, Song TK, et al. Computer-aided diagnosis of lung nodules on CT scans: ROC study of its effect on radiologists' performance. Acad Radiol 2010;17:323–32. doi:10.1016/j.acra.2009.10.016.

[60] Li F, Aoyama M, Shiraishi J, Abe H, Li G, Suzuki K, et al. Radiologists' performance for differentiating benign from malignant lung nodules on high-resolution CT using computer-estimated likelihood of malignancy. Am J Roentgenol 2004;183:1209–15.

[61] Okumura M, Ota T, Kainuma K, Sayre JW, McNitt-Gray M, Katada K. Effect of edge-preserving adaptive image filter on low-contrast detectability in CT systems: application of ROC analysis. Int J Biomed Imaging 2008;2008:379486. doi:10.1155/2008/379486.

[62] Bushberg JT, Seibert JA, Leidholdt EM, Boone JM. The essential physics of medical imaging. Philadelphia (PA): Wolters Kluwer Health; 2011.

[63] Wunderlich A, Abbey CK. Utility as a rationale for choosing observer performance assessment paradigms for detection tasks in medical imaging. Med Phys 2013;40:111903. doi:10.1118/1.4823755.

[64] Chakraborty DP. A brief history of FROC paradigm data analysis. Acad Radiol 2013;20:915–19. doi:10.1016/j.acra.2013.03.001.

[65] Popescu LM. Nonparametric signal detectability evaluation using an exponential transformation of the FROC curve. Med Phys 2011;38:5690–702. doi:10.1118/1.3633938.

[66] Burgess A. Comparison of receiver operating characteristic and forced-choice observer performance-measurement methods. Med Phys 1995;22:643–55. doi:10.1118/1.597576.

[67] Macmillan NA, Creelman CD. Detection theory: a user's guide. Mahwah (NJ): Lawrence Erlbaum Associates; 2005.

[68] Bochud FO, Abbey CK, Eckstein MP. Visual signal detection in structured backgrounds. III. Calculation of figures of merit for model observers in statistically nonstationary backgrounds. J Opt Soc Am A Opt Image Sci Vis 2000;17:193–205. doi:10.1364/JOSAA.17.000193.

