UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE MEDICINA

DEPARTAMENTO DE RADIOLOGÍA Y MEDICINA FÍSICA

TESIS DOCTORAL

Model observers applied to low contrast detectability in computed tomography

Modelos de observador aplicados a la detectabilidad de bajo contraste en tomografía computarizada

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

Irene Hernández Girón

DIRECTORES

Alfonso Calzado Cantera
Wouter J.H. Veldkamp

Madrid, 2017

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE MEDICINA

Programa de doctorado en Ciencias Biomédicas

TESIS DOCTORAL

Model observers applied to low contrast detectability in Computed Tomography

Modelos de observador aplicados a la detectabilidad de bajo contraste en Tomografía Computarizada

Memoria para optar al grado de doctor presentada por

Irene Hernández Girón

Directores

Alfonso Calzado Cantera

Wouter J. H. Veldkamp

Madrid, 2015

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE MEDICINA

Programa de doctorado en Ciencias Biomédicas

Departamento de Radiología y Medicina Física

TESIS DOCTORAL

Model observers applied to low contrast detectability in Computed Tomography

Memoria para optar al grado de doctor presentada por

Irene Hernández Girón

Madrid, 2015

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE MEDICINA

Programa de doctorado en Ciencias Biomédicas

Departamento de Radiología y Medicina Física

TESIS DOCTORAL

Model observers applied to low contrast detectability in Computed Tomography

Modelos de observador aplicados a la detectabilidad de bajo contraste en Tomografía Computarizada

Memoria para optar al grado de doctor presentada por

Irene Hernández Girón

Directores

Alfonso Calzado Cantera

Wouter J. H. Veldkamp

Madrid, 2015

A mis padres y a Marcos

La máquina, la hace el hombre,
y es lo que el hombre hace con ella.

Hay manos capaces de fabricar herramientas
con las que se hacen máquinas para hacer
ordenadores que a su vez diseñan máquinas que
hacen herramientas para que las use la mano.

Jorge Drexler

Visie. Initiatief. Volharding.

Acknowledgements/Agradecimientos

Ésta, ha sido una tesis viajera. Nació en Madrid en 2009, se mudó a Reus y ha terminado de crecer en Leiden. Ha sido un largo camino, en el tiempo y el espacio. Mis amigos saben que si, cuando estaba acabando Física, alguien me hubiera dicho que más de una década más tarde, habría escrito una tesis, trabajaría en un hospital y estaría mirando la lluvia holandesa desde mi ventana, habría pensado que eso nunca pasaría, ni tan siquiera en una extraña realidad paralela. Mi abuelo Valeriano solía decir que en la vida a veces es mejor no hacer planes, y creo que tenía mucha razón.

No estaría aquí, ni podría dedicarme profesionalmente a algo que me llena, si no fuera por la ayuda de mucha gente, maestros, compañeros, amigos y mi familia. Es una suerte teneros en mi vida, o que os hayáis cruzado en mi camino.

Ahora viene la parte complicada, en que los agradecimientos empezarán a ser en distintos idiomas, así que allá vamos:

First, I would like to give my most sincere thanks to my supervisors, Alfonso Calzado and Wouter J. H. Veldkamp.

Alfonso, has sido mi mentor, desde aquel lejano trabajo de máster de 2008, cuando ni pensaba dedicarme a la investigación. Gracias por empujarme siempre un poco más allá, por tu exigencia y rigor y por haberme traído por primera vez a Leiden a visitar el LUMC. Gracias también por las charlas sobre libros, películas y la vida en general, y por ser además de maestro, amigo.

To Wouter, I can only express gratitude for making it so easyto work with you, even with almost 1800 km in between during my thesis (thanks to Skype), for the brainstorms we have had around a cup of coffee, for being so open to new ideas, and for having trusted me as a researcher.

I would like to warmly thank Koos Geleijns, head of Medical Physics at LUMC, for opening the door of the hospital for me years ago, for sharing your knowledge about medical physics so readily and encourage discussion and new ideas. Also, for being my 'landlord' during my visits to Leiden and your generosity and kindness. Gracias, de corazón.

A Marçal Salvadó, compañero y jefe durante cinco años en la Unitat de Física Mèdica de la URV, quiero agradecerle su apoyo, comprensión y confianza, especialmente en los momentos difíciles. Porque mis primeros pasos en investigación los di en la Rovira, con las simulaciones de Monte Carlo, y por todo lo aprendido estos años.

A Miguel López, fundador de la Unitat de Física Mèdica, agradecerle que confiara en mí para unirme a la unidad en Reus, y me permitiera trabajar por primera vez en lo que espero que sea mi profesión de por vida. Por su cercanía, y por compartir conmigo su experiencia, sobre el trabajo y la vida.

A Esteban Velasco, por haberme ofrecido mi primer trabajo en ASIGMA, ya que aunque solo fuera durante un año, me ayudó a foguearme en el control de calidad de los equipos

y a valorar que el conocimiento técnico de los equipos es esencial para poder hacer investigación.

A Juan José Morant y a Pili, por haber sido mi familia en Reus, desde el primer día que aterricé allí, por vuestra amabilidad y generosidad, y las charlas hasta altas horas en vuestra casa. A Juanjo por compartir tu experiencia en física médica con tanta facilidad, por llevarme de "excursión" a medir a los hospitales y por conseguir no solo que aprendiera sino que me lo pasara bien en el trabajo de campo.

A mis antiguos compañeros de la Unitat de Física Mèdica por haber hecho el trabajo tan fácil y agradable estos años. A Maria Cros, por ser mi amiga, en los momentos buenos y en los no tan buenos, y por enseñarme el Riudoms medieval. A Ramon Casanovas por su buen humor y porque aunque hacíamos cosas totalmente distintas, siempre estabas dispuesto a discutir sobre ciencia. Ramon, Maria, prometo financiar las gafas que necesitéis por los malditos megachis. A Elena Prieto, por ser mi compañera de cueva, ahora vacía, y por todos los buenos momentos que hemos pasado en ella juntas. El café en Leiden no es comparable con nuestros cafés en el Caracas, chicos.

A Margarita Chevalier, por haberme enseñado lo que sé de mamografía durante mi año en la UCM y porque conocí Japón gracias a ti. A los compañeros y profesores del departamento de Radiología de la UCM, especialmente a María Castillo y Diego García Pinto, por su ayuda con las imágenes, y del máster de Física Biomédica, por todo lo aprendido. A José Luis, Mercedes, Toñi, Susana y Rashi, por haber hecho que el año que trabajé en el departamento fuera tan agradable, y por recibirme con una sonrisa siempre que voy de visita.

To Raoul Joemai, for sharing your experience about CT with me all these years, and for the weirdest trip I could ever imagine to Zion Park, truly legendary. To Paul de Bruin for taking me to scan the strangest creatures on Earth, from crocodile hearts to sharks, and the good conversations sharing drinks. To Jan Wondergem and Dirk Zweers for making me feel so welcome when I was a guest in Wouter's desk corner during my visits. To my colleagues at K4-44 for having welcome me so easily in this new period.

To Ilya and Nicole, for the good times spent at Lebkov and Lemmy's. To Ece, Sanneke, Thijs, Itamar, Andrew, Wouter, Naj and all the former and current MR people from LUMC, for accepting me as one of your kind, though I belong to the CT faction.

A Juan Carlos, porque aunque cada vez nos separen más kilómetros (ahora mismo casi 12000, sí lo he mirado) estás siempre ahí y porque cada vez que escucho a alguien con acento extremeño no puedo evitar sonreír. A Antonio porque podemos pasar de hablar de los temas más serios o los más absurdos en cuestión de segundos, y por aquellos días en Dublín con Jenny, en que reímos tanto que al día siguiente me dolía todo el cuerpo. A José Alberto e Irma, porque fuisteis los primeros compañeros que conocí en la puerta de atrás de Físicas y mis primeros amigos, allí en la quinta fila del graderío, mientras mirábamos el cogote a Juan Carlos y Antonio.

A Bego, David, Sandra, Cris, Ire, Pedro, Miguel, Luis, Diana, Esther y todos los amigos de Físicas. Por los buenos ratos en las mesas del hall, por las risas en las clases, por las cervecitas de después y por la tradicional cena de Navidad, gracias chicos.

A Ángela, Sara y Elena, porque somos familia, hemos crecido juntas, desde los catorce años, porque echo de menos ir a llamar al telefonillo de vuestras casas para ir al cine o a tomar algo, por todos estos años, gracias.

A mis padres, Carmen y Manolo, por su amor y apoyo, por haberme enseñado lo poco que sé de la vida, por enseñarme a no ponerme límites y a aprender a levantarme. A ti, mamá, porque te necesito cada día, porque me hubiera gustado que vieras esta tesis acabada después de tantos años. A ti, papá por enseñarme lo importantes que son la honestidad y la responsabilidad y a que hiciera lo que hiciera en la vida, pusiera todo mi empeño en ello. A los dos por comprarme todos los libros del mundo, fomentar mi curiosidad y a enseñarme a buscar las respuestas cuando no las sabíais, por llevarme al campo a ver bichos y buscar piedras, y ver las estrellas conmigo en Ribatejada.

A Marcos, mi hermano, mi amigo y mi socio, por haberme enseñado a andar, los números en inglés, la sabiduría encerrada en los tebeos de Spiderman, a jugar al fútbol y con menor éxito, a que me tirara de cabeza en la piscina. Por haberme tratado como una igual, aunque nos llevemos siete años, y porque solo con mirarnos, sabemos lo que pensamos, casi siempre. Por ser una roca a la que asirme en los malos ratos y en los buenos, y porque guardas mi espalda como yo guardo la tuya, gracias.

A Mayte, por hacer feliz a Marcos, porque aunque no tengo hermanas, para mí lo eres, porque cada vez que estoy con vosotros en Pelegrina, me siento en casa, gracias niña.

A ti, Bruno, que por ser el más pequeño, te he dejado para el final. Todo esto empezó cuando no eras más que un puñadito de células dentro de mamá y yo estudiaba el máster. Ahora, me llegas al hombro. Gracias por querer jugar siempre conmigo, porque soy tu tía-tía-tía-tía, y porque a veces, sin darte cuenta, me llamas mamá.

Finally, I would like to thank the Sociedad Española de Física Médica (SEFM) for both scholarships I was awarded. One allowed me to join the First CT meeting in Utah in 2010, which was the first international congress I ever attended. And the other, to visit the LUMC for three months, which boosted my research and this thesis. I would also like to thank the Medical Imaging Perception Society (MIPS) for the scholarships that enabled me to join the meetings in Dublin, Washington DC and Ghent and learn from the experience of the attendees.

Index

List of contributions

PhD Thesis papers

This PhD thesis is based on the following publications, which will be referred to as follows, using capital Roman numerals in the text:

[I] I. Hernández-Girón, J. Geleijns, A. Calzado, M. Salvadó, R. M. S. Joemai, W. J. H. Veldkamp. Objective assessment of low contrast detectability for real CT phantom and in simulated images using a model observer. IEEE Nuclear Science Symposium Conference Record 2011;3477-3480 (doi:10.1109/NSSMIC.2011.6152637)*

[II] I. Hernández-Girón, J. Geleijns, A. Calzado, W. J. H. Veldkamp. Automated assessment of low contrast sensitivity for CT systems using a model observer. Med Phys 2011;38:S25-S35 (doi:10.1118/1.3577757)

[III] I. Hernández-Girón, A. Calzado, J. Geleijns, R. M. S. Joemai, W. J. H. 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. Br J Radiol 2014;87:20140014 (doi: 10.1259/bjr.20140014)

[IV] I. Hernández-Girón, A. Calzado, J. Geleijns, R. M. S. Joemai, W. J. H. Veldkamp. Low contrast detectability performance of model observers based on CT phantom images: kVp influence. Phys Medica 2015 Corrected proof in press (doi: 10.1016/j.ejmp.2015.04.012)

Model observers applied to low contrast detectability in Computed Tomography

SUMMARY

Introduction

Medical imaging has become one of the cornerstones in modern healthcare. Computed tomography (CT) is a widely used imaging modality in radiology worldwide. This technique allows to obtain three-dimensional volume reconstructions of different parts of the patient with isotropic spatial resolution. Also, to acquire sharp images of moving organs, such as the heart or the lungs, without artifacts. The spectrum of indications which can be tackled with this technique is wide, and it comprises brain perfusion, cardiology, oncology, vascular radiology, interventionism and traumatology, amongst others.

CT is a very popular imaging technique, widely implanted in healthcare services worldwide. The amount of CT scans performed per year has been continuously growing in the past decades, which has led to a great benefit for the patients. At the same time, CT exams represent the highest contribution to the collective radiation dose. Patient dose in CT is one order of magnitude higher than in conventional X-ray studies.

Regarding patient dose in X-ray imaging the ALARA criteria is universally accepted. It states that patient images should be obtained using a dose as low as reasonably achievable and compatible with the diagnostic task. Some cases of patients' radiation overexposure, most of them in brain perfusion procedures have come to the public eye and had a great impact in the USA media. These cases, together with the increasing number of CT scans performed per year, have raised a red flag about the patient imparted doses in CT. Several guidelines and recommendation for dose optimization in CT have been published by different organizations, which have been included in European and National regulations and adopted by CT manufacturers.

In CT, the X-ray tube is rotating around the patient, emitting photons in beams from different angles or projections. These photons interact with the tissues in the patient, depending on their energy and the tissue composition and density. A fraction of these photons deposit all or part of their energy inside the patient, resulting in organs absorbed dose. The images are generated using the data from the projections of the X-ray beam that reach the detectors after passing through the patient. Each projection represents the total integrated attenuation of the X-ray beam along its path.

A CT protocol is defined as a collection of settings which can be selected in the CT console and affect the image quality outcome and the patient dose. They can be acquisition parameters such as beam collimation, tube current, rotation time, kV, pitch, or reconstruction parameters such as the slice thickness and spacing, reconstruction filter and method (filtered back projection (FBP) or iterative algorithms).

All main CT manufacturers offer default protocols for different indications, depending on the anatomical region. The user can frequently set the protocol parameters selecting

amongst a range of values to adapt them to the clinical indication and patient characteristics, such as size or age. The selected settings in the protocol affect greatly image quality and dose. Many combinations of scan parameters can render an appropriate image quality for a particular study. Protocol optimization is a complex task in CT because most scan protocol parameters are intertwined and affect image quality and patient dose.

One of the reasons of the popularity of CT is its capacity to reveal lesions with attenuation properties very similar to those of the surrounding tissue (i.e.: with low contrast), that cannot be detected with other medical imaging techniques. Thus, low contrast detectability (LCD) is a relevant image quality parameter to investigate. LCD is highly affected by the selected acquisition and reconstruction parameters in CT. This parameter is critical in CT as small low contrast lesions can be masked by noise and also varies with spatial resolution.

LCD is frequently determined in human observer studies, scoring the visibility of low contrast objects in phantom images acquired with different protocols, assessing the smallest object of a given contrast level that can be detected. Human observer studies are complex, expensive and time consuming, and they have to be carefully planned and performed. In the outcomes, a great intra- and inter-observer variability may appear. Besides, the results of these studies can be biased, as the observer normally knows the distribution of the objects in the phantom in advance.

There is a need for automated methods for the analysis of image quality, especially in modalities such as CT in which many acquisition and reconstruction parameters, that in turn can take a range of values, affect image quality. Model observers stand as an alternative to human observers studies. They are mathematical models that aim to predict human performance for certain detection and discrimination tasks, in particular, in medical images.

Motivation and goals

The motivation of this PhD thesis was to develop a framework to assess image quality in CT images in an objective way based on model observers, in particular low contrast detectability. The starting hypothesis of this thesis is that model observers can be applied for certain detection and discrimination tasks in CT phantom images and predict human observer performance for LCD, selecting different protocols.

The goals and milestones were as follows:

1. To develop a software to automatically extract samples from phantom images, containing objects or background. In particular in a phantom containing distributions of low contrast objects.

2. To implement a model observer (non-prewhitening matched filter with an eye filter, NPWE) to assess LCD in CT phantom images. To compare the model

performance with results in the literature based on the detection of objects in simulated Gaussian white noise backgrounds.

3. To investigate the effect of selecting different acquisition and reconstruction parameters in LCD performance for the model observers and humans in simple detection tasks in phantom images. In particular, to study the influence of selecting a range of kV, tube charge per rotation and reconstruction kernel settings.

4. To develop a software to perform 2-alternative forced choice experiments with human observers to enable a quantitative comparison between humans and model observers.

5. To implement the channelized Hotelling observer (CHO) in the framework as an alternative for NPWE. To investigate the influence of the selected kVp in LCD when dose is kept constant, comparing CHO and NPWE performance with human observers.

6. To study the influence of iterative reconstruction algorithms in LCD with the model observer and human observers compared to FBP algorithms.

Results

This PhD thesis is comprised by four papers. The first paper **[I]** was focused on the implementation of the NPWE model observer and its validation. To this end, the detectability of objects or different signal values and diameters in Gaussian white noise simulated backgrounds was analysed. The detectability values increased with object size and contrast.

An in-house software was developed to automatically extract samples from phantom images, in particular from the Catphan phantom, widely used in quality control in CT, which contains three distributions of low contrast objects with different diameters. The NPWE model was integrated in the software to assess LCD automatically, analysing samples with object present or absent extracted from the phantom images. Sets of images of the phantom were acquired in a CT scanner varying the tube charge per rotation (mAs). For NPWE, LCD increased as a function of object diameter, object contrast and dose, as expected.

The second paper **[II]**, analysed the influence of a range of acquisition (kVp and mAs) and reconstruction parameters (different reconstruction filters) in the model LCD performance in images of the same phantom. A human observer study, in which observers scored the number of visible objects in the phantom images, was carried out to validate the model performance in a qualitative way. These results might be biased as the observers know beforehand the distributions of the objects in the phantom. The NPWE model reproduced the human performance trends, showing an improvement in LCD as a

function of increasing object diameter, contrast level, kV, mAs and for *soft* reconstruction kernels.

The next step was to develop an in-house software to perform 2-alternative forced choice (2-AFC) studies and overcome the possible bias in human observers LCD assessment with the method applied in **[II]**, for which the object distribution in the phantom is known beforehand.

This software was used in paper **[III]** to analyse the influence of using iterative reconstruction algorithms in LCD compared to FBP for a range of dose levels, with the NPWE model observer and humans. The model obtained higher LCD scores than the human observers and its results were normalized applying an efficiency factor. Model and humans showed the same trends and a high correlation in their performance, checked with Pearson's correlation coefficients and Bland-Altman plots. LCD improved with increasing object diameter, contrast and dose. The selected iterative algorithm improved the detectability of the low contrast objects compared to FBP, especially for low dose and low contrast objects.

The last paper of the thesis **[IV]** analysed the influence of kV in the detectability of low contrast objects, applying two model observers, NPWE and CHO to analyse CT phantom images. The CHO model was implemented with the same set of channels proposed by other authors which had been validated in simple detection tasks in CT phantom images. The models were modified applying efficiency factors and internal noise. The results obtained with both models were compared with human scores in a 2-AFC study. The NPWE model showed better correlation with LCD human performance than the CHO model for this particular task. Selecting lower kV values lead to an increase in LCD in the phantom images for both NPWE and human observers.

Conclusions

This PhD thesis presents a framework for CT image quality assessment in phantom images using model observers, in particular for low contrast detectability. They are an objective and fast alternative to human observer studies as it has been proved that they can predict human performance in simple detection and discrimination tasks. The use of these mathematical models has grown in the past few years. They are especially interesting in medical imaging modalities, such as CT, because a wide range of parameters affect image quality. The initial intended goals of this PhD were covered with the methodology and results presented in the papers **[I-IV]** that constitute the core of this PhD thesis.

The use of computed tomography has expanded in the past decades and will continue to do so, as it has recently been approved for screening in certain indications in the US (lung cancer) and others are under study (colorectal cancer). Automatic methods to assess
image quality in an objective and fast way, such as those proposed in this thesis based on model observers, are needed in medical imaging. Model observers can be used to investigate different strategies in protocol optimization, to compare the LCD performance of different CT manufacturers for similar indications or applied in other medical imaging modalities.

Model observers are not intended to be a substitute of clinical validation of medical imaging systems, based on patient images assessed by radiologists. Further research is needed to investigate the correlation between humans and model observers in more complex tasks and in anatomical backgrounds. Anthropomorphic phantom images, including 3D printed phantoms, can be a good thread to follow for these goals.

Modelos de observador aplicados a la detectabilidad de bajo contraste en Tomografía Computarizada

RESUMEN

Introducción

La imagen médica se ha convertido en uno de los pilares en la atención sanitaria actual. La tomografía computarizada (TC) es una modalidad de imagen ampliamente extendida en radiología en todo el mundo. Esta técnica permite adquirir imágenes de órganos en movimiento, como el corazón o los pulmones, sin artefactos. También permite obtener reconstrucciones de volúmenes tridimensionales de distintas partes del cuerpo de los pacientes. El abanico de indicaciones que pueden abordarse con esta técnica es amplio, e incluye la perfusión cerebral, cardiología, oncología, radiología vascular, intervencionismo y traumatología, entre otras.

La TC es una técnica de imagen muy popular, ampliamente implantada en los servicios de salud de hospitales de todo el mundo. El número de estudios de TC hechos anualmente ha crecido de manera continua en las últimas décadas, lo que ha supuesto un gran beneficio para los pacientes. A la vez, los exámenes de TC representan la contribución más alta a la dosis de radiación colectiva en la actualidad. La dosis que reciben los pacientes en un estudio de TC es un orden de magnitud más alta que en exámenes de radiología convencional.

En relación con la dosis a pacientes en radiodiagnóstico, el criterio ALARA es aceptado universalmente. Expone que las imágenes de los pacientes deberían obtenerse utilizando una dosis tan baja como sea razonablemente posible y compatible con el objetivo diagnóstico de la prueba. Algunos casos de sobreexposición de pacientes a la radiación, la mayoría en exámenes de perfusión cerebral, se han hecho públicos, lo que ha tenido un gran impacto en los medios de comunicación de EEUU. Estos accidentes, junto con el creciente número de exámenes TC anuales, han hecho aumentar la preocupación sobre las dosis de radiación impartidas a los pacientes en TC. Varias guías y recomendaciones para la optimización de la dosis en TC han sido publicadas por distintas organizaciones, y han sido incluidas en normas europeas y nacionales y adoptadas parcialmente por los fabricantes de equipos de TC.

En TC, el tubo de rayos-X rota en torno al paciente, emitiendo fotones en haces desde distintos ángulos o proyecciones. Estos fotones interactúan con los tejidos en el paciente, en función de su energía y de la composición y densidad del tejido. Una fracción de estos fotones depositan parte o toda su energía dentro del paciente, dando lugar a la dosis absorbida en los órganos. Las imágenes se generan usando los datos de las proyecciones del haz de rayos-X que alcanzan los detectores tras atravesar al paciente. Cada proyección representa la atenuación total del haz de rayos-X integrada a lo largo de su trayectoria.

Un protocolo de TC se define como una colección de opciones que pueden seleccionarse en la consola del equipo y que afectan a la calidad de las imágenes y a la dosis que recibe

el paciente. Pueden ser parámetros de adquisición, tales como la colimación del haz, la intensidad de corriente, el tiempo de rotación, el kV, el factor de paso parámetros de reconstrucción como el espesor y espaciado de corte, el filtro y el método de reconstrucción (retroproyección filtrada (FBP) o algoritmos iterativos).

Los principales fabricantes de equipos de TC ofrecen protocolos recomendados para distintas indicaciones, dependiendo de la región anatómica. El usuario con frecuencia fija los parámetros del protocolo eligiendo entre un rango de valores disponibles, para adaptarlo a la indicación clínica y a las características del paciente, tales como su tamaño o edad. Las condiciones seleccionadas en el protocolo tienen un gran impacto en la calidad de imagen y la dosis. Múltiples combinaciones de los parámetros pueden dar lugar a un nivel de calidad de imagen apropiado para un estudio en concreto. La optimización de los protocolos es una tarea compleja en TC, ya que la mayoría de los parámetros del protocolo están relacionados entre sí y afectan a la calidad de imagen y a la dosis que recibe el paciente.

Una de las razones por las que la TC es tan popular es su capacidad de mostrar lesiones con una atenuación muy parecida a la del tejido circundante, es decir con bajo contraste, que no pueden ser detectadas utilizando otras técnicas de imagen médica. Por tanto, la detectabilidad de bajo contraste (LCD) es un parámetro relevante en calidad de imagen que hay que investigar. La LCD varía mucho en función de los parámetros de adquisición y reconstrucción seleccionados en TC. Este parámetro es crítico en TC, ya que las pequeñas lesiones de bajo contraste, pueden quedar enmascaradas por el ruido, y también varía con la resolución espacial.

La LCD es normalmente medida en estudios con observadores humanos, que valoran la visibilidad de objetos de bajo contraste en imágenes de maniquíes adquiridas usando distintos protocolos, determinando el objeto de menor tamaño y de cierto valor de contraste, que puede ser detectado. Los estudios con observadores son complejos, caros y se necesita mucho tiempo para realizarlos. Tienen que ser cuidadosamente planeados y llevados a cabo. Puede aparecer una gran variabilidad intra- e inter-observador en los resultados. Además, los resultados de estos estudios pueden presentar sesgos, dado que el observador normalmente conoce de antemano la distribución de los objetos en el maniquí.

Se necesitan métodos automáticos para el análisis de la calidad de imagen, especialmente en modalidades como la TC en la cual muchos parámetros de adquisición y reconstrucción, que además pueden tomar distintos valores, afectan a la calidad de imagen.

Los modelos de observador son una alternativa a los estudios con observadores humanos. Son modelos matemáticos que están diseñados para predecir los resultados de los observadores humanos en ciertas tareas de detección y discriminación de objetos, en particular en imagen médica.

Motivación y objetivos

La motivación de esta tesis es desarrollar un método para evaluar la calidad de las imágenes en TC de manera objetiva basado en modelos de observador, en particular la detectabilidad de bajo contraste. La hipótesis de partida de esta tesis es que los modelos de observador pueden usarse para ciertas tareas de detección y discriminación de objetos en imágenes de maniquíes adquiridas en equipos de TC y que pueden predecir los resultados de los observadores humanos, seleccionando distintos protocolos.

Los objetivos son:

1. Desarrollar un software para extraer de manera automática muestras de imágenes de maniquíes, conteniendo objetos o el fondo circundante. En particular, en un maniquí que contiene distribuciones de objetos de bajo contraste.

2. Implementar un modelo de observador (*non-prewhitening matched filter with an eye filter*, NPWE) para evaluar la LCD en imágenes de maniquíes. Comparar los resultados del modelo con los publicados por otros autores basados en la detección de objetos en imágenes simuladas generadas con ruido gaussiano blanco.

3. Investigar el efecto de seleccionar distintos parámetros de adquisición y reconstrucción en la respuesta LCD del modelo de observador y observadores humanos en tareas de detección sencillas en imágenes de maniquíes. En particular, estudiar la influencia de variar el kVp, la carga del tubo por rotación y el filtro de reconstrucción.

4. Desarrollar un *software* para realizar estudios de 2-alternativas forzadas con observadores humanos para poder llevar a cabo una comparación cuantitativa entre los resultados de los observadores humanos y el modelo.

5. Implementar el modelo de observador *channelized Hotelling* (CHO) en el método como una alternativa al modelo NPWE. Investigar la influencia del valor de kVp seleccionado en la LCD cuando la dosis se mantiene constante, comparando los resultados de los modelos CHO y NPWE con observadores humanos.

6. Estudiar la influencia de los algoritmos de reconstrucción iterativa en la LCD con el modelo de observador y con observadores humanos comparado con reconstrucción FBP.

Resultados

Esta tesis está constituida por cuatro artículos científicos. El primer artículo **[I]** se centra en la implementación del modelo de observador NPWE y su validación. Para ello se analizó la detectabilidad de objetos con diferentes valores de señal y diámetros en imágenes simuladas con ruido blanco Gaussiano. Los valores del índice de detectabilidad aumentaron con el tamaño de los objetos y el valor de contraste.

Se desarrolló un *software* propio para extraer de manera automática muestras de las imágenes de un maniquí, en particular del maniquí Catphan, de uso frecuente en control de calidad en TC, que contiene tres distribuciones de objetos de bajo contraste con diferentes diámetros.

El modelo de observador NPWE se integró en el programa para evaluar la LCD de manera automática, analizando muestras con el objeto presente o ausente, extraídas de las imágenes del maniquí. Se adquirieron varias series de imágenes del maniquí en un escáner TC variando la carga del tubo por vuelta (mAs). Para el modelo NPWE, la LCD aumentaba en función del diámetro de objeto, su contraste y la dosis seleccionada.

El segundo artículo **[II]**, analaliza la influencia de un rango de valores de parámetros de adquisición (kVp y mAs) y reconstrucción (distintos filtros) en los resultados de LCD del modelo en imágenes del mismo maniquí. Un estudio con observadores humanos, en que estos evaluaban el número de objetos visibles en las imágenes del maniquí, se llevó a cabo para validar el funcionamiento del modelo de manera cualitativa. Estos resultados pueden presentar sesgos, ya que los observadores conocen de antemano la distribución de los objetos en el maniquí. El modelo NPWE reprodujo las tendencias observadas en los resultados de los observadores, con una mejora de la LCD en función de diámetros de objeto, nivel de contraste, kV y mAs crecientes, y también en las imágenes reconstruidas con filtros de reconstrucción *soft*.

El siguiente paso consistió en desarrollar un *software* para realizar estudios de 2-alternativas forzadas (2-AFC) con observadores humanos y así evitar el sesgo que puede aparecer cuando la LCD se evalúa con métodos como el aplicado en el artículo **[II]**, en los que se conoce de antemano la distribución de objetos en el maniquí.

Este *software* se utilizó en el artículo **[III]** para investigar la influencia de seleccionar algoritmos de reconstrucción iterativa en la LCD comparado con FBP para un rango de valores de dosis, con el modelo NPWE y observadores humanos. Los valores de LCD del modelo fueron mayores que para los observadores humanos y sus resultados se normalizaron aplicando un factor de eficiencia. Las tendencias del modelo y los observadores humanos fueron equivalentes, con una correlación alta en sus resultados, estimada con factores de correlación de Pearson y gráficos de Bland-Altman. La LCD mejoró con valores crecientes de tamaño de objeto, contrate y dosis. El algoritmo iterativo estudiado mejoró la detectabilidad de bajo contraste de los objetos comparada con la reconstrucción FBP, especialmente para dosis bajas y objetos de menor contraste.

El último artículo de esta tesis **[IV]** se centra en el análisis de la influencia del valor de kV en la detectabilidad de los objetos de bajo contraste, utilizando dos modelos de observador, NPWE y CHO para evaluar imágenes de un maniquí adquiridas en un escáner TC. El modelo CHO se implementó aplicando una selección de canales de Gabor propuesta por otros autores que fue validada en tareas de detección de objetos sencillos en imágenes de maniquíes en TC. Los modelos fueron modificados aplicando factores de eficiencia y ruido interno. Los resultados de ambos modelos se compararon con los valores obtenidos por observadores humanos en un estudio de 2-AFC. El modelo NPWE mostró mejor correlación con los resultados de LCD de los observadores humanos que el modelo CHO para la tarea de detección planteada. Para valores bajos de kV, la LCD

mejoró tanto para el modelo NPWE como para los observadores humanos, analizando las imágenes del maniquí.

Conclusiones

Esta tesis doctoral presenta una metodología para evaluar la calidad de imagen en TC en imágenes de maniquíes usando modelos de observador, en particular para la detectabilidad de bajo contraste. Estos son una alternativa objetiva y rápida a los estudios con observadores humanos y se ha probado que pueden predecir los resultados de los observadores para tareas de detección y discriminación de objetos sencillos. El uso de estos modelos matemáticos ha aumentado en los últimos años. Su aplicación es especialmente interesante en modalidades de imagen, como la TC, ya que en ella un amplio rango de parámetros afectan a la calidad de imagen. Los objetivos iniciales de esta tesis se han cubierto con la metodología y resultados presentados en los artículos que acompañan a esta tesis **[I–IV]**.

El uso de la tomografía computarizada va a continuar creciendo, a tenor de indicios tales como su empleo en estudios de cribado para indicaciones como el cáncer de pulmón, autorizado por administraciones sanitarias en EEUU y otras propuestas en la misma dirección, como el cáncer colorrectal.

En imagen médica se necesitan métodos automáticos que permitan evaluar la calidad de imagen de manera objetiva y rápida, como los propuestos en esta tesis, basados en modelos de observador. Los modelos de observador pueden emplearse para investigar distintas estrategias de optimización de protocolos, para comparar la resolución de bajo contraste de distintos fabricantes de TC para indicaciones similares o ser aplicados en otras modalidades de imagen médica.

Los modelos de observador no están destinados a sustituir la validación clínica de los sistemas de imagen médica basados en imágenes de pacientes analizadas por radiólogos. Se necesita continuar investigando la correlación entre los resultados de los modelos de observador y observadores humanos en tareas de detección más complejas y en imágenes con fondo anatómico. Los maniquíes antropomórficos, incluyendo aquellos basados en la impresión 3D, pueden ser una buena línea de investigación a seguir con este fin.

List of acronyms and abbreviations

AUC	Area under the ROC curve
CHO	Channelized Hotelling observer
CNR	Contrast to noise ratio
CT	Computed Tomography
$CTDI_{vol}$	Volumetric computed tomography dose index
d'	Detectability index
FBP	Filtered back projection
HU	Hounsfield units
HVS	Human visual system
IR	Iterative reconstruction
LCD or LCDet	Low contrast detectability
mAs	Tube charge per rotation
MTF	Modulation transfer function
M-AFC	Multi-alternative forced choice
NPWE	Non-prewhitening matched filter with an eye filter model observer
PC	Proportion correct
PSF	Point spread function
r	Pearson's product-moment correlation coefficient
ROC	Receiver operating characteristic curve
ROI	Region of interest
SKE/BKE	Signal known exactly/background known exactly
2-AFC	2-alternative forced choice experiment
WL	Window level
WW	Window width
α	Internal noise
Δ	Mean difference in Bland-Altman plot
$[\Delta \pm 2\sigma]$	Range of the differences in Bland-Altman plot
η	Efficiency
λ	Visibility threshold
σ	Statistical deviation

Introduction

1. Computed tomography

Medical imaging modalities provide the physicians crucial information to perform an adequate diagnostic or to decide the patient's treatment. One of the cornerstones of healthcare is understanding the different imaging techniques available and the information about the patient anatomy or physiological functioning that each of them can render[1,2].

The group of Godfrey N. Hounsfield performed the first clinical computed tomography (CT) study in London in 1971. Two contiguous axial images of a patient's head, in which a brain cyst was visible, were acquired in over four minutes with a single detector CT and reconstructed taking seven minutes per image. A. M. Cormack and G. N. Hounsfield were awarded the Nobel prize in Physiology or Medicine for the development of computer tomography in 1979. The first CT systems were only used in neuroradiology studies due to their detector limitations[3].

The technological evolution of the CT systems, which started with the introduction of multiple detectors, helical acquisition and multi-slice CT, enables nowadays to obtain images with an isotropic submillimetre spatial resolution and a temporal resolution below 10^{-2} seconds[3,4]. This allows for acquiring images of organs in movement, like the heart or the lungs, analysing different phases in perfusion studies using contrast media, or reconstructing 3D volumes of different regions in the patient. Dual energy CT, based on the acquisition of two scans with different kVp, uses the energy dependence of the attenuation coefficients of the different tissues to generate images in which only certain tissues are visible, to differentiate materials with similar density or composition as the surrounding tissue and reduce beam hardening artifacts[4,5].

In the past decades, CT has turned into a versatile medical imaging specialty with a wide range of indications, including cardiology, brain perfusion, oncology, vascular radiology, interventionism and traumatology, amongst others. Software developments have corrected metal artifacts in patients images with certain metallic prosthesis. In the past, these artifacts hindered an adequate diagnosis due to the image streaks they caused. Some devices hybridise CT with other imaging techniques, such as positron emission tomography (PET-CT), single photon emission computed tomography (SPECT-CT) or magnetic resonance (CT-MRI) blending anatomical and functional information in the same study[3-6].