[69] Swets JA. Signal detection and recognition by human observers: contemporary readings. New York: Wiley; 1964.
[70] Craven BJ. A table of d′ for M-alternative odd-man-out forced-choice procedures. Percept Psychophys 1992;51:379–85.
[71] Hacker MJ, Ratcliff R. A revised table of d′ for M-alternative forced choice. Percept Psychophys 1979;26:168–70. doi:10.3758/BF03208311.
[72] Dahlquist G, Björck Å. Numerical methods. New York: Courier Corporation; 2012.
[73] Green DM, Swets JA. Signal detection theory and psychophysics. New York: John Wiley and Sons; 1966.
[74] Burgess AE. Visual perception studies and observer models in medical imaging. Semin Nucl Med 2011;41:419–36. doi:10.1053/j.semnuclmed.2011.06.005.
[75] Tapiovaara M. Efficiency of low-contrast detail detectability in fluoroscopic imaging. Med Phys 1997;24:655–64. doi:10.1118/1.598076.
[76] Eckstein MP, Abbey CK, Bochud FO. Visual signal detection in structured backgrounds. IV. Figures of merit for model performance in multiple-alternative forced-choice detection tasks with correlated responses. J Opt Soc Am A Opt Image Sci Vis 2000;17:206–17. doi:10.1364/JOSAA.17.000206.
[77] Zhang Y, Leng S, Yu L, Carter RE, McCollough CH. Correlation between human and model observer performance for discrimination task in CT. Phys Med Biol 2014;59:3389–404. doi:10.1088/0031-9155/59/13/3389.
[78] 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.
[79] 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.
[80] Monnin P, Marshall NW, Bosmans H, Bochud FO, Verdun FR. Image quality assessment in digital mammography: part II. NPWE as a validated alternative for contrast detail analysis. Phys Med Biol 2011;56:4221–38. doi:10.1088/0031-9155/56/14/003.
[81] Borasi G, Samei E, Bertolini M, Nitrosi A, Tassoni D. Contrast-detail analysis of three flat panel detectors for digital radiography. Med Phys 2006;33:3580. doi:10.1118/1.2337636.
[82] Rivetti S, Lanconelli N, Bertolini M, Acchiappati D. A new clinical unit for digital radiography based on a thick amorphous selenium plate: physical and psychophysical characterization. Med Phys 2011;38:4480–8. doi:10.1118/1.3605471.
[83] Wagner RF, Brown DG, Pastel MS. Application of information theory to the assessment of computed tomography. Med Phys 1979;6:83–94.
[84] Myers KJ, Rolland JP, Barrett HH, Wagner RF. Aperture optimization for emission imaging: effect of a spatially varying background. J Opt Soc Am A 1990;7:1279–93.
[85] Burgess AE, Ghandeharian H. Visual signal detection. II. Signal-location identification. J Opt Soc Am A 1984;1:906–10. doi:10.1364/JOSAA.1.000906.
[86] Burgess A. Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation. J Opt Soc Am A 1985;2:1498–507.
[87] Wagner RF, Myers KJ, Tapiovaara MJ, Brown DG, Burgess AE. Schneider RH, editor. Maximum a-posteriori detection and figures of merit for detection under uncertainty. 1990. p. 195–204 doi:10.1117/12.18797.
[88] Chesters MS. Human visual perception and ROC methodology in medical imaging. Phys Med Biol 1992;37:1433. doi:10.1088/0031-9155/37/7/001.
[89] Brown DG, Insana MF, Tapiovaara M. Detection performance of the ideal decision function and its McLaurin expansion: signal position unknown. J Acoust Soc Am 1995;97:379–98.
[90] Bochud FO, Abbey CK, Bartroff J, Vodopich D, Eckstein MP. Effect of the number of locations in MAFC experiments performed with mammograms 1999.
[91] Burgess AE. Evaluation of detection model performance in power-law noise, vol. 4324. 2001. p. 123–32 doi:10.1117/12.431180.
[92] Kotre CJ. The effect of background structure on the detection of low contrast objects in mammography. Br J Radiol 1998;71:1162–7. doi:10.1259/bjr.71.851.10434911.
[93] Bochud FO, Valley JF, Verdun FR, Hessler C, Schnyder P. Estimation of the noisy component of anatomical backgrounds. Med Phys 1999;26:1365–70.
[94] Marshall NW, Kotre CJ, Robson KJ, Lecomber AR. Receptor dose in digital fluorography: a comparison between theory and practice. Phys Med Biol 2001;46:1283–96.
[95] 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.
[96] Samei E, Richard S. Assessment of the dose reduction potential of a model-based iterative reconstruction algorithm using a task-based performance metrology. Med Phys 2015;42:314–23. doi:10.1118/1.4903899.
[97] Chen B, Ramirez Giraldo JC, Solomon J, Samei E. Evaluating iterative reconstruction performance in computed tomography. Med Phys 2014;41:121913. doi:10.1118/1.4901670.
[98] Tapiovaara MJ, Wagner RF. SNR and noise measurements for medical imaging: I. A practical approach based on statistical decision theory. Phys Med Biol 1993;38:71. doi:10.1088/0031-9155/38/1/006.
[99] Rolland JP, Barrett HH. Effect of random background inhomogeneity on observer detection performance. J Opt Soc Am A 1992;9:649–58. doi:10.1364/JOSAA.9.000649.
[100] Tapiovaara MJ. SNR and noise measurements for medical imaging. II. Application to fluoroscopic X-ray equipment. Phys Med Biol 1993;38:1761. doi:10.1088/0031-9155/38/12/006.
[101] Boedeker KL, Cooper VN, McNitt-Gray MF. Application of the noise power spectrum in modern diagnostic MDCT: part I. Measurement of noise power spectra and noise equivalent quanta. Phys Med Biol 2007;52:4027–46. doi:10.1088/0031-9155/52/14/002.
[102] Burgess AE. Statistically defined backgrounds: performance of a modified nonprewhitening observer model. J Opt Soc Am A Opt Image Sci Vis 1994;11:1237–42.
[103] Loo LN, Doi K, Metz CE. A comparison of physical image quality indices and observer performance in the radiographic detection of nylon beads. Phys Med Biol 1984;29:837–56.
[104] Zhang Y, Pham BT, Eckstein MP. Evaluation of internal noise methods for Hotelling observer models. Med Phys 2007;34:3312–22. doi:10.1118/1.2756603.
[105] Brankov JG. Evaluation of channelized Hotelling observer with internal-noise model in a train-test paradigm for cardiac SPECT defect detection. Phys Med Biol 2013;58:7159–82. doi:10.1088/0031-9155/58/20/7159.
[106] 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.
[107] Brankov JG. Optimization of the internal noise models for channelized Hotelling observer. 2011 IEEE Int Symp Biomed Imaging Nano Macro 2011;1788–91. doi:10.1109/ISBI.2011.5872753.
[108] Hernandez-Giron I, Calzado A, Geleijns J, Joemai RMS, Veldkamp WJH. Comparison between human and model observer performance in low-contrast detection tasks in CT images: application to images reconstructed with filtered back projection and iterative algorithms. Br J Radiol 2014;87:20140014. doi:10.1259/bjr.20140014.
[109] 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.
[110] Abbey CK, Barrett HH. Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability. J Opt Soc Am A Opt Image Sci Vis 2001;18:473–88.
[111] 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.
[112] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[113] Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat 2007;17:571–82. doi:10.1080/10543400701329422.

4. Conclusions and perspectives

SIR algorithms were released a few years ago onto the CT market. They are now widely used in clinical routines in order to reduce the dose delivered to patients during CT examinations and/or improve image quality. By investigating image quality using both a clinical and physical approach, this work has shown that SIR are efficient solutions to reduce dose in CT procedures while maintaining the information necessary to perform an accurate diagnosis.

Previous work already showed that simple scalar metrics based on noise measurements in the spatial domain (like SD, CNR or SNR) was not sufficient to optimise CT protocols when dealing with IR algorithms [Miéville PhD, 2012]. Indeed, the outcome of these simple metrics is systematically improved with IR, because they integrate knowledge of the noise statistics to penalise pixel variations and reduce noise. Miéville's work [Miéville PhD, 2012] therefore suggests relying on Fourier-based metrics to get more information, especially about the way the texture of the noise is modified when using SIR. In this context, our clinical studies showed that Fourier-space tools (mostly NPS and MTF for the respective assessment of noise and spatial resolution) are useful for helping radiologists select the best option when performing a clinical examination. However, these studies also highlight the need for developing more elaborate tools, which are in compliance with the requirements of IR techniques. If only small adaptations are necessary regarding the noise estimation using the NPS (like using small ROIs or focusing the measurement on particular regions of the image in order to overcome noise stationarity problems), things are different for the spatial resolution estimation using the MTF since this parameter might also be signal dependent.

Therefore, we developed and tested the TTF metric, the new spatial resolution metric resulting from the above-mentioned considerations, in the second part of our study (i.e. physical approach). It managed to overcome the linearity problems of images reconstructed using IR algorithms and to yield reliable resolution measurements. In spite of the complexity of the detection processes performed by radiologists, the use of the TTF enabled us to better understand radiologist impressions when dealing with the examination of relatively high-contrast structures. This metric was in particular quite useful to investigate the spatial resolution reduction that could be associated to multi-plane image reformatting (MPR).