As a consequence, CT has become a very popular imaging technique, widely available in healthcare services. Compared to a conventional X-ray study, in CT, patient dose is one order of magnitude higher. The amount of CT scans performed per year has been continuously growing worldwide in the past two decades, which has resulted in a great benefit for the patients. In parallel, of all the diagnostic imaging techniques using X-rays, CT has turned into the biggest contributor to the collective radiation dose[7,8]. In the CT console, two dose parameters are shown which are saved together with the patient images: the volume CT dose index ($CTDI_{vol}$) and the dose length product (DLP). They are a useful reference for the scan output performance and their definition is worldwide accepted

although they do not represent directly the radiation risk related to a particular CT exam and patient[9].

1.1. Image acquisition in CT

CT is an imaging technique in which the X-ray tube is rotating around the patient, emitting photons in thin X-ray beams with an intensity I_0 from different angles or projections, as shown in **Fig. 1.** In this example, the Shepp-Logan head phantom, which is a test image representing a section containing different materials was used. The photons traverse the patient and part of them reach the detectors with an intensity I, which is registered at each of them.

The photons emitted by the X-ray tube in a CT scanner in each X-ray tube position, reach the patient after passing through the so-called bowtie filters, which change the beam spectrum and intensity depending on the anatomical region characteristics. The X-ray photons interact with the different tissues in the patient, depending on their composition, density and the energy of the photons. These dependencies are expressed by the linear attenuation coefficients (μ) for each material or tissue, which represent the exponential probability that a photon of the X-ray beam is absorbed or scattered. A fraction of these photons deposit all or part of their energy inside the patient, resulting in the organs absorbed doses. In each tube rotation around the patient up to 800-1500 projections can be taken[4,10].

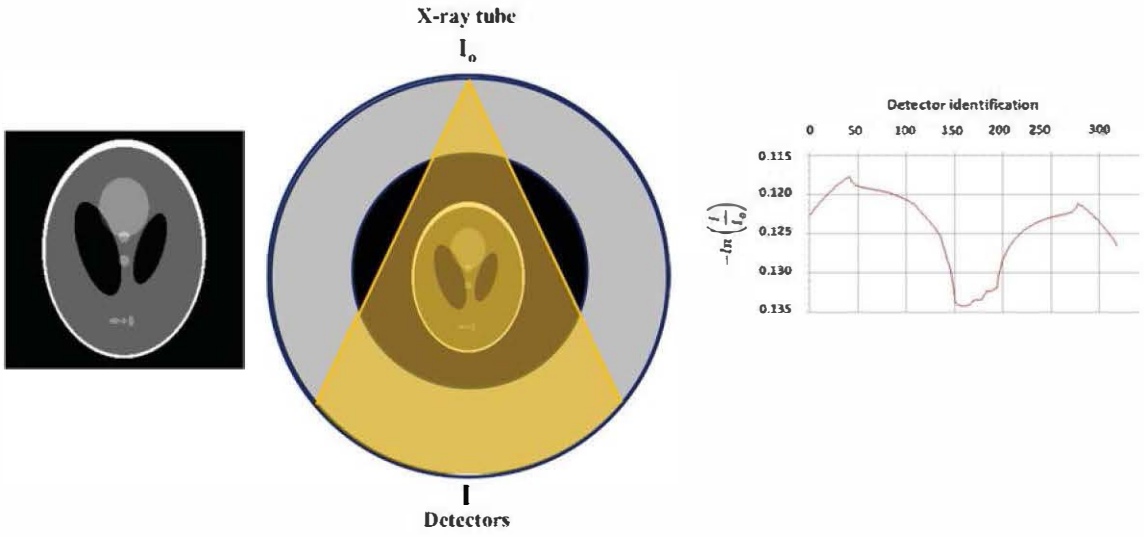


Fig. 1. Shepp-Logan head phantom image (left) placed in the isocenter of a CT scanner (center). The X-ray beam is represented in yellow for one of the positions of the tube. The detection system compares the initial intensity of the beam (I_0) with the intensity that arrives to each detector (I). The plot on the right represents the line attenuation for each detector, measured as $-\ln(I/I_0)$.

1.2. Image reconstruction in CT

CT images are generated using the data from the projections of the X-ray beam that reach the detectors after passing through the patient. Each projection represents the total integrated attenuation of the X-ray beam along its path. All the projections related to one acquisition are represented in the sinogram or raw data, which shows the detector elements readings plotted as a function of the acquisition angle. During the reconstruction process, these projections are transformed into the image, in which each pixel is related the X-ray attenuation at the equivalent position in the patient[4,10]. **Figure 2** depicts an example of attenuation profile, sinogram and reconstructed image based on a simulated CT acquisition with the Shepp-Logan phantom.

The unit of the CT number scale is the Hounsfield Unit (HU), which represents the relative difference between the attenuation coefficient of a material with water, multiplied by 1000. In CT, even materials with similar attenuation properties, appear with a different grey level in the image[3].

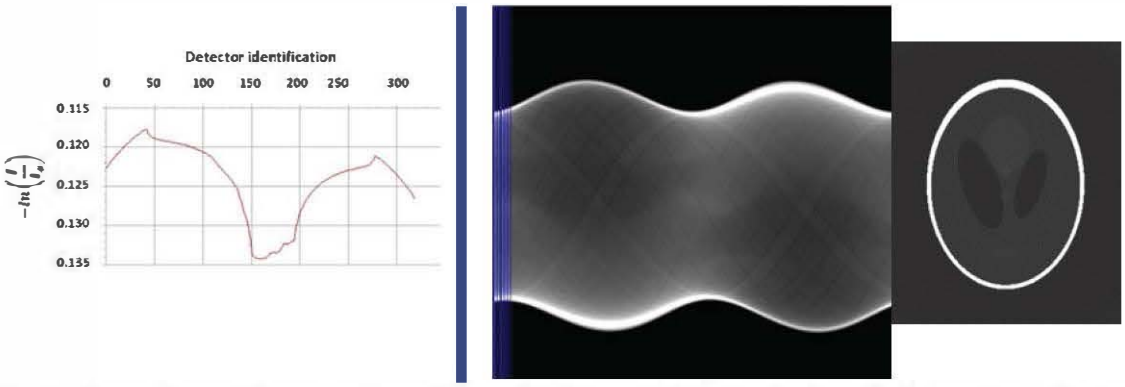


Fig. 2. Shepp- Logan phantom attenuation profile for one of the projections (left) together with the generated sinogram (center), which represents the readings of the detectors (Y-axis) as a function of the acquisition angle (X-axis) and the final reconstructed image, applying the inverse Radon transform (right).

Image reconstruction has a great impact in the appearance of the CT images, determining how defined of sharp structures and boundaries appear, the noise texture or how certain artifacts are more or less visible in the image. The attenuation information contained in the raw data or sinogram is used as input for the image reconstruction[3].

The 'traditional' reconstruction techniques are based on the properties of the Fourier transform, and the standard used to reconstruct CT images is called filtered back projection (FBP). If only simple back projection is used, the resulting image is blurred. In general, to reconstruct the images, given a sinogram, the inverse Radon transform is applied which comprises the filtration and back projection of the data to generate the images. The reconstruction algorithm or kernel modifies the spatial content of the noise and it can also change image appearance. Certain image aspects can be enhanced, improving the edges definition or the visibility of low contrast structures[3-5,10].

1.2.1. Iterative reconstruction algorithms

Iterative reconstruction (IR) algorithms are nowadays available in all major CT manufacturers systems and different studies have proved that image noise and artifacts can be reduced to different degrees, when comparing images reconstructed with FPB for similar dose levels[10-13]. Thus, they have the potential of obtaining the patient images with a substantial dose reduction without losing relevant diagnostic information for certain indications. A drawback of IR is that the appearance of the images may be quite different than in FBP, as the noise texture changes[10].

These techniques have long been applied in PET or SPECT and were also used in the first CT systems. Due to the big amount of data that had to be analysed in the multi-slice CT, these reconstruction methods were too time consuming to be applied in daily clinical practice. IR methods have been re-introduced in CT imaging in the early 2000s thanks to the improved computational power that allows to perform the image reconstruction in a reasonable time[10,13].

Each CT manufacturer has its own iterative reconstruction algorithm and little is disclosed about how each system actually performs. Some IR algorithms use X-ray spectrum information and model the acquisition geometry of the scanner in the iterative process that is performed applying non-linear algorithms[10,11,13]. In the reconstruction process, a first assumption is made about the object attenuation values, which can be a matrix full of zeros, random values or the original filtered back projection reconstructed image. Then, for each X-ray tube location, the IR algorithm simulates the beam and its propagation through the patient and reaching the detectors. These simulated attenuation profiles are compared to the original raw data, and a correction is generated to update the estimated solution for the patient attenuation values. In the reconstruction process, this correction is repeated in a loop until either of the following conditions is met: a number of iterations is reached, the updated image is very similar to the one obtained in the previous iteration or an image quality criterion defined beforehand is achieved[10,13]

1.3. CT protocols

A given CT protocol is defined as a collection of settings which can be selected in the CT console and affect both, the image quality outcome and the patient dose. They can be classified in three main groups:

a) Acquisition parameters: beam collimation, tube charge per rotation, kV, pitch, activation of dose reduction techniques, field of view (FOV), among others.
b) Reconstruction parameters: slice thickness and spacing, reconstruction filter, reconstruction method (filtered back projection or the manufacturer iterative algorithm)
c) Patient related parameters: size, positioning, contrast administration…

All main CT manufacturers offer default protocols for different indications, depending on the anatomical region. The user can frequently set the protocol parameters selecting amongst a range of values to adapt them to the patient characteristics, such as size or age. The selected settings in the protocol affect greatly image quality and dose. Many of the

scan protocol parameters are intertwined. Therefore, if one of the scan parameters is changed, others have to be adjusted to keep the image quality up to the necessary level for the diagnostic task, and the patient dose down to a reasonable value[10]. There are no gold standards for CT protocols. Many combinations of scan parameters can render an appropriate image quality for a particular study and ultimately they have to be adjusted to the patient characteristics.

1.4. Protocol optimization in CT

The general rule in X-ray imaging is to follow the ALARA criteria, to obtain the patient images using a dose as low as reasonably achievable and compatible with the diagnostic task. Some cases of patients' radiation overexposure, most of them in brain perfusion procedures have come to the public eye and had a great impact in the USA media. According to the US Food and Drug Administration (FDA), undesired radiation effects such as erythema, epilation, dizziness, headache, and other neurological disorders were observed in up to 385 patients until October 2010. Those symptoms could indicate that the deterministic threshold for acute radiation injury was exceeded, which is established by the ICRP-60 report guidelines in a peak skin dose of 2 Gy and 3 Gy for temporary erythema and epilation, respectively[14].

These incidents, together with the increasing number of CT scans performed yearly has raised an international concern about the patient imparted doses in CT[7,8]. Guidelines and recommendations for dose optimization in CT have been developed in different initiatives carried out by either official organizations or scientific societies[15-17]. Some of them have been included in national protocols or embraced by CT manufacturers[18].

The Medical Imaging and Technology Alliance (MITA), comprising the five main CT manufacturers (Toshiba, Philips, General Electric, Siemens and Hitachi) have started the *dose check* initiative, which sets up an alert before the scan is performed if the settings selected by the technician lead to an estimated dose index exceeding certain value for that particular protocol[15].

The Heads of the European Radiological Protection Competent Authorities (HERCA) and the European Coordination committee of the Radiological, Electromedical and Healthcare IT Industry (COCIR) have task forces related to CT which have led to several documents in which manufacturers have shown a voluntary compromise to reducing patient dose and encourage protocol optimization[18].

In some hospitals, the patient imparted doses in CT are being recorded. These data, combined with information from the images DICOM headers, enables to perform population studies based on the patients' age, sex and body mass index[19]. The analysis of these databases can help to compare different CT units performance, evaluate the used protocols for each indication and to study the cumulative patient doses over time.

The American Association of Physicists in Medicine (AAPM) has released recommended settings for certain indications such as brain perfusion, lung cancer screening and routine head, chest, abdomen-pelvis and chest-abdomen-pelvis adult CT exams for the main manufacturers and some CT models[17].

Dose optimization in CT can be performed using several approaches, based on the image acquisition and reconstruction phases. The goal is to reduce patient dose without compromising the image quality level needed for the diagnosis. Regarding the image acquisition, the most common strategies are tube current modulation, kV adjustment and adaptive section collimation.

All manufacturers offer tube current modulation systems based on the morphology of the patient, the studied anatomical region or reducing the tube current in the projections that are not used to reconstruct the images[10]. Some systems have tube current modulation synchronized with the electrocardiogram of the patient to perform cardiac-CT and a quite recent approach is based on an adapted modulation based on the location of different critical organs such as eyes, breast or thyroid[20]. Adaptive kV selection can also be applied depending on the patient size, imaged region or if iodinated contrast is used, and it is specially recommended in small patients and children. Dose reductions in the range 25–40% have been observed when selecting 80 kV instead of the ≥120 kV in paediatric CT[21]. Adaptive section collimation, reduces the beam collimation at the beginning and at the end of the programmed scan range. This reduces the dose related to the extra rotations (overranging) that CT systems perform to reconstruct the images in the limit of the programmed range[10].

2. Image quality assessment in CT: Physical measurements

There exist several figures of merit based on physical measurement that have been traditionally used to assess CT scanners performance in an objective way. Some of them can be applied either in the direct space or in the frequency domain. In CT, some acquisition or reconstruction parameters affect in the same fashion all the pixels in the image whereas others are spatially correlated. Some of these parameters are noise, contrast-to-noise ratio (CNR), noise power spectrum (NPS), low contrast resolution, pixel value uniformity or spatial resolution. Phantoms or test objects, containing several structures and cast on different materials are frequently used to assess various parameters related to image quality[10,22].

2.1. Noise in CT

Noise represents the stochastic fluctuations of pixel values, and appears as a graininess in the image. There are different sources of noise in CT, such as electronic noise, quantum noise and structural noise. Quantum noise represents the most important contribution to image noise and it is related to the discrete number of photons reaching the detector, N, which follows a Poisson's distribution. Quantum noise is inversely proportional to the square root of N, which rises with increasing tube charge (mAs), kV and slice thickness values[22]. This relationship is always valid in the raw data and in FBP reconstructed images, but not necessarily in iterative reconstruction. In **Fig. 3** the effect of varying the mAs value is depicted with CT phantom images. **Figure 4** depicts the effect of the selected kV value in image noise.

	Mean pixel value (MPV)	Noise (σ)
(A)	44 HU	33 HU
(B)	44 HU	21 HU
(C)	44 HU	15 HU
(D)	45 HU	10 HU

Fig. 3. Cropped sections from CT images of a phantom, cast on a uniform material, acquired with different tube charge values (mAs). The rest of the acquisition and reconstruction parameters were kept constant.

	Mean pixel value (MPV)	Noise (σ)
(A)	14 HU	5.6 HU
(B)	35 HU	4.0 HU
(C)	45 HU	3.1 HU
(D)	50 HU	2.8 HU

Fig. 4. Cropped sections from CT images of a phantom, cast on a uniform material, acquired with different kV values and keeping the rest of the protocol parameters unchanged.

2.2. Noise power spectrum

The noise power spectrum (NPS) represents the noise amplitude for each frequency value in an image. CT noise is non-stationary in FBP reconstruction meaning that its value is not uniform in the image, existing a noise radial dependency. The formulation of NPS assumes that noise is stationary in the image. Despite some limitations, such as noise being not stationary for CT images reconstructed with iterative algorithms, NPS provides more information than pixel noise or contrast-to-noise ratio and can be useful to analyse FBP reconstructed images[10,11,23]. **Figure 5** depicts how the acquisition parameters, in this example, the tube charge per rotation can modify the magnitude of the NPS curve, whereas its shape, including the lp/cm value for which the NPS reaches a maximum and the cut-off frequency are the same.

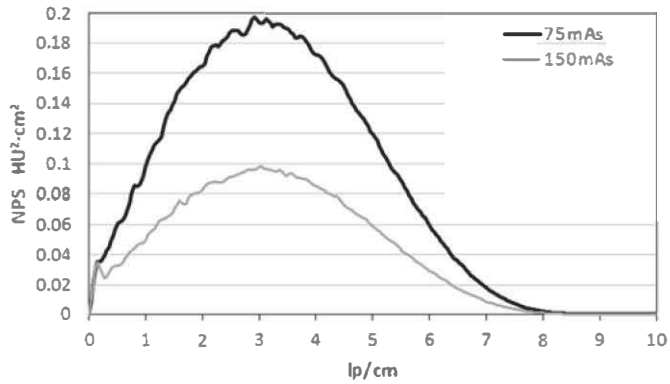


Fig. 5. Noise power spectrum (NPS) measured in images of a uniformity phantom in a CT unit reconstructed with FBP and two dose levels, varying the tube charge (mAs).

The NPS is affected when applying iterative algorithms compared to FBP, as shown in **Fig. 6** for one of the main CT manufacturers.

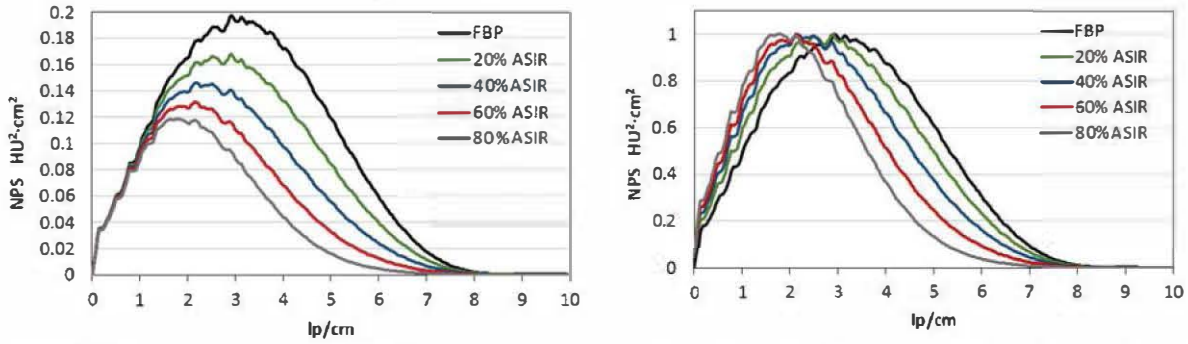


Fig. 6. Noise power spectrum (NPS) measured in images of a uniformity phantom in a CT unit reconstructed with FBP and different levels of iterative reconstruction (left) and the same curves normalized at the maximum NPS values (right).

2.3. Contrast and contrast-to noise ratio

Object contrast in the image is the difference between the pixel value in the object and the surrounding background. It also represents the difference in X-ray attenuation between the object and the background. The visibility of an object in a CT image is determined by its inherent contrast, size and shape and the image noise. The quotient between the contrast of an object and image noise or contrast-to-noise ratio (CNR) is related to its visibility. The attenuation coefficient of a material depends on the effective energy of the X-ray beam which varies with the selected kV and thus can affect object contrast[10,22]. Contrast and contrast-to-noise ratio do not take the frequency dependency into account.

2.4. Spatial resolution

The spatial or high contrast resolution in a CT system is the ability to reproduce small features in both, the image slice plane and through the z-axis (along the patient). CT images are reconstructed applying different reconstruction filters, which can enhance a range of the frequencies in the image. Spatial resolution can be measured in the spatial domain or in the frequency domain. In the spatial domain, it can be determined acquiring images of a phantom containing a small metallic bead and measuring the point spread function (PSF). Other alternatives are the line spread function (LSF) and the edge spread function (ESF)[10]. In the frequency domain, spatial resolution is assessed with the modulation transfer function (MTF), given at some modulation percentage, usually MTF (50%), which represents the image frequency at which the MTF is reduced to 50%. Many factors determine the spatial resolution, either related to the scanner itself or the selected acquisition and reconstruction parameters. Traditionally in CT, when only FBP was available, MTF was calculated based on the ESF of high contrast disks. With the introduction or iterative algorithms, it is recommended to measure MTF for different materials, covering a wide range of attenuation properties[24]. An additional method to evaluate spatial resolution consists on human observers scoring the visibility of patterns of line pairs in phantoms.

Figure 7 shows the effect of selecting different reconstruction kernels in spatial resolution.

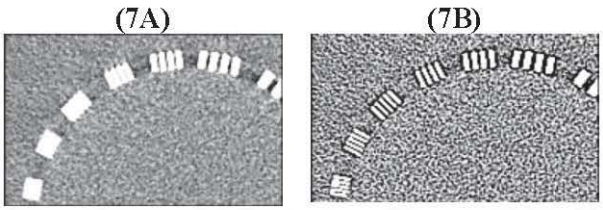


Fig. 7 . Cropped sections from CT images of a phantom's module to assess spatial resolution, acquired with different kV values keeping the rest of the acquisition and reconstruction parameters constant. Image **(7A)** was reconstructed applying a *soft* kernel, and **(7B)** with a *sharp* kernel.

2.5. Low contrast detectability

One of the reasons of the popularity of CT is its capacity to reveal lesions with attenuation properties very similar to those of the surrounding tissue (i. e.: with low contrast), that cannot be detected with other medical imaging techniques. Noise can mask lesions, especially if they are small and with low contrast. LCD is influenced both by NPS and MTF, that makes it an important image quality test. It takes into account the frequency dependencies of contrast and noise. The system blur may deteriorate the contrast of small objects and thereby may influence detectability.

Low contrast detectability (LCD) is frequently determined in human observer studies, scoring the visibility of low contrast objects in phantom images acquired with different protocols, assessing the smallest object of a given contrast level that can be detected[10,22].

As an example, **Fig. 8** shows the low contrast module of the Catphan phantom, used in quality control in CT, which contains three groups of objects with different contrast levels and diameters.

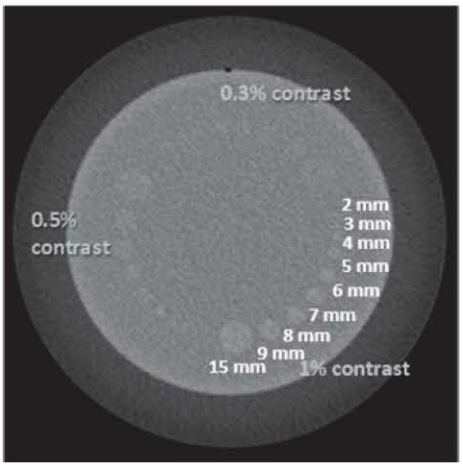


Fig. 8. CT image of the low contrast module of the Catphan phantom.

Low contrast detectability improves when the image noise is decreased. Some examples are shown in **figures 9** and **10**, which reflect how increasing slice thickness and tube charge, decrease the noise level and boost LCD.

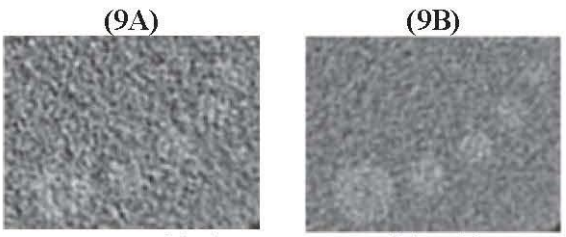


Fig. 9. Images of the 1% contrast group of the low contrast module of the Catphan phantom, reconstructed with slice thicknesses of 1.25 mm **(9A)** and 5 mm **(9B)**, respectively.

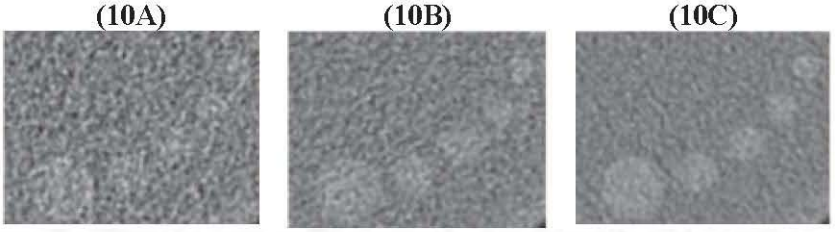


Fig. 10. Images of the 1% contrast group of the low contrast module of the Catphan phantom, acquired with tube charge values of 50 mAs **(10A)**, 100 mAs **(10B)** and 200 mAs **(10C)**, respectively.

In the dose optimization process it is very important to take into account what the radiologists need to detect for that particular indication. That will determine the lowest dose level (depending on the patient anatomical characteristics) that can be used to reach a confident diagnosis of presence or absence of abnormalities and determine the protocol parameter values. For this goal, the appropriate assessment of low contrast detectability is essential.

CT noise is textured, which means that there is a spatial frequency dependence of the noise in an image. Reconstruction filters are applied to the attenuation profiles to generate the images enhancing certain aspects of the image, reducing the noise in a limited frequency range, depending on the diagnostic task and the scanned region. Thus, some of these filters or kernels create smoother images, improving detail visibility, decreasing noise at low frequencies and degrading the edge definition of structures. Others, offer higher spatial resolution, improved edge definition and a sharper image, at the expense of a higher noise level[10,22]. Examples of the effect of reconstruction filter in spatial resolution are shown in **Fig. 7**.

Low contrast resolution is highly dependent on the reconstruction algorithm, as shown in **Fig. 11**. The value of image noise and CNR in image quality assessment is limited because they do not reflect the image frequency correlations, which can affect the objects detectability.

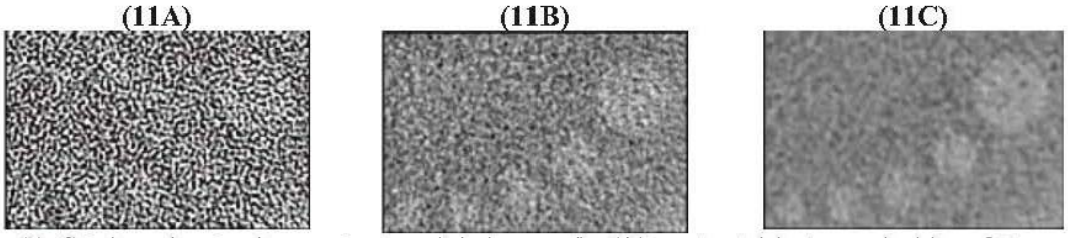


Fig. 11. Catphan phantom low contrast module images for 1% contrast objects acquired in a CT scanner selecting different reconstruction filters: in **(11A)** (sharp kernel) and **(11C)** (soft kernel) high and low frequencies in the image are enhanced, respectively, and **(11B)** depicts the effect of a standard kernel.

Low contrast detectability can be affected if the images are reconstructed selecting iterative algorithms, which can have an impact in lesion detection, as shown in **Fig. 12**[25,26].

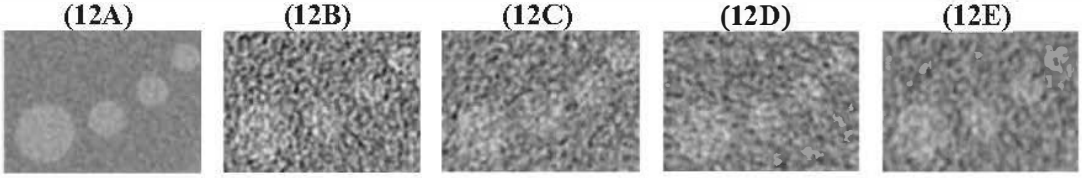


Fig. 12. Catphan phantom low contrast module images for 1% contrast objects acquired in a CT scanner selecting different reconstruction algorithm levels: **(12B)** (FPB), **(12C)** (IR 40%), **(12D)** (IR 60%) and **(10E)** (IR 80%) for the same kernel. **(12A)** is shown as a reference of the objects array.

3. Human observer studies in medical imaging

Perception studies with human observers are one of the pillars of medical image quality assessment. For these experiments, a group of observers, score sets of images based on the assigned task, for example, assessing if abnormalities are present in the images or not. Depending on the goal of the study they can be radiologists, experts or naïve observers. These results can be used to investigate the observers' performance, rank them or to determine if a certain imaging system or protocol is suitable for a given diagnostic purpose.

The first human observer studies in Radiology images were carried out in the 1940s to obtain contrast-detail (CD) diagrams based on the scoring of the visibility of disks embedded in a phantom. The objects, of different diameters and X-ray attenuation properties, were arranged in rows of decreasing attenuation and in columns of decreasing object diameter. The CD curves represented the threshold contrast that the observer could detect in the images as a function of the object diameter, both in logarithmic scale[27].

The methodologies that are more widely used for human observer studies are the receiver operating characteristic (ROC) and the multi-alternative forced choice (M-AFC). For both, two sets of images are analysed by the observers, one containing abnormality (or positive) and another without it (negative), defined depending on the diagnostic task. The goal of these methods is to determine if the imaging system or the protocol used to acquire the images can be used to discriminate between normal and abnormal cases[1,28].

3.1. Receiver operating characteristic (ROC) studies

The receiver operating characteristic (ROC) analysis represents human performance in detection or classification tasks. It is based on the signal detection theory (STD)[29,30]. STD was ignited in the World War II when different mathematical methods were developed to detect signals in the presence of noise in radar communications, such as flocks of birds which could collide with the planes. ROC analysis is widely used in Medicine, when the task is to decide if the investigated case is 'normal' or 'abnormal', which is a binary task. This methodology is useful for qualitative performance comparisons, for example between observers, or the diagnostic utility of two imaging modalities for a certain indication [1,27,28].

When a case is investigated to decide if it is 'normal' or 'abnormal', it is a binary task that can be represented in a 2x2 table displaying all the possible outcomes, as shown in **Table 1**.

Table 1. Decision matrix for a ROC study based on the classification of abnormal and normal cases

	Abnormality present	Abnormality absent
Diagnosis: abnormal	True positive (TP)	False positive (FP)
Diagnosis: normal	False negative (FN)	True negative (TN)

Based on the variables shown in **Table 1**, two quantities can be defined, the true-positive fraction (TPF), also called sensitivity and the false positive fraction (TPF), as follows:

$$TPF = \frac{TP}{TP + FN} = Sensitivity \quad (1) \qquad FPF = \frac{FP}{TN + FP} = 1 - \frac{TN}{TN + FP} = 1 - Specificity \quad (2)$$

The ROC curve is built representing the TPF (sensitivity) as a function of the FPF (1−sensitivity) for the analysed cases. The area under the ROC curve (AUC) is the summation of the observer sensitivities at all specificity values and can represent the accuracy of the observer or the imaging system for the assigned task in the analysed images.

3.2. Multi-alternative forced choice (M-AFC) experiments

In multi-alternative forced choice experiments, several images or alternatives are shown simultaneously and the observer is 'forced' to select the image that meets the task, for example, the one that contains the lesion[28]. An example of an interface used to display the images in a M-AFC experiment is shown in **Fig. 13**. It represents a 4-AFC experiment for a signal known exactly and background known exactly (SKE/BKE) task. Information about the target size, shape, signal and location in the image are given to the observer. The quantity that is measured in M-AFC studies is the proportion correct (PC) ratio obtained by the observer dividing the number of correct decisions between the number of scored images.

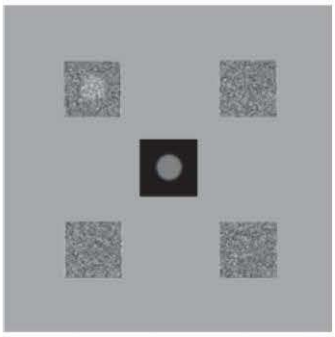
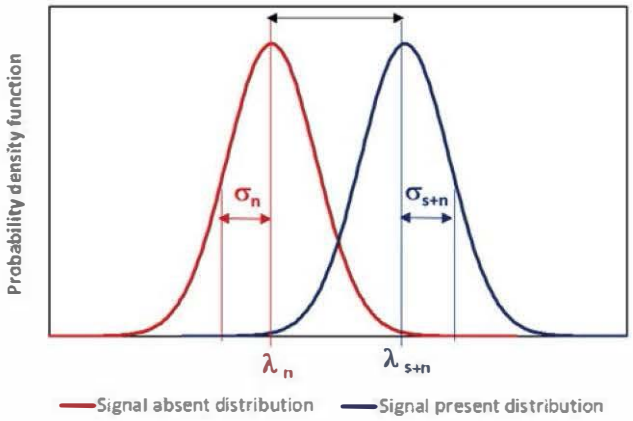


Fig. 13. Example of a software interface used to perform 4-alternative forced choice detection experiments.

The core of M-AFC experiments is to quantify the observer's ability to distinguish between two distributions, one related to the object present (which can be an abnormality) and another to the abnormality absent set, measured by a parameter called detectability (d')[28]. The observer has no information about the origin of each of the samples, which can be the set of signal present or absent images. A decision criterion is applied to score the images and determine from which distribution the scored images come from.

The observer decision process can be modeled considering that the probability of the presence of the object in each of the displayed images is pondered or measured internally, assigning a value to a decision variable, λ for each image. These decision variables will have a variance due to the presence of noise in the images. The d' is defined as the difference between the decision variables of the abnormality present (s+n) and absent (n) images divided by the squared root of the average variance (σ^2) of both distributions (**Fig. 14**), as shown in Eq. (**3**)[1,27,28,30].



$$d' = \frac{\lambda_{s+n} - \lambda_n}{\sqrt{0.5\left[\sigma_{s+n}^2 + \sigma_n^2\right]}} \quad \textbf{(3)}$$

Fig. 14. Probability density functions of two classes of images, one with the signal present (s+n) and one with signal absent (n). The mean values (λ) and the statistical deviation σ of each distribution are also shown. The detectability index d' for the detection task appears on the right.

Assuming that the decision variables follow Gaussian distributions, for 2-AFC studies, PC represents the area under the ROC curve. The AUC can be related to the detectability index with the following equation[1,27,28]:

$$d' = 2\, erf^{-1}[2(AUC) - 1] \quad \textbf{(4)}$$

where erf $^{-1}(\cdot)$ is the inverse error function, being erf (x): $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-x^2} dx$ **(5)**

3.3. Designing perception studies: practical considerations

The design of human observer studies is a complex issue as it is necessary to consider different aspects such as defining the study goals, finding an appropriate setup and planning the experiments, and the posterior analysis of results. Protocol optimization based on the analysis of CT images with human observers is even more arduous due to the wide range of parameters that affect image quality and patient dose and the existing inter-relations between them.

Some image quality parameters are still assessed by human observers scoring phantom images. For example, to measure low contrast detectability the observer has to determine the smallest object of lowest contrast that can be detected in phantom images acquired with different protocols (**Fig. 8**). The observers know beforehand the distribution of the objects in the phantom, which can introduce a bias in the results[1].

The experience of the selected sample of observers for the intended task has to be taken into consideration in perception studies[1,28]. For instance, the same degree of experience is not necessary for simple detection tasks in Gaussian white noise backgrounds than to detect lung nodules in patients CT images.

Human performance can be unreliable and difficult to reproduce. There exist an intra-observer variability, as there is a probability that the same observer scoring the same case twice reach a different decision. To reduce this effect, redundant readings or session repetitions can be programmed with a time gap between them to avoid learning effects. The images should be displayed at random in the different sessions too. The inter-observer variability has to be taken into account too, as different observers analysing the same sets of images can get different results. Typically, the number of observers in perception tests is between four and six, a number that can be increased in large studies with patient clinical images[1,27,31].

The reading environment affects human performance in perception studies. In particular, to score medical images, the ambient illumination has to be approximately constant and kept low following recommended visualization conditions and in a monitor calibrated to display DICOM images[1,32]. Visual and acoustic distractions should be avoided during the image reading sessions.

Human observer studies based on radiology images are time consuming. Even when the aim is to study a simple task, for example, the detection of objects in a uniform background, the amount of images to be analysed and the number of observers involved can make the resulting data difficult to handle. If the amount of images to analyse is big, care has to be taken to perform the study in different sessions to avoid undesired biases related to the observers' fatigue. Training sessions and pilot studies are necessary before any perception human study to make the observers familiar with the task and the handling of the images[1,28].

4. Objective assessment of low contrast detectability in CT

4.1. Methods based on grids and uniformity phantoms

Different approaches to assess LCD in an objective way, not based in model observers, have been published. They are based on making measurements on CT images of a mono-material or uniformity phantom[33,34]. A group of ROIs of equal size distributed over the images and the mean pixel value is measured in them. This method assumes that in CT images, the means of a group of low contrast objects of the same size follow a Gaussian distribution. The measured means of the ROIs taken in the uniformity phantom will also follow a Gaussian distribution with the same standard deviation as the distribution related to the objects. Both distributions only differ in their mean. The middle point between both distributions can be taken as a visibility threshold for the objects to be detected in the surrounding background. The contrast that an hypothetical object of the same size as the selected ROI should have to be detected with a 95% confidence interval, can be calculated as 3.29 times the standard deviation of the ROIs mean values. Repeating the measurements with different ROI sizes, contrast-detail graphs can be obtained.

The number of ROIs taken and the number of images has to be high, especially for noisy images. The method proposed by Chao et al is based on a distribution of ROIs in a matrix centered in the phantom[33]. Torgensen et al use circular ROIs randomly placed and have developed software to automatically create contrast-detail curves based on the measured values using a phantom specifically developed for image quality control in dental cone beam CT (CBCT) devices[34]. These methods, though useful for quality control and LCD constancy measurements, have not been checked with human performance and due to the assumptions for the mean pixel values distributions they might not be applicable in the case of images reconstructed with iterative algorithms or with complex backgrounds.

4.2. Model observers

Medical imaging systems are becoming more complex in the sense that more parameters can be set up by the user to acquire, reconstruct or visualize the images. The clearance of a new system or technological improvement is based on the results of clinical trials involving a subpopulation of patients and radiologists. These studies are not justified for systems that are at an early stage of development or undergoing through validation[1,35].

Human observer studies are useful in that case, analysing geometric or anthropomorphic phantom images. The drawbacks of these studies were stated in the previous section, especially being time-demanding, complex and expensive to conduct[35]. Simplified perception studies, in which skilled observers, such as medical physicists, perform simple tasks can be also selected to perform image quality analysis. Their usefulness is limited by the range of conditions and images analysed, which rarely represent all the available options in the device.

Human visual system is based on several complex stages. Photons from the objects reach the eye surface and pass through the pupil, they travel inside the eye globe and interact with the sensitive detectors, cones and rods in the retina. The response of them to the

visual stimulus is transformed into nervous impulses that reach the visual cortex through the optic nerves, where the visual interpretation takes place.

Model observers are mathematical algorithms, which were introduced in medical imaging in the 1950s and aim to predict the human observer performance, especially in technology validation. They are usually applied to simple detection and discrimination tasks with noisy images containing objects with different characteristics, for example phantom images acquired with different protocols[36]. Two model observers widely used in different medical imaging modalities are the non-prewhitening matched filter with an eye filter (NPWE) and the channelized Hotelling model observer (CHO)[1,36].

The first model applied in medical imaging was the Bayesian or ideal model observer. It uses all the available information in the image, calculating the likelihood ratio. This model outperforms human observers and it does not include any type of internal uncertainty to the decision process, only the intrinsic variability in the analysed images. For detection tasks with uncorrelated Gaussian white noise, the ideal model observer overestimates humans but can reproduce the performance trends. When the noise is frequency dependent in the image, as for example in CT with the reconstruction algorithms, human performance is highly dependent on the noise spectrum and the object characteristics, so that the ideal observer is a poor predictor in this case. Besides, for signals embedded in complex backgrounds the likelihood ratio can be difficult or even impossible to calculate, and this model cannot be applied[37,38].

The implementation of model observers for detection tasks is based on performing different transformations to the classes of images that are to be compared (abnormality present, I_1 and abnormality absent, I_2, images) to finally calculate decision variables and reach a decision by comparison with a threshold. The model observers that are used in medical imaging analysis are in general linear. For these models and 2-AFC experiments, the two classes of images are multiplied by a template, which varies depending on the model observer. This template represents the strategy followed by the model to deal with the possible objects in the image.

The process of applying the template to the images is described in Eq. (**6**) and results in two decision variables, one forthe samples with abnormality present (T_1) another one for the samples without it (T_2)[1]:

$$T_i = w^t I_i = \sum_{n=1}^{N^2} w_n I_n \quad \textbf{(6)}$$

where $\boldsymbol{w}^t \boldsymbol{I_i}$ is an inner product between the column vectors of the template ***(w)*** and the image ***(I)*** and the subindex i can take values 1 (abnormality present) or 2 (abnormality absent). Finally, from the test statistics, of the resulting distributions, a detectability index can be calculated using Eq. (7):

$$d' = \frac{\langle T \rangle_1 - \langle T \rangle_2}{\sqrt{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2}} \quad (7)$$

where $\langle \cdot \rangle$ refers to the mean of the decision variables, $\sigma(\cdot)$ is their standard deviation and subindexes 1 and 2 denote the image class, respectively.

The detectability index can be transformed into proportion correct (PC) using Eq. (8)[1,37]:

$$PC = 0.5 + 0.5\, erf\left(\frac{d'}{2}\right) \quad (8)$$

4.3. Tuning the model observer results

Model observers can reproduce human observers' performance for certain detection or discrimination tasks but, in general, they outperform them. There exist different approaches to tune the models output to obtain results closer to humans.

4.3.1. Internal noise calibration

Internal noise (α) is frequently added to the model observer decision variable (T) to reproduce certain aspects of the human performance in visual detection, in particular, that the observer can reach different decisions for the same images, as shown in Eq. (9)[1,38,39]:

$$T'_i = T_i + \alpha\, x \quad (9)$$

where the subindex i denotes the image class (i = 1, 2), x is a variable that follows a normal distribution with zero mean and a standard deviation σ that can be obtained as the square root of the variance of the decision variable in the abnormality absent images. In general, in CT detectability studies, a range of object contrasts and sizes is involved, and also images are acquired with different dose levels. There are different approaches to calibrate α. One of them is to take the model results for one of the analysed objects from the images acquired at an intermediate dose level in the studied range and apply Eq. (9), obtaining new d' and PC values as a function of α[40]. These recalculated PC values are compared with the human results for the same condition, to find the α value that makes them match. That internal noise level is then applied to recalculate the model observer performance in all the study.

4.3.2. Efficiency

One of the methods to compare observers is to calculate the efficiency (η). This parameter was proposed by Tanner and Birdsall in 1958 for psychometric experiments based on acoustic signals[41]. The performance of the human observer (for example represented by d'_{human}) can be set against a model observer ($d'_{model\ observer}$) as follows[1,41]:

$$\eta = \frac{(d'_{human})^2}{(d'_{model\ observer})^2} \quad (10)$$

After the efficiency is calculated for the given task, the model results are corrected by its value and detectability indexes recalculated.

4.4. Model observers in CT

4.4.1. NPWE model observer

This model is a modification of the non-prewhitening matched filter (NPW) with the addition of a so called eye filter (***E***), which is a function representing the human eye contrast sensitivity function (CSF)[42]. This function, obtained experimentally, represents the measured contrast detectability threshold for a range of spatial frequencies. Different studies have been carried out to assess these curves based on human observers' data and there are different functions published[1,43,44]. Among them, the eye filter proposed by Burgess has been widely used in combination with the NPW model for assessing image quality in mammography and CT, amongst others[45,46].

$$E(f) = f e^{-bf} \quad (11)$$

where f represents the spatial frequency and b is chosen to make E(f) reach the maximum at 4 cycles per visual angle degree, which corresponds to the middle frequencies in the image.

Figure 15 represents this eye filter, depending on the distance to the monitor.

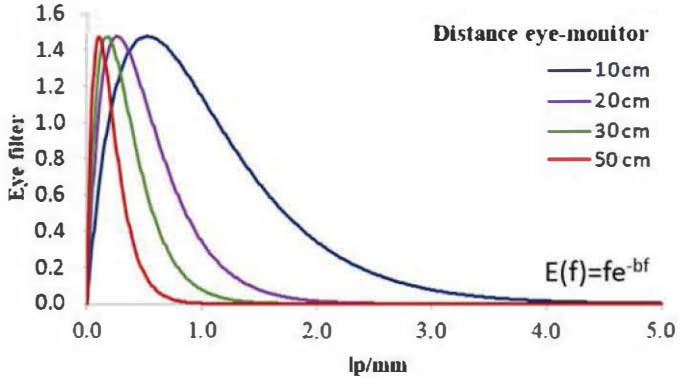


Fig. 15. Eye filter proposed by Burgess which peaked at 4 cycles per degree, for different distances eye-monitor.

The flowchart in **Fig. 16** illustrates a possible implementation of NPWE. The expected signal is modified to take into account the imaging system and it is convolved with the eye filter, which results in the template. The sets of images with or without abnormality are also filtered by ***E***. Cross-correlations are performed between the template and the filtered sets of images and test statistics are calculated to derive a detectability index.

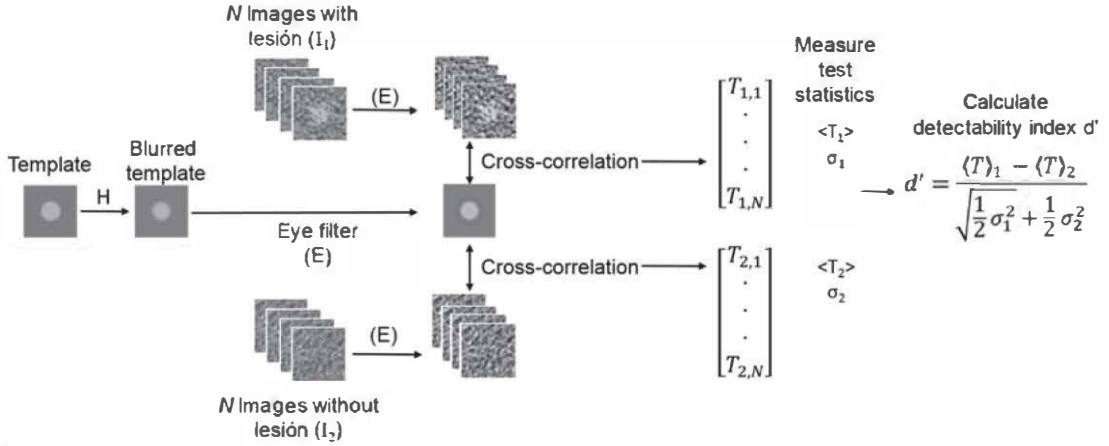


Fig. 16. Flowchart describing a possible implementation of the NPWE model observer. H represents the CT system blur, < > represents the mean value and σ represents the statistical deviation of the cross-correlations of the template with the N images with (1) and without abnormality (2).

4.4.2. CHO model observer

In the 1950s and 1960s several perception studies were performed to study the human performance based on the visualization of patterns of luminance or gratings following different distributions like sinusoids, saw-tooth or rectangular waves. The results suggested that the visual cortex interpretation process can be modelled as a group of independent receptors, which were sensitive only for a narrow range of spatial frequencies of the visual stimulus. Detection is triggered when a certain threshold is reached in one of these receptors which are called channels[47,48].

The channelized Hotelling observer (CHO) aims to mimic this behavior using channels to filter the images (abnormality present or absent, for example). The test variables (T_1 and T_2) for both classes of images are in this case (Eq. (**12**))[1]:

$$T = w^t_{CHO}\, I_{ch} \quad \sum_{m=1}^{M} w_{CHO}^{m} I_{ch}^{m} \quad (\mathbf{12})$$

where the total number of channels is M, I_{ch} is the transformed image after being filtered by the channels and w_{CHO} is the template given by Eq. (**13**):

$$w_{CHO} = \overline{K_c}^{-1}(\langle I_{1\,ch}\rangle - \langle I_{2\,ch}\rangle) \quad (\mathbf{13})$$

where $\overline{K_c}^{-1}$ is inverse of the average of the covariance matrices of the signal present and absent classes after being filtered by the channels, and <I_{ch}> represents the mean of each class (1 abnormality present, 2 abnormality absent) after being filtered by the channels.

Different implementations of these channels are available in the literature, such as square band-pass radial frequency filters, differences of Gaussians or Laguerre-Gauss channels, among others. One of these implementations was proposed by Wunderlich et al, based on Gabor channels (Eq. (**14**))[49]:

$$Ga(x,y) = \exp\left[-\frac{4(ln2)((x-x_o)^2+(y-y_o)^2)}{\omega_s^2}\right] \cdot \cos\left[2\pi f_c\left((x-x_o)cos\theta + (y-y_o)sin\theta\right) + \beta\right] \quad (\mathbf{14})$$

where ω_s is the channel width, f_c is the central frequency and β is a phase factor.

A CHO model observer based on this implementation and extended by adding extra channel passbands has been applied to detection and discrimination tasks in CT phantom images[40]. To illustrate the channels used in this particular model, **Figure 17** shows the 60 generated channels.

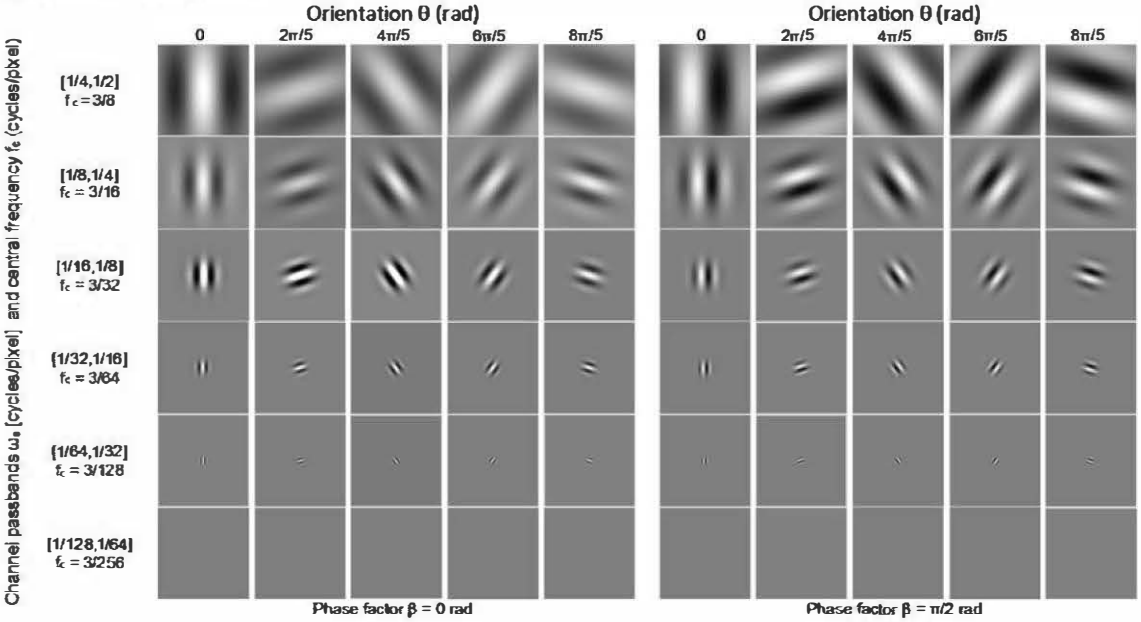


Fig. 17. Images of the 60 Gabor channels used for a CHO model applied in CT phantom images detection tasks, together with the parameter values (channel passbands, ω_s, central frequency, f_c, orientation, θ, and phase factor, β) selected to generate the channels.

Motivation, hypothesis and objectives

There is a need for automated methods for the analysis of image quality, especially in modalities such as CT in which many acquisition and reconstruction parameters, that in turn can take a range of values, affect image quality.

The motivation of this thesis was to develop a framework to assess image quality in CT images in an objective way based on model observers. By the beginning of this work, in 2010, the use of model observers for detection and discrimination tasks in CT images, had not been explored.

In particular, low contrast detectability (LCD) was of interest as it is a parameter that can be highly affected by the selected acquisition and reconstruction options. This parameter is critical in CT as small low contrast lesions can be masked by noise and also varies with spatial resolution. Traditionally, it has been assessed by human observers scoring the visibility of low contrast objects in phantom images. These studies are complex and expensive and they have to be carefully planned and performed.

The starting hypothesis of this thesis is that model observers can be applied for certain detection and discrimination tasks in CT phantom images and predict human observer performance for LCD, selecting different protocols. If this hypothesis is corroborated model observers can be a fast and objective tool to assess low contrast detectability in CT and used in technology validation and protocol optimization.

The goals and milestones of this thesis are (not necessarily in chronological order):

1. To develop a software to automatically extract samples from phantom images, containing objects or background. In particular in a phantom containing distributions of low contrast objects.

2. To implement a model observer (NPWE) to assess LCD in CT phantom images. To compare the model performance with results in the literature based on the detection of objects in simulated Gaussian white noise backgrounds.

3. To investigate the effect of selecting different acquisition and reconstruction parameters in LCD performance for the model observers and humans in simple detection tasks in phantom images. In particular, to study the influence of selecting a range of kV, tube charge per rotation and reconstruction kernel settings.

4. To develop a software to perform 2-alternative forced choice experiments with human observers to enable a quantitative comparison between humans and model observers.

5. To implement the channelized Hotelling observer (CHO) in the framework as an alternative for NPWE. To investigate the influence of the selected kVp in LCD when dose is kept constant, comparing CHO and NPWE performance with human observers.

6. To study the influence of iterative reconstruction algorithms in LCD with the model observer and human observers compared to FBP algorithms.

PhD thesis outline

This PhD thesis is a collection of four papers, published in the fields of medical imaging and radiology. They are organized in chronological order of publication and will be referred to using capital Roman numerals in the text:

[I] I. Hernández-Girón, J. Geleijns, A. Calzado, M. Salvadó, R. M. S. Joemai, W. J. H. Veldkamp. Objective assessment of low contrast detectability for real CT phantom and in simulated images using a model observer. IEEE Nuclear Science Symposium Conference Record 2011;3477-3480 (doi:10.1109/NSSMIC.2011.6152637)

[II] I. Hernández-Girón, J. Geleijns, A. Calzado, W. J. H. Veldkamp. Automated assessment of low contrast sensitivity for CT systems using a model observer. Med Phys 2011;38:S25-S35 (doi:10.1118/1.3577757)

[III] I. Hernández-Girón, A. Calzado, J. Geleijns, R. M. S. Joemai, W. J. H. 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. Br J Radiol 2014;87:20140014 (doi: 10.1259/bjr.20140014)

[IV] I. Hernández-Girón, A. Calzado, J. Geleijns, R. M. S. Joemai, W. J. H. Veldkamp. Low contrast detectability performance of model observers based on CT phantom images: kVp influence. Phys Medica 2015 Corrected proof in press (doi: 10.1016/j.ejmp.2015.04.012)

Material and methods and Results

This section gathers the following four papers, published in the fields of medical imaging and radiology. They are organized in chronological order of publication and will be referred to using capital Roman numerals in the text. Each of them constitutes a subsection of this section, which is chapter 4 in this thesis.

[I] I. Hernández-Girón, J. Geleijns, A. Calzado, M. Salvadó, R. M. S. Joemai, W. J. H. Veldkamp. Objective assessment of low contrast detectability for real CT phantom and in simulated images using a model observer. IEEE Nuclear Science Symposium Conference Record 2011;3477-3480 (doi:10.1109/NSSMIC.2011.6152637)

[II] I. Hernández-Girón, J. Geleijns, A. Calzado, W. J. H. Veldkamp. Automated assessment of low contrast sensitivity for CT systems using a model observer. Med Phys 2011;38:S25-S35 (doi:10.1118/1.3577757)

[III] I. Hernández-Girón, A. Calzado, J. Geleijns, R. M. S. Joemai, W. J. H. 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. Br J Radiol 2014;87:20140014 (doi: 10.1259/bjr.20140014)

[IV] I. Hernández-Girón, A. Calzado, J. Geleijns, R. M. S. Joemai, W. J. H. Veldkamp. Low contrast detectability performance of model observers based on CT phantom images: kVp influence. Phys Medica 2015 Corrected proof in press (doi: 10.1016/j.ejmp.2015.04.012)

4.1. Implementation of a model observer for low contrast detection tasks in simulated and CT images.

[I] I. Hernández-Girón, J. Geleijns, A. Calzado, M. Salvadó, R. M. S. Joemai, W. J. H. Veldkamp.

Objective assessment of low contrast detectability for real CT phantom and in simulated images using a model observer.

IEEE Nuclear Science Symposium Conference Record 2011;3477-3480 (doi:10.1109/NSSMIC.2011.6152637)

Abstract

The variability in doses and image quality used by different manufacturers of scanner models to reach similar diagnostic tasks has been proved to be wide. Image quality is frequently assessed performing human observer studies scoring the visibility of objects on CT images. These studies may become time consuming and expensive due to the high number of observers and observations required. Besides a bias can appear as the objects are arranged in patterns the observer knows beforehand. A great inter and intra-observer variability may exist too. Computer model observers attempt to objectively predict human performance on the images and seem useful in investigating the influence of acquisition and reconstruction parameters and object size or shape on CT images.

We have developed an objective statistical method with a model observer (non-prewhitening matched filter with an eye filter, NPWE) for detection tasks on CT, implementing characteristics of the selected kernel in the method. Images of the Catphan low contrast module (containing low contrast objects distributions) were acquired under different dose settings. Detectability (d') and proportion correct (PC) values were obtained for each object in the phantom. The results showed that d' increased with object size and mAs, and higher values were obtained as object contrast increased. Psychometric fits were performed and a visibility threshold of PC≥75% was established. In this way, the smallest visible object for each condition was obtained. To validate the model it was also applied to detection tasks on simulated white noise background images.

Objective assessment of low contrast detectability for real CT phantom and in simulated images using a model observer

I. Hernández-Girón, J. Geleijns, A. Calzado, M. Salvadó, R. M. S. Joemai and W. J. H. Veldkamp

Abstract–**The variability in doses and image quality used by different manufacturers for scanner models for each similar diagnostic tasks has been proved to be wide. Image quality is frequently assessed performing human observer studies scoring the visibility of objects on CT images. These studies may become time consuming and expensive due to the high number of observers and observations required. Besides a bias can appear as the objects are arranged in patterns the observer knows beforehand. Besides, a great inter and intra-observer variability may exist. Computer model observers attempt to objectively predict human performance on the images and seem useful in investigating the influence of acquisition and reconstruction parameters and object size or shape on CT images.**

We have developed an objective statistical method with a model observer (non-prewhitening matched filter with an eye filter, NPWE) for detection tasks on CT, implementing characteristics of the selected kernel in the method. Images of the Catphan low contrast module (containing low contrast objects distributions) were acquired under different dose settings. Detectability (d') and proportion correct (PC) values were obtained for each object in the phantom. The results showed that d' increased with object size and mAs, and higher values were obtained as object contrast increased. Psychometric fits were performed and a visibility threshold of PC≥75% was established. In this way, the smallest visible object for each condition was obtained. To validate the model it was also applied to detection tasks on simulated white noise background images.

I. Hernández-Girón is with the Unitat de Física Mèdica, Universitat Rovira i Virgili, 43201 Reus, Spain; and with the Departamento de Radiología, Universidad Complutense de Madrid, 28040 Madrid, Spain (email: irene.debroglie@gmail.com)

J. Geleijns is with the Radiology Department, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands (email: K.Geleijns@lumc.nl)

A. Calzado is with the Departamento de Radiología, Universidad Complutense de Madrid, 28040 Madrid, Spain (email: calzado@med.ucm.es)

M. Salvadó is with the Unitat de Física Mèdica, Universitat Rovira i Virgili, 43201 Reus, Spain (email: m.salvado@urv.cat)

R. M. S. Joemai is with the Radiology Department, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands (email: R.M.S.Joemai@lumc.nl)

W. J. H. Veldkamp is with the Radiology Department, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands (email: W.J.H.Veldkamp@lumc.nl)

I. Introduction

Evaluating image quality is essential to investigate new acquisition protocols or new technical developments in CT. Receiver operating characteristics (ROC) studies with human observers are a frequent approach to assess image quality but they may become time consuming and expensive due to the high number of observers and observations required.

Low contrast detectability (LCD) on CT is frequently determined by human observers scoring the visibility of low contrast objects within phantom images. As these objects are arranged in patterns the observer knows beforehand, this method might be biased. Besides, a substantial inter and intra-observer variability may exist.

As an alternative, computer model observers attempt to objectively predict human performance on the images. They seem useful in investigating the influence of acquisition and reconstruction parameters and object size or shape on CT images.

We have developed an objective statistical method with a model observer (non-prewhitening matched filter with an eye filter, NPWE) to automatically investigate the influence of some CT parameters related to dose on LCD. The model was also applied to study the detectability on simulated images of low contrast objects in white noise backgrounds.

II. Materials and Methods

The low contrast module of the Catphan 600 phantom was used in our study, especially the supra-slice region. It contains three groups of low contrast objects (each consisting of 9 circular objects with diameter 2-15 mm and contrast 0.3, 0.5 and 1.0%, respectively). Images were acquired in a 16-detector row CT scanner (Aquilion 16, Toshiba, Japan) selecting the parameters shown in Table I with different values for the tube charge per rotation.

The reconstruction was performed with 0.5 mm as slice thickness and 0.5 mm as reconstruction interval with FC12 (smooth convolution kernel; body) as reconstruction filter. A set of 75 images was available for each selected mAs value.

978-1-4673-0120-6/11/$26.00 ©2011 IEEE

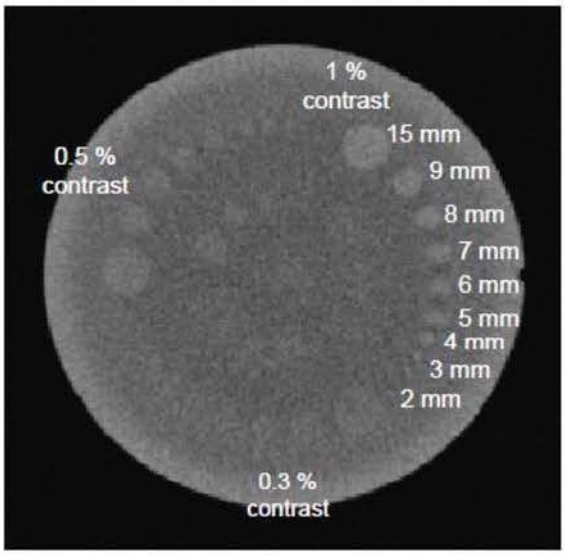


Fig. 1. A constructed 40 mm thick slice of the Catphan low contrast module. The contrast groups and the object diameters are tagged for the supraslice region.

TABLE I. ACQUISITION CONDITIONS FOR REAL CT IMAGES

Collimation 16x0.5mm	Helical acquisition	120kV		
FOV400mm	Pitch factor 0.94	25 mAs	50 mAs	100 mAs

The software automatically calculated low contrast detectability using a NPWE model observer for each object and the three contrast groups present in the low contrast module of the phantom [1]. Polar coordinates were used to locate the disks (signals) using their relative position to the module centre. Background (no signal) samples were obtained from an area located at the same polar angle as the smallest disk of each contrast group but positioned further from the module centre. The model performs a two-alternative forced choice (2-AFC) detection task comparing test statistics related to signal (T1) and background (T2) samples [2]. These values are obtained by cross-correlation between the expected signal (template for each object), the selected background region and the appropriate known signal region in the image, respectively [3]. This procedure was performed for all low contrast objects in the three groups in all the CT images. Before correlation, the templates are blurred using the kernel MTF value.

The point spread function (PSF) was measured acquiring images of a metallic bead of 0.18 mm of diameter. The MTF was modelled taking the corresponding value for the full width at half maximum (FWHM) of the measured PSF [4].

Finally, both, the image regions (signal and background) and the templates were filtered by the human visual-response function (E) to account for the frequency response of the human eye. The applied eye filter was $E(f) = fe^{-bf}$, with b chosen such that E(f) peaked at 4 cycles per degree [5]. A fixed viewing distance of 40 cm from the monitor was assumed.

From the distribution of test statistics, a discrimination index d' was calculated as shown in (1):

$$d' = \frac{\langle T \rangle_1 - \langle T \rangle_2}{\sqrt{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2}}, \quad (1)$$

where $\langle \bullet \rangle$ is the mean and $\sigma(\bullet)$ is the standard deviation of the respective distributions of signal/background samples [2]. This index can be used as a measure of detection performance and related to object diameter.

When considering normally distributed test statistics, d' can be also related to the area under the receiver operating characteristic (ROC) curve (AUC). This AUC is equal to the proportion correct (PC) in a two-alternative forced choice (2-AFC) experiment [2]. PC values were determined using (2):

$$PC = 0.5 + 0.5erf\left(\frac{d'}{2}\right), \quad (2)$$

where *erf(x)* is the Gaussian error function (3):

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2} dx \quad (3)$$

Additionally, to validate the performance of our method in other noise distributions, simulated images were created by adding computer-generated white noise to images with disk shaped signals. These objects were of the same size as the supra-slice low contrast objects in the real Catphan CT images and were located at the same positions. The difference in signal of these circular objects with their background was 7.65. White noise was finally added to these images. The noise was generated from a Gaussian distribution with a mean of 0 and a standard deviation of 30 [3]. A total of 1000 images were generated in this way and d' and PC values were calculated for each disk size as in real CT images.

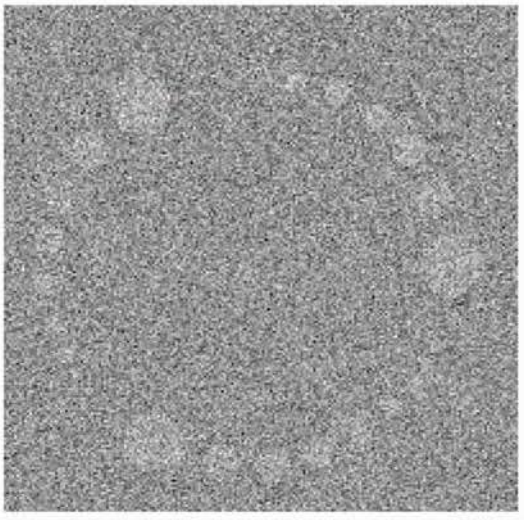


Fig. 2. Simulated image of the low contrast module of the Catphan phantom objects in white noise.

As, just by chance in a 2-AFC experiment a PC=50% may be obtained, we propose a visibility threshold of PC=75% for the software to decide whether objects were visible or not. Only results for the 1% contrast group are shown in this work for the real T images.

Finally, psychometric fits were performed for the calculated P for real T and simulated images (4) [6]:

$$PC = \frac{0.5}{1 + e^{-flog\left(\frac{d'}{\lambda}\right)}} + 0.5 \qquad (4)$$

In this way the smallest object visible, λ, (related to PC=75%) was obtained in each case.

III. Results and Discussion

Real CT images

In Fig. 3 detectability values (d') are shown as a function of object diameter for the 100 mAs and the three contrast groups. It can be seen that as object contrast increases, higher values are obtained for all object sizes.

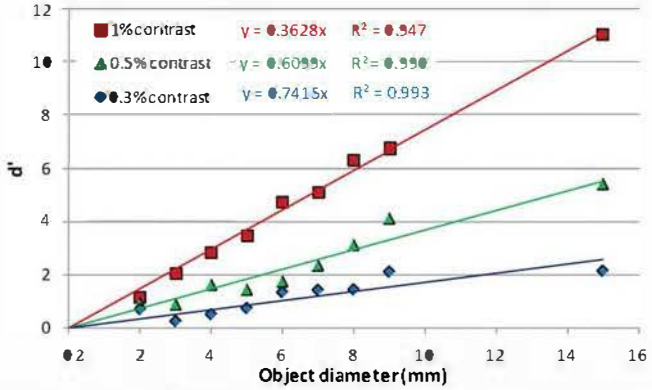


Fig. 3. Detectability values as a function of object diameter for the three contrast groups in the 100 mAs series.

In Fig.4 the influence of the change in mAs in LCD for our model is shown for the 1% contrast group. This index increased linearly with object diameter (R^2 in the range 0.94-0.99). The slopes of the linear fits increased with the selected mAs value, as expected.

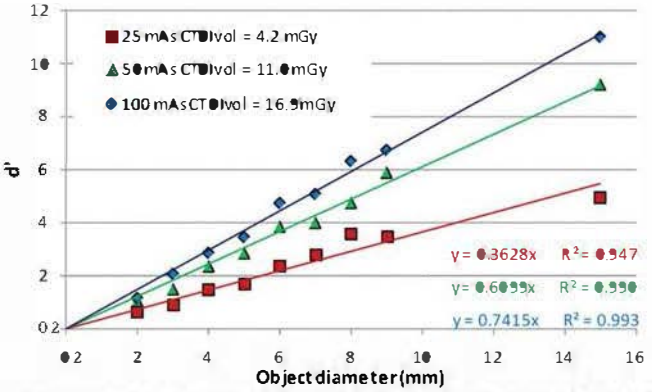


Fig. 4. Detectability values (d') as a function of object diameter for real CT images and different mAs for the 1% contrast group.

The transformation of the d's into P values is shown in Fig. 5 for the 1% contrast group. Higher P values where obtained with increasing objects diameter and dose. The performed psychometric fits gave values for the parameter λ between 1.8 and 2.8 mm. The just visible object size was smaller as dose increased.

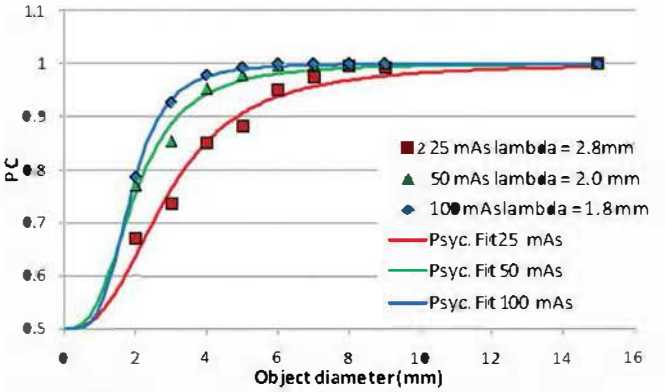


Fig. 5. PC values as a function of object diameter for real CT images together with the psychometric fits.

For the other contrast groups λ values were larger: for example, for the 100 mAs series, λ was 2.6 and 5.2 mm for the 0.5% and 0.3% contrast groups, respectively.

Simulated white noise images

The d' results obtained with our model for the simulated white noise images are depicted in Fig. 6. Detectability values increase with object size and are comparable to those obtained by other authors [3]. It can be seen that the slopes are similar with a relative difference lower than 5%. Thus, our method is validated after reproducing the results obtained by a similar model in detection tasks with white noise.

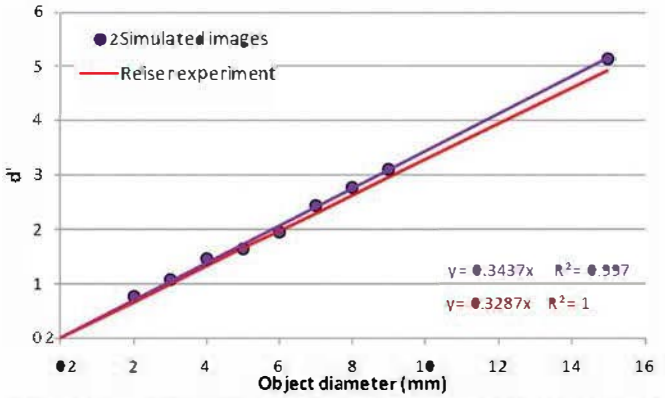


Fig. 6. Detectability values (d') obtained using a set of 1000 simulated images for disk shaped signals on white noise background.

In Fig. 7 the related P values are shown together with the psychometric fit. The visibility threshold under the selected conditions is the 2.6 mm. It is noticeable that the differences between the experimental P values and the psychometric fit are small.

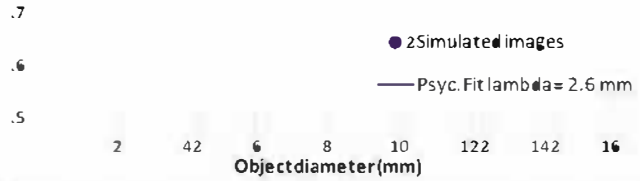


Fig. 7. PC values and psychometric fit for the set of simulated images for disk shaped signals on white noise background.

IV. CONCLUSIONS

The proposed automated method for low contrast detectability using the NPWE observer has proved to be a useful tool for investigating CT images acquired under different dose conditions. Our method was successfully validated when applied to computer-generated simulated images in white noise, being the results similar to those obtained by other authors.