Having investigated the problem of spatial resolution characterisation when dealing with IR, we then focused on the detection of low-contrast structures, a task that is strongly dependent on the noise aspect of the image. If the use of NPS makes it possible to explain certain changes in image texture, it cannot be directly linked with the detection of structures on radiological images. To get some information concerning the detectability of structures one needs to use task-based type metrics. In this context, model observers are particularly appealing and can be seen as a reliable solution for objective image quality assessment in relation to a radiologist's task. In this work we were able to optimise CT protocols by establishing links between an objective image quality parameter measured by model observers on phantoms and the actual detection of simple structures by human observers. Several types of model observers were used regarding their properties in relation to our experimental paradigm. In a clinical paradigm where model and human findings on image quality had to be compared, we relied on an anthropomorphic model observer, whereas a more physical paradigm in which only purely physical image quality parameters had to be estimated meant the use of ideal model observers. Similarly, our method also relied on the use of model observers in the Fourier space for high-contrast lesions, whereas the detectability of low-contrast lesions was assessed using image space model observers. Using this methodology, we were able to demonstrate that among the different SIR, MBIR constitutes an efficient type of reconstruction algorithm, since its use enables a drastic dose reduction while preserving the detectability of low-contrast structures. As a matter of fact, some of our results indicate that in certain conditions, a dose level between 1 and 5 mGy would still allow for an accurate diagnosis [Ott, 2015 a; Ott, 2015 b].

Although this work has led to several answers regarding the use of SIR algorithms in the clinic and their potential dose benefits without impacting image quality, there are still several open questions on the topic. Some work remains to be done in order to ensure that the outcome of the model observer matches human detection. To get a reasonable match between the CHO and human observers it is necessary to introduce a noise function into the mathematical model observer. One should ensure that the noise function introduced into the model is robust enough to the change of reconstruction algorithm. Current model observers produce reasonable results when dealing with the detection of simple structures in homogeneous backgrounds. This is obviously an oversimplification of the actual radiological problem. Thus, a great amount of work remains to be done by medical physicists to better understand the reliability of the model observer approach. Furthermore, we need to know the type of signals that are very important to transfer to the image to ensure an adequate diagnosis. This information is the basis of the optimisation process. This work has shown that a close collaboration between the world of medical physics and radiology is necessary to take full advantage of the tools offered by manufacturers. In this sense, we suggest that the work here helps open a new type of collaboration between medical physicists and radiologists.

Lastly, if all patients undergoing CT examinations could take advantage of the above-described optimisation scheme, we believe that children and young adults could nevertheless retrieve an even bigger benefit out of it. Indeed, those populations are particularly sensitive to radiations effects and should therefore be handled cautiously regarding the dose amount they receive when undergoing medical examinations.

References

[AAPM 96, 2008] American Association of Physicists in Medicine, “*The measurement, reporting, and management of radiation dose in CT*.” 2008 Report No. 96 of AAPM Task Group 23 of the Diagnostic Imaging Council CT Committee, AAPM College Park.

[Abbey and Barrett, 2001] C. K. Abbey and H. H. Barrett, *“Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability”* 2001 J. Opt. Soc. Am. **18(3)**:473-88.

[Aroua, 2002] A. Aroua, B. Burnand, I. Decka, J. P. Vader, and J. F. Valley, *“Nationwide Survey on Radiation Doses in Diagnostic and Interventional Radiology in Switzerland in 1998”* 2002 Health Phys **83**, 46–55.

[Barrett, 1993] H. H. Barrett, J. Yao, J. P. Rolland and K. J. Myers, *"Model observer for assessment of image quality"* 1993 Proc Natl Acad Sci **90** 9758-65

[Barrett, 2004] H. H. Barrett and K. J. Myers, “*Foundations of Image Science*.” Wiley, New York, 2004.

[Berrington de González, 2004] A. Berrington de González and S. Darby, *“Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries.”* 2004 Lancet **31; 363(9406)**:345-51.

[Beutel, 2000] J. Beutel, H. L. Kundel and R. L. Van Metter, *"Handbook of medical imaging Vol 1. Physics and psychophysics"* 2000 SPIE press.

[Boone, 2001] J. M. Boone, *"Determination of the presampled MTF in computed tomography."* 2001 Med Phys **28**:356–60.

[Brenner, 2008] D. J. Brenner and E. J. Hall, *“Cancer risks from diagnostic radiology”* 2008 Br. J. Radiol. **81(965)**:362–378

[Brunner, 2013] C. C. Brunner and I. S. Kyprianou, *"Material-specific transfer function model and SNR in CT"* 2013 PhysMed Biol **58** 7447-61.

[Burgess, 1994] A. E.Burgess, *"Statistically defined backgrounds: performance of a modified non prewhitening observer model"* 1994 JOSA A **11(4)**:1237-42.

[Buvat, 2011] Consulted on August, 21st 2015. Available from: http://www.guillemet.org/irene/coursem/Lyon_Recon2011.pdf

[Deak, 2009] P. D. Deak, O. Langner, M. Lell, and W. A. Kalender, *“Effects of adaptive section collimation on patient radiation dose in multisection spiral CT.”* 2009 Radiology **252**, 140–147.

[den Harder, 2015] A. M. den Harder, M.J. Willemink, Q. M. de Ruiter, A. M. Schilham, G.P. Krestin, T. Leiner, P.A. de Jong and R. P. Budde, *“Achievable dose reduction using iterative reconstruction for chest computed tomography: A systematic review.”* 2015 Eur J Radiol. **S0720-048X(15)**:30050-4.

[Dixon, 2003] R. L. Dixon, *“A new look at CT dose measurement: Beyond CTDI”* 2003 Med Phys **30**, 1272–1280.

[Durand, 2011] D. J. Durand, *"A rational approach to the clinical use of cumulative effective dose estimates"* 2011 Am J Roentgenol **197**, 160–162.