The next step of the research should be the study of the efficiency of the automated model related to human observer performance. An equivalent 2-AFC experiment with human observers considering samples of real CT or simulated images can be performed. Other model observers (hotelling, PWE...) can be implemented in the method as well.

REFERENCES

[1] I. Hernández-Girón, J. Geleijns, A. Calzado and W. J. H. Veldkamp, "Automated assessment of low contrast sensitivity for CT systems using a model observer", Med. Phys., vol. 38, pp. S25-S35, 2011.

[2] H. H. Barrett and K. J. Myers, *Foundations of image science*. Hoboken, USA: Wiley-Blackwell, 2004.

[3] I. Reiser and R. M. Nishikawa, "Identification of simulated microcalcifications in white noise and mammographic backgrounds", Med. Phys., vol. 33, pp. 2905-2911, 2006.

[4] S. Mori, M. Endo, K. Nishizawa, K. Murase, H. Fujiwara and S. Tanada, "Comparison of patient doses in 256-slice CT and 16-slice CT scanners", Br. J. Radiol., vol. 79, pp. 56-61, 2006.

[5] E. Burgess, F. L. Jacobson and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise", Med. Phys., vol. 28, pp. 419-437, 2001.

[6] W. J. H. Veldkamp, M. A. O. Thijssen, N. Karssemeijer, "The value of scatter removal by a grid in full field digital mammography", Med. Phys., vol. 30, pp. 1712-1718, 2003.

4.2. Automated analysis of the influence of acquisition and reconstruction parameters in low contrast detectability in CT phantom images based on a model observer.

[II] I. Hernández-Girón, J. Geleijns, A. Calzado, W. J. H. Veldkamp.

Automated assessment of low contrast sensitivity for CT systems using a model observer.

Med Phys 2011;38:S25-S35 (doi:10.1118/1.3577757)

Abstract

Purpose: Low contrast sensitivity of CT scanners is regularly assessed by subjective scoring of low contrast detectability within phantom CT images. Since in these phantoms low contrast objects are arranged in known fixed patterns, subjective rating of low contrast visibility might be biased. The purpose of this study was to develop and validate a software for automated objective low contrast detectability based on a model observer.

Methods: Images of the low contrast module of the Catphan 600 phantom were used for the evaluation of the software. This module contains two subregions: the supraslice region with three groups of low contrast objects (each consisting of nine circular objects with diameter 2–15 mm and contrast 0.3, 0.5 and 1.0%, respectively) and the subslice region with three groups of circular objects each (diameter 3–9 mm; contrast 1%). The software method offered automated determination of low contrast detectability using a NPWE (non-prewhitening matched filter with an eye filter) model observer for the supraslice region. The model observer correlated templates of the low contrast objects with the acquired images of the Catphan phantom and a discrimination index d' was calculated. This index was transformed into a proportion correct (PC) value. In the two-alternative forced choice (2-AFC) experiments used in this study, a PC≥75% was proposed as a threshold to decide whether objects were visible. As a proof of concept, influence of the kVp (between 80 and 135 kV), mAs (25–200 mAs range) and reconstruction filter (four filters, two *soft* and two *sharp*) on low contrast detectability was investigated. To validate the outcome of the software in a qualitative way, a human observer study was performed.

Results: The expected influence of kV, mAs and reconstruction filter on image quality are consistent with the results of the proposed automated model. Higher values of d' (or PC) were found with increasing mAs or kV values and for the *soft* reconstruction filters. For the highest contrast group (1%), PC values were fairly above 75% for all the object diameters >2 mm, and all the conditions. For the 0.5% contrast group, the same behaviour was observed for object diameters >3 mm for all conditions. For the 0.3% contrast group, PC values were higher than 75% for object diameters >6 mm except for the series acquired at the lowest dose (25 mAs), which gave lower PC values. In the human observer study, similar trends were found.

Conclusions: We have developed an automated method to objectively investigate image quality using the NPWE model in combination with images of the Catphan phantom low contrast module. As a first step, low contrast detectability as a function of both acquisition and reconstruction parameter settings was successfully investigated with the software. In future work, this method could play a role in image reconstruction algorithms evaluation, dose reduction strategies or novel CT technologies, and other model observers may be implemented as well.

Automated assessment of low contrast sensitivity for CT systems using a model observer

I. Hernandez-Giron[a)]
Física Mèdica, Facultat de Medicina i Ciències de la Salut, Universitat Rovira i Virgili, 43201 Reus, Spain and Departamento de Radiología, Universidad Complutense de Madrid, 28040 Madrid, Spain

J. Geleijns
Radiology Department, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands

A. Calzado
Departamento de Radiología, Universidad Complutense de Madrid, 28040 Madrid, Spain

W. J. H. Veldkamp
Radiology Department, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands

(Received 7 September 2010; revised 22 November 2010; accepted for publication 27 December 2010; published 20 July 2011)

Purpose: Low contrast sensitivity of CT scanners is regularly assessed by subjective scoring of low contrast detectability within phantom CT images. Since in these phantoms low contrast objects are arranged in known fixed patterns, subjective rating of low contrast visibility might be biased. The purpose of this study was to develop and validate a software for automated objective low contrast detectability based on a model observer.

Methods: Images of the low contrast module of the Catphan 600 phantom were used for the evaluation of the software. This module contains two subregions: the supraslice region with three groups of low contrast objects (each consisting of nine circular objects with diameter 2–15 mm and contrast 0.3, 0.5, and 1.0%, respectively) and the subslice region with three groups of four circular objects each (diameter 3–9 mm; contrast 1.0%). The software method offered automated determination of low contrast detectability using a NPWE (nonprewhitening matched filter with an eye filter) model observer for the supraslice region. The model observer correlated templates of the low contrast objects with the acquired images of the Catphan phantom and a discrimination index d' was calculated. This index was transformed into a proportion correct (PC) value. In the two-alternative forced choice (2-AFC) experiments used in this study, a PC $\geq$ 75% was proposed as a threshold to decide whether objects were visible. As a proof of concept, influence of kVp (between 80 and 135 kV), mAs (25–200 mAs range) and reconstruction filter (four filters, two soft and two sharp) on low contrast detectability was investigated. To validate the outcome of the software in a qualitative way, a human observer study was performed.

Results: The expected influence of kV, mAs and reconstruction filter on image quality are consistent with the results of the proposed automated model. Higher values for d' (or PC) are found with increasing mAs or kV values and for the soft reconstruction filters. For the highest contrast group (1%), PC values were fairly above 75% for all object diameters >2 mm, for all conditions. For the 0.5% contrast group, the same behavior was observed for object diameters >3 mm for all conditions. For the 0.3% contrast group, PC values were higher than 75% for object diameters >6 mm except for the series acquired at the lowest dose (25 mAs), which gave lower PC values. In the human observer study similar trends were found.

Conclusions: We have developed an automated method to objectively investigate image quality using the NPWE model in combination with images of the Catphan phantom low contrast module. As a first step, low contrast detectability as a function of both acquisition and reconstruction parameter settings was successfully investigated with the software. In future work, this method could play a role in image reconstruction algorithms evaluation, dose reduction strategies or novel CT technologies, and other model observers may be implemented as well. © *2011 American Association of Physicists in Medicine.* [DOI: 10.1118/1.3577757]

Key words: image quality, CT, low contrast, dose

I. INTRODUCTION

Evaluating image quality is essential when investigating new acquisition protocols or new technical developments in CT. A first approach is to look at individual physical properties of the image such as image contrast, resolution, and noise. Together, these aspects play an important role in the detection, classification, and estimation tasks in medical imaging.

The dramatic increase in the number CT scans performed per year in the last decades has raised an important concern about the radiation dose involved in this practice.[1] Its progressive introduction in healthcare services around the world has lead to major benefits in the diagnosis of patients but at the cost of an increase in the doses received by the population. Moreover, new CT applications such as vascular, brain perfusion, or cardiac studies entail higher doses.[2] Besides, the variability in doses and image quality used by different manufacturers of scanner models to reach similar diagnostic tasks has been proved to be wide. Recently, several patients' overdose cases due to an inappropriate CT protocol practice (most in brain-perfusion procedures) have come to the public eye.[3,4] The Working Group of Standardization of CT Nomenclature and Protocols of the AAPM is currently publishing a set of reasonable scan protocols for frequently performed CT examinations with examples for different models and manufacturers. Dose and image quality are intertwined. In some applications, it is not necessary to have best quality possible images to perform an adequate diagnosis task.[3] There is still certain leeway to reduce dose without losing relevant diagnostic image information.[5] Therefore, there is the necessity to develop methods and tools that can lead to fair comparisons and improvements in CT protocols.

Following on physical measurements, receiver operating characteristics (ROC) studies involving human observers are a well known method of evaluating the impact of a particular image manipulation on clinical diagnosis.[6] However, ROC studies may become time consuming and costly because they require a significant number of human observers and a large number of observations. Moreover, the number of possible conditions to be investigated may be large.[7] Frequently, low contrast (LC) sensitivity is assessed in ROC studies using phantom CT images. This subjective rating of image quality might be biased because the spatial pattern distribution of low contrast objects in the phantom is normally known beforehand by the observer. Furthermore, recent studies have shown that a great intra- and interobserver variability exists in ROC tests for CT phantom images.[8]

As an alternative to human observers, computer-model observers can be considered. These are algorithms that attempt to predict human visual performance in noisy images. These models seem very useful in investigating different conditions of the CT imaging procedures, such as the influence of acquisition and reconstruction parameters and object size or shape, in detection tasks.[9,10]

Some attempts have been made to objectively assess low contrast detectability (LCD) in CT. Chao *et al.* proposed a statistical method to characterize CT scanner performance in low contrast detectability from measurements within a uniform water-equivalent phantom.[11] Some authors proposed to measure the contrast-to-noise ratio (CNR) as a figure of merit on phantom images.[12] As low contrast detectability is related to dose, dose efficiency indices have also been proposed to determine the probability to detect a target of defined size and contrast using a reference dose.[13] One of the limitations of using exclusively pixel noise to study low contrast detectability comes from the influence of parameters not related to dose, such as the reconstruction kernel. For this reason, other quality indices combining the effects of spatial resolution, low noise, and slice thickness have been also defined to compare the performance of different scanners.[14–16]

In this work, we propose an objective statistical method with a model observer to investigate the influence of different acquisition and reconstruction CT parameters on low contrast detectability and dose. As a starting point, we chose the nonprewhitening matched filter with eye filter (NPWE) model.[17] This method could be helpful in investigating image reconstruction algorithms, dose reduction strategies, and novel CT technologies. In a next stage, it could enable fair comparisons between scanners of different vendors especially when image quality as a function of dose could be established.

II. MATERIALS AND METHODS

The Catphan 600 Phantom (Phantom Laboratories, New York) is dedicated to perform quality control on CT scanners and is made of different modules. The CTP515 module consists of several groups of cylindrical rods of various diameters and three contrast levels to measure low contrast performance as shown in Fig. 1. The module contains two subregions: the supraslice region (outer circle) shows three contrast patterns of nine low contrast objects each (diameter 2–15 mm; nominal contrast 0.3, 0.5, and 1.0%; 40 mm length) and the subslice region (inner circle) with three patterns of four objects each (size 3–9 mm; nominal contrast 1.0%; 3, 5, and 7 mm length). Only the supraslice region was chosen for this study.

Images of this module were taken with the phantom aligned with the axis of rotation of the scanner (z-axis) and used for the evaluation of the software. The acquisition was

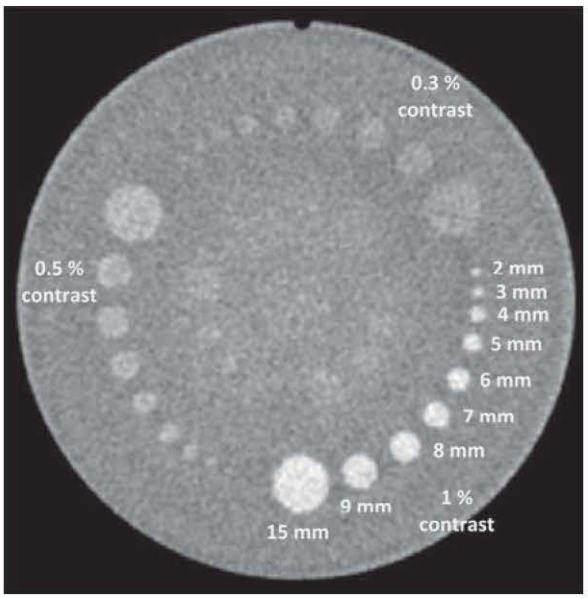


FIG. 1. A constructed 38 mm CT slice of the CTP515 module. Each contrast group and the low contrast object diameters have been tagged in the figure for the supraslice region.

performed with a 16-detector row CT scanner (Aquilion 16, Toshiba, Japan), selecting 16 × 0.5 mm as beam collimation, scan field of view of 400 mm, helical acquisition (0.94 as pitch factor), and different combinations of tube voltage (80–135 kV range) and tube charge per rotation (25–200 mAs interval). Image reconstruction was performed with 0.5 mm as slice thickness and 0.5 mm reconstruction interval using four different reconstruction filters: FC12 (*soft* convolution kernel; body), FC50 (*soft* convolution kernel; lung), FC53 (*sharp* convolution kernel; lung), and FC81 (*sharp* convolution kernel; bone). For those series with varying kV or mAs, as the objective is to detect low contrast objects, a *soft* reconstruction filter (FC12) was chosen, whereas for comparing different reconstruction filters, tube voltage and tube charge were kept constant (120 kV and 100 mAs). Table I gives an overview of the acquisition and reconstruction parameters used. Each image series consisted of 76 images to cover the total 40 mm length of the low contrast module (two images at both boundaries were discarded).

Image series in Table I were used to test the software. Detectability results were obtained as a function of tube charge per rotation (mAs), tube voltage (kV), and reconstruction filter, respectively, taking image series where all acquisition and reconstruction parameters were constant but one, respectively. Then, analysis for each series and contrast groups were performed.

The first step in the image data processing performed by the software was the phantom detection in the images using a fixed threshold value. The detection of the low contrast module was based on thresholding and verifying the absence of high contrast objects. Using a distance transform in combination with the known size of the circular module and the ring around it, module and ring were separated by the program. Median pixel values of the module and the ring were determined in each slice. The median values of the module should fall between the fixed threshold values (30 and 70 HU, respectively) in a selected slice. Furthermore the module median value should be at least 30 HU higher than the ring value and the module area in the selected slices should not contain any high contrast objects (those related to HU > 500). Information concerning the exact position of the module within the phantom was not used: in this way, the module could be detected whether the entire phantom, or only a part of it, was scanned. The software is implemented in MATLAB.

The next step was the creation of a template mask of the distribution of the low contrast patterns in the phantom. An initial mask image was constructed (Fig. 2), containing templates to match all low contrast objects (with respect to size, shape, and position) in the actual images. Position of the templates within the mask was derived from the manufacturer's specifications. The 2-D mask was scaled and positioned automatically to optimize its application on the acquired phantom image series shown in Table I. It was applied to each slice individually. Polar coordinates (angle, θ, radius, ρ) values were predefined to locate the LC disks positions with respect to the module's center. LC object radius r was defined for each disk relative to the radius of the phantom. LC angle θ was defined for each object relative to the high contrast reference object present in the outer rim of the Catphan phantom. Using a distance transform, the actual phantom's radius value is determined in the image and expressed as a number of pixels. The relative radius r of the LC object is scaled with respect to this value. The special high contrast reference object present in the outer rim of the Catphan phantom was automatically detected and used to estimate the orientation of the phantom in the images. An angle of rotation was successively determined and taken into account concerning the predefined LC angles (θ). A thick constructed slice [38 mm thickness; to obtain a high signal-to-noise ratio (SNR)] of the low contrast module was used for visual verification of the mask's final matching accuracy (Fig. 2).

Low contrast object templates were blurred to model the modulation transfer function (MTF) of the CT system.[18] The MTF was modeled assuming a full width half maximum of the point spread function (PSF) of 0.68 mm, based on previous work for a CT scanner with similar MTF as the one used in this study.[19] Regarding the model observer implementation in the software, the NPWE was applied. Its strategy consists of correlating the image with the shape of the expected signal profile filtered by the visual-response function (E).

TABLE I. Overview of the acquisition and reconstruction parameters.

Series	Reconstruction filter	Tube voltage (kV)	Tube charge per rotation (mAs)
1–4	FC12	80 100 120 135	100
5–8	FC12	120	25 50 100 200
9	FC50	120	100
10	FC53	120	100
11	FC81	120	100

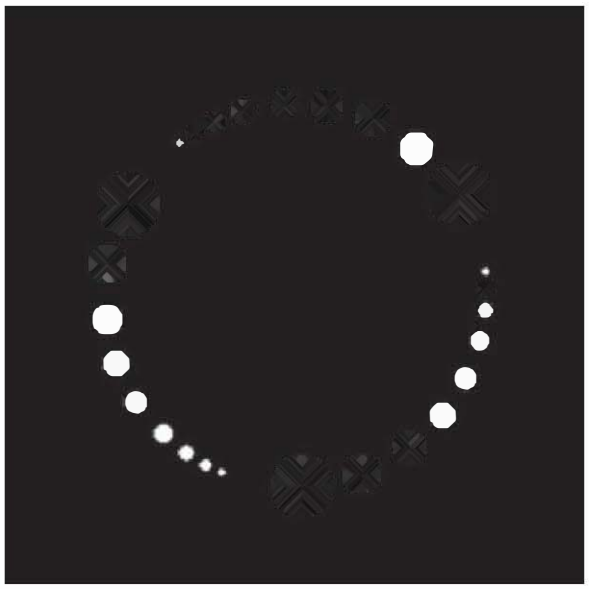


FIG. 2. Mask consisting of templates with respect to each low contrast object in the Catphan module.

In a two-alternative forced choice (2-AFC) detection task, the model reaches a decision by comparing test statistics T_1 (test statistics related to signal) and T_2 (test statistics related to background). These values are obtained by cross-correlation between the expected signal (template) with a background region and with the appropriate known signal region in the image.[18,20]

The process of location of the disks (signal) was based on their relative position with respect to the module's center using polar coordinates, as has been already stated. The images without object for each series were obtained from an area located at the same polar angle as the smallest disk of the series but at a radius 1.2 times larger. This ensured that sample area did not contain any signal and that background samples were taken from a distance to the center that was similar to the LC objects.

This procedure is performed for all low contrast objects in a large number of images. Before correlation, both the image regions and the templates are filtered by the human visual-response function (E) to account for the frequency response of the human eye.

The human visual-response function used in the model was radially symmetric and defined by $E(s) = se^{-bs}$, being s the spatial frequency and with b chosen such that $E(s)$ peaked at four cycles per degree.[17,18] In the experiments, a fixed viewing distance of 500 mm from the monitor was assumed.[21]

From the distribution of test statistics, a discrimination index d' (Refs. 9 and 18) was computed as shown in Eq. (1), where $\langle\cdot\rangle$ is the mean and $\sigma(\,)$ is the standard deviation of the respective distributions

$$d' = \frac{\langle T\rangle_1 - \langle T\rangle_2}{\sqrt{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2}}. \tag{1}$$

This index can be used as a measure of detection performance. The discrimination index can be determined as a function of the low contrast signal energy (SE), (i.e., the squared expected signal value integrated over all pixels in the observer template) as described in the literature[18] and can also be related to object size (diameter) as was done in this work.

In this study, the software derived the distributions T_1 and T_2 by performing cross-correlations in 76 consecutive slices of 0.5 mm slice thickness (for each series in Table I). The discrimination index d' was expressed as a function of object diameter for all the image series acquired.

When considering normally distributed test statistics, d' can be related to the area under the ROC curve (AUC). This AUC is equal to the proportion correct (PC) in a 2-AFC experiment. Thus, PC values can be determined according to[18]

$$\mathrm{PC} = 0.5 + 0.5 erf\left(\frac{d'}{2}\right), \tag{2}$$

where $erf(x)$ is the Gaussian error function given by

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-x^2} dx. \tag{3}$$

In this way, d' values were transformed in AUC and consequently in PC values for all the conditions proposed in Table I. As in a 2-AFC experiment, a default PC = 50% value can be obtained just by chance, we assumed an additional detectability threshold in PC = 75%. In this way, when PC 75% in the analysis, objects were considered visible.

Finally, we used a model-based interpolation scheme to fit a curve through PC values as a function of object diameter. The applied scheme is based on a method described by Karssemeijer and Thijssen who related PC to object contrast.[22] In our work, the probability of detecting an object (PC) as a function of its diameter is described by means of a psychometric curve with probability range from 0.5 (chance) to 1.00 and so, for this 2-AFC experiment an expression as shown in Eq. (4) was used

$$\mathrm{PC} = \frac{0.5}{1 + e^{-f\log\left(\frac{d}{\lambda}\right)}} + 0.5, \tag{4}$$

where d represents the low contrast object diameter and f and λ are the parameters that are fitted to the data. The value of f determines the steepness of the psychometric curve. Using f and λ values obtained with the psychometric fittings, we determined the smallest object diameter that would achieve the visibility criterion proposed (PC 75%) and that, just by working it out on Eq. (4), is λ itself. A least-squares procedure was applied[23] independently for fitting the psychometric curves to the data, for each image series from Table I. Thus, a reliable estimate of λ is obtained.

To compare PC curves (PC against object diameter) obtained at different mAs with constant kV value, they were normalized at a reference dose using the equivalent diameter (d_{ref}) concept proposed by Ishida *et al.*[13] In our case, as for a particular phantom, the volume computed tomography dose index $\mathrm{CTDI_{vol}}$ (Ref. 24) is proportional to the tube charge per rotation, we used mAs values for each image series to obtain the normalized object diameters, d_{ref}, with

$$d_{\mathrm{ref}} = \sqrt{\frac{\mathrm{mAs}}{\mathrm{mAs_{ref}}}} \cdot d, \tag{5}$$

where d represents object diameter for each contrast group in the phantom (2–15 mm), mAs is the tube charge per rotation value for the image series which is being normalized, and $\mathrm{mAs_{ref}}$ represents the reference image series value. This expression gives an estimate of the normalized object diameters (d_{ref}) for each mAs series that would score the same PC values as if the series were acquired at the reference dose.[25,26]

In this way, all PC profiles were transformed to a common reference dose by rescaling their abscissas (object diameter). This normalization was applied to each contrast group (1, 0.5, and 0.3% contrast) separately. The series 3 from Table I (120 kV, 100 mAs, FC12 filter, $\mathrm{CTDI_{vol}} = 16.9$ mGy) was selected as the reference. For each mAs series, a

new PC curve was obtained with d'_{ref} values and a psychometric fit was performed for all data applying Eq. (4). The values of parameter f obtained in this way for each contrast group were used to recalculate the psychometric fits for each mAs series individually. Finally, to check the goodness of this approximation, a comparison between the obtained λ values (λ_{Norm}) and those of the individual psychometric fits for each mAs series (λ) was performed.

To study the influence of kV on the intrinsic contrast on the Catphan images, it was measured using the 38 mm thickness images created for each image series. Identical size ROIs were taken on the largest circle (15 mm) for each contrast series, and six more were distributed in the image to measure the background. In this way, a mean pixel value was obtained for each ROI and an average value for the background as well. Finally, contrast was found using the measured HU differences between background and the corresponding ROI for each contrast series. The CNR with respect to the background was also obtained.

Finally, to validate the trends shown by the software as a function of kV, mAs, and reconstruction filter, a human observer study was carried out. Six observers scored the images, two of them being radiologists with several years of experience in medical imaging diagnosis (experts) and four observers with different levels of experience in image quality assessment (non-experts).[25] Image visualization was performed in calibrated monitors under appropriate viewing conditions according to international recommendations.[27]

Each observer scored 220 images of the Catphan low-contrast module, 20 images per each series shown in Table I. Care was taken to select equivalent images (i.e., at the same locations) for all series. The scoring was carried out in two sessions (separated by a minimum of two weeks). Any image identification was removed and image order was randomly arranged, so observers were unaware of which image they were scoring. Only the visibility of the 1% contrast series objects was assessed.

Observers were asked to record how many objects they were able to see for each image. Appropriate fixed values of the window level and width were selected for each image series from Table I. The total number of Catphan's scored images was 1320 being 440 scored by experts and 880 by non-experts.

An analysis of the intraobserver consistency in both sessions was performed using the Wilcoxon signed rank test[28] for matched-pair samples (p 0.05) by comparing the scores for each series from Table I separately. Thus, for each observer, we were able to select those scorings showing no significant differences between both sessions for each image series considered.[29] Using these depurated scoring results, detectability profiles (PC curves as a function of object size) were obtained for each observer and image series. PC values were determined as the ratio between the number of times that the observer detected one object and the total number of images. For example, if the 5 mm object was scored in 15 out of the 20 evaluated images which form the image series, a PC value of 75% was assigned.[25] Finally, mean PC values were obtained and assigned to an average nonexpert and an average expert observer, respectively, for each image series shown in Table I.

As this observer study is not a 2-AFC experiment we will not perform a quantitative comparison between PC values obtained by the observers and the software. Instead, we will perform a qualitative comparison between the trends of PC as a function of object size for varying kV, mAs, and reconstruction filter obtained with the LC detectability software and the average expert observer.

III. RESULTS

In Fig. 3, graphs for the 1% contrast group LCD software results are presented for all the acquisition conditions from Table I. The left column of Fig. 3 represents the influence on d' of kV, mAs (both for the FC12 filter) and reconstruction filter [Figs. 3(A)–3(C), respectively] as a function of object diameter. The right column shows the influence of the same parameters on PC values [Figs. 3(D)–3(F)]. In all cases, d' and PC values increase with object size, as expected. As a dose reference, $CTDI_{vol}$ values have been included next to the corresponding acquisition series in the graphs.

Regarding kV influence, d' increases linearly with object size ($R^2 > 0.97$) for each series and for higher kV values. In all cases PC values were above the visibility criterion PC 75%, except for the smallest circle, which is slightly lower for 80 and 120 kV (PC = 70 and 73%, respectively).

Analyzing the influence of tube charge per rotation, d' is higher when high mAs values are selected and a linear dependency between d' and object diameter is observed ($R^2 > 0.93$). All low contrast objects were visible in the 200 mAs series (PC > 90%). For the 100 and 50 mAs series, the 2 mm diameter circle is not visible (PC = 73%). When the lowest mAs was selected (25 mAs), the 2 and 3 mm objects fail the criterion (PC = 59 and 70%, respectively).

Concerning the reconstruction filter study, d' increased linearly with object diameter ($R^2 > 0.96$) as expected. The *soft* filters (FC12 and FC50) gave higher d' and PC results, showing the *soft* body filter slightly better results. PC values were higher than the 75% threshold (at least 84%) except for the smallest 2 mm diameter objects. In the best case considered, this object size was related to PC = 73% (FC12, *soft* body filter).

Since the behavior of d' and the observed trends in the graphs when varying kV, mAs or reconstruction filter were qualitatively similar for all contrast groups, those figures were not included in the paper. A summary of all contrast groups results appears in Table II. The values given as representative for each series, called D_1 and PC_1, correspond to the largest object diameter which in each case does not accomplish the PC 75% visibility criterion and its corresponding PC value.

Psychometric fits based on Eq. (4) were performed for the three contrast groups PC results as a function of object size for all acquisition and reconstruction conditions considered. As a result of this analysis, λ values ($\lambda_{Indiv.\ Fit}$), [the smallest

object diameter visible according to our threshold (PC 75%)], are shown in Table III.

In Figs. 4–6, λ is plotted against kV, mAs, and reconstruction filter, respectively, for the three contrast groups (1, 0.5, and 0.3%) of the low contrast module in the Catphan phantom.

Finally, d_{ref} values used in the normalization of PC curves to the reference $CTDI_{vol}$ (16.9 mGy) were obtained for all contrast groups and different mAs settings. Figure 7 shows the fitting curves obtained applying Eq. (4) for the 1% contrast group and different mAs values. Figure 8 shows the same PC curves normalized to the reference dose value together with the global psychometric fit for all data. Table IV gives an overview of the values obtained for the smallest visible object diameter applying the normalization (λ_{Norm}) and without it ($\lambda_{Indiv.Fit}$) for all contrast groups. The relative differences between both values are also shown. In Table V, the results for the measured intrinsic contrast and the CNR appear. Measured contrast value was maximum for 100 kV in this CT scanner for all contrast series being the effect more evident for the 0.5 and 0.3% contrast series.

Regarding the human observer study, from the 66 pairs of scoring series considered, 12 of them, corresponding to reconstruction *sharp* filters (FC53 and FC81), were not analyzed because human observers were not able to score them. In the statistical analysis of the remaining 54 pairs of scoring series, 7 of them, showing statistically significant differences ($p < 0.05$), were discarded. All the observers had one scoring series discarded, but two of them, who failed in two series. The comparison between the average expert and nonexpert observer image scoring results as a function of kV, mAs, and reconstruction filter is shown in Figs. 9(A)–9(C), respectively.

Figures 10(A)–10(F) illustrate the qualitative comparison between expert human observer and the software performance. In them, PC curves are plotted as a function of the acquisition or reconstruction parameter considered (kV, mAs, and reconstruction filter) for each object diameter. To validate the software, only the experts scoring results were used. We chose this option because trends showed by experts and non-experts were alike and the former showed lower intraobserver variation between scoring sessions. Although observers only scored the 1% contrast series, graphs for the 0.5% contrast series results are also included for the software model observer. As the 0.3% contrast series PC curves trends were very similar to the 0.5% contrast, they are omitted. Note that for these graphs in the case of human observers PC values run from 0 to 1 and in the case of the software from 0.5 to 1.

IV. DISCUSSION AND CONCLUSIONS

A software for automated assessment of low contrast detectability in CT Catphan phantom images, based on a model observer, has been developed and validated for this study. The software, and the implemented model observer

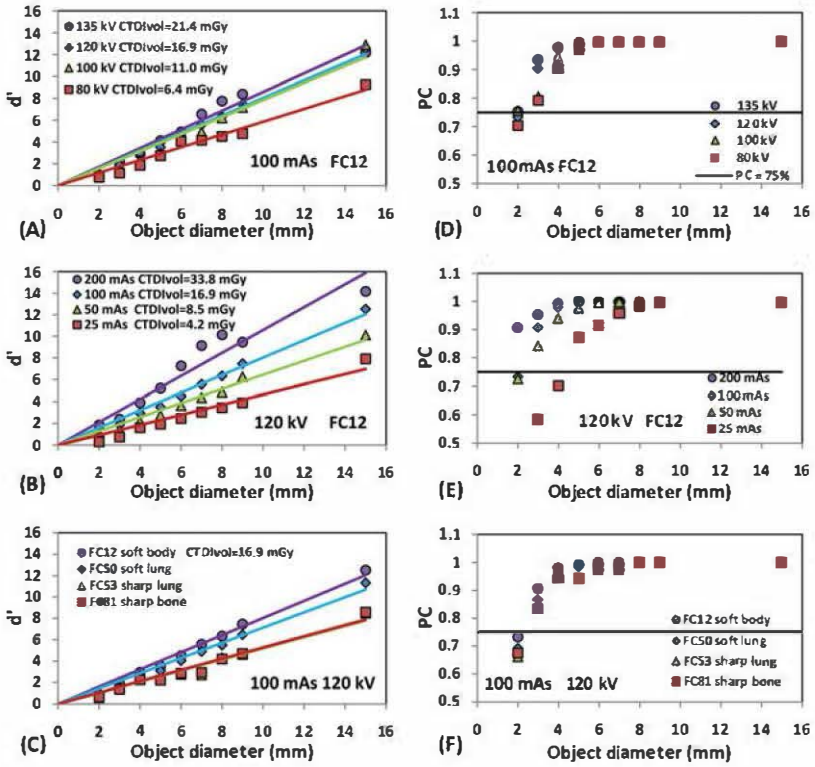


FIG. 3. Detectability index d' and proportion correct PC (left and right column, respectively) as a function of low contrast circle diameters of the Catphan phantom for the 1% contrast group for different kV (A, D), mAs (B, E), and reconstruction filters (C, F). The parameters kept constant in each case appear as a legend on the figures to identify them as image series in Table I. The black line on PC graphs (D, E, F) represents the visibility threshold criterion applied (PC = 75%).

TABLE II. Values of D_1 (largest diameter object) and its corresponding PC value (PC_1) which fail the PC 75% criteria to be considered visible, for the three contrast groups and all image series.

			(D_1 (mm), PC_1 (%))		
		$CTDI_{vol}$ (mGy)	1% contrast	0.5% contrast	0.3% contrast
kV	80	6.4	(2 mm, 70%)	(3 mm, 66%)	(6 mm, 69%)
	100	11.0	(-, -)	(3 mm, 71%)	(3 mm, 64%)
	120	16.9	(2 mm, 73%)	(3 mm, 71%)	(4 mm, 69%)
	135	21.4	(-, -)	(3 mm, 71%)	(4 mm, 68%)
mAs	25	4.2	(3 mm, 70%)	(5 mm, 71%)	(9 mm, 71%)
	50	8.5	(2mm, 72%)	(4 mm, 73%)	(5 mm, 74%)
	100	16.9	(2 mm, 73%)	(3 mm, 71%)	(4 mm, 69%)
	200	33.8	(-, -)	(2 mm, 59%)	(3 mm, 72%)
Rec. Filter	FC12	16.9	(2 mm, 73%)	(3 mm, 71%)	(4 mm, 69%)
	FC50	16.9	(2mm, 69%)	(3 mm, 66%)	(4 mm, 61%)
	FC53	16.9	(2 mm, 66%)	(3 mm, 62%)	(5 mm, 67%)
	FC81	16.9	(2 mm, 68%)	(3 mm, 63%)	(5 mm, 68%)

*(-,-) means that all objects were visible (PC 75%). $CTDI_{vol}$ values have been included for each series.

(NPWE), are appropriate for objectively assessment of low contrast detection in CT images. Human observers showed similar performance compared to the software model observer.

In the selection of the NPWE as a model observer, some considerations were made. The ideal observer (Rose model) can remove any correlations that are present in the noise (prewhiten) so that it can be treated as white noise.

Human observers show an imperfect prewhitening noise ability in detection tasks and perform cross-correlations with the expected displayed signal.[7] Their results have fallen between the two extremes of complete prewhitening and nonprewhitening.[17,30] The best predictions of human performance are made with partially prewhitening models, which include arrays of spatial frequency filters called channels (hotelling models).[31]

TABLE III. Summary of the results of the individual psychometric fittings for all contrast groups and acquisition conditions considered. Parameter λ represents the smallest object diameter which reached PC 75% and so, considered visible. The error associated with this parameter is also shown as a (%).

		λ (mm)		
		1% contrast	0.5% contrast	0.3% contrast
kV	80	2.4 ± 4%	4.1 ± 3.5%	6.1 ± 11%
	100	2.1 ± 7%	2.7 ± 7%	4.0 ± 4%
	120	2.1 ± 1%	3.0 ± 4%	3.9 ± 7%
	135	1.8 ± 0.5%	3.3 ± 2%	4.4 ± 6%
mAs	25	3.2 ± 2%	5.5 ± 7%	9.9 ± 10%
	50	2.2 ± 3%	3.7 ± 6%	5.8 ± 4%
	100	2.1 ± 1%	3.0 ± 4%	3.9 ± 7%
	200	1.2 ± 8%	2.9 ± 2%	2.8 ± 10%
Rec.Filter	FC12	2.1 ± 1%	3.0 ± 4%	3.9 ± 7%
	FC50	2.3 ± 1%	3.4 ± 3%	4.8 ± 6.5%
	FC53	2.4 ± 2%	3.8 ± 2%	5.7 ± 8%
	FC81	2.4 ± 2%	4.0 ± 3%	6.4 ± 8%

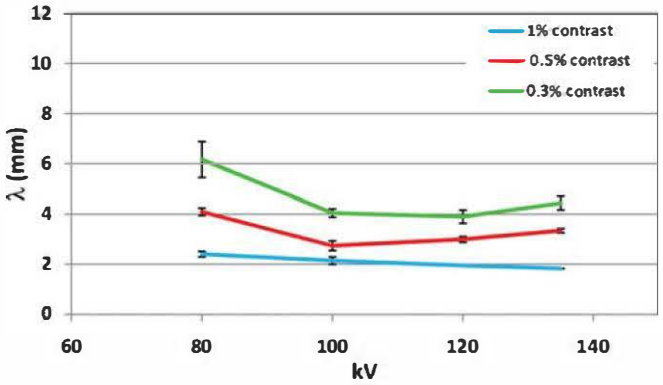


FIG. 4. Smallest object diameters visible (with a PC = 75%), λ, obtained with the psychometric fitting for all contrast groups for different kV values. Lines are a mere data connector in the graph.