[EURATOM, 2013] *Council Directive – EURATOM 2313/59*. Consulted on August, 6th 2015. Available from: http://eur-x.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2014:013:0001:0073:EN:PDF

[Escalon, 2015] J. G. Escalon, M.B. Chatfield, D. Sengupta and M. L. Loftus, *"Dose Length Products for the 10 Most Commonly Ordered CT Examinations in Adults: Analysis of Three Years of the ACR Dose Index Registry."* 2015 J Am Coll Radiol **12(8)**:815-23.

[Fessler, 2009] Consulted on September, 6th 2015. Available from: http://web.eecs.umich.edu/~fessler/book/deman-10-sir.pdf

[Frush, 2002] D. P. Frush, B. Soden, K. S. Frush and C. Lowry *"Improved pediatric multidetector body CT using a size-based color-coded format."* 2002 AJR Am J Roentgenol **178(3)**:721-6.

[Geyer, 2015] L. L. Geyer, U. J. Schoepf, F. G. Meinel, J. W. Nance Jr, G. Bastarrika, J. A. Leipsic, N. S. Paul, M. Rengo, A. Laghi and C. N. De Cecco, *"State of the Art: Iterative CT Reconstruction Techniques."* 2015 Radiology **276(2)**:339-57.

[Hara, 2009] A. K. Hara, R. G. Paden, A. C. Silva, J. L. Kujak, H. J. Lawder, and W. Pavlicek, *"Iterative Reconstruction Technique for Reducing Body Radiation Dose at CT: Feasibility Study"* 2009 Am J Roentgenol **193**, 764–771.

[He, 2013] X. He and S. Park, *"Model observer in medical imaging research"* 2013 Theranostics **3(10)** 774–86.

[Herman, 1980] G. T. Herman, *"Image Reconstruction from Projections: The Fundamentals of Computerised Tomography"* 1980 Academic Press, New York, USA.

[Hernandez-Giron, 2011] I. Hernandez-Giron, J. Geleijns, A. Calzado and W. J. Veldkamp, *"Automated assessment of low contrast sensitivity for CT systems using a model observer"* 2011 Med. Phys. **38** Suppl 1:S25

[Hernandez-Giron, 2014] I. Hernandez-Giron, A. Calzado, J. Geleijns, R. M. Joemai and W. J. Veldkamp, *"Comparison between human and model observer performance in low-contrast detection tasks in CT images: application to images reconstructed with filtered back projection and iterative algorithms."* 2014 Br J Radiol **87(1039)**:20140014.

[Howard, 2014] M. E. Howard, C. H. McCollough, S. Leng, L. Yu and M. R. Bruesewitz, *"Use of CT dose notification and alert values in routine clinical practice."* 2014 J Am Coll Radiol. **11 (5)**:450-5.

[Hsieh, 2009] J. Hsieh. *"Computed Tomography: Principles, Design, Artifacts, and Recent Advances."* 2009 SPIE, Bellingham, WA.

[Hsieh, 2013] J. Hsieh, B. Nett, Z. Yu, K. Sauer, J.-B. Thibault and C. A. Bouman, *"Recent Advances in CT Image Reconstruction"* 2013 Curr Radiol Rep.

[IAEA, 2014] Radiation Protection and Safety of Radiation Sources: International Basic Safety Standard. IAEA Safety Standard. Vienna.

[ICRU 54, 1996] International Commission on Radiation Units and Measurements, "*Medical Imaging: The Assessment of Image Quality*." 1996 Report 54 ICRU Publications, Bethesda, MD.

[ICRU 87, 2012] International Commission on Radiation Units and Measurements, *"Radiation dose and image-quality assessment in computed tomography."* 2012 Report 87 ICRU Publications, Bethesda, MD.

[ICRP 103, 2007] International Commission on Radiological Protection (ICRP), *"Recommendations of the International Commission on Radiological Protection."* 2007 Publication 103 Ann. ICRP 37.

[Imanishi, 2015] Y. Imanishi, A. Fukui, H. Niimi, D. Itoh, K. Nozaki , S. Nakaji, K. Ishizuka, H. Tabata, Y. Furuya, M. Uzura, H. Takahama, S. Hashizume, S. Arima and Y. Nakajima, *"Radiation-induced temporary hair loss as a radiation damage only occurring in patients who had the combination of MDCT and DSA."* 2005 Eur Radiol **15**:41-46.

[Judy, 1976] P. F. Judy, *"The line spread function and modulation transfer function of a computed tomographic scanner."* 1976 Med Phys **3**:233–6.

[Kaza, 2014] R. K. Kaza, J. F. Platt, M. M. Goodsitt, M. M. Al-Hawary, K. E. Maturen,A. P. Wasnik and A. Pandya, *"Emerging techniques for dose optimization in abdominal CT."* 2014 Radiographics **34(1)**:4-17.

[Kofler, 2014] J. M. Kofler, D. D. Cody and R. L. Morin, *"CT protocol review and optimization."* 2014 J Am Coll Radiol **11(3)**:267-70.

[Kofler, 2015] J. M. Kofler, L. Yu, S. Leng, Y. Zhang, Z. Li, R. E. Carter and C. H. McCollough, *"Assessment of Low-Contrast Resolution for the American College of Radiology Computed Tomographic Accreditation Program: What Is the Impact of Iterative Reconstruction?"* 2015 J Comput Assist Tomogr **39(4)**:619-23.

[Le Coultre, 2016] R. Le Coultre, J. Bize, M. Champendal, D. Wittwer, N. Ryckx, A. Aroua, P. Trueb and F. R. Verdun. *"Exposure of the Swiss population by radiodiagnostics: 2013 review."* Rad. Prot. Dosim. In press.

[MacGregor, 2015] K. MacGregor, I. Li, T. Dowdell, B. G. Gray. "Identifying Institutional Diagnostic Reference Levels for CT with Radiation Dose Index Monitoring Software." K. MacGregor, I. Li, T. Dowdell and B. G. Gray. 2015 Radiology **276(2)**:507-17.