The NPWE is a suboptimal observer that is unable to undo any correlations in the data. It uses a template matched to the expected difference image to form a test statistic, regardless the sources of variability in the data. This mathematical model is a good predictor of human performance in both filtered anticorrelated and in colored noise images (e.g., CT images)[9,30] on uniform backgrounds.

The results obtained using the blurred template of the low contrast section of the Catphan phantom were compared with results obtained when blurring of templates was omitted. The analysis of the influence of kV and tube charge (with all other acquisition parameters being constant) gave small relative differences in PC values between both templates (around 5% in the worst case, considering all contrast groups). With the blurred template, PC values were slightly higher in all cases. The higher discrepancies were observed for the smallest object sizes and lowest contrast group (0.3% contrast). In the reconstruction filter influence study, higher relative differences were found. The maximum discrepancies were obtained for the *sharp* filters (FC53 and FC81, 14% and 11%, respectively) and for the 0.3% contrast group. The lower differences were obtained with FC12 (*soft* body) followed by FC50 (*soft* lung), with relative differences below 1 and 4%, respectively. Thus, the consideration of the CT

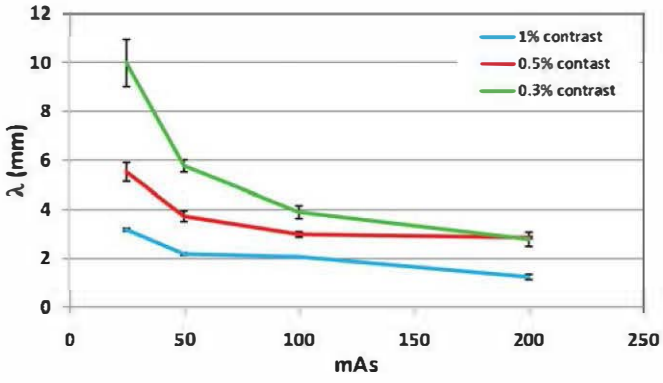


FIG. 5. Smallest object diameters visible (with a PC = 75%), λ, obtained with the psychometric fitting for all contrast groups for different mAs values. Lines are a mere data connector in the graph.

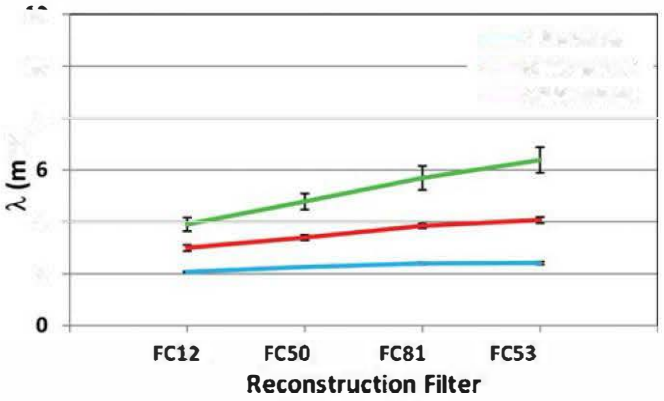


FIG. 6. Smallest object diameters visible (with a PC = 75%), λ, obtained with the psychometric fitting for all contrast groups for different reconstruction filters. Lines are a mere data connector in the graph.

system MTF is an important fact when studying reconstruction filter's influence in low contrast image quality.

The study of kV and tube charge influence on low contrast patterns detectability showed that higher PC values were scored with increasing kV or mAs values for all contrast groups, being the differences in the latter much larger. This was expected to some extent, as image quality increases when higher detector doses are involved and thereby noise is decreasing. It was found that d' was linearly dependent with object diameter for all contrast groups. PC values reached the highest values for the highest contrast groups considered (1% contrast), progressively decreasing for the lower contrast groups (0.5 and 0.3% contrast).

The visibility criterion proposed (PC 75%) granted that objects are visible not just by chance, in our 2-AFC experiment. This threshold was reached in all cases for the 1% contrast group except for the smallest circle (2 mm), which was not visible for the 80, 120 kV and 50 and 25 mAs series. In the first three cases, PC values were very close to the proposed threshold (values above PC = 70%). For the last one, which corresponded with the lowest dose considered, even the 3 mm circle was not visible due to high image noise. However, this low tube charge is not usually selected in ordinary practice.

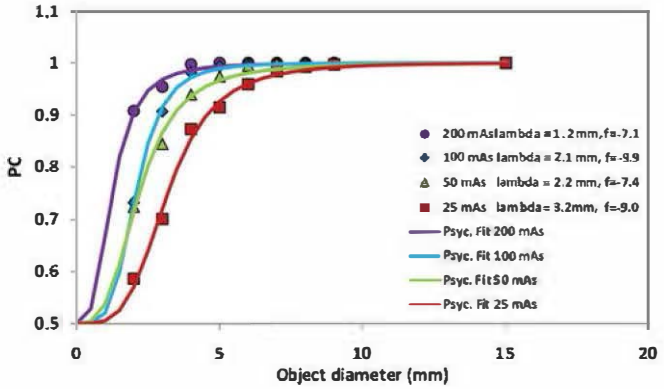


FIG. 7. Psychometric fitting functions of PC as a function of object size for different mAs and 1% contrast group. Fitting parameters, λ and f, have been included in the legend for each image series.

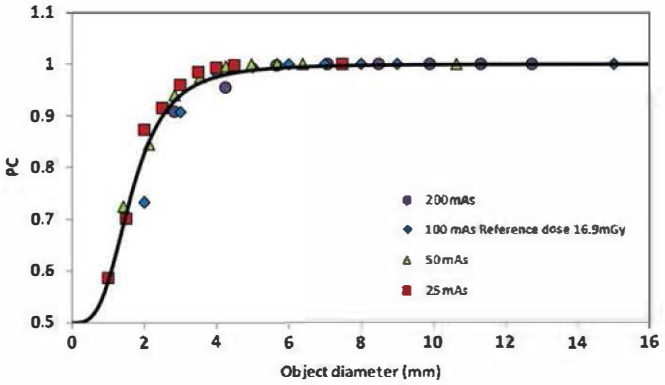


FIG. 8. PC curves normalized to the reference dose (16.9 mGy) for different tube charge per rotation values. The black line represents the global psychometric fit for all data.

As expected for the two lower contrast groups, PC values were lower compared to the 1% contrast group. For the 0.3% contrast group, the 2 and 3 mm objects were not visible in all cases according to our PC threshold value.

The FC12 (*soft* body) reconstruction filter gave the highest PC values compared to the other reconstruction filters. The FC50 filter performance was very similar. These filters (*soft*) suppress the high frequencies (noise in the images) that strengthen the visibility of low contrast objects. Considering the 1% contrast group, all objects were visible with all filters except for the smallest one (2 mm). For 0.5% contrast, the smallest visible diameter was 4 mm and for the 0.3% groups, 5–6 mm for *soft* and *sharp* filters, respectively.

The psychometric curve model proposed for this 2-AFC experiment was applied to PC values as a function of pattern diameter for all contrast groups and acquisition conditions proposed. One of the fitting parameters, λ, tallies with the diameter of the object, which exactly meets the visibility criteria (PC 75%). For the 1% contrast group, when increasing mAs, an improvement in visibility of smaller objects was clearly shown. This trend was slightly showed for increasing kV (PC curves were much closer in this case). In both cases, with increasing dose, smaller objects become visible, as expected. For the lower contrast groups (0.5 and 0.3%), the

TABLE IV. Lambda values (λ_{Norm}) obtained performing the psychometric fits taking as f the value obtained with the psychometric fit of the normalized data to the reference dose ($CTDI_{vol\ ref} = 16.9$ mGy) for all contrast groups and different mAs values. The relative differences ($\varepsilon_{relative}$) between lambda values obtained for the individual fits ($\lambda_{Indiv.Fit}$) for each series (with lambda and f unbound) and this method, are also shown.

	mAs	25	50	100	200
1% contrast	λ_{Norm} (mm)	3.1 ± 3%	2.2 ± 2%	2.0 ± 0.3%	1.3 ± 3%
	$\lambda_{Indiv\ Fit}$ (mm)	3.2 ± 2%	2.2 ± 3%	2.1 ± 1%	1.2 ± 8%
	$\varepsilon_{relative}$ (%)	1.9%	0.6%	4.3%	3.7%
0.5% contrast	λ_{Norm} (mm)	5.7 ± 7%	3.7 ± 5%	2.8 ± 5%	2.6 ± 8%
	$\lambda_{Indiv.Fit}$ (mm)	5.5 ± 7%	3.7 ± 6%	3.0 ± 4%	2.9 ± 2%
	$\varepsilon_{relative}$ (%)	2.5%	1.2%	5.6%	9.7%
0.3% contrast	λ_{Norm} (mm)	9.3 ± 7.5%	5.7 ± 7%	3.9 ± 6%	3.0 ± 8%
	$\lambda_{Indiv.Fit}$ (mm)	9.9 ± 10%	5.8 ± 4%	3.9 ± 7%	2.8 ± 10%
	$\varepsilon_{relative}$ (%)	6.5%	0.4%	1.3%	7.9%

TABLE V. Measured and nominal contrast values (%) and contrast-to-noise ratio (CNR) for the low contrast module of the Catphan phantom.

	80 kV		100 kV		120 kV		135 kV	
Nominal contrast	Measured contrast	CNR	Measured contrast	CNR	Measured contrast	CNR	Measured contrast	CNR
1%	1.15%	0.44	1.20%	0.64	1.11%	0.74	1.08%	0.95
0.5%	0.60%	0.24	0.66%	0.34	0.56%	0.37	0.54%	0.41
0.3%	0.31%	0.13	0.34%	0.19	0.28%	0.20	0.27%	0.20

same trend is apparent for increasing mAs though just visible objects are larger.

With increasing kV, both noise and contrast are decreasing (Table V). Thus, CNR will not increase in the same way as with changing mAs. For the 0.3 and 0.5% contrast series, CNR does not increase substantially as a function of kV for 100 kV and higher. PC curves are closer in Fig. 3(D) (kV variation) than in Fig. 3(E) (mAs variation). As an example of this, the 25–50 mAs curves are more separated than the 80–100 kV curves, being dose practically doubled in both cases. Thus, the statistical variation will become more apparent with kV, so we cannot exclude that kV results are biased by statistical variation. As a result, when the variation of kV is considered, λ does not change substantially between 100 and 135 kV (Table III, Fig. 4). Regarding the mAs analysis, the steepness of the curves will dominate statistical variations and a clear decreasing trend for λ as a function on mAs appears (Fig. 5).

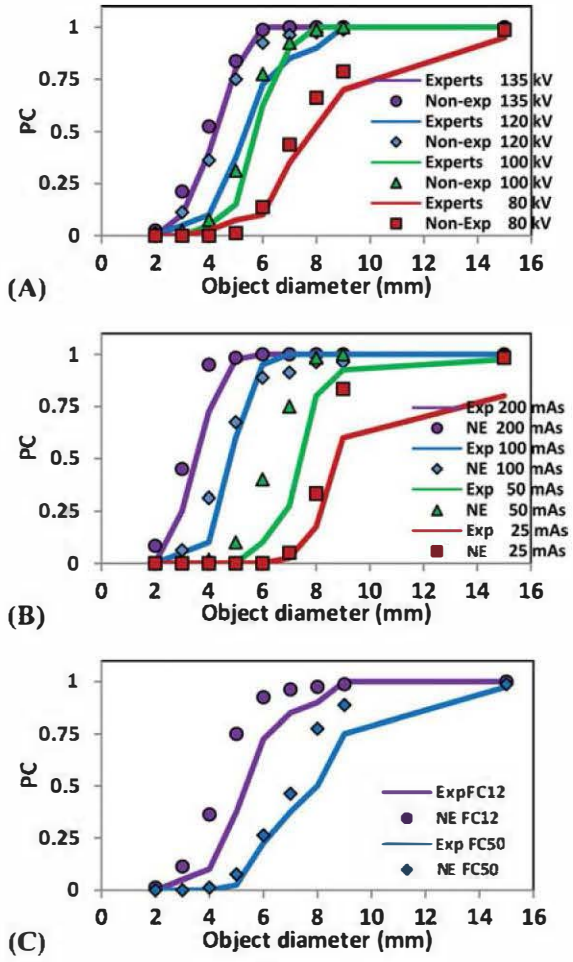


FIG. 9. PC curves as a function of object size for the average expert (lines, *Exp*) and nonexpert (dots, *NE*) observers for varying kV (A), mAs (B), and reconstruction filter (FC12 and FC50, *soft* body and lung, respectively, (C). Observers were not able to score images related to the *sharp* filters (FC81, FC53). Note that PC values run from 0 to 1.

Regarding reconstruction filter influence, it is shown to be critical for the lower contrast small objects. Lambda (λ) varies in a range of 0.3, 1, and 2.5 mm, respectively, for the 1, 0.5, and 0.3% contrast groups, depending on the chosen reconstruction filter. Smaller objects were visible when selecting the *soft* body filter FC12 in all cases.

The human observer study results showed a considerable interobserver variability. The statistical tests performed allowed us to discard those image series scorings that showed high intraobserver variation. PC values were slightly higher in the non-experts group and they scored smaller objects as the least visible one.

The PC trends of the expert and nonexpert average observers showed higher values for increasing kV, mAs (Fig. 9), and object diameter. In the kV analysis, non-experts obtained similar PC values for the 120 and 135 kV series, and experts obtained slightly better values for the latter. Observers PC curves are much closer for kV than for mAs and showed higher PC values for the FC12 (*soft* body) filter. All these trends appeared for the software PC curves in a similar fashion (Fig. 3).

The software proved to be more sensitive than the expert average observer. Experts PC values are lower, especially for low dose and small objects, as expected (Fig. 10). It is important to note that as slice thickness is so narrow in this experiment (0.5 mm), SNR was small in some images, which made the human observer's task, difficult especially for the lowest mAs series. In Fig. 10(A), the average expert human observer scored the 1% contrast objects with diameters in the range (7–9 mm) with slightly higher PC values for 100 kV than for 120 kV. This can be related to the maximum intrinsic contrast value measured for this kV (Table V).

The proposed approximation method of normalizing PC curves to a reference dose gives acceptable results for λ. Relative differences between the values obtained with both methods were in the worst case below 10% considering the three contrast series results. This approximation would be a good approach to save some calculations (individual psychometric fits can be optionally not made for each mAs setting) when studying the influence of tube charge on low contrast detectability keeping the other acquisition parameters constant. This normalization might allow to study the effect of

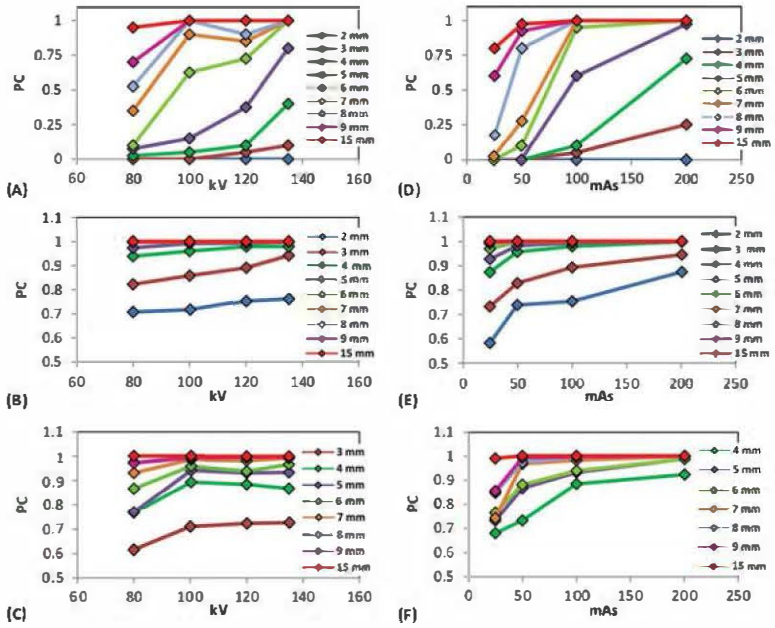


FIG. 10. PC curves as a function of kV (left column) and mAs (right column) for different object sizes for the average expert observer (first row, (A) and (D), 1% contrast series) and the LCD software (second and third row, 1% and 0.5% contrast series, respectively). Note that PC values run from 0 to 1 for human observer (first row) and from 0.5 to 1 for the software (second and third row).

acquisition and reconstruction parameters other than mAs on the low-contrast visibility as dependence on the latter is suppressed.

In conclusion, we have validated a method for investigating low contrast detectability in CT images. The implemented model observer (NPWE) seems appropriate for objectively investigating low contrast detection in CT images. As a limitation in this study, the performance of the model observer used (and possibly of other model observers') may differ from observers. Nonetheless, the trends showed by both, the LCD software and expert observer were similar for scanning at different mAs and kV and for the *soft* reconstruction filters as well. As a next step in this research, other model observers may be implemented while relating their performance may differ from human observers. In the current version of the software we have simulated an observer that is not trained specifically with respect to different reconstruction filters. In this way, our model gives a first objective judgment of image quality. For future work, it will be taken into account the frequency response of the different reconstruction filters as well.

The proposed method can be considered a reasonable aid for investigating trends regarding diagnostic image quality as a function of dose reduction, acquisition, and reconstruction parameter settings and new CT technologies.

ACKNOWLEDGMENTS

We are deeply indebted to Dr. R. Rodríguez and Dr. R. Méndez, radiologists at the Hospital Clínico San Carlos in Madrid, for their help in the performance of the observer study.

[a]Electronic mail: irene.debroglie@gmail.com.

[1]D. J. Brenner, E. J. Hall, "Computed tomography—An increasing source of radiation exposure," N. Engl. J. Med. **357**, 2277–2284 (2007).
[2]A. Berrington de González, M. Mahesh, K. P. Kim, M. Bhargavan, R. Lewis, F. Mettler, and C. Land, "Projected cancer risks from computed tomography scans performed in the United States in 2007," Arch. Intern. Med. **169**, 2071–2077 (2009).
[3]R. Smith-Bindman, "Is computed tomography safe?," N. Engl. J. Med. **363**, 1–4 (2010).
[4]R. F. Redberg, "Cancer risks and radiation exposure from computed tomographic scans: How can we be sure that the benefits outweight the risks?," Arch. Intern. Med. **169**, 2049–2050 (2009).
[5]A. C. T. Martinsen, H. K. Saether, D. R. Olsen, P. A. Wolff, and P. Skaane, "Improved image quality of low-dose thoracic CT examinations with a new postprocessing software," J. Appl. Clin. Med. Phys. **11**, 250–258 (2010), Available from: http://www.jacmp.org/index.php/jacmp/article/viewArticle/3242/1948.
[6]International Commission on Radiation Units and Measurements, "Receiver operating characteristic analysis in medical imaging," ICRU Report No. 79 (International Commission on Radiation Units and Measurements, Bethesda, MD, 2008).
[7]M. S. Chesters, "Human visual perception and ROC methodology in medical imaging," Phys. Med. Biol. **37**, 1433–1476 (1992).
[8]A. Thilander-Klang, K. Ledenius, J. Hansson, P. Sund, and M. Bath, "Evaluation of subjective assessment of the low-contrast visibility in constancy control of computed tomography," Radiat. Prot. Dosimetry **139**, 449–454 (2010).
[9]International Commission on Radiation Units and Measurements, "Medical imaging—The assessment of image quality," ICRU Report No. 54 (International Commission on Radiation Units and Measurements, Bethesda, MD, 1996).
[10]M. P. Eckstein, C. K. Abbey, and F. O. Bochud, "A practical guide to model observers for visual detection in synthetic and natural noisy images," in *Handbook of Medical Imaging. Physics and Psychophysics*, Vol. 1, edited by J. Beutel, H. L. Kundel, and R. L. Van Metter (SPIE, 2000), pp. 595–629.
[11]E. H. Chao, T. L. Toth, N. B. Bromberg, E. C. Williams, S. H. Fox, and D. A. Carleton, "A statistical method of defining low contrast detectability," Radiology **217**, 162 (2000).
[12]F. R. Verdun, A. Denys, P. S. Valley, R. A. Meuli, "Detection of low-contrast objects: Experimental comparison of single-and multi-detector row CT," Radiology **223**, 426–431 (2002).
[13]T. Ishida, S. Tsukagoshi, K. Kondo, K. Kainuma, M. Okumura, and T. Sasaki, "Evaluation of dose efficiency index compared to receiver operating characteristics for assessing CT low-contrast performance," Proc. SPIE. **5368**, 527–533 (2004).
[14]S. J. Riederer, N. J. Pelc, and D. A. Chesler, "The noise power spectrum in computed x-ray tomography," Phys. Med. Biol. **23**, 446–454 (1978).
[15]R. Brooks and G. Di Chiro, "Statistical limitations in X-ray reconstructive tomography," Med. Phys. **3**, 237–240 (1976).

[16]Imaging Performance and Assessment of CT scanners, "32 to 64 slice CT scanner comparison report version 14," ImPACT Report 06013 (NHS Purchasing and supply agency, NHS PASA, 2005).
[17]E. Burgess, F. L. Jacobson, and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise," Med. Phys. **28**, 419–437 (2001).
[18]I. Reiser and R. M. Nishikawa, "Identification of simulated microcalcifications in white noise and mammographic backgrounds," Med. Phys. **33**, 2905–2911 (2006).
[19]B. M. Verbist, R. M. S. Joemai, W. M. Teeuwisse, W. J. H. Veldkamp, J. Geleijns, and J. H. M. Frijns, "Evaluation of 4 multisection CT systems in postoperative imaging of a cochlear implant: A human cadaver and phantom study," AJNR **29**, 1382–1388 (2008).
[20]A. Wunderlich and F. Noo, "Estimation of channelized hotelling observer performance with known class means or known difference of class means," IEEE Trans. Med. Imag. **28**, 1198–1207 (2009).
[21]W. J. H. Veldkamp, L. J. M. Kroft, J. P. Van Delft, and J. Geleijns, "A technique for simulating the effect of dose reduction on image quality in digital chest radiography," J. Digit. Imag. **2**, 1114–1125 (2009).
[22]N. Karssemeijer and M. A. O. Thijssen, "Determination of contrast-detail curves of mammography systems by automated image analysis," in *Digital Mammography*, edited by K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt (Elsevier, Amsterdam, 1996), pp. 155–160.
[23]W. J. H. Veldkamp, M. A. O. Thijssen, and N. Karssemeijer, "The value of scatter removal by a grid in full field digital mammography," Med. Phys. **30**, 1712–1718 (2003).
[24]International Electrotechnical Commission, "Medical electrical equipment. Part 2–44: Particular requirements for the safety of x-ray equipment for computed tomography," IEC publication No. 60601–2-44 (IEC, Geneva, Switzerland, 2002).
[25]R. Fahrig, R. Dixon, T. Payne, and R. L. Morin, "Dose and image quality for a cone-beam C-arm CT system," Med. Phys. **33**, 4541–4550 (2006).
[26]A. Ganguly, S. Yoon, and R. Fahrig, "Dose and detectability for a cone-beam C-arm system revisited," Med. Phys. **37**, 2264–2268 (2010).
[27]E. Samei, A. Badano, D. Chakraborty, K. Compton, C. Cornelius, K. Corrigan, M.J. Flynn, B. Hemminger, N. Hangiandreou, J. Johnson, D. M. Moxley-Stevens, W. Pavlicek, H. Roehrig, L. Rutz, J. Shepard, R. A. Uzenoff, J. Wang, and C. E. Willis, "Assessment of display performance for medical imaging systems: Executive summary of AAPM TG18 report," Med. Phys. **32**, 1205–1225 (2005).
[28]R. F. Woolson, *Statistical Methods of Analysis of Biomedical Data*, 1sted. (Wiley, New York, 1987), pp. 172–187.
[29]L. M. Morán, R. Rodríguez, A. Calzado, A. Turrero, A. Arenas, A. Cuevas, B. García-Castaño, N. Gómez, and P. Morán, "Image quality and dose evaluation in spiral chest CT examinations of patients with lung carcinoma," Br. J. Radiol. **77**, 839–846 (2004).
[30]A. E. Burgess, "Statistically defined backgrounds: Performance of a modified non-prewhitening observer model," J. Opt. Soc. Am. A **11**, 1237–1242 (1994).
[31]A. Wunderlich and F. Noo, "Image covariance and lesion detectability in direct fan-beam x-ray computed tomography," Phys. Med. Biol. **53**, 2471–2493 (2008).

4.3. Studying the effect of iterative reconstruction algorithms in low contrast detectability performance of a model observer and human observers analysing CT phantom images.

[III] I. Hernández-Girón, A. Calzado, J. Geleijns, R. M. S. Joemai, W. J. H. 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.

Br J Radiol 2014;87:20140014 (doi: 10.1259/bjr.20140014)

Abstract

Objective: To compare low-contrast detectability (LCDet) performance between a model [non-pre-whitening matched filter with an eye filter (NPWE)] and human observers in CT images reconstructed with filtered back projection (FBP) and iterative [adaptive iterative dose reduction three-dimensional (AIDR 3D; Toshiba Medical Systems, Zoetermeer, Netherlands)] algorithms.

Methods: Images of the Catphan® phantom (Phantom Laboratories, New York, NY) were acquired with Aquilion ONE™ 320-detector row CT (Toshiba Medical Systems, Tokyo, Japan) at five tube current levels (20–500 mA range) and reconstructed with FBP and AIDR 3D. Samples containing either low-contrast objects (diameters, 2–15 mm) or background were extracted and analysed by the NPWE model and four human observers in a two-alternative forced choice detection task study. Proportion correct (PC) values were obtained for each analysed object and used to compare human and model observer performances. An efficiency factor (η) was calculated to normalize NPWE to human results.

Results: Human and NPWE model PC values (normalized by the efficiency, η = 0.44) were highly correlated for the whole dose range. The Pearson's product-moment correlation coefficients (95% confidence interval) between human and NPWE were 0.984 (0.972–0.991) for AIDR 3D and 0.984 (0.971–0.991) for FBP, respectively. Bland-Altman plots based on PC results showed excellent agreement between human and NPWE [mean absolute difference 0.5±0.4%; range of differences (−1.7%, 5.6%)].

Conclusion: The NPWE model observer can predict human performance in LCDet in phantom CT images reconstructed with FBP and AIDR 3D algorithms at different dose levels.

Advances in knowledge: Quantitative assessment of LCDet in CT can accurately be performed using software based on a model observer.

BJR

Received: 31 December 2013 | Revised: 9 April 2014 | Accepted: 14 May 2014

doi: 10.1259/bjr.20140014

Cite this article as:
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.

FULL PAPER

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

[1,2]I HERNANDEZ-GIRON, MSc, [2]A CALZADO, PhD, [3]J GELEIJNS, PhD, [3]R M S JOEMAI, PhD and [3]W J H VELDKAMP, PhD

[1]Física Mèdica, Universitat Rovira i Virgili, Tarragona, Spain
[2]Departamento de Radiología, Universidad Complutense de Madrid, Madrid, Spain
[3]Radiology Department, Leiden University Medical Center, Leiden, Netherlands

Address correspondence to: Miss Irene Hernandez-Giron
E-mail: *irene.debroglie@gmail.com*

Objective: To compare low-contrast detectability (LCDet) performance between a model [non-pre-whitening matched filter with an eye filter (NPWE)] and human observers in CT images reconstructed with filtered back projection (FBP) and iterative [adaptive iterative dose reduction three-dimensional (AIDR 3D; Toshiba Medical Systems, Zoetermeer, Netherlands)] algorithms.

Methods: Images of the Catphan® phantom (Phantom Laboratories, New York, NY) were acquired with Aquilion ONE™ 320-detector row CT (Toshiba Medical Systems, Tokyo, Japan) at five tube current levels (20–500 mA range) and reconstructed with FBP and AIDR 3D. Samples containing either low-contrast objects (diameters, 2–15 mm) or background were extracted and analysed by the NPWE model and four human observers in a two-alternative forced choice detection task study. Proportion correct (PC) values were obtained for each analysed object and used to compare human and model observer performances. An efficiency factor (η) was calculated to normalize NPWE to human results.

Results: Human and NPWE model PC values (normalized by the efficiency, $\eta = 0.44$) were highly correlated for the whole dose range. The Pearson's product-moment correlation coefficients (95% confidence interval) between human and NPWE were 0.984 (0.972–0.991) for AIDR 3D and 0.984 (0.971–0.991) for FBP, respectively. Bland–Altman plots based on PC results showed excellent agreement between human and NPWE [mean absolute difference $0.5 \pm 0.4\%$; range of differences (−4.7%, 5.6%)].

Conclusion: The NPWE model observer can predict human performance in LCDet tasks in phantom CT images reconstructed with FBP and AIDR 3D algorithms at different dose levels.

Advances in knowledge: Quantitative assessment of LCDet in CT can accurately be performed using software based on a model observer.

CT has become one of the most used techniques in radiology departments. Its progressive introduction in healthcare services and the increasing number of CT scans performed worldwide per year has raised the concern about the related radiation dose.[1,2] Several improvements have been incorporated in the scanners to obtain images at the lowest achievable dose without losing relevant diagnostic information. Among them, iterative reconstruction techniques are promising. Several studies have shown that, with these algorithms, the image noise can be decreased and that higher contrast-to-noise ratios (CNRs) can be obtained compared with traditional filtered back projection (FBP) and thus a significant dose reduction can be achieved.[3–6]

A wide variability in dose and image quality has been found between different CT scanners to perform similar diagnostic tasks.[7] To assess image quality, low-contrast detectability (LCDet) is determined as the smallest object visible for certain contrast value at a given dose level. LCDet can be subjectively assessed by several observers scoring the visibility of objects on CT phantom images. These studies are time consuming and expensive owing to the large required number of observers and observations.[8] The range of available protocols and custom parameters for each application adds complexity to optimization too.[9] Furthermore, the results might be biased if the observers know beforehand the location of the objects in the phantom. Tests of statistical significance are controversial to

obtain average results based on human observer studies, as a great inter- and intra-observer variability may appear.[10,11] Computer model observers, intended to predict the performance of human observers in image analysis, can be an alternative to objectively assess image quality. They can be a useful tool when investigating the influence of acquisition and reconstruction parameters on image quality or the effect of object size, shape and contrast in detection tasks.[12–15]

In a previous work, an objective statistical method using a specific model observer [non–pre-whitening matched filter with an eye filter (NPWE)] was presented to investigate the influence of different CT acquisition parameters on LCDet.[16]

The main goal of this work is to compare the model observer LCDet performance in CT images acquired at different dose levels with human observers. Images reconstructed with two algorithms (FBP and iterative) were used in this study. Two-alternative forced choice (2-AFC) experiments, in which the observers scored samples containing signals or background (Bg) extracted from the images, were carried out. The results were presented at the Medical Imaging Perception Society XV Conference held in Washington DC during 14–16 August 2013, which is focused on observer performance analysis and diagnostic quality of imaging technique improvements.

METHODS AND MATERIALS

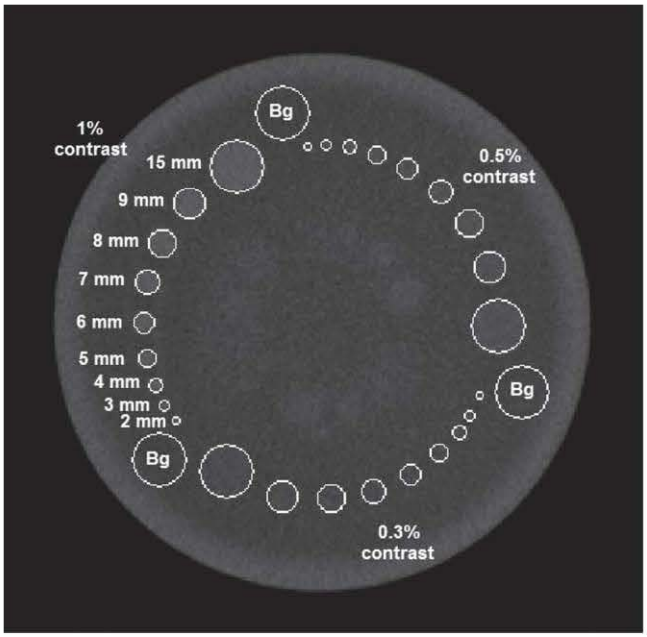
Image acquisition

Throughout this study, images of the Catphan® 500 phantom (Phantom Laboratories, New York, NY), which is dedicated to quality control tasks on CT scanners, were used. The low-contrast module (CTP515) contains three groups of cylindrical rods of various diameters (2–15mm) and three contrast levels (0.3%, 0.5% and 1.0% nominal contrast), as shown in Figure 1. The nominal contrast (expressed as a percentage) is defined by the Catphan manufacturer as the difference in CT number between the target object and the background divided by 10.

CT images of the phantom were acquired on a 320-detector row CT scanner (Aquilion ONE™; Toshiba Medical Systems, Tokyo, Japan) by selecting the following parameters: 64 × 0.5-mm beam collimation, 240-mm field of view, helical acquisition (pitch, 0.828), 120 kVp tube voltage, 0.5 s rotation time and five different tube current levels (20, 40, 80, 300 and 500 mA). Images of 0.5-mm slice thickness were reconstructed with a soft-body kernel (FC13), which enhances low frequencies in the image, reduces high-frequency noise and smooths the appearance of the image in general. Two reconstruction algorithms were selected: FBP and an iterative algorithm [adaptive iterative dose reduction three dimensional (AIDR 3D); Toshiba Medical Systems]. The latter is an iterative algorithm that performs calculations in the raw data domain using statistical models, scanner characteristics and projection noise estimation to decrease the electronic noise and, afterwards, applies an iterative technique in the image domain to decrease image noise.[5]

The phantom was scanned two times for each tube current–time product (mA) value. To avoid possible artefacts owing to the nearby modules, only the 42 central axial images of the LC module were taken into account from each scan. Thus, ten image series (considering the five mA values and two reconstruction algorithms used), composed by 84 images each, were available for the model and human observer tests in this study.

Figure 1. A constructed 40-mm thick slice of the Catphan® (Phantom Laboratories, New York, NY) low-contrast module. The contrast groups and the object diameters are tagged for the supraslice region. The mask for the objects and the background (Bg) sample locations are overlaid in the figure.



Model observer (NPWE) and low-contrast detectability software

A software program dedicated to automated LC objects detection on CT, implemented in MATLAB® (MathWorks®, Natick, MA), was described in a previous work.[16] The improvements implemented in the methodology are explained in detail in this section.

To locate the LC objects in the CT images, a mask of the distribution of the disks in the phantom was created. The manufacturer specifications (size, shape, position and contrast) were used to generate templates to match the objects in the real CT images (Figure 1). The object templates were blurred to model the modulation transfer function in each case, which was obtained as the full width at half maximum of the point spread function (PSF).[17] Images of a phantom containing a 0.18-mm diameter tungsten bead were acquired for the different mA values and reconstruction algorithms to measure the PSF values. A thick slice is automatically created for each mA set by averaging all the available related images. To optimize the detection of the objects in the CT images, the templates were individually shifted 3 × 3 pixels around the initial location estimated using Catphan specifications.

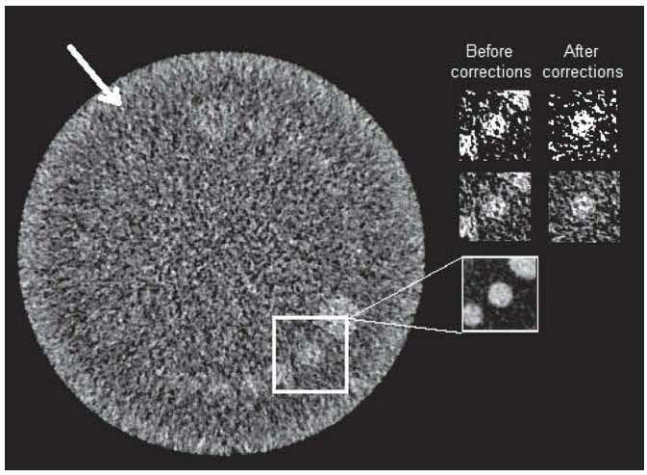
A circular white band was found close to the outer rim of the phantom images (Figure 2). These background inhomogeneities

may affect LCDet. To correct them, in the thick slices previously created, for each of the image sets individually, an annular-shaped region of interest (ROI) was taken around the 15-mm object of each contrast group, and another circular ROI was taken on these objects. The signal difference in Hounsfield units was measured between these regions. Based on these values, artificial signals were created, blurred by the measured PSF value and subtracted from the thick slice image. The resulting thick slice (equivalent to the LC module without objects in it) was then subtracted from the individual CT images.

To avoid any bias in the human observer study, the samples taken from the CT images should have the same size, independently of object diameter. The geometrical distribution of the LC objects in the Catphan phantom was a limitation for this purpose, as nearby objects could be included in the samples. To overcome this, an additional image correction was performed, using the templates previously created, to wipe out, from each object sample, the nearby objects in its corners.

The effect of these corrections (Bg inhomogeneities and object wipe out) in the images was analysed comparing the noise and contrast in the original and corrected images. For each mA and FBP/AIDR 3D series (for either the corrected or original set), the mean pixel value and the standard deviation σ (used as a measure of noise) were measured in ROIs of size $26.7 \times 26.7 mm^2$ taken in the Bg sample locations (Figure 1). A relative difference value (%) was calculated for each condition as $(\sigma_{original\ images} - \sigma_{corrected\ images})/\sigma_{corrected\ images}$. Regarding the effect on contrast, a ROI was defined at the exact location of the 15-mm object for the three contrast groups. Contrast (C) was measured, averaged for each set (original or corrected image), and relative difference values were obtained as $(C_{original\ images} - C_{corrected\ images})/C_{corrected\ images}$.

Figure 2. An example of the wiping out of nearby object processes in the Catphan phantom CT images for the 150 mA filtered back projection series. Inside the white square, a crop of the thick slice is shown before the correction. For one of the images in the set, the object samples are shown with different window settings before and after the corrections. The arrow highlights the band background inhomogeneities.



For the 2-AFC experiment, Bg samples were extracted from an area located close to the smallest disk of each contrast group but positioned farther from the module centre (Figure 1). Object (signal) samples were extracted following the process explained above. Both types of samples had the same size ($26.7 \times 26.7\ mm^2$) for all the objects in the module with independence of their diameter.

The software automatically calculated LCDet using an NPWE model observer for each object and the three contrast groups present in the LC module of the phantom. This model is based on the assumption that the human observer uses templates of the expected signals for cross-correlation in the images and that it is unable to modify the template to pre-whiten correlated noise. The addition of an eye filter (E) takes into account the spatial frequency (f) response of the human eye. We selected the eye filter proposed by Burgess $E(f) = fe^{-bf}$, with b chosen such that $E(f)$ peaked at four cycles per degree and assuming a fixed viewing distance of 50 cm from the monitor.[18] Different studies have shown that human performance lies between pre-whitening and non–pre-whitening, depending on the spectral distribution of the image noise.[19,20]

For each object in the phantom, the model cross-correlates the samples (signal or Bg) taken from the 84 images of the set with the appropriate template (blurred expected signal), after filtering them by an eye filter (E).[18] This results in T_1 (correlations of the template and object samples) and T_2 (correlations of the template and Bg samples). Based on distributions of the test statistics of the correlation results, a discrimination index d' was calculated applying Equation (1):[18]

$$d' = \frac{\langle T \rangle_1 - \langle T \rangle_2}{\sqrt{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2}}, \quad (1)$$

where $< >$ refers to the mean and $\sigma(\ )$ is the standard deviation; subindexes 1 and 2 are related to the object and to the Bg distributions of test statistics, respectively.

This procedure was performed for all the contrast groups in the phantom and repeated for the five selected mA values and two reconstruction techniques sets. The detectability index d' was expressed as a function of object diameter for the three contrast groups and each condition. Then, d' values were transformed into proportion correct (PC) using Equation (2):[16,18]

$$PC = 0.5 + 0.5\,\mathrm{erf}\left(\frac{d'}{2}\right) \quad (2)$$

where erf (x) is the error function given by Equation (3):

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-x^2} dx \quad (3)$$

This method was applied for the 10 CT image series, and thus, d' and PC profiles as a function of the object diameter were obtained for each mA and FBP or AIDR 3D sets. As, just by chance, in a 2-AFC experiment, a default PC = 50% value can

be obtained, the detectability threshold (λ) was fixed at PC = 75%. Thus, when PC $\geq$ 75% in the analysis, the related object diameter was considered visible.

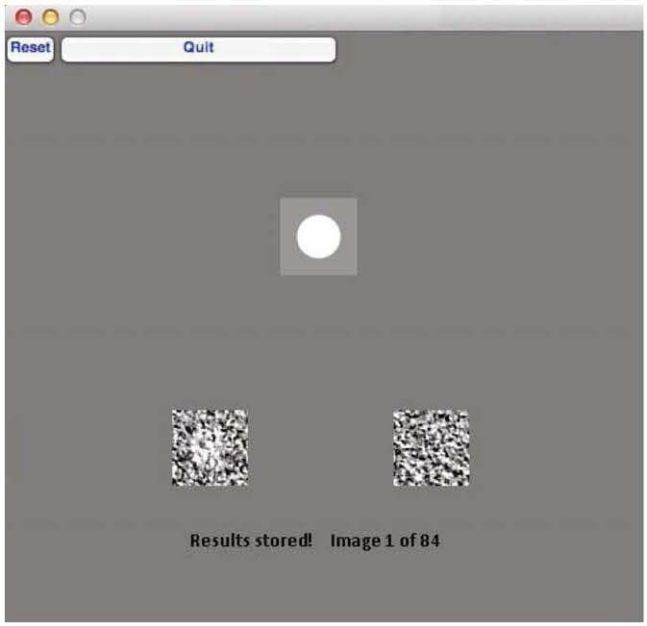
Human observer two-alternative forced choice study

To validate the trends shown by the NPWE, a 2-AFC human observer study was carried out by four medical physicists, each of them scoring pairs of ROIs (signal or Bg samples) extracted from the different sets of images for the 1% contrast group. To analyse intra-observer variability, each observer scored twice the 84 pairs of images (the same used for the NPWE model) related to a given object diameter acquired at certain mAs and reconstructed with FBP or AIDR 3D. Thus, each observer scored 84 (pairs of images) × 9 (diameters) × 5 (mA values) × 2 (FBP or AIDR 3D) × 2 (intra-observer variability), which makes 15,120 images in total.

For the signal known exactly and background known exactly (SKE/BKE) task performed in the human observer study, an application was created in MATLAB. In Figure 3, an example of the 2-AFC software interface is shown: two images are displayed together with the template on a grey canvas; the one which contains the object must be clicked on and scoring results are automatically stored in an output file. The object always appears in the centre of the sample, as shown in the template. The images with or without object were displayed randomly at left or right. Each set of images (for a given diameter, mA and reconstruction method) was independently scored and images related to different conditions were not mixed in this study.

The scoring was performed on an i-MAC 27″ (Apple Inc., cupertino, CA) monitor using recommended visualization conditions, with

Figure 3. Interface of the two-alternative forced choice software used for the human observer experiment. The template (above) is shown together with signal and background samples extracted from the images.



fixed values for window level and width (taken as 3σ, where σ is the average standard deviation of pixel values of the Bg samples for each series). The quotient between the maximum and minimum luminance that the monitor can deliver or luminance ratio was 491, and the measured ambient luminance was kept <10 lux.[21]

One training session was programmed for the observers to get used to the software features and the task. All observers scored the images twice (without any time limitation to review them) in four different sessions (two for each reconstruction method to analyse the intra-observer variability), which lasted approximately 2 hours each. There was a gap of at least 2 weeks between them, to avoid learning effects. The viewing distance was fixed at 50 cm, and the observers were allowed to rest whenever they wanted to avoid fatigue.

An analysis of the intra-observer consistency was performed using the Wilcoxon signed rank test for matched-pair samples (consistent results if $p \geq 0.05$) by comparing the scores for each object size and mA separately obtained in each session for AIDR 3D and FBP.[22] If one observer was inconsistent in his results between both sessions for a given condition, that scoring was ruled out.[16] The average human observer performance was obtained as the mean of the PC values that passed the intra-observer tests for each condition. Finally, PC curves, as a function of the object diameter, were obtained for each mA and either FBP or AIDR 3D.

Efficiency (η) calculation and agreement between human and model observer

To obtain an efficiency (η) between the human observers and the model in our experiments, PC values had to be transformed into d' applying Equation (4):[12,23,24]

$$d' = \sqrt{2}\Phi^{-1}(\mathrm{PC}) \quad (4)$$

where Φ^{-1}(PC) is the inverse of the standard cumulative normal distribution function.

Finally, η could be calculated to relate the average human observer performance (d'_{human}) to the model observer (d'_{NPWE}) by applying Equation (5) and using a least-squares procedure to fit the data.[12,19] The error bars used as weights in the linear fit were estimated as 2σ, where σ is the standard deviation of the d'_{human} squared values. The efficiency η tallied the linear fit slope.

$$\left(d'_{\mathrm{human}}\right)^2 = \eta \left(d'_{\mathrm{NPWE}}\right)^2 \quad (5)$$

To study the agreement of the NPWE and human observers, their related PC values were compared using Bland–Altman plots using EpiDat software.[25] Additionally, Pearson's product–moment correlation coefficients (r) between human and model PC scorings were calculated for both reconstruction methods and each mA separately (perfect correlation if the absolute value of $r = 1.0$).[14]

Psychometric fits and visibility thresholds

Psychometric fits were performed for the obtained PC profiles as a function of the object diameter.[26,27] For this 2-AFC experiment, fitting curves according to Equation (6) were applied for

each mA and reconstruction set independently, for both the average human and model observer.[16] For the average human observer, the error bars related to the PC values, previously calculated, were used as weights in the fitting process based on a least-squares procedure. The range of the fitting curves runs from 0.5 (pure guessing) and 1.00 (certain detection).

$$PC = \frac{0.5}{1 + e^{f\log\left(\frac{d}{\lambda}\right)}} + 0.5 \quad (6)$$

where d represents the object diameter and f and λ are the fitting parameters. The steepness of the psychometric curve is determined by f. The smallest object diameter, which matches the proposed visibility threshold (PC = 75%) is λ itself.

In the case of the NPWE, additional psychometric fits were performed using the PC values corrected by the efficiency value η.

Image quality comparison between both reconstruction algorithms

To analyse the effect of selecting FBP or AIDR 3D in LCDet performance, two-tailed paired t-tests ($\alpha = 0.05$) were performed comparing d' values obtained with NPWE model and all contrast groups for the different mAs. Similar tests were also performed using the PC values obtained for the 1% contrast group and all mAs, by the human observers and the model observer, respectively.[22]

Additionally, an estimation of the average noise value was obtained for each mA and reconstructed image set. Pixel noise was measured as the standard deviation (σ) of the pixel values, in three circular ROIs taken at the same locations as the Bg samples, and the average noise value was calculated. A relative difference value (%) between FBP and AIDR 3D sets was obtained for each mA as $(\sigma_{FBP} - \sigma_{AIDR\,3D})/\sigma_{FBP}$. A repeated measures analysis of variance (ANOVA) test was performed between the noise measurements calculated for both algorithms and each mA separately (significant differences if $p \leq 0.05$) in the original images.

RESULTS

Analysing the signal and Bg samples before and after applying the Bg corrections (to suppress undesired Bg trends and to wipe out nearby objects), it was found that contrast varied <5% in all cases. The standard deviation of pixel values, which reflects the combined effect of inhomogeneities and noise in the images, was also reduced after applying these corrections in the range 4–10%. To depict the effect on the images, in Figure 2, it can be seen on one of the signal samples before and after this correction.

Model observer results

The NPWE model observer obtained higher d' values with increasing object contrast. Detectability also increased approximately linearly with object diameter. In Table 1, the slopes for the linear fits performed for all the sets of d' as a function of the object diameter and the three contrast groups are summarized [95% confidence interval (CI)]. The range of R_2 for the linear fits was 0.907–0.995 for FBP and 0.890–0.993 for AIDR 3D sets, respectively.

The influence of contrast and mAs in LCDet is shown in Table 1: higher slopes are obtained with increasing contrast and mAs for both FBP and AIDR 3D. Two-tailed paired t-tests ($\alpha = 0.05$) were performed comparing the d' values related to the contrast groups for both reconstruction methods and each mA separately. A significant improvement in the detection of objects as contrast increased was found ($p \leq 0.05$ in all cases). Similar tests were performed to determine the differences in d', with increasing mAs indicating that NPWE showed a significant improvement in LCDet as tube current increased for all contrast groups and both reconstruction algorithms ($p \leq 0.05$).

Human observer results

To study human observers LCDet performance, 60,480 pairs of images (for 1% contrast group in the Catphan phantom) were analysed [15,120 (images scored by 1 observer) × 4 (4 observers)]. From now on we will use the term "scoring" to refer to the series of results for a given diameter, mAs and reconstruction obtained by an observer.

The intra-observer variability test led to discard ($p < 0.05$) eight individual pairs of scorings, three for FBP and five for AIDR 3D (2.2% of all the scorings). The distribution of discarded scorings by the four observers was 4, 3, 1 and 0, respectively. After filtering the results, removing the inconsistent data, no significant differences were found between the human scorings ($p \geq 0.05$).

The psychometric fits obtained for the average human observer based on the AIDR 3D scoring data (1% contrast) are illustrated in Figure 4. The related R^2 fitting values were in the ranges 0.743–0.945 for FPB reconstruction and 0.710–0.955 for AIDR 3D. The error of the mean PC value for the average human observer for the different mA series (10, 20, 40, 150 and 250 mAs) were in the ranges 0.2–18%; 0.8–12.5%; 0.8–17.8%; 0.7–16.7%; and 0.5–5% for AIDR 3D and 6.5–12.8%; 1.1–8.2%; 0.8–9.8%; 0.6–16.5% and 0.7–6.7% for FBP, respectively.

Efficiency calculation

Owing to the shape of the curve of d' as a function of PC, it is difficult to measure d' when its value is above three, approximately (PC ≈ 0.98) in a 2-AFC experiment.[28,29] Only the human PC values below this threshold were used to determine the efficiency of the NPWE model observer. In Figure 5, the d' values for the average human observer are plotted as a function of NPWE models (both squared). The data related to all the mA series for AIDR 3D and FBP for 1% contrast were taken into account in this graph. The linear fit slope, which tallies the efficiency, η, was 0.44 (0.42–0.46, 95% CI).

Visibility thresholds non-pre-whitening matched filter with an eye filter and average human observer

The visibility thresholds λ (related to PC = 75%, 95% CI) for the 1% contrast group obtained by the average human observer in the 10–250 mA range are depicted in Table 2 together with the NPWE model values, after correcting them by the efficiency ($\eta = 0.44$). It can be seen that smaller objects could be detected as mAs increased for both reconstruction algorithms by the human and model observer.

Table 1. Slopes of the linear fits of detectability index (d') as a function of object diameter for the tube current–time product (mA) range and filtered back projection (FBP)/adaptive iterative dose reduction three dimensional (AIDR 3D) algorithm reconstructed sets of images for the three contrast groups and non-pre-whitening matched filter with an eye filter model observer (values for confidence interval = 95%). The results of two-tailed paired t-tests (significant differences for $p \leq 0.05$) comparing FBP and AIDR 3D d' values for each condition are also shown

Contrast	Tube current–time product	10 mA	20 mA	40 mA	150 mA	250 mA
1%	FBP	0.21 (0.20–0.22)	0.37 (0.35–0.40)	0.48 (0.46–0.50)	0.81 (0.80–0.83)	1.03 (0.99–1.07)
	AIDR 3D	0.27 (0.25–0.28)	0.40 (0.37–0.43)	0.56 (0.54–0.58)	0.85 (0.83–0.86)	1.08 (1.03–1.13)
	p-value	0.005	0.052	<0.001	<0.001	0.002
0.5%	FBP	0.10 (0.09–0.10)	0.19 (0.18–0.19)	0.22 (0.22–0.23)	0.38 (0.37–0.38)	0.60 (0.58–0.62)
	AIDR 3D	0.11 (0.11–0.12)	0.18 (0.17–0.18)	0.25 (0.25–0.26)	0.39 (0.39–0.40)	0.63 (0.60–0.66)
	p-value	<0.001	<0.001	<0.001	<0.001	<0.001
0.3%	FBP	0.05 (0.05–0.05)	0.11 (0.11–0.11)	0.12 (0.11–0.12)	0.20 (0.19–0.20)	0.38 (0.37–0.39)
	AIDR 3D	0.06 (0.06–0.06)	0.09 (0.08–0.09)	0.13 (0.13–0.14)	0.20 (0.20–0.21)	0.39 (0.38–0.40)
	p-value	0.004	0.002	0.002	0.002	<0.001

For NPWE, the visibility threshold λ (related to PC = 75%) increased dramatically with decreasing contrast (Table 3, 95% CI) for both AIDR 3D and FBP. This effect was more evident below <150 mA.

Analysis of agreement between non-pre-whitening matched filter with an eye filter and human observer

The normalization of the NPWE results by the efficiency led to a high correlation with the average human observer, for all mAs and both reconstruction methods. The overall Pearson's product–moment correlation coefficients (considering all mAs) calculated for 95% CI were 0.984 (0.972–0.991) and 0.984 (0.971–0.991) for AIDR 3D and FBP, respectively. The correlations for 10, 20, 40, 150 and 250 mA are shown in Table 2. Figure 6 depicts the psychometric fits for the human observer and the NPWE model (after the efficiency correction) as a function of mAs for the FBP reconstructed sets.

Figure 7 shows the Bland–Altman plot performed for the PC values obtained by the average human observer and the NPWE model (after correction by efficiency) for AIDR 3D and FBP altogether. It showed an excellent agreement with a mean absolute difference, Δ, of 0.5 ± 0.4%. The range of the differences, given by $(\Delta - 2\sigma, \Delta + 2\sigma)$ was (−4.7%, 5.6%), where Δ is the mean absolute difference and σ is the standard deviation of the differences between NPWE and human observers. For AIDR 3D images, the mean absolute difference (Δ) and the range of the differences were 0.4 ± 0.4% and −4.8%, 5.2%, respectively, whereas for FBP sets they were 0.4 ± 0.2% and −3.9%, 5.0%.

Image quality comparison between both reconstruction algorithms

The repeated measures ANOVA test performed to analyse the differences in the image noise when applying FBP or AIDR 3D in the original images showed significant differences for all the mA values ($F > 113{,}985$; $p < 0.001$). AIDR 3D produced a significant reduction of noise compared with FBP of 51%, 43%,

Figure 4. Psychometric fits [proportion correct (PC) as a function of the object diameter] for the average human observer and all tube current–time product (mA) for the images reconstructed with adaptive iterative dose reduction three dimensional algorithm and 1% contrast. The dots represent the average human observer PC values.

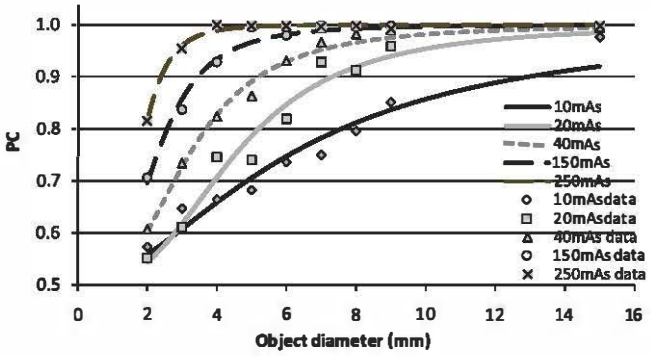


Figure 5. Squared detectability index (d'^2) for the average human observer as a function of the model observer [non-pre-whitening matched filter with an eye filter (NPWE)]. The efficiency η is given by the slope of the linear fit [95% confidence interval (CI)].

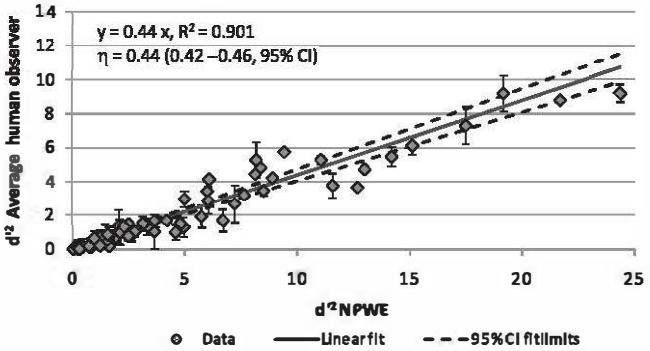


Table 2. Visibility thresholds [proportion correct (PC) = 75%] for the average human and non-pre-whitening matched filter with an eye filter (NPWE) (corrected by the efficiency) model observers and Pearson's product-moment correlation coefficients (r) for their PC values for all tube current–time products (mAs) and both reconstruction algorithms [filtered back projection (FBP) and adaptive iterative dose reduction three dimensional (AIDR 3D) algorithm] (95% confidence interval)

	λ (mm) FBP		Pearson coefficient (r)	λ (mm) AIDR 3D		Pearson coefficient (r)
	Average human	NPWE		Average human	NPWE	
10 mA	6.5 (6.0–7.0)	6.8 (6.6–7.0)	0.969 (0.857–0.993)	6.8 (6.6–7.0)	6.0 (5.7–6.3)	0.988 (0.943–0.997)
20 mA	4.5 (4.4–4.6)	4.6 (4.4–4.7)	0.983 (0.921–0.996)	4.6 (4.4–4.7)	4.3 (4.1–4.5)	0.978 (0.897–0.995)
40 mA	3.6 (3.5–3.7)	3.8 (3.6–3.9)	0.984 (0.925–0.996)	3.8 (3.6–3.9)	3.2 (3.1–3.3)	0.991 (0.953–0.998)
150 mA	2.9 (2.8–3.0)	2.5 (2.3–2.8)	0.996 (0.978–1.000)	2.5 (2.3–2.8)	2.3 (2.2–2.4)	0.989 (0.946–0.997)
250 mA	1.9 (1.8–2.0)	2.0 (1.9–2.0)	0.997 (0.986–1.000)	2.0 (1.9–2.0)	1.8 (1.7–1.9)	0.994 (0.971–0.998)

34%, 25% and 23% relative to FBP for 10, 20, 40, 150 and 250 mA, respectively.

For the NPWE model, two-tailed paired t-tests ($\alpha = 0.05$) were performed comparing the d' values obtained for FBP and AIDR 3D, each mA and all contrast groups. The related p-values for each mA are shown in Table 1. Significant improvement ($p \leq 0.05$) was shown with AIDR 3D for all mAs and contrast groups.

Figure 8 depicts the overall effect of selecting each reconstruction method on the NPWE LCDet performance showing the psychometric fits for the 0.3% contrast group. R^2 values were in the range 0.995–0.960 for FPB and 0.993–0.953 for AIDR 3D for all the contrast groups. This trend was the same for the human observer (1% contrast).

For NPWE, the results of the two-tailed paired t-tests ($\alpha = 0.05$) performed for the PC values related to each mA comparing both algorithms showed significant differences in all cases ($p \leq 0.05$). For the human observer, significant differences ($p < 0.05$) appeared for the lower mA series (10, 20 and 40 mA). No significant differences between both reconstruction methods were found for the 150 and 250 mA series (p-values of 0.05 and 0.06, respectively).

DISCUSSION

The selected model observer NPWE reproduced the LCDet performance trends of the average human observer as a function of mAs. In this study, the model and human observers scored the same sets of images (corrected to suppress undesired background trends). The model was more efficient than the human observer to detect LC objects in FPB and AIDR 3D reconstructed CT images. The calculated efficiency (0.44) is in the range obtained by other authors ($\eta \approx 0.5$) when applying the same model observer to other types of images.[14,18,29] The agreement between the model and human observer was excellent at the dose range considered in this work (10–250 mA) for both reconstruction algorithms after applying the η factor, as shown in Figure 6.

The efficiency was also calculated using all the human scorings (without discarding any values owing to intra-observer inconsistency), obtaining a slightly smaller η of 0.41 (0.39–0.43, 95% CI) in this case.

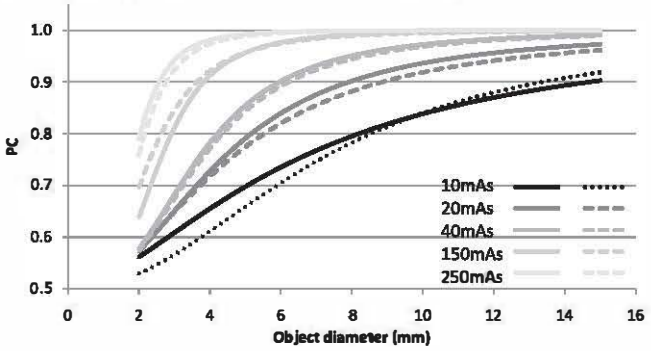
The Bland–Altman plot showed an excellent agreement ($\Delta = 0.5 \pm 0.4\%$) between the human and NPWE, the range of the differences being about ±5%. This analysis was also performed taking into account all the original human PC values to study the effect or discarding data (owing to intra-observer inconsistency) on the correlation between human and model. In this case, the differences increased on average $\Delta = -1.0\% \pm 0.7\%$ and also in range −11.2% to 9.1%.

By analysing the slopes of d' as a function of object diameter fits (Table 1), it was shown that the NPWE model LCDet

Table 3. Visibility thresholds (related to proportion correct = 75%) for the non-pre-whitening matched filter with an eye filter model and both reconstructions [filtered back projection (FBP) and adaptive iterative dose reduction three dimensional (AIDR 3D) algorithm] for all the tube current–time product (mA) series and contrast groups (confidence interval = 95%)

	λ (mm) 1% contrast		λ (mm) 0.5% contrast		λ (mm) 0.3% contrast	
	FBP	AIDR 3D	FBP	AIDR 3D	FBP	AIDR 3D
10 mA	5.1 (5.0–5.3)	4.3 (4.2–4.4)	9.6 (9.3–9.9)	8.2 (7.9–8.5)	18.4 (17.8–18.9)	14.7 (14.3–15.1)
20 mA	3.2 (3.2–3.3)	3.1 (3.0–3.1)	5.4 (5.2–5.6)	5.6 (5.5–5.8)	8.5 (8.2–8.7)	10.9 (10.5–11.4)
40 mA	2.8 (2.7–2.9)	2.6 (2.5–2.7)	4.7 (4.5–4.9)	4.0 (3.9–4.1)	8.2 (7.9–8.5)	7.3 (7.1–7.5)
150 mA	1.9 (1.9–1.9)	1.8 (1.8–1.8)	2.8 (2.7–2.9)	2.6 (2.5–2.7)	4.9 (4.7–5.1)	4.6 (4.5–4.8)
250 mA	1.7 (1.7–1.7)	1.7 (1.7–1.7)	2.1 (2.0–2.1)	1.9 (1.9–2.0)	2.9 (2.9–3.0)	2.8 (2.7–2.8)

Figure 6. Psychometric fits for the human (lines) and the non-pre-whitening matched filter with an eye filter model (dashed lines) based on the results for the filtered back projection reconstructed images and all tube current-time products (mAs) for 1% contrast objects. PC, proportion correct.



performance significantly improved for all mAs and contrast groups with AIDR 3D ($p \leq 0.05$). These trends were also reflected in the psychometric fits for both, humans and model (Figure 8), obtaining higher PC values with AIDR 3D. In general, AIDR 3D showed an overall improvement in detectability as object diameter increased, compared with FBP for the entire dose range. The two-tailed *t*-tests performed for the PC values and each mA showed significant improvement ($p \leq 0.05$) for the NPWE when using AIDR 3D in all the dose range. For the human observer, significant improvement was found only in the range 10, 20 and 40 mA when applying the iterative algorithm.

The visibility thresholds for 1% contrast showed differences between both reconstruction methods, with the same trends for the model and human observers, but they were very subtle for high mAs. It has been noted that for the human observers, no significant differences between the algorithms were found between the PC values obtained for the higher mAs (150–250 mA).

Figure 7. Bland–Altman plot of proportion correct (PC) difference between human and model observer (after correcting by efficiency) for filtered back projection (FBP) (○) and the adaptive iterative dose reduction three dimensional (AIDR 3D) algorithm (◇). The black line represents the average absolute difference Δ ($0.5 \pm 0.4\%$); the two dash lines represent $\Delta \pm 2\sigma$, where σ is the standard deviation of the differences, which are −4.7%, 5.6%. NPWE, non-pre-whitening matched filter with an eye filter.

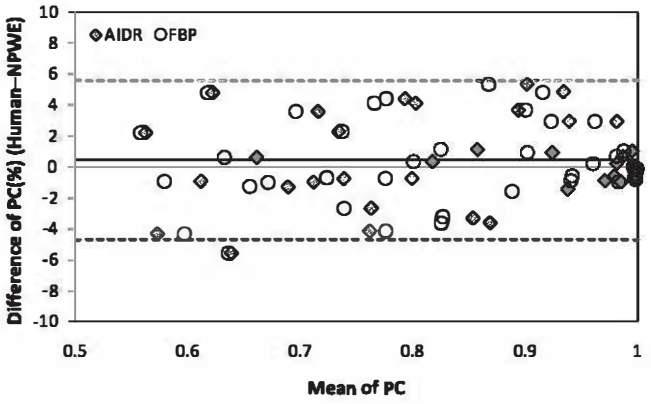
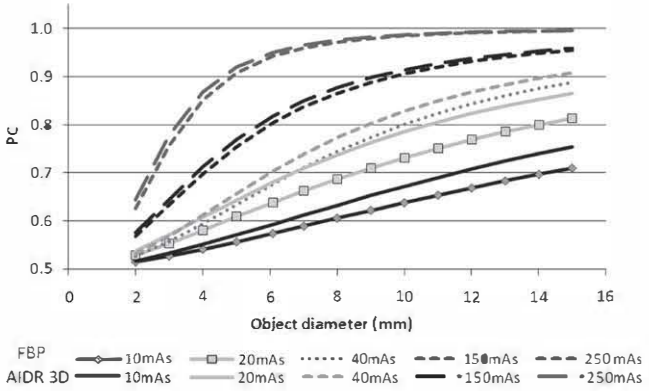


Figure 8. Psychometric fits of proportion correct (PC) as a function of object diameter for non-pre-whitening matched filter with an eye filter and both reconstructions filtered back projection (FBP)/adaptive iterative dose reduction three dimensional (AIDR 3D) algorithm in all the tube current-time products (mAs) range for 0.3% contrast.



Selecting only one threshold value may lead to missing relevant information related to LCDet performance, although it can be helpful as a rough estimate to compare different protocols where dose is changed significantly. As an alternative, the profiles shown in Figure 8 and Bland–Altman plots represent a good tool to study LCDet performance in CT.

In a previous study, a different and smaller set of images reconstructed with AIDR 3D was compared with FBP for the same mA range.[5] The visibility thresholds obtained with NPWE were slightly different then, but it has to be noted that a different psychometric fit was used. In the present work, the selected psychometric curve was a good candidate to be applied to both sets of data (human and NPWE model).[16]

The undesired Bg trends in the images (white band) were suppressed by applying a correction based on the creation of a thick slice image. The transformations applied to the images to correct these trends and to wipe out the nearby objects (to enable taking samples of a reasonable and equal size for the human observer study) did not affect substantially the CNR of the objects owing to the low noise level of the thick slice. Despite being promising, the effect was not of the same order for all the mA sets, and these processes can still be optimized. Other studies opted for a different strategy to perform 2-AFC human observer experiments based on the entire Catphan image and covering the objects that were not being scored by crops taken from the nearby background regions in the image.[30]

The performed human observer study has some limitations. The first one is the reduced number of observers (only four). To obtain a good average of the human LCDet performance for the proposed task, a statistical analysis was performed to remove the inconsistent data. Even so, the study was quite complex to carry out, owing to the high number of images and conditions analysed, although it was restricted only to the 1% contrast group.

The results shown in this work are based on geometrical phantom images, which were modified (Bg correction), and its

conclusions have to be taken cautiously and cannot be extrapolated directly to patient images. Model observers can be helpful tools to analyse image quality in an objective and fast way and to compare different CT scanners, protocols or reconstruction algorithms in terms of image quality. The increasing complexity and variety in the available CT protocols and reconstruction algorithms makes the development of these automated methods even more necessary.

CONCLUSIONS

The LCDet performance of human and a model observer (NPWE) has been compared in this study analysing phantom images reconstructed with AIDR 3D and FBP algorithms and a range of mAs. The A 2-AFC study was carried out to estimate the average human observer performance for an SKE/BKE task. The NPWE model was more efficient than the average human ($\eta = 0.44$) and showed an excellent agreement after the correction by the efficiency factor. Other alternatives to match the model observer results in order to reproduce the human observer performance are based on internal noise, which will be explored in the near future. The iterative algorithm (AIDR 3D) showed an overall improvement in LCDet, especially for low mAs and low-contrast objects. The methodology that we have developed for the human study can be used to perform analysis with different types of medical images, not necessarily CT. The proposed method can be adapted to other phantoms and other model observers will be implemented to assess image quality in an objective way. Applying the model observer to more realistic diagnostic images based on anthropomorphic phantoms or real patients will be one of the future applications to investigate.

FUNDING

The authors would like to warmly thank the Medical Imaging Perception Society (MIPS) for both scholarships that enabled the first author to attend the XIV and XV MIPS conferences. We would also like to acknowledge the Spanish Medical Physics Society (SEFM) for their funding and support under its international grant program. One of the authors RMSJ held a research grant from Toshiba Medical Systems.

ACKNOWLEDGMENTS

We would like to thank Maria Cros and Ramon Casanovas, from the Unitat de Física Mèdica at the Universitat Rovira i Virgili for their help in scoring the images in the human observer study.