[McCollough, 2012] C. H. McCollough, G. H. Chen, W. Kalender, S. Leng, E. Samei, K. Taguchi K, G. Wang, L. Yu and R. I. Pettigrew, *"Achieving routine submillisievert CT scanning: report from the summit on management of radiation dose in CT."* 2012 Radiology **264(2)**:567-80.

[Metz, 2006] C. E. Metz, *"Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems Journal of the American College of Radiology?"* 2006 J Am Coll Radiol **3(6)**:413-22.

[Miéville, 2011] F. A. Miéville, F. Gudinchet, E. Rizzo, P. Ou, F. Brunelle, F. Bochud, and F R. Verdun, *"Pediatric cardiac CT examinations: impact of the iterative reconstruction method (ASIR) on image quality – a preliminary study"* 2011 Pediatr Radiol **41**, 1154–1164.

[Miéville PhD, 2012] F. A. Miéville, *"Assessment of statistical iterative reconstructions to optimize CT examination protocols for children and young adults"* 2012 Unil.

[Myers, 1987] K. J. Myers and H. H. Barrett H H, *"Addition of a channel mechanism to the ideal-observer model"* 1987 J Opt Soc Am **4(12)**:2447-57.

[Nakaya, 2012] Y. Nakaya, Y. Kawata, N. Niki, K. Umetatni, H. Ohmatsu and N. Moriyama, *"A method for determining the modulation transfer function from thick microwire profiles measured with x-ray microcomputed tomography."* 2012 Med Phys **39**:4347–64.

[Naoum, 2015] C. Naoum, P. Blanke and J. Leipsic, *“Iterative reconstruction in cardiac CT.”* 2015 J Cardiovasc Comput Tomogr. **9(4)**:255-63.

[NHS, 2012] Consulted on August 6th 2015. Available from:
http://www.nhs.uk/news/2012/06june/Pages/ct-scan-child-cancer-brain-tumours.aspx

[Nickoloff, 1985] E. L. Nickoloff and R. Riley, *"A simplified approach for modulation transfer function determinations in computed tomography"* 1985 Med Phys **12** 437-42.

[Noda, 2015] Y. Noda, M. Kanematsu, S. Goshima, H. Kondo, H. Watanabe, H. Kawada, N. Kawai, Y. Tanahashi, T. R. Miyoshi and K. T. Bae, *"Reducing iodine load in hepatic CT for patients with chronic liver disease with a combination of low-tube-voltage and adaptive statistical iterative reconstruction."* 2015 Eur J Radiol. **84(1)**:11-8.

[Omoumi, 2014] P. Omoumi, F. R. Verdun, Y. B. Salah, B. C. Berg , F. E. Lecouvet, J. Malghem, J. G. Ott, R. Meuli and F. Becce, Low-dose multidetector computed tomography of the cervical spine: optimization of iterative reconstruction strength levels. Acta Radiologica **55(3)** pp. 335-344.

[Omoumi, 2015] P. Omoumi, F. R. Verdun, J. G. Ott, D. Racine, F. Becce, Optimization of radiation dose and image quality in musculoskeletal CT: Emphasis on iterative reconstruction techniques (part 1). Semin Musculoskelet Radiol **19**: 415-21

[Ott, 2014] J. G. Ott, F. Becce, P. Monnin, S. Schmidt, F.O. Bochud, F. R. Verdun, *“Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms.”* 2014 Phys Med Biol **59(15)** pp. 4047-4064.

[Ott, 2015 a] J. G. Ott, A. Ba, D. Racine, A. Viry, F. O. Bochud and F. R. Verdun, *“Assessment of low contrast detection in CT using model observers: Developing a clinically-relevant tool for characterizing adaptive statistical and model-based iterative reconstruction.”* Submitted to Zeitschrift für Medizinische Physik in 2015.

[Ott, 2015 b] J. G. Ott, A. Ba, D. Racine, N. Ryckx, F. O. Bochud, H. Alkadhi and F. R. Verdun, *“Patient exposure optimisation through task-based assessment of a new model-based iterative reconstruction technique.”* 2015 Accepted for publication in Radiation Protection Dosimetry.

[Ott, 2015 c] J. G. Ott, P. Omoumi, E. Dugert, F. O. Bochud, F. R. Verdun and F. Becce, *“Task-based assessment of an MBIR algorithm: A study in all reconstruction planes.”* Submitted to investigative radiology.

[Padole, 2015] A. Padole, S. Singh, D. Lira , M. A. Blake, S. Pourjabbar, R. D. Khawaja, G. Choy, S. Saini, S. Do and M. K. Kalra, *“Assessment of Filtered Back Projection, Adaptive Statistical and Model-Based Iterative Reconstruction for Reduced Dose Abdominal Computed Tomography.”* 2015 J Comput Assist Tomogr. **39(4)**:462-7.

[Patino, 2015] M. Patino, J. M. Fuentes, S. Singh, P. F. Hahn and D. V. Sahani, *"Iterative Reconstruction Techniques in Abdominopelvic CT: Technical Concepts and Clinical Implementation."* 2015 AJR **205(1)**:W19-31.

[Pearce, 2012] M. S. Pearce, J. A. Salotti, M. P. Little, K. McHugh, C. Lee, K. P. Kim, N. L. Howe, C. M. Ronckers, P. Rajaraman, A. W. Craft, L. Parker and A. Berrington de Gonzálezc, *"Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study"* 2012 Lancet. **380(9840)**: 499–505.

[Pineda, 2008] A. R. Pineda, J. H. Siewerdsen, and D. J. Tward, *"Analysis of Image Noise in 3D Cone-Beam CT: Spatial and Fourier Domain Approaches under Conditions of Varying Stationarity" 2008* Proc. SPIE **6913**, 69131Q-1–69131Q-10.

[Rehani, 2015] M. M. Rehani, R. Gupta, S. Bartling, G. C. Sharp, R. Pauwels, T. Berris and J. M. Boone, *"Radiological Protection in Cone Beam Computed Tomography (CBCT)."* 2015 Ann ICRP. **44(1)**:9-127.

[Richard, 2012] S. Richard, D. B. Husarik, G. Yadava, S. N. Murphy and E. Samei, *"Towards task-based assessment of CT performance: System and object MTF across different reconstruction algorithms"* 2012 Med Phys **39** 4115-22.

[Samara, 2012] E. T. Samara, A. Aroua, F. O. Bochud, B. Ott, T. Theiler, R. Treier, P. R. Trueb, J. P. Vader, and F. R. Verdun, *"Exposure of the Swiss population by medical X-rays: 2008 review"* 2012 Health Phys **102(3)**:263-70.

[Samei, 2015] E. Samei and S. Richard, *"Assessment of the dose reduction potential of a model-based iterative reconstruction algorithm using a task-based performance metrology."* 2015 Med Phys. **42(1)**:314-23.

[Schindera, 2011] S. T. Schindera, L. Diedrichsen, H. C. Müller, O. Rusch, D. Marin, B. Schmidt, R. Raupach, P. Vock and Z. Szucs-Farkas, *"Iterative reconstruction algorithm for abdominal multidetector CT at different tube voltages: assessment of diagnostic accuracy, image quality, and radiation dose in a phantom study."* 2011 Radiology **260(2)**:454-62.

[Schindera, 2013] S. T. Schindera, D. Odedra, S. A. Raza, T. K. Kim, H. J. Jang, Z. Szucs-Farkas and P. Rogalla, *"Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved?"* 2013 Radiology **269(2)**:511-8.

[Singh, 2011] S. Singh, M. K. Kalra, J. H. Thrall and M. Mahesh, *"Pointers for optimizing radiation dose in head CT protocols."* 2011 J Am Coll Radiol. **8(8)**:591-3.

[Smedby, 2013] O. Smedby, M. Fredrikson, J. De Geer, L. Borgen and M. Sandborg, *"Quantifying the potential for dose reduction with visual grading regression."* 2013 Br J Radiol. **86(1021)**:31197714.

[Solomon, 2015] J. Solomon, A. Mileto, J. C. Ramirez-Giraldo and E. Samei, *"Diagnostic Performance of an Advanced Modeled Iterative Reconstruction Algorithm for Low-Contrast Detectability with a Third-Generation Dual-Source Multidetector CT Scanner: Potential for Radiation Dose Reduction in a Multireader Study."* 2015 Radiology **275(3)**:735-45.

[Suh, 2013] Y. J. Suh, Y. J. Kim, S. R. Hong, Y. J. Hong, H. J. Lee, J. Hur and B. W. Choi. *"Combined use of automatic tube potential selection with tube current modulation and iterative reconstruction technique in coronary CT angiography."* 2013 Radiology **269(3)**:722-9.

[Taylor 2015] S. Taylor, D. E. Litmanovich, M. Shahrzad, A. A. Bankier, P .A. Gevenois and D. Tack. *“Organ-based tube current modulation: are women's breasts positioned in the reduced-dose zone?”* Radiology **274(1)**:260-6.

[Thibault, 2007] J. B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, *“A three-dimensional statistical approach to improved image quality for multislice helical CT” 2007* Med Phys **34**, 4526–4544.

[Tobalem & Dugert, 2014] F. Tobalem, E. Dugert, F. R. Verdun, V. Dunet, J. G. Ott, H. A. Rudiger, S. Cherix, R. Meuli, F. Becce 2014 MDCT Arthrography of the Hip: Value of the Adaptive Statistical Iterative Reconstruction Technique and potential for Radiation Dose Reduction. AJR **203(6)**:W665-73

[Tseng, 2014] H. W. Tseng, J. Fan, M. A. Kupinski, P. Sainath and J. Hsieh, *"Assessing image quality and dose reduction of a new x-ray computed tomography iterative reconstruction algorithm using model observers"* 2014 Med. Phys. **41(7)**.

[Vaishnav, 2014] J. Y. Vaishnav, W. C. Jung, L. M. Popescu, R. Zeng and K. J. Myers, *"Objective assessment of image quality and dose reduction in CT iterative reconstruction" 2014 Med Phys* **41(7)**:071904.

[Yu, 2013] L. Yu, S. Leng, L. Chen, M. Kofler, R. E. Carter and C. H. McCollough, *"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"* 2013 Med. Phys. **40(4)**.

[Zarb, 2015] F. Zarb, M. F. McEntee and L. Rainford, *"Visual grading characteristics and ordinal regression analysis during optimisation of CT head examinations."* 2015 Insights Imaging **6(3)**:393-401.

[Zhang, 2014] Y. Zhang, S. Leng, L. Yu, R. E. Carter and C. H. McCollough, *"Correlation between human and model observer performance for discrimination task in CT.*" 2014 Phys. Med. Biol. **59**:3389.

Appendix: Details of FBP reconstruction

This appendix presents the main steps for the derivation of the filtered back projection formula. Developments are based on [Hsieh, 2009].