REFERENCES

1. Brenner DJ, Hall EJ. Computed tomography–an increasing source of radiation exposure. *N Engl J Med* 2007; **357**: 2277–84.
2. Berrington de González A, Mahesh M, Kim KP, Bhargavan M, Lewis R, Mettler F, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. *Arch Intern Med* 2009; **169**: 2071–7.
3. Bittencourt MS, Schmidt B, Seltmann M, Muschiol G, Ropers D, Daniel WG, et al. Iterative reconstruction in image space (IRIS) in cardiac computed tomography: initial experience. *Int J Cardiovasc Imaging* 2011; **27**: 1081–7. doi: 10.1007/s10554-010-9756-3
4. Leipsic J, Labounty TM, Heilbron B, Min JK, Mancini GB, Lin FY, et al. Adaptive statistical iterative reconstruction: assessment of image noise and image quality in coronary CT angiography. *AJR Am J Roentgenol* 2010; **195**: 649–54. doi: 10.2214/AJR.10.4285
5. Joemai RM, Veldkamp WJ, Kroft LJ, Hernandez-Giron I, Geleijns J. Adaptive iterative dose reduction 3D versus filtered back projection in CT: evaluation of image quality. *AJR Am J Roentgenol* 2013; **201**: 1291–7. doi: 10.2214/AJR.12.9780
6. Beister M, Kolditz D, Kalender WA. Iterative reconstruction methods in X-ray CT. *Phys Med* 2012; **28**: 94–108.
7. Imaging Performance and Assessment of CT scanners. 32 to 64 slice CT scanner comparison report version 14. ImPACT. Report 06013. NHS Purchasing and Supply Agency. NHS PASA, 2005.
8. International Commission on Radiation Units and Measurements. Receiver operating characteristic analysis in medical imaging. ICRU Report No. 79. Bethesda, MD: International Commission on Radiation Units and Measurements; 2008.
9. Ogden K, Huda W. Applications of AFC methodology in optimization of CT systems. In: Samei E, Krupinski E, eds. *Medical image perception and techniques*. New York, NY: Cambridge University Press; 2010. pp. 356–63.
10. Chesters MS. Human visual perception and ROC methodology in medical imaging. *Phys Med Biol* 1992; **37**: 1433–76.
11. Klein Zeggelink WF, Hart AA, Gilhuijs KG. Assessment of analysis-of-variance-based methods to quantify the random variations of observers in medical imaging measurements: guidelines to the investigator. *Med Phys* 2004; **31**: 1996–2007.
12. 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
13. Popescu LM, Myers KJ. CT image assessment by low contrast signal detectability evaluation with unknown signal location. *Med Phys* 2013; **40**: 111908. doi: 10.1118/1.4824055
14. 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
15. Eckstein MP, Abbey CK, Bochud FO. A practical guide to model observers for visual detection in synthetic and natural noisy images. In: Van Metter RL, Beutel J, Kundel HL, eds. *Handbook of medical imaging. Physics and psychophysics*. Vol. 1. Bellingham, WA: SPIE-The International Society for Optical Engineering; 2000. pp. 595–628.
16. Hernandez-Giron I, Geleijns J, Calzado A, Veldkamp WJ. Automated assessment of low contrast sensitivity for CT systems using a model observer. *Med Phys* 2011; **38**: S25–35. doi: 10.1118/1.3577757
17. Mori S, Endo M, Nishizawa K, Murase K, Fujiwara H, Tanada S. Comparison of patient doses in 256-slice CT and 16-slice CT scanners. *Br J Radiol* 2006; **79**: 56–61. doi: 10.1259/bjr/39775216
18. Reiser I, Nishikawa RM. Identification of simulated microcalcifications in white noise and mammographic backgrounds. *Med Phys* 2006; **33**: 2905–11.
19. Burgess AE. Prewhitening revisited. In: Kundel HL, ed. *Proceedings of SPIE 3340,*

medical imaging 1998: image perception, 55; 21 April 1998; San Diego, CA.
20. Burgess AE, Jacobson FL, Judy PF. Human observer detection experiments with mammograms and power-law noise. *Med Phys* 2001; **28**: 419–37.
21. Samei E, Badano A, Chakraborty D, Compton K, Cornelius C, Corrigan K, et al. Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report. *Med Phys* 2005; **32**: 1205–25.
22. Woolson RF. Comparison of two groups: t-tests and rank tests. In: *Statistical methods of analysis of biomedical data*. New York, NY: Wiley; 1987. pp. 172–87.
23. MacMillan NA, Creelman CD. Comparison (two-distribuion) designs for discrimination. In: *Detection theory: a user's guide*. Mahwah, NJ: Lawrence Eribaum Associates; 2005. pp. 165–85.
24. Prins N, Kingdom FAA. (2009) Palamedes: matlab routines for analysing psychophysical data. [Cited 27 May 2014] Available from: http://www.palamedestoolbox.org
25. Hervada Vidal X, Santiago Pérez MI, Vázquez Fernández E, Castillo Salgado C, Epidat 3.0: Programme for epidemiological analysis of tabulated data. *Rev Esp Salud Pública* 2004; **78**: 277–80. [Cited 27 May 2014] Available from: http://www.sergas.es/MostrarContidos_N3_T01.aspx?IdPaxina=62714
26. Karssemeijer N, Thijssen MAO. Determination of contrast-detail curves of mammography systems by automated image analysis. In: Doi K, Giger ML, Nishikawa RM, Schmidt RA, eds. *Digital mammography*. Amsterdam, Netherlands: Elsevier; 1996. pp. 155–60.
27. Klein SA. Measuring, estimating, and understanding the phsychometric function: a commentary. *Percept Phsychophys* 2001; **63**: 1421–55.
28. Kingdom FAA, Prins N. Psychophysics: a practical introduction. London, UK: Academic press imprint of Elsevier; 2010.
29. Tapiovaara MJ. Efficiency of low-contrast detail detectability in fluoroscopic imaging. *Med Phys* 1997; **24**: 655–64.
30. Fan J, Madhav P, Sainath P, Cao X, Wu H, Nilsen R, et al. Evaluation of low contrast detectability performance using the two-alternative forced choice method on computed tomography dose reduction algorithms. In: Abbey CK, Mello-Thoms CR, eds. *Proceedings of the 2012 SPIE medical imaging: image perception, observer performance, and technology assessment*. San Diego, CA: Proc. SPIE 8318, 2012. 86731F1–7.

4.4. Investigating the kVp influence in the detection of low contrast objects in CT phantom images with two model observers.

[IV] I. Hernández-Girón, A. Calzado, J. Geleijns, R. M. S. Joemai, W. J. H. Veldkamp.

Low contrast detectability performance of model observers based on CT phantom images: kVp influence.

Phys Medica 2015 Corrected proof in press (doi: 10.1016/j.ejmp.2015.04.012)

Abstract

This paper studies low contrast detectability (LCD) performance of two model observers in CT phantom images acquired at different kVp levels and compares the results with humans in a 2-alternative forced choice experiment (2-AFC). Images of the Catphan phantom with objects of different contrasts (0.5 and 1%) and diameters (2–15 mm) were acquired in an Aquilion ONE 320-detector row CT (Toshiba Medical Systems, Tokyo, Japan), in two experiments, selecting (80–100–120–135 kV) with fixed mAs and varying the mAs to keep the dose constant, respectively. Four human observers evaluated the objects visibility obtaining a proportion correct (PC) for each case. LCD was also analysed with two model observers (non-prewhitening matched filter with an eye filter, NPWE and channelized Hotelling observer with Gabor channels, CHO).

Object contrast was affected by kV, with differences up to 17% between the lowest and the highest kV. Both models overestimated human performance and were corrected by efficiency and internal noise factors. The NPWE model reproduced better the human PC values trends showing Pearson's correlation coefficients ≥0.976 (0.954–0.987, 95% CI) for both experiments, whereas for CHO they were ≥0.706 (0.493–0.839, 95% CI). Bland–Altman plots showed better agreement between NPWE and humans, being the average difference Δ and the range of the differences $\Delta \pm 2\sigma$ (σ, standard deviation) of $\Delta = -0.3\%$, $\Delta \pm 2\sigma = [-4.0\%, 4.5\%]$. For CHO, $\Delta = -1.2\%$, $\Delta \pm 2\sigma = [-10.7\%, 8.3\%]$. The NPWE model can be a useful tool to predict human performance in CT low contrast detection tasks in a standard phantom and be potentially used in protocol optimization based on kV selection.

Physica Medica xxx (2015) 1–10

Contents lists available at ScienceDirect

Physica Medica

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

Low contrast detectability performance of model observers based on CT phantom images: kVp influence

I. Hernandez-Giron [a, b, *], A. Calzado [c], J. Geleijns [b], R.M.S. Joemai [b], W.J.H. Veldkamp [b]

[a] Unitat de Física Mèdica, Universitat Rovira i Virgili (URV), Spain
[b] Radiology Department, Leiden University Medical Center (LUMC), The Netherlands
[c] Departamento de Radiología, Universidad Complutense de Madrid (UCM), Spain

ARTICLE INFO

Article history:
Received 31 December 2014
Received in revised form
16 March 2015
Accepted 18 April 2015
Available online xxx

Keywords:
Model observer
Computed tomography
Image quality
Low contrast
kV

ABSTRACT

This paper studies low contrast detectability (LCD) performance of two model observers in CT phantom images acquired at different kVp levels and compares the results with humans in a 2-alternative forced choice experiment (2-AFC). Images of the Catphan phantom with objects of different contrasts (0.5 and 1%) and diameters (2–15 mm) were acquired in an Aquilion ONE 320-detector row CT (Toshiba Medical Systems, Tokyo, Japan), in two experiments, selecting (80–100–120–135 kV) with fixed mAs and varying the mAs to keep the dose constant, respectively. Four human observers evaluated the objects visibility obtaining a proportion correct (PC) for each case. LCD was also analyzed with two model observers (non-prewhitening matched filter with an eye filter, NPWE, and channelized Hotelling observer with Gabor channels, CHO).

Object contrast was affected by kV, with differences up to 17% between the lowest and highest kV. Both models overestimated human performance and were corrected by efficiency and internal noise factors. The NPWE model reproduced better the human PC values trends showing Pearson's correlation coefficients 0.976 (0.954–0.987, 95% CI) for both experiments, whereas for CHO they were 0.706 (0.493–0.839). Bland–Altman plots showed better agreement between NPWE and humans being the average difference Δ and the range of the differences Δ±2σ (σ, standard deviation) of Δ 0.3%, Δ±2σ [4.0%,4.5%]. For CHO, Δ 1.2%, Δ± 2σ [10.7%,8.3%]. The NPWE model can be a useful tool to predict human performance in CT low contrast detection tasks in a standard phantom and be potentially used in protocol optimization based on kV selection.

Introduction

The use of computed tomography (CT), since its introduction in clinical practice in the 1970s, has been continuously increasing. The progressive technological evolution of these devices has expanded the range of indications for this medical imaging technique, and in parallel the number of scans performed per year has been growing. Dose concern still exists in CT, especially for certain indications and more vulnerable patients, like pregnant women or children [1]. Different approaches have been taken by the manufacturers to decrease the patient doses related to this practice, obtaining dramatic dose drops for some indications, without compromising the diagnostic information [2,3]. Together with technical improvements, it is essential to optimize the CT protocols to adapt them to the patient characteristics. One possible strategy is based in selecting the tube potential (kVp), depending on the patient size and application [3–6]. Object contrast is affected by kVp, as the attenuation coefficients of the materials are dependent with the X-ray beam energy. Many CT exams are performed using intravenous iodinated contrast agents, which show an increase in CT number at lower kVp [7,8]. There are many studies investigating how these properties can be used to reduce patient dose, selecting low kVp, without losing relevant diagnostic information, for high contrast objects. The effect of selecting low kVp on the detectability of low contrast objects is studied less frequently [9,10].

One of the parameters of interest in CT image quality is low contrast detectability (LCD), i. e.: the ability to differentiate between materials with similar attenuation properties. The detection of small objects in CT can be compromised by noise, especially if

* Corresponding author. Radiology Department, Leiden University Medical Center (LUMC), Albinusdreef 2, 2333 ZA Leiden, The Netherlands. Tel.: +34 660376634.
E-mail address: irene.debroglie@gmail.com (I. Hernandez-Giron).

http://dx.doi.org/10.1016/j.ejmp.2015.04.012
1120-1797/

their contrast is low. LCD is measured, in general, using phantom images, containing objects of different contrasts and sizes. These studies are frequently carried out with human observers scoring the images to determine the smallest object of the lowest contrast that they are able to detect. Human observer studies in CT are very complex to carry out and time consuming, due to the amount of CT acquisition or reconstruction parameters that can affect image quality and the high number of images to analyze. Besides, subjective studies can be affected by intra- and inter-observer variability, which can be wide [11,12].

Model observers are an objective alternative, to analyze certain parameters related to image quality, attempting to mimic human performance. These models have been applied to lesion detection and discrimination tasks in medical imaging, in uniform and anthropomorphic backgrounds, considering different types of noise [13–15].

The use of model observers has been spreading in the past few years in CT. In particular, the channelized Hotelling model observer (CHO) and the non prewhitening matched filter with an eye filter (NPWE), which were selected for this study. There exist different modified versions of them in the literature, which have been validated for different detection and discrimination tasks in CT phantom images [16–20].

In this work, the effect of selecting different kVp levels in the detectability of low contrast objects has been investigated in phantom images, analyzing different object sizes and contrast levels. A human observer study was carried out to investigate the LCD response depending on the selected tube voltage. Two model observers, CHO and NPWE were used to analyze the same sets of images. There are few studies in which model observer performance has been addressed in low contrast detection tasks for different kVp in CT. Finally, the suitability of both models to reproduce human low contrast detectability performance as a function of tube voltage was addressed.

Materials and methods

Image acquisition and reconstruction

The Catphan phantom (Phantom Laboratories, New York, NY), which was selected for this work, is used in image quality assessment in CT and consists of different modules. In particular, the one used in this study, the lowcontrast module (CTP 515) contains three series of 9 cylindrical rods each, with diameters ranging between 2 and 15 mm and three contrast levels (1%, 0.5% and 0.3%), respectively, as shown in Fig. 1. According to the manufacturer specifications, these contrast levels, expressed as percentages, are defined as the difference, in Hounsfield units (HU), between the mean pixel value measured on a ROI placed in the 15 mm object and a nearby background region of equivalent size, divided by 10. All the objects for each contrast series are cast on the same material. The objects for 0.5% and 1% contrast were analyzed in this study.

Images of the Catphan 500 phantom were acquired with a 320-detector row CT scanner (Aquilion ONE; Toshiba Medical Systems, Tokyo, Japan) in two different experiments, for the whole range of available kVp values in the scanner (80–100–120–135 kV). In both, the phantom was scanned selecting 80 × 0.5 mm as beam collimation and helical acquisition (pitch, 0.828). Then the images were reconstructed using a filtered back projection method (QDS+), for a field of view of 210 mm, with 5 mm slice thickness, applying FC02 soft body kernel, which enhances low contrast frequencies in the images.

In experiment 1, the goal was to investigate the influence of kVp in the contrast to noise ratio, and in LCD, for the human observer and the models. The tube charge per rotation was fixed, selecting

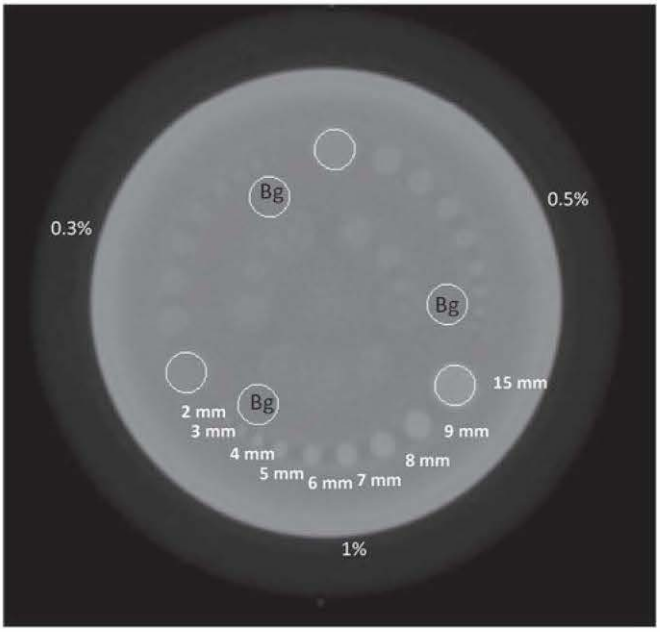


Figure 1. Distribution of the low contrast objects in the Catphan phantom shown on a CT image acquired at 120 kV ($CTDI_{vol}$ of 24.4 mGy) and averaged over 80 images.

200 mA and 1 s as rotation time, with kVp as the only varying parameter, in the range 80–100–120–135 kV. To verify the CT dose index ($CTDI_{vol}$) dosimetric values displayed in the console, and to analyze the dependence between dose and kVp for this particular CT device, dose measurements were previously performed within a body phantom (32 cm diameter), cast on PMMA, using a 10 cm long Capintec CT ionization chamber, connected to a Keithley 35050A electrometer, both treaceable to the National Physical Laboratory.

In experiment 2, the goal was to investigate how object contrast (C), contrast to noise ratio (CNR), and LCD were affected by selecting different kVp values, but keeping the dose level, and subsequently, image noise, constant. For this task, the tube charge per rotation (mAs) was varied for each series, to achieve $CTDI_{vol}$ values as similar as possible for all the acquisitions. The selected values, for 80–100–120–135 kV, were 500 mAs, 262.5 mAs, 165 mAs and 120 mAs, respectively.

In both experiments, the phantom was scanned 20 times, for each of the selected kVp values. The images in the boundaries of the low contrast module were discarded to avoid possible artifacts caused by nearby modules, saving the four central images from each scan. Thus, for each kVp and experiment, 80 phantom images were available to be analyzed by the human observers and the selected model observers.

For the detection tasks involved in this work, samples with signal present or absent were extracted from the Catphan images, using an in house software, implemented in Matlab (MathWorks, Natick, MA). The objects were located based on a mask of templates, created using the phantom manufacturer specifications regarding the size, shape and position of the objects. The templates were blurred to model the modulation transfer function (MTF) in each case. The MTF was obtained as the full width at half maximum of the point spread function (PSF) and measured with a phantom containing a 0.18 mm diameter tungsten bead. The samples of the background (signal absent) were taken from specific locations (Fig. 1), to avoid the inclusion of the so called supra-slice objects (inner circle of objects in Fig. 1) which were not used in this study.

The extracted samples had the same size (2.5 × 2.5 cm^2) independently of object diameter. The distribution of the objects in the

Please cite this article in press as: Hernandez-Giron I, et al., Low contrast detectability performance of model observers based on CT phantom images: kVp influence, Physica Medica (2015), http://dx.doi.org/10.1016/j.ejmp.2015.04.012

Catphan phantom is a limitation, as nearby objects could be included inside the samples. A correction was applied to the acquired images, to wipe out, from each sample containing an object, the nearby objects in its corners, as follows [18]. The mean pixel value was measured on the 15 mm object and in an annular shaped region of interest (ROI) around it, of the same area, for the two analyzed contrast series. Based on the signal difference between both regions (measured contrast), artificial signals were created and, after being blurred by the measured MTF, subtracted from the samples containing signals with the exception of the object of interest in each case. Examples of the appearance of the samples before and after this correction are depicted in Fig. 2 for all the kVp values and both experiments.

The dependence of C and CNR with kVp, was analyzed for the 0.5% and 1% contrast groups and all the image series, for both experiments. Object contrast was measured as the difference between the mean pixel value inside a ROI taken on the 15 mm object and an identical size area in the background sample (see Fig. 1). The noise was estimated as the statistical deviation of pixel values in the background samples for each kVp series. The CNR was calculated dividing the measured contrast by pixel noise. This parameter was measured for each contrast group in the 80 images related to each kVp, and averaged. To study the impact of selecting different conditions (kV value for experiment 1, kV and mAs value for experiment 2) in the measured C and CNR in the acquired images, Wilcoxon matched pairs signed rank tests were performed (significant differences for C or CNR between kV series, if p-value ≤ 0.05) [21].

Additionally, these measures were repeated on the extracted samples from the images, used for the human and the model observer studies, to check that the wipe out correction applied to delete the nearby objects did not affect CNR significantly.

Human observer study

The human observer study is based on a two alternative forced choice (2-AFC) detection experiment, in which the observer compares two classes of images, one with the object present (g_1) and another with the object absent (g_2), to a given template of the object, and reaches a decision.

To perform this 2-AFC study, an application was developed in Matlab (MathWorks, Natick, MA) which showed, on a grey canvas, the pairs of images, with or without signal, side by side, together with a template of the expected signal, for that particular task, on top of them [18].

This is a signal known exactly, background known exactly (SKE/BKE) experiment, as the location of the object, in the center of the samples in this study, is known beforehand by the observers. The images with object present or absent, were randomly displayed at left or right in the interface and also shuffled from the original folder. Proportion correct (PC) values were obtained for each observer and task, as the quotient between the number of correct decisions and the total number of images analyzed for each case.

Our work analyzes the influence of different parameters in low contrast detectability: object size (9 diameters), object contrast (2 contrast groups) and tube voltage (4 kVp values) which makes the human observer study quite arduous to carry out. To make this task more affordable, it was decided to analyze in this part of the study only a selection of the acquired images. Thus, from the 80 images available for each kVp and experiment, each human observer evaluated 50 images, scoring all the objects and contrast groups, in independent 2-AFC tests, twice to analyze human observer variability. In this way, each observer scored 50 images by 9 diameters by 2 contrast values by 4 kVp levels by two, which makes a total of 7200 pairs of images for both experiments. Four medical physicists participated in the tests for experiment 1. For experiment 2, the two more experienced observers, with several years of practice in image quality assessment in CT, analyzed the images.

The image scoring was performed in an i-MAC 27" DICOM calibrated monitor, according to recommended visualization conditions in a darkened room, with a fixed distance between observer and monitor of 50 cm. Images were displayed selecting as window level (WL), the measured mean pixel value in the samples containing the 15 mm object for each kVp series, and the window width (WW) was taken as 4σ, being σ the statistical deviation of the pixel values. The tests were performed in different sessions, over several days, lasting no longer than 2 h each, to reduce fatigue. All the observers scored all the images twice, with a gap of at least two weeks, to avoid learning effects [18].

The average human observer performance was obtained as the mean of the PC values for each object diameter, contrast, and kVp analyzed in this study. This parameter can be used as an estimator of the area under the ROC curve (AUC). The AUC can be related to a detectability index d' in a 2-AFC experiment [22]:

$$d' = 2\,\mathrm{erf}^{-1}[2(\mathrm{AUC}) - 1] \quad (1)$$

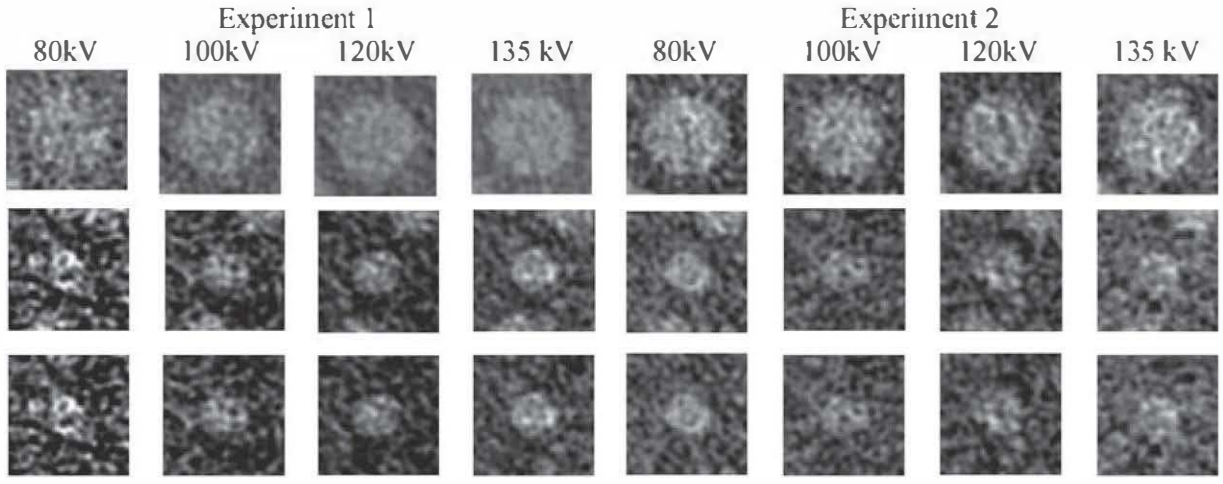


Figure 2. Samples extracted from the Catphan phantom CT images, containing the 15 mm object (top row of images) and the 8 mm object (central row), both with 1% contrast. The bottom row shows the latter, after applying the correction to wipe out the nearby objects. The displayed WL and WW are the same used for the human observer study.

Please cite this article in press as: Hernandez-Giron I, et al., Low contrast detectability performance of model observers based on CT phantom images: kVp influence, Physica Medica (2015), http://dx.doi.org/10.1016/j.ejmp.2015.04.012

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-x^2} dx \quad (2)$$

Finally, human LCD performance was estimated as curves representing either PC or d′ as a function of object size for each contrast level or kVp, in both experiments. The intra-observer variation was estimated performing McNemar analyses and the inter-observer agreement was analyzed calculating Fleiss' kappa for experiment 1 and Cohen's kappa for experiment 2 [21].

Model observers

Model observers performance, in a 2-AFC detection test, is based on applying different transformations to the two classes of images analyzed (signal present, g_1, and signal absent, g_2) and calculating a scalar decision variable or test statistic, $r(g_i)$, being $i = 1, 2$ a subindex that represents each image class ($i = 1$ means signal present and $i = 2$ means signal absent). This decision variable can be expressed, in a general way, following the formulation of Abbey and Eckstein, as [22]:

$$r_i = \boldsymbol{w}(g_i) \quad (3)$$

where subindex i denotes the image class ($i = 1, 2$), and $\boldsymbol{w}$ is a scalar-valued function that transforms the image (g_i) and depends on the considered model.

For linear model observers, this $\boldsymbol{w}$ function can be described as a matrix of weights of the same size as the image (which we will call N^2) and it is also known as the observer template. Thus, Eq. (3) can be expressed for linear models, as [22]:

$$r_i = \boldsymbol{w}^t g_i = \sum_{n-1}^{N^2} w_n g_n \quad (4)$$

where $\boldsymbol{w}^t \boldsymbol{g_i}$ is an inner product between the column vectors $\boldsymbol{w}$ (template) and $\boldsymbol{g}$ (image).

Cross-correlations are performed with both classes of images and the template applying this equation and statistical analysis are worked out in the resulting sets. One of the figures of merit used in detection tests is the detectability index or d′, which can be calculated as [15].

$$d' = \frac{\langle r \rangle_1 - \langle r \rangle_2}{\sqrt{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2}} \quad (5)$$

where $<\cdot>$ refers to the mean of the decision variables, $\sigma(\cdot)$ is the standard deviation

In this study, detectability indices were calculated for each object diameter and contrast as a function of kVp. Finally, these detectability indices can be transformed into proportion correct (PC) using Eq. (6) [15]:

$$PC = 0.5 + 0.5\, erf\left(\frac{d'}{2}\right) \quad (6)$$

A. NPWE model observer

The NPWE model includes the human contrast sensitivity function, as an eye filter, into the template. From the different eye filters in the literature, the one proposed by Burgess was selected, given by $E(f) = fe^{bf}$, being f the spatial frequency and with b chosen such that E(f) peaked at 4 cycles per degree [23].

The template, for the NPWE model, is the expected signal filtered by the square of the eye filter. The same distance from the monitor used in the human study (50 cm), was considered in this case [17]. The samples extracted from the images (signal present or absent) and the template are filtered by the eye filter E, before performing the cross-correlations between the template and the images (Eq. (5)) to obtain the d' values.

To estimate the error bars related to the calculated d' values obtained with NPWE and CHO model observers a bootstrapping method was applied [18]. For this task, the correlations related to the signal present or signal absent sets related to each kVp and contrast, were saved and combined randomly, to calculate d' values. This process was repeated 100 times for each condition and thus an average d' value and the related standard deviation was calculated. The standard error of the mean for each case, was obtained as the quotient between the standard deviation and the square root of the number of iterations. These average d′ values were transformed into PC using Eq. (6), and the error bars were estimated by propagation of errors.

The efficiency (η) was the parameter selected to relate the performance of the human observer (d'_{human}) and the NPWE model. It was calculated applying Eq. (7) [22]:

$$d'^{\,2}_{human} = \eta \; d'^{\,2}_{NPWE} \quad (7)$$

Linear fits were performed applying a least-squares procedure, using the error bars related to the data as weights in both axis. Only the human detectability values below 3 (PC ≈ 0.98), were taken into account in this calculation (together with their equivalent model observer values), because the shape of the curve representing d' as a function of PC, saturates above this threshold in 2-AFC experiments [18]. From now on, the NPWE model values corrected by the efficiency will be called $NPWE_\eta$.

B. CHO model observer

The human vision process in the visual cortex can be described as multiple channels, each of them sensitive only to a narrow range of spatial frequencies. The CHO model observer uses channels to filter the data, before making a decision. The test variable is [19]:

$$r_i = \boldsymbol{w}_{CH}^t g_{i,ch} = \sum_{m-1}^{M} w_{CHO\,m} g_{ch\,m} \quad (8)$$

where the total number of channels is $\mathbf{M}$; $\boldsymbol{g_c}$ is the transformed image after being filtered by the channels; subindex i denotes the image class ($i = 1, 2$); and $\boldsymbol{w_{CHO}}$ is the template given by Eq. (9):

$$w_{CH} = \overline{K_c}^{-1}(\langle g_{1\,ch}\rangle - \langle g_{2\,ch}\rangle) \quad (9)$$

where $\overline{K_c}^{1}$ is inverse of the average of the covariance matrices of the signal present and absent classes after being filtered by the channel matrix, and $<g_{i\,ch}>$ represents the mean of each class after being filtered by the channels.

Gabor channels were selected for this work, using those proposed by Wunderlich et al., as shown in Eq. (10) [24]:

$$G_a(x,y) = \exp\left[-4(ln2)\left((x-x_o)^2 + (y-y_o)^2\right)\Big/\omega_s^2\right] \cdot \cos[2\pi f_c \times ((x-x_o)cos\theta + (y-y_o)sin\theta) + \beta] \quad (10)$$

Please cite this article in press as: Hernandez-Giron I, et al., Low contrast detectability performance of model observers based on CT phantom images: kVp influence, Physica Medica (2015), http://dx.doi.org/10.1016/j.ejmp.2015.04.012

where ω_s is the channel width, f_c is the central frequency and β is a phase factor.

The values selected to implement the channels (summarized in Table 1) were those used by Yu et al., which were an extension of those published by Wunderlich et al., by adding two extra channel passbands [19,24]. A total of 60 channels were created in this way. This model observer has been successfully applied to different detection and discrimination tasks in CT, over a limited range of object sizes, contrasts, and tube charge values but not for a kVp range [19,20,25].

Human observers can reach different decisions for the same images in repeated tests. To try to mimic this behaviour, that arise from the performance of the visual neuronal system, internal noise (α) can be applied to the model observer calculations. This parameter can be added to the model decision variable (r), as shown in Eq. (10) [19,22]:

$$r'_i = r_i + \alpha x \tag{10}$$

where the subindex i denotes the image class (i = 1, 2), x is a variable that follows a normal distribution with zero mean and a standard deviation σ, given by the square root of the variance of r_2 (signal absent class) and α is the internal noise value.

The calibration of α was performed for each contrast level as follows. One of the objects (3 mm diameter) was taken as a reference for each kV and experiment [19]. A range of α values, between 0 and 20, was applied to the calculations of d' for the CHO model. The d' values modified by α were transformed into PC and compared to the equivalent PC human values for the 3 mm object. The α value producing the closest PC to human value was stored. The most frequent α value for all kV and each contrast was applied to all the analyzed images. Thus, two α values were obtained, one for 0.5% contrast objects and another one for 1% contrast objects. From now on, the CHO model observer with internal noise will be called CHO_α.

Comparison between human and model observers

Bland–Altman plots were obtained to study the degree of agreement between the model observers and the humans, for the different detection tasks, using their related PC values [26]. The mean difference (Δ) and the range of the differences [Δ ± 2σ], where σ is the standard deviation of the differences, between the PC values obtained by the average human and NPWE or CHO (after being corrected by the efficiency and internal noise, respectively), were calculated.

Pearson's product-moment correlation coefficients (r), between human and the model observers PC scorings, were also calculated for each condition (perfect correlation if the absolute value of r = 1.0) [21].

Psychometric fits and visibility thresholds

Detectability profiles (PC as a function of object diameter) were obtained for each contrast and kVp based on the model observers and the human PC values, respectively. Psychometric fitting curves according to Eq. (11) were applied in each case [18]. The error bars related to each PC, were used as weights in the fitting process, based on a least-squares procedure.

$$PC = \frac{0.5}{1 + e^{-f\log\left(\frac{d}{\lambda}\right)}} + 0.5 \tag{11}$$

where d represents the object diameter and f and λ are the fitting parameters.

The range of the fitted curves runs between 0.5 (pure guessing) and 1 (certain detection) in this 2-AFC study. A visibility threshold of PC = 75% was determined, which tallies λ itself.

Results

Physical parameters

The measured $CTDI_{vol}$ with the body phantom in experiments 1 and 2, showed a good agreement with the values given by the CT console, with relative differences between 3.6 and 7.0%. The latter were higher than the measured values for all kV values. The $CTDI_{vol}$ was plotted as a function of kV and fitted, showing a power dependence ($CTDI_{vol} = A \cdot kV^B$), being A ($3 \cdot 10^{-5}$) for both measured and console values, and B, 2.73 and 2.77, respectively. The R-squared fitting values were 0.999 in both cases. As a reference, the $CTDI_{vol}$ values retrieved from the CT console related to each acquisition and experiment are included in Table 2. As expected, the mean pixel values were affected by kVp level, being for the 1% contrast 15 mm object of 24 HU, 44 HU, 54 HU and 59 HU (±2% in all cases) for 80–100–120–135 kV, respectively. The measured contrast values, noise and CNR values, for both experiments are summarized in Table 2. It can be seen that object contrast varied when changing the kVp value and that the higher measured contrast values were obtained with the lowest kVp (80 kV), followed by 120 kV, 100 kV and 135 kV, respectively for 1% and 0.5% low contrast objects. Selecting 80 kV produced in most cases, significant improvements in the measured contrast for the objects analyzed in this work (p-value < 0.05). The highest kV level (135 kV) produced the lowest contrast values in the images, being down to 12% and 17% lower than with 80 kV for 1% and 0.5% contrast, respectively.

Regarding the measured image noise, for 80–100–120–135 kV and experiment 1, it was of 5.5 HU, 3.7 HU, 3.0 HU and 2.7 HU. For experiment 2, as mAs was selected to obtain similar $CTDI_{vol}$ values independently of the selected kV, image noise was approximately constant for all kV (≈3 HU). The analysis of the CNR values in the samples extracted from the images, after applying the wipe-out correction, showed that the effect on the images was small, with CNR variations ≤5% in all cases.

Wilcoxon matched pairs signed rank tests were performed comparing the contrast and CNR measured in the acquired images at the different kV levels for each experiment and contrast level. Significant differences were found in most cases (p-value < 0.05), with some exceptions. For experiment 1 not significant differences were found for the following: 1% contrast (100–135 kV, p-value = 0.09 for contrast) and 0.5% contrast (80–120 kV, p-value = 0.21 for contrast and 120–135 kV, with p-value = 0.31, for CNR). For experiment 2, the series were: 1% contrast (100–135 kV,

Table 1
Parameters used in the implementation of the Gabor channels for the CHO model observer.

Channel passbands ω_s (cycles/pixel)	Central frequency f_c (cycles/pixel)	Orientation θ (rad)	Phase factor β (rad)
[1/128–1/64] [1/64–1/32] [1/32–1/16] [1/16–1/8] [1/8–1/4] [1/4–1/2]	3/256 3/128 3/64 3/32 3/16 3/8	0 2π/5 4π/5 6π/5 8π/5	0 π/2

Please cite this article in press as: Hernandez-Giron I, et al., Low contrast detectability performance of model observers based on CT phantom images: kVp influence, Physica Medica (2015), http://dx.doi.org/10.1016/j.ejmp.2015.04.012

Table 2
Measured contrast and contrast to noise ratio (CNR) averaged over the sets of 80 images obtained for both experiments and each of the analyzed kVp values.

Contrast group		Experiment 1				Experiment 2			
		80 kV	100 kV	120 kV	135 kV	80 kV	100 kV	120 kV	135 kV
1%	Contrast (HU)	9.5 ± 2.2%	8.7 ± 2%	9.0 ± 1.6%	8.4 ± 2%	9.3 ± 1.5%	8.4 ± 1.4%	8.9 ± 1.1%	8.2 ± 1.3%
	CNR	1.7 ± 2.5%	2.4 ± 2.5%	3.0 ± 2.1%	3.2 ± 2.1%	2.7 ± 1.7%	2.6 ± 2.1%	2.7 ± 1.5%	2.4 ± 1.7%
0.5%	Contrast (HU)	5.4 ± 3.7%	4.8 ± 3.1%	5.2 ± 2.9%	4.6 ± 3.3%	5.2 ± 2.8%	4.7 ± 2.4%	5.0 ± 2.3%	4.3 ± 2.7%
	CNR	1.0 ± 4%	1.3 ± 3.3%	1.7 ± 3.2%	1.7 ± 3.5%	1.5 ± 3%	1.3 ± 2.7%	1.5 ± 2.5%	1.2 ± 3.1%
	$CTDI_{vol}$ (mGy)	7.3	14.9	24.4	33.7	18.2	19.5	20.1	20.2

with p-value = 0.06 for contrast and p-value = 0.07 for CNR) and 0.5% contrast (80–120 kV, with p-value = 0.2 for contrast and p-value = 0.08 for CNR).