First step entails the demonstration of the Fourier slice theorem (FST).
This theorem stats that, considering a 2D function f(x,y): the Fourier transform of a parallel projection of $f(x,y)$ done at an angle θ is equivalent to the 2D Fourier transform of $f(x,y)$ along a line going through the origin and taken at the same angle θ in the Fourier space.
This situation can be summarised according to Figure A1:

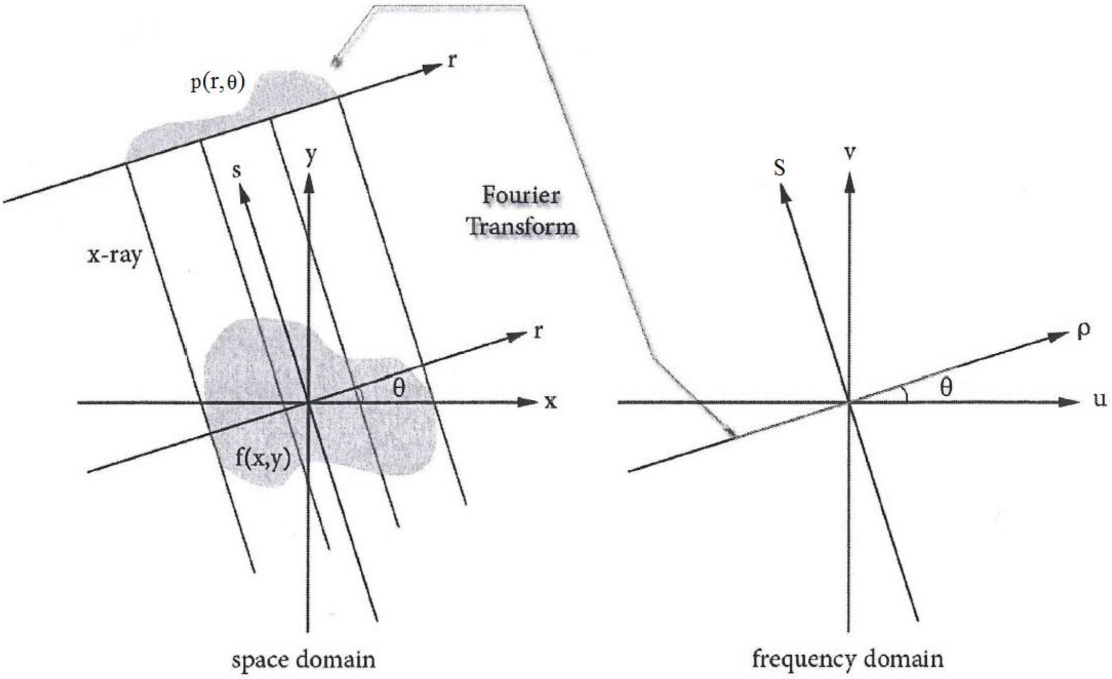


Figure 1A: Illustration of the Fourier slice theorem.

The transformations between the two reference systems $(x, y) \rightarrow (r, s)$ can be summarised as:

$$\begin{cases} s = x\cos\theta - y\sin\theta \\ r = x\sin\theta + y\cos\theta \end{cases} \quad and \quad drds = Jdxdy = \underbrace{\begin{vmatrix} \frac{\partial r}{\partial x} & \frac{\partial s}{\partial x} \\ \frac{\partial r}{\partial y} & \frac{\partial s}{\partial y} \end{vmatrix}}_{1} dxdy \quad and \quad f(x,y) = f'(r,s) \qquad (\boldsymbol{A.1})$$

The projection of $f(x,y)$ at an angle θ, along the s axis can be written as:

$$p(r,\theta) = \int_{-\infty}^{+\infty} f'(r,s)ds \qquad (\boldsymbol{A.2})$$

The Fourier transform of $p(r,\theta)$ over r can then be written as:

$$P(\rho,\theta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f'(r,s)ds\, e^{-2i\pi\rho r} dr \qquad (\boldsymbol{A.3})$$

Using the elements stated in Eq. A.1 to perform the coordinate changing $(r,s) \to (x,y)$in Eq. A.3 we obtain:

$$P(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f\,(x,y)e^{-2i\pi\rho(x\sin\theta + y\cos\theta)}dxdy \qquad (\boldsymbol{A.4})$$

On the other hands, the 2D Fourier transform of $f(x,y)$ can be calculated as:

$$F(u,v) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)e^{-2i\pi(xu+yv)}\,dxdy \qquad (\boldsymbol{A.5})$$

If we let: $u = \rho\cos\theta\ \ and\ v = \rho\sin\theta$ in Eq. A.5, then the right hand side of Eq. A.4 and A.5 are the same and we have:

$$F(\rho\cos\theta\,,\rho\sin\theta) = P(\rho,\theta) \qquad (\boldsymbol{A.6})$$

Since $u = \rho\cos\theta\ \ and\ v = \rho\sin\theta$ also define a straight line tilted from θ regarding the u axis in the Fourier space, then Eq. A.6 demonstrates the FST.

From there, we can determine the formula of the filtered back projection (that is to say Eq. 1).

Starting from the identity:

$$f(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(u,v)e^{2i\pi(xu+yv)}\,dudv \qquad (\boldsymbol{A.7})$$

The transformations between the two reference systems $(u,v) \to (\rho,\theta)$ can be summarised as:

$$\begin{cases} u = \rho\cos\theta \\ v = \rho\sin\theta \end{cases}\ \ and\ \ dudv = Jd\rho d\theta = \underbrace{\begin{vmatrix} \dfrac{\partial u}{\partial\rho} & \dfrac{\partial u}{\partial\theta} \\ \dfrac{\partial v}{\partial\rho} & \dfrac{\partial v}{\partial\theta} \end{vmatrix}}_{|\rho|} d\rho d\theta \qquad (\boldsymbol{A.8})$$

Using those elements to perform a coordinate changing in Eq. A.7 yields:

$$f(x,y) = \int_{\theta=0}^{2\pi}\int_{\rho=0}^{+\infty} F(\rho\cos\theta\,,\rho\sin\theta)|\rho|e^{2i\pi\rho\underbrace{(x\cos\theta + y\sin\theta)}_{r}}d\rho d\theta \qquad (\boldsymbol{A.9})$$

Using the FST on Eq. A.9 and using the properties of the projection in polar coordinates ($P(-\rho,\theta) = P(\rho,\theta+\pi)$) yields:

$$f(x,y) = \int_{\theta=0}^{\pi}\int_{\rho=-\infty}^{+\infty} P(\rho,\theta)|\rho|e^{2i\pi\rho r}\,d\rho d\theta \qquad (\boldsymbol{A.10})$$

With $P(\rho,\theta)$ the Fourier transform of the projection acquired at an angle θ.

This therefore demonstrates Eq. 1.

NB: Assuming $h(r) = TF^{-1}[|\rho|]$ and using the properties of the convolution product Eq. A.10 could also be written as:

$$f(x,y) = \int_{\theta=0}^{\pi} p(r,\theta) * h(r)\,d\theta \qquad (\boldsymbol{A.11})$$

Curriculum Vitae

Julien Ott

📫 Route Cantonale, 104
CH-1025 St-Sulpice
☎ +41 76 615 26 96
✉ ottjulien1@gmail.com

Born April 30th, 1988
French & German
Single

Education

• **PhD in Medical Physics,** Université de Lausanne & Institut de radiophysique, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland	2015
• **Radiation protection expert**, Institut de radiophysique, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland	2013
• **Entrance examination of French medical physicist certification (DQPRM)** INSTN, Saclay, France	2012
• **MSc in Medical Physics,** Université Joseph Fourier, Grenoble, France	2012
• **BSc in Physics,** Université Joseph Fourier, Grenoble, France	2010

List of publications

- [**Omoumi *et al.* 2014**] P. Omoumi, F. R. Verdun, Y. B. Salah, B. C. Berg , F. E. Lecouvet, J. Malghem, J. G. Ott, R. Meuli and F. Becce, Low-dose multidetector computed tomography of the cervical spine: optimization of iterative reconstruction strength levels. 2014 Acta Radiologica 55(3) pp. 335-344.

- [**Tobalem & Dugert *et al.* 2014**] F. Tobalem, E. Dugert, F. R. Verdun, V. Dunet, J. G. Ott, H. A. Rudiger, S. Cherix, R. Meuli and F. Becce, MDCT Arthrography of the Hip: Value of the Adaptive Statistical Iterative Reconstruction Technique and potential for Radiation Dose Reduction. 2014 AJR **203(6)**:W665-73

- [**Ott *et al.* 2014**] J. G. Ott, F. Becce, P. Monnin, S. Schmidt, F.O. Bochud and F. R. Verdun, Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms. 2014 Phys Med Biol **59(15)** pp. 4047-4064.

- [**Ott *et al.* 2015 a**] J. G. Ott, A. Ba, D. Racine, A. Viry, F. O. Bochud and F. R. Verdun, Assessment of low contrast detection in CT using model observers: Developing a clinically-relevant tool for characterizing adaptive statistical and model-based iterative reconstruction. Submitted to Zeitschrift für Medizinische Physik in 2015.

- [**Ott *et al.* 2015 b**] J. G. Ott, A. Ba, D. Racine, N. Ryckx, F. O. Bochud, H. Alkadhi and F. R. Verdun, Patient exposure optimisation through task-based assessment of a new model-based iterative reconstruction technique. 2015 Accepted for publication in Radiation Protection Dosimetry.

- [**Ott *et al.* 2015 c**] J. G. Ott, P. Omoumi, E. Dugert, F. O. Bochud, F. R. Verdun and F. Becce, *"Task-based assessment of an MBIR algorithm: A study in all reconstruction planes."* Submitted to investigative radiology.

- [**Verdun & Racine *et al.* 2015**] F.R. Verdun & D. Racine, J. G. Ott, M. J. Tapiovaara, P. Toroi, F. O. Bochud, W. J. H. Veldkamp, A. Schegerer, R. W. Bouwman, I. Hernandez Giron, N. Marshall and S. Edyvean, Characterising clinically relevant image quality in CT: From physical measurements to model observers. 2015 Phys. Med. **31(8)**:823-43.

- [**Omoumi *et al.* 2015**] P. Omoumi, F. R. Verdun, J. G. Ott, D. Racine and F. Becce, Optimization of radiation dose and image quality in musculoskeletal CT: Emphasis on iterative reconstruction techniques (part 1). Semin Musculoskelet Radiol **19**: 415-21.

Conferences

- Annual congress of the **Swiss society of radiobiology and medical physics (SSRPM)**, Neuchatel, Switzerland, 2013

  A simple method to assess spatial resolution of moving structures in fluoroscopy
  J. G. Ott, A. Viry, N. Ryckx, A. Ba and F. R. Verdun

- Annual congress of the **French society of medical physics (SFPM)**, Deauville, France, 2014

  Impact of the reconstruction plan on image quality for CT images (Award of the best presentation)
  J. G. Ott, E. Dugert, F. Becce, P. Omoumi and F. R. Verdun

- International congress of **International Radiation Protection Association (IRPA)**, Geneva, Switzerland, 2014

  Towards a CT benchmarking in the framework of patient exposure optimisation
  J. G. Ott, A. Viry, I. Diaz, A. Ba, F. O. Bochud and F. R. Verdun

- International congress **Optimisation in X-ray and Medical Imaging (OXMI)**, Göteborg, Sweden, 2015

  Patient exposure optimisation through task-based assessment of a new model-based iterative reconstruction technique
  J. G. Ott, A. Ba, D. Racine, N. Ryckx, F. Bochud, H. Alkadhi and F. R. Verdun

- International congress of **Medical Image Perception Society (MIPS)**, Gent, Belgium, 2015

  Assessment of CT image quality in all reconstruction planes using an updated NPWE model observer
  J. G. Ott, F. Becce, E. Dugert, F. O. Bochud and F. R. Verdun

- Annual congress of the **Swiss society of radiobiology and medical physics (SSRPM)**, Fribourg, Switzerland, 2015

  Task-based assessment of the new ADMIRE algorithm
  J. G. Ott, A. Ba, D. Racine, N. Ryckx, F. Bochud and F. R. Verdun