Human observer study

Human observer performance was affected by the selected kV in both experiments. For experiment 1, even though the highest dose level was related to the 135 kV, human PC values were not the highest for this series. It has to be noted that this trend was observed for both contrast series. Figure 3 summarizes the results of the human observer study for both experiments, depicting PC values as a function of object diameter for all the kVp values and object contrasts, after performing the psychometric fits. The left column represents the results from experiment 1 and the right one, those related to experiment 2, and the top and low row the 1% and 0.5% contrast results, respectively.

The intra-observer analysis, which was made for each object size, kV and contrast level individually showed no significant variations (p-value 0.05) between both sessions in most scorings. For experiment 1, a total of 282 series (9 diameters by 2 contrast levels by 4 kV values by 4 observers) were compared for both sessions and 2.5% of them (7 out of 282) showed significant differences (p-value < 0.05). These series were evaluated again by the observers. For experiment 2, performed by two observers, 2 out of 144 series (1.4%) did not pass the McNemar test and were repeated.

Regarding the inter-observer variability analysis, only the smaller diameters in the range (2–5 mm) were considered. For diameters 6 mm all observers obtained scorings near PC = 100% (Fig. 3) and they were not included in this analysis because Cohen's kappa can lead to unreliable results when the prevalence is high. In experiment 1, for 1% contrast, the Fleiss' kappa ranged between 0.830 (0.828–0.833, 95% CI) and 0.832 (0.829–0.834, 95% CI). For 0.5% contrast they varied between 0.706 (0.702–0.709, 95% CI) and 0.811 (0.808–0.813, 95% CI). In experiment 2, for 1% contrast, Cohen's kappa ranged between 0.875 (0.872–0.878, 95% CI) and 0.890 (0.897–0.893). For 0.5% contrast the values were in the interval 0.871 (0.868–0.874, 95% CI) and 0.888 (0.885–0.891, 95% CI).

Model observer results: CHO and NPWE

Both model observers obtained higher detectability index (d') values with increasing object size and contrast in both experiments.

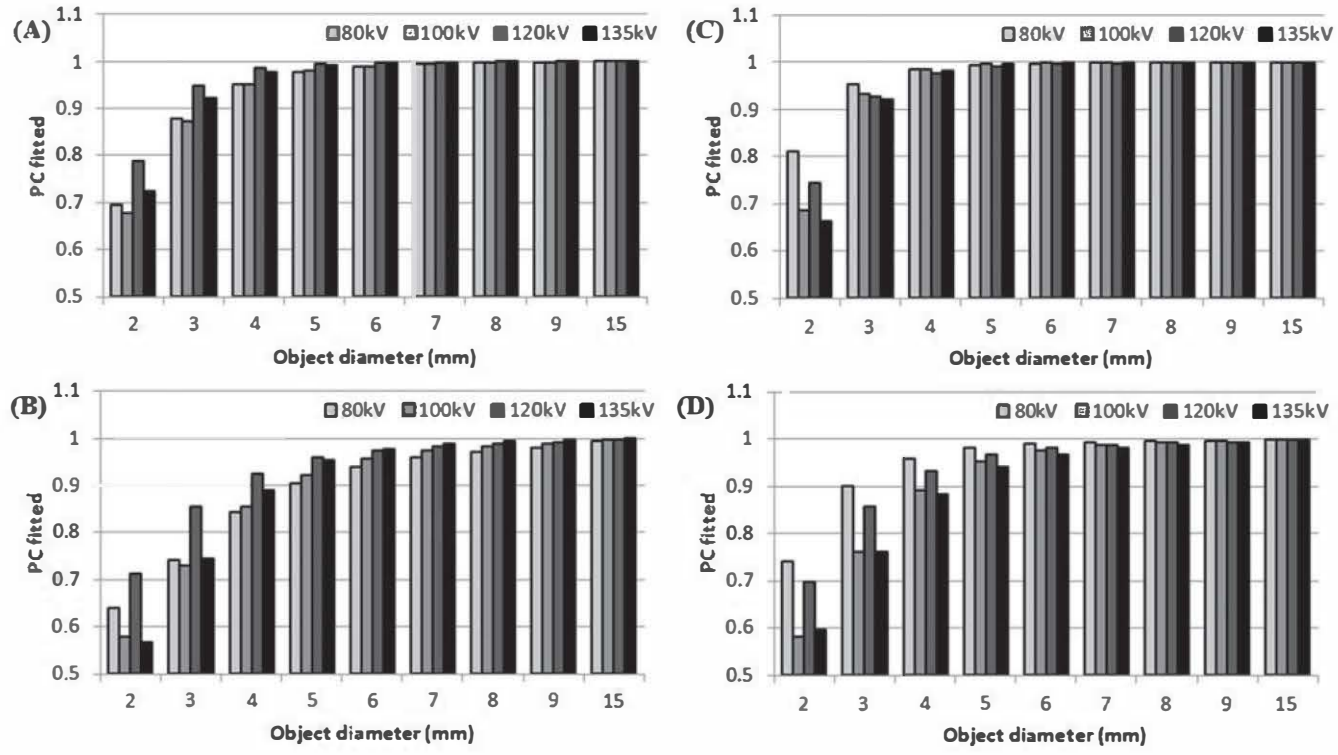


Figure 3. Human observers proportion correct (PC) as a function of object diameter and kVp values after the psychometric fits. The left column shows the graphs related to experiment 1 (A–B) and the right column to experiment 2 (C–D). The top row is related to 1% contrast and the bottom row to 0.5% contrast levels.

Please cite this article in press as: Hernandez-Giron I, et al., Low contrast detectability performance of model observers based on CT phantom images: kVp influence, Physica Medica (2015), http://dx.doi.org/10.1016/j.ejmp.2015.04.012

The CHO model obtained much higher d' values than NPWE for all the analyzed conditions. In experiment 1, in general, the detectability increased with kV for both models. The detectability values obtained by both model observers, before applying any efficiency or internal noise correction, in experiment 2 are depicted in Fig. 4. The left column corresponds to NPWE and the right column to CHO model. The first row is related to 1% contrast objects and the second row to 0.5% contrast.

Comparison between human and model observers

The selected model observers overestimated human performance in both experiments, especially the CHO model. Figure 5 shows the efficiency calculation process for the NPWE model, which tallies the slope of the linear fit. As previously stated, only d' below 3 (PC ≈ 0.98) for humans and the related model observers values were used. The d' values related to all kV values for both experiments and contrasts were included in this calculation. The efficiency for the NPWE model was 0.45 (0.42–0.47, 95% CI). This analysis was also performed considering each contrast level data separately, which gave efficiencies of 0.44 (0.41–0.46) for 1% contrast and 0.46 (0.43–0.49) for 0.5% contrast. Regarding the CHO model, it was modified adding internal noise (α) to the decision variable in the detection process of 4 and 6, for 1% and 0.5% contrast objects, respectively.

The models results were modified applying the efficiency ($NPWE_\eta$) or internal noise (CHO_α) values. The Pearson's product-moment correlation coefficients (r) between the average human and the modified model observers were the following, with 95% CI between brackets. For $NPWE_\eta$ and experiment 1, r was 0.976 (0.954–0.987) and 0.979 (0.954–0.987) for 1% and 0.5% contrast, respectively. For experiment 2, r was 0.986 (0.973–0.992) and 0.983 (0.967–0.991). For CHO_α model and experiment 1, r was 0.706 (0.493–0.839) and 0.861 (0.743–0.927) for 1% and 0.5% contrast. For experiment 2, r was 0.829 (0.689–0.910) and 0.818 (0.671–0.903).

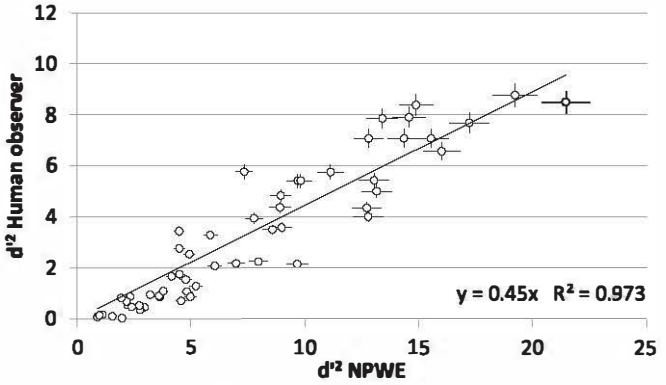


Figure 5. Squared detectability index (d'^2) for the average human observer as a function of the NPWE model observer. All the data related to experiments 1 and 2, 1% and 0.5% contrast levels and four kV values are included. The efficiency η is given by the slope of the linear fit.

Bland–Altman plots were performed comparing the PC values obtained by the average human observer and the models ($NPWE_\eta$ and CHO_α), considering all the kV values and both contrast levels and experiments, as depicted in Fig. 6. Both the average difference (Δ) and the range of the differences ($\Delta \pm 2\sigma$) were lower with NPWE than with CHO.

Psychometric fits

The visibility thresholds (λ) obtained with the psychometric fits for the human observer and both models are summarized in Table 3. It can be seen that smaller objects were detected when object contrast increased in all cases.

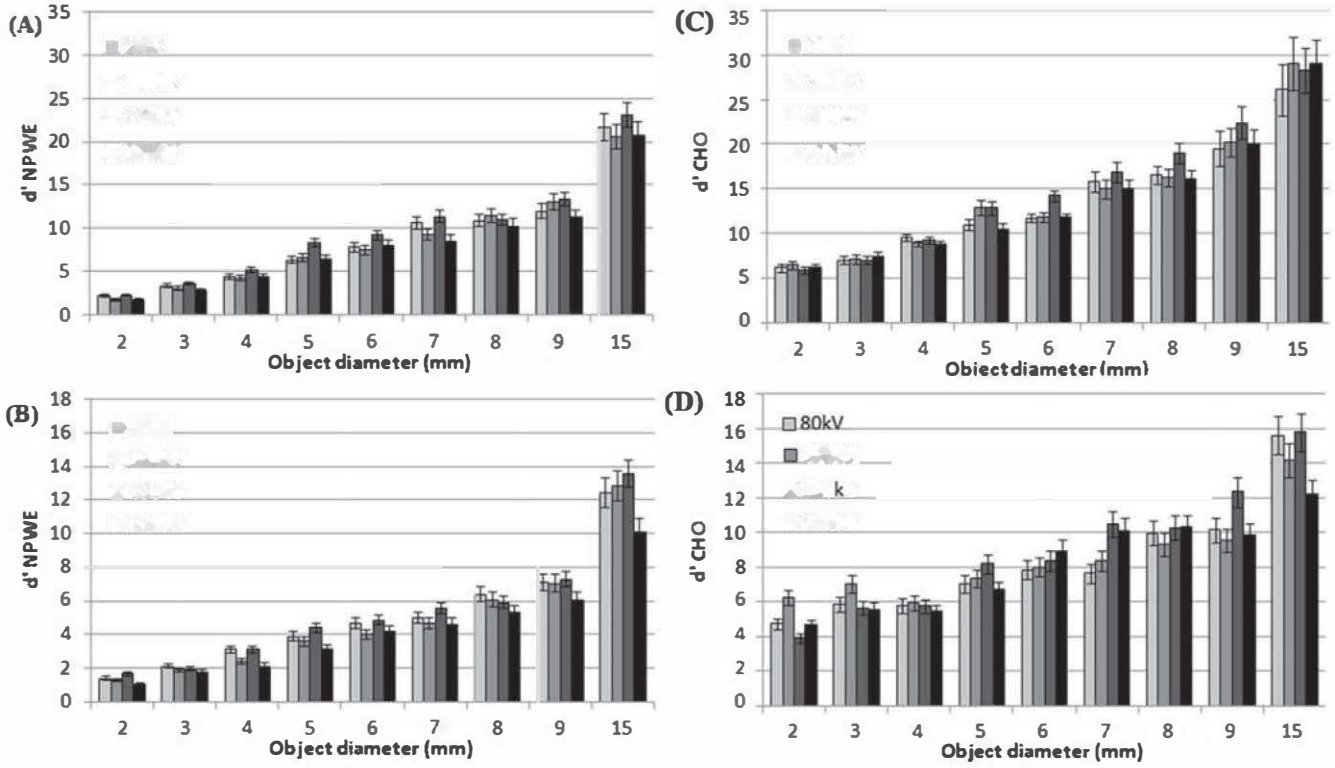


Figure 4. Detectability index (d') as a function of object diameter for 1% (top row) and 0.5% contrast groups (bottom row) and kV levels in experiment 2. The left column represents the NPWE model results (A–B) and the right column the CHO results (C–D).

Please cite this article in press as: Hernandez-Giron I, et al., Low contrast detectability performance of model observers based on CT phantom images: kVp influence, Physica Medica (2015), http://dx.doi.org/10.1016/j.ejmp.2015.04.012

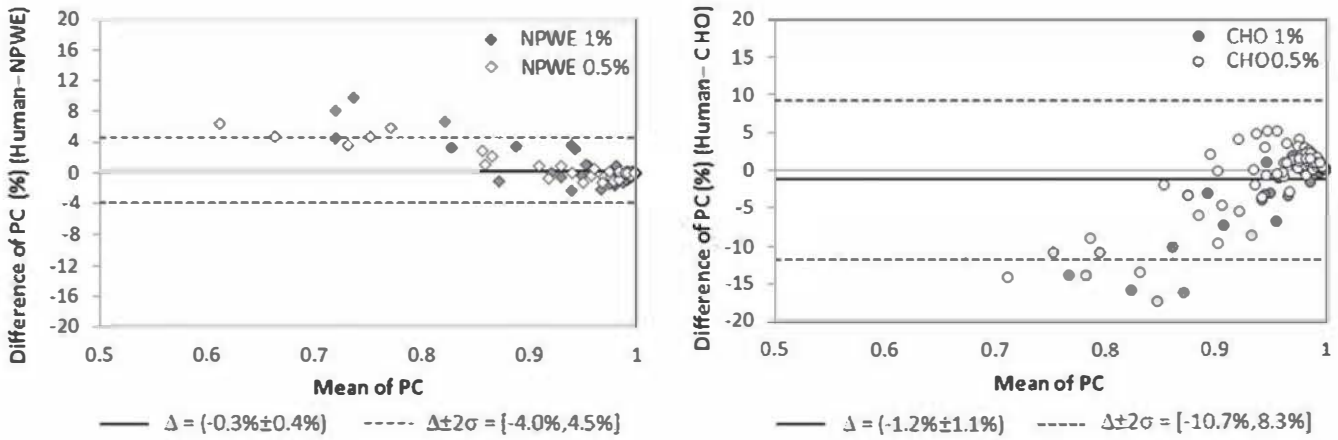


Figure 6. Bland–Altman plots of percent correct (PC) difference between human and NPWE model (A, left) and CHO model (B, right), for all contrast groups, kV and both experiments. The NPWE$_\eta$ model was corrected by the efficiency ($\eta = 0.45$) and the CHO$_\alpha$ model by internal noise ($\alpha = 4$ a.u. for 1% contrast, $\alpha = 6$ a.u. for 0.5% contrast). The straight lines represent the average difference, Δ, and the dash lines represent the range of the differences [$\Delta \pm 2\sigma$], where σ is the standard deviation of the differences.

Discussion

Image quality assessment in medical imaging has been traditionally performed by human observers. In the past few years, the application of model observers for these tasks has become more common in CT, because with them fast objective analysis of big sets of images can be performed.

Human observer studies can be very laborious and difficult to carry out, in particular in CT. For this work, analyzing each set of 9 diameters for a given kV and contrast value, took on average 12 min for each observer, after which a resting pause was performed. Considering the four kV values, two contrast levels, and the repetition of sessions (to check intra-observer variability), the study took around 5 h per observer (without taking into account the resting pauses) for each experiment. The computer calculations performed to obtain the equivalent results for the models and all the analyzed conditions, lasted less than 20 min per model, including the extraction of image samples, which took most of the time, using an ordinary computer. The intra-observer variability analysis showed that there were significant differences between the readings of both sessions for 2.5% and 1.4% of the scorings in experiments 1 and 2, respectively, which were related to small objects for which PC was close to chance. The inter-observer variability tests showed good agreement between the human observers for all the analyzed conditions in this study (Fleiss' kappa 0.710 (0.708–0.712, 95% CI) and Cohen's kappa 0.810 (0.807–0.813, 95% CI), for the analysis of each kV and contrast results separately).

The distribution of the objects in the Catphan phantom, packed closely together, makes it difficult to extract samples, in order to perform image quality assessment in an objective way. These geometrical characteristics are also present in other commercial phantoms (ACR CT phantom (Gammex, Middleton, WI)). To assess LCD with humans, the phantom images are shown to the observers, which can introduce a bias in the results as they know the distribution of the objects beforehand. To overcome this possible bias, the presented method to wipe-out nearby objects in the images seems a good alternative, as no significant differences in the measured contrast or CNR were found between the original and the corrected images. Some authors propose different phantom designs to overcome these issues [16,19].

The kV influence in the contrast and CNR was analyzed. Comparing the lowest and highest selected kV measurements, for experiment 1, image noise was two times lower and CNR was increased by a 1.8 factor with 135 kV, but the $CTDI_{vol}$ was more than 4.5 times higher (Table 2). Comparing the results for 120 kV and 135 kV, CNR was only 6.3% higher in the latter, whereas $CTDI_{vol}$ was a 38% higher. Using higher kV in the CT device increases the mean energy of the X-ray beam photons and also the photon intensity along the energy spectrum. The use of high kV settings is normally restricted to special cases, such as obese patients, when the high patient attenuation can lead to an unacceptable noise level in the images. The contrast of many human tissues relative to water decreases with high kVp [1]. In experiment two, when the dose level and thus image noise were kept approximately constant, it was seen that the contrast of the objects analyzed in this study was lower for 135 kV (12% for 1% contrast objects and a 17% for 0.5% contrast level, respectively). Different studies show that selecting lower kVp and increasing the tube charge per rotation can lead to

Table 3
Visibility thresholds (λ, related to PC = 75%) obtained with the psychometric fits of the average human observer, the NPWEη model (η = 0.45) and the CHOα model (α = 4 a.u. for 1% contrast, α = 6 a.u. for 0.5% contrast) for experiments 1 and 2.

Contrast group	Experiment 1				Experiment 2			
	80 kV	100 kV	120 kV	135 kV	80 kV	100 kV	120 kV	135 kV
λ(mm) Human observer								
1%	2.2 ± 1%	2.3 ± 2%	1.9 ± 2%	2.1 ± 1%	1.8 ± 1%	2.2 ± 2%	2.0 ± 1%	2.3 ± 2%
0.5%	3.0 ± 2%	3.1 ± 3%	2.2 ± 3%	3.0 ± 3%	2.0 ± 2%	2.9 ± 2%	2.3 ± 2%	2.9 ± 3%
λ(mm) NPWE$_\eta$								
1%	2.1 ± 5%	2.0 ± 2%	1.6 ± 2%	1.7 ± 2%	1.6 ± 2%	1.8 ± 2%	1.6 ± 2%	1.9 ± 3%
0.5%	2.5 ± 4%	2.8 ± 4%	2.0 ± 3%	2.1 ± 4%	2.1 ± 3%	2.3 ± 4%	2.0 ± 5%	2.6 ± 4%
λ(mm) CHO$_\alpha$								
1%	1.8 ± 7%	1.4 ± 6%	1.5 ± 9%	0.9 ± 6%	1.4 ± 8%	1.2 ± 7%	1.5 ± 6%	1.4 ± 7%
0.5%	2.7 ± 6%	2.1 ± 5%	2.7 ± 6%	1.9 ± 4%	1.8 ± 4%	2.3 ± 5%	3.5 ± 6%	2.4 ± 4%

Please cite this article in press as: Hernandez-Giron I, et al., Low contrast detectability performance of model observers based on CT phantom images: kVp influence, Physica Medica (2015), http://dx.doi.org/10.1016/j.ejmp.2015.04.012

important dose savings without compromising diagnostic information for thin patients or children [3–8]. In our study, the phantom size (20 cm) is a limitation to reproduce the effect of kV in image quality in the case of obese patients.

Regarding the human observer results, for experiment 1, it was found that with 80 kV and 120 kV higher PC values were obtained for both contrast groups, especially for small objects. With 135 kV, humans obtained poorer scorings even though the dose related to that study was 4.5 times higher than for 80 kV and 1.4 times higher than for 120 kV, respectively. For experiment 2, the highest scorings were obtained again with 80 kV and 120 kV which were the series related to higher contrast and CNR measured values. The psychometric fits of the human results (Fig. 3) showed that the visibility threshold (related to PC = 75%) decreased with object contrast.

Before applying a certain model observer to protocol optimization, it is necessary to check the level of agreement between the model and human observers, for the tasks of interest. It is frequent to compare human and model observer performance for a range of dose levels, varying the mAs value in the scanner [17–20,25]. In this work, a less trodden path was taken analyzing the effect of selecting different kV in low contrast objects using phantom CT images and studying the LCD performance of humans compared to two model observers.

The selected models, NPWE and CHO had been previously validated to perform simple detection tasks in phantom CT images for a range of dose levels, varying the mAs. The NPWE model reproduced the trends of the human observers depending on the selected kVp for both experiments. The Pearson's correlation coefficients showed good agreement between the human and this model, after being corrected by the efficiency ($\eta = 0.45$). The Bland–Altman plots showed good agreement between model and human observer (Fig. 6). Even after applying the efficiency correction, the model overestimated human performance, as it detected smaller objects (Table 3). The η value calculated in this study is similar to those published for this same model ($\eta \approx 0.5$) applied to different detection tasks [15,18]. There is still leeway to tune this model to improve its agreement with human performance. There are other eye filters in the literature that can be implemented and the addition of internal noise is also possible. The choice of efficiency for NPWE seems appropriate for this particular study and offers the advantage of being easy to calculate.

The selected version of the CHO model showed poorer agreement than NPWE with human results. For experiment 2, it did not reproduce the human trends, especially for small objects and 0.5% contrast. The Pearson correlation coefficients were lower than those obtained with for NPWE and the Bland Altman plots showed a bias, as the model, in general, overestimated human performance for lower PC values and underestimated it for higher PC values (Fig. 6). This is also reflected in the visibility thresholds, as the CHO model detected smaller objects than NPWE and humans (Table 3). Even without applying any correction, this model did not accurately reproduce the general trends with kV obtained in the human study (Figs. 3 and 4). Other versions of the CHO model have been published, considering different sets of channels or including a border detection feature to detect the objects in the images and they can be possible alternatives [22,25]. The wide range of object sizes, different contrast levels and kV settings considered made it difficult to fix an appropriate internal noise level for all the conditions. Some studies propose to add different levels of noise, depending on the lesion size [25].

If a model observer can predict human performance, then it can be used on further image quality tests, assuming that the model will reproduce the human results. This assumption has to be taken carefully, because if the model has to be applied to more complicated tasks, for example, lesion detection in real patient images, other studies are necessary to analyze the correlation with human observers.

Conclusion

We have studied the influence of selecting different kV in the measured contrast and the contrast to noise ratio for low contrast objects in CT phantom images. For the study, images of simple objects (disks) of different sizes and contrasts, which were embedded in a uniform background, were extracted from the phantom images acquired at different kV levels and analyzed. Human observer LCD performance was improved with 80 kV compared to the other kV values, when dose was kept constant, for a 20 cm diameter phantom. The non-prewhitening matched filter with an eye filter (NPWE) model observer reproduced the human trends for all the range of kV considered, though it overestimated human performance. The version of the channelized Hotelling observer (CHO), implemented with a particular set of channels did not reproduce the human performance to the same stand. There is still leeway to tune the models to human performance, investigating different eye filters in the case of NPWE, or channel definition for the CHO model.

Further research has to be done before using these models for clinical practice applications. An intermediate step could be performing lesion detection tests in anthropomorphic phantoms, which reproduce the patient anatomy more realistically.

Acknowledgments

We would like to thank Maria Cros, from the Unitat de Física Mèdica at the Universitat Rovira i Virgili and Maria Castillo from the Departamento de Radiología at the Universidad Complutense de Madrid for their help in scoring the images in the human observer study.

Abbreviations

AUC	area under the ROC curve
C	contrast
CHO	channelized Hotelling model observer
CNR	contrast to noise ratio
$CTDI_{vol}$	volumetric computed tomography dose index
d'	detectability index
d	object diameter
HU	Hounsfield units
LCD	low contrast detectability
mAs	tube charge per rotation
MTF	modulation transfer function
NPWE	non-prewhitening matched filter with an eye filter model observer
PC	proportion correct
PSF	point spread function
r	Pearson's product-moment correlation coefficients
ROI	region of interest
SKE/BKE	signal known exactly and background know exactly experiment
2-AFC	2-alternative forced choice experiment
WL	window level
WW	window width
Δ	mean difference in Bland–Altman plot
$[\Delta \pm 2\sigma]$	range of the differences in Bland–Altman plot
η	efficiency
λ	visibility threshold
σ	statistical deviation

Please cite this article in press as: Hernandez-Giron I, et al., Low contrast detectability performance of model observers based on CT phantom images: kVp influence, Physica Medica (2015), http://dx.doi.org/10.1016/j.ejmp.2015.04.012

References

[1] Tack D, Mannudeep KK, Genevois PA, editors. Radiation dose from multi-detector CT. 1st ed. Heidelberg: Springer-Verlag Berlin; Heidelberg; 2012.
[2] Beister M, Kolditz D, Kalender WA. Iterative reconstruction methods in X-ray CT. Phys Medica 2012;28:94–108.
[3] Schindera ST, Diedrichsen L, Müller HC, Rusch O, Marin D, Schmidt B, et al. Iterative reconstruction algorithm for abdominal multidetector CT at different tube voltages: assessment of diagnostic accuracy, image quality, and radiation dose in a phantom. Radiology 2011;260:454–62.
[4] Winklehner A, Goetti R, Baumueller S, Karlo C, Schmidt B, Raupach R, et al. Automated attenuation-based tube potential selection for thoracoabdominal computed tomography angiography: improved dose effectiveness. Invest Radiol 2011;46:767–73.
[5] Noël PB, Köhler T, Fingerle AA, Brown KM, Zabic S, Münzel D, et al. Evaluation of an iterative model-based reconstruction algorithm for low-tube-voltage (80kVp) computed tomography angiography. 033501 J Med Imaging 2014;1:1–7.
[6] Mathieu KB, Turner AC, Khatonabadi M, McNitt-Gray MF, Cagnon CH, Cody DD. Varying kVp as a means of reducing CT breast dose to pediatric patients. Phys Med Biol 2013;58:4455–69.
[7] Marin D, Nelson RC, Barnhart H, Schindera ST, Ho LM, Jaffe TA, et al. Detection of pancreatic tumors, image quality, and radiation dose during pancreatic parenchymal phase: effect of a low-tube-voltage, high-tube-current CT technique–Preliminar results. Radiology 2010;256:450–9.
[8] Lee KH, Lee JM, Moon SK, Baek JH, Park JH, Flohr TG, et al. Attenuation-based automatic tube voltage selection and tube current modulation for dose reduction at contrast-enhanced liver CT. Radiology 2012;265:437–47.
[9] Funama Y, Awai K, Nakayama Y, Kakei K, Nagasue N, Shimamura M, et al. Radiation dose reduction without degradation of low-contrast detectability at abdominal multisection CT with a low-tube voltage technique: phantom study. Radiology 2005;237:905–10.
[10] Nakaura T, Nakamura S, Maruyama N, Funama Y, Awai K, Harada K, et al. Low contrast agent and radiation dose protocol for hepatic dynamic CT of thin adults at 256-detector row CT: effect of low tube voltage and hybrid iterative reconstruction algorithm on image quality. Radiology 2012;264:446–54.
[11] Klein Zeggelink WFA, Hart AAM, Gilhuijs KGA. Assessment of analysis-of-variance-based methods to quantify the random variations of observers in medical imaging measurements: guidelines to the investigator. Med Phys 2004;31:1996–2007.
[12] 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 Medica 2013;29:99–110.
[13] Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. Proc Natl Acad Sci U. S. A 1997;90:9758–65.
[14] Gifford HC, King MA, de Vries DJ, Soares EJ. Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging. J Nucl Med 2000;41:514–21.
[15] Reiser I, Nishikawa RM. Identification of simulated microcalcifications in white noise and mammographic backgrounds. Med Phys 2006;33:2905–11.
[16] 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.
[17] 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.
[18] 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.
[19] 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.
[20] 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.
[21] Woolson RF. Statistical methods for the analysis of biomedical data. 1st ed. New York: John Wiley & Sons, Inc; 1987.
[22] Samei E, Krupinski E, editors. The handbook of medical image perception and techniques. 1st ed. New York: Cambridge University Press; 2010.
[23] Burgess AE, Jacobson FL, Judy PF. Human observer detection experiments with mammograms and power-law noise. Med Phys 2001;28:419–37.
[24] Wunderlich A, Noo F. Image covariance and lesion detectability indirect fan-beam x-ray computed tomography. Phys Med Biol 2008;53:2471–93.
[25] 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.
[26] Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. J Biopharm Statistics 2007;17:571–82.

Please cite this article in press as: Hernandez-Giron I, et al., Low contrast detectability performance of model observers based on CT phantom images: kVp influence, Physica Medica (2015), http://dx.doi.org/10.1016/j.ejmp.2015.04.012

Discussion

The discussion is divided in two sections. The first section consists on a general discussion on the papers which constitute the core of this thesis. The second section dives separately into some subjects studied in the thesis, emphasizing the comparison and links with related studies in the literature.

1. General discussion

There is a need for automated methods for the analysis of image quality, especially in modalities such as CT in which many acquisition and reconstruction parameters, that in turn can take a range of values, affect image quality.

The main contribution of this PhD thesis is to have developed a framework for the objective assessment of low contrast detectability in computed tomography in phantom images, applying model observers. The validation of the model observers' results is based on human observers performance obtained with tools developed during this thesis. Low contrast detectability is of interest in CT as it takes into account the frequency dependencies of contrast and noise, including the imaging system blur.

At the beginning of this thesis, model observers had been applied in other medical imaging modalities, such as mammography, SPECT and in simulated objects embedded in CT images[1,28,46,50-52]. The proposed method can be applied as an alternative to human observers in simple detection and discrimination tasks in CT phantom images, analysing the influence of reconstruction and acquisition parameters in low contrast detectability in a fast and objective way.

The first step to achieve the proposed milestones in the thesis was to explore which model observers had been investigated for the assessment of image quality in medical applications up to date. Among them, the NPWE model observer was considered a good candidate to predict human performance as it had been previously applied to detect objects in images with mammographic and simulated backgrounds[46].

The model implementation was validated in **[I]**. For this task, Gaussian white noise background images were simulated embedding objects of different diameters in them, and the detectability results were successfully benchmarked with those published in the literature for the same model and similar experiments[46]. White noise is uncorrelated, in both the spatial and the frequency domains, which means that the noise level is independent of the image frequencies. There are many studies analysing human performance in detection and discrimination tasks in this type of backgrounds, as a function of the noise level[1,28,27,46].

The NPWE model observer was integrated in an in-house software which was developed to automatically assess the low contrast detectability of objects in images of a particular phantom (Catphan) widely used in image quality assessment in CT. The software is divided in two parts. The first one is focused on the automatic subtraction of samples from the phantom (or simulated) images. A template of the objects distribution in the

phantom is created and automatically fitted to the acquired phantom images. Once the objects are located, samples are cropped, containing either each object or the surrounding background, which completes the first goal of this thesis. In the second part of the software, the selected model observer is used to analyse the samples for each object diameter and contrast level and detectability indexes are calculated.

Images of the phantom were acquired in a CT scanner at different dose levels, varying the tube charge per rotation. The effect of object contrast, diameter and tube charge per rotation in the NPWE low contrast detectability performance was investigated. LCD increased as a function of object diameter, object contrast and dose, as expected. The model observer performance was not compared with human observers at this stage.

In the second paper of this thesis **[II]**, LCD performance in a different CT scanner was analysed as a function of a range of acquisition (kV and mAs) and reconstruction parameters (selected reconstruction filter) applying the model observer to images of the same phantom. To validate the trends obtained with the model observer, a human observer study was carried out. The observers had to assess the number of visible objects in the phantom images, which is the most frequent way to assess LCD. As the distribution of the objects in the phantom is usually known beforehand, a bias in the outcomes can exist. A careful setup is needed, including the selection of observers, their training for the task and dividing the study in different sessions to avoid fatigue. The visualization environment and the image display have to be suitable for medical imaging assessment and the observers carefully selected, having an adequate experience for the intended task. The detection tasks were not exactly the same for the model and the human observers in this study, and thus their results could not be quantitatively compared. The NPWE model reproduced the human performance trends, showing an improvement in LCD as a function of increasing object diameter, contrast, kV, mAs and for *soft* reconstruction kernels. Up to this point, goals **1–3** had been accomplished.

The next milestone (**4**) was to develop the necessary tools to perform 2-AFC experiments with human observers to enable a quantitative comparison with the model observer results. For this, an in-house software was created to display the images in pairs, one with signal present and another without it. The scoring of each observer was automatically stored in an output file and proportion correct values calculated for each experiment. With the proposed setup the bias that can appear when humans know beforehand the distribution of objects in the phantom is overcome. The detection task is transformed into a 2-AFC experiment in which the visibility of each object is determined individually. The software that was designed can also be adapted to perform M-AFC studies and score images which can be either simulated or from other medical imaging techniques.

The focus of the third paper of this thesis **[III]** was to analyse the influence of using iterative reconstruction algorithms in low contrast detectability compared to FBP for a range of dose levels, with a model observer and humans. The human observer study was carried out using the 2-AFC software. The model obtained higher LCD scores than the human observers. An efficiency factor was calculated to normalize the model results to humans, which was similar to other published in the literature ($\eta \approx 0.5$)[46,53,54]. Model and humans showed a high correlation in their performance, checked with Pearson's correlation coefficients and Bland-Altman plots. They also showed the same trends, with higher detectability values with increasing object diameter, contrast and dose. LCD was

improved with the iterative algorithm compared to FBP, especially for low dose and low contrast objects. This study corroborated previous results where the NPWE model was applied to a smaller set of images of the Catphan phantom acquired in the same CT scanner[55]. With this study, goal **(6)** was covered.

Planning and performing human observer studies is laborious and complex. The amount of data analysed, considering the different object sizes and all the observers involved was considerable in **[III]** (more than 60000 pairs of images were scored). The observers scored the images in several sessions to avoid fatigue and learning effects, and it took several hours for each of them to end the study. It is well known that there can appear an intra- and inter-observer variability in human studies. These effects were assessed in **[III]** and **[IV]** performing Wilcoxon signed rank test for matched-pair samples for the intra-observer variability and Fleiss' kappa to investigate the inter-observer consistency[56].

The fourth paper of the thesis **[IV]** analysed the influence of kV in the detectability of low contrast objects, applying two model observers, NPWE and CHO to analyse CT phantom images (goals **3** and **5**). Images were acquired in two experiments, one in which kV varied in a range of values and so did dose, and another one in which dose was kept constant in all the acquired sets, modifying the tube charge per rotation. The CHO model was implemented with the same set of channels proposed by other authors which had been validated in simple detection tasks in CT phantom images[40]. The models were modified applying efficiency factors and internal noise. The results obtained with both models were compared with human outcomes in a 2-AFC study. The selected kV affected the LCD performance of the NPWE and the human observers in a similar way, with an improvement in detectability for the lowest kV. The NPWE model showed better correlation with LCD human performance than the CHO model with the selected set of channels for this particular task. Selecting lower kV values lead to an increase in LCD in the phantom images for both NPWE and human observers. The model observer results can be used to investigate in protocol optimization based on selecting lower kV and increasing the mAs, which can entail important dose reductions, especially for thin patients and children[21,57-61].

2. Discussion on the state of the art

2.1. Human observer studies

A careful setup is needed for human observer studies in image quality assessment. This includes the selection of observers, their training for the task and dividing the study in different sessions to avoid fatigue. The visualization environment and the image display have to be suitable for medical imaging assessment and the observers carefully selected, having an adequate experience for the intended task. Despite the complexity, expensiveness and all the aforementioned considerations about human observer studies, they are essential in image quality evaluation at the clinical stage.

About tackling the observers response variability, a pragmatic approach was selected, checking the intra-observer and inter-observer consistency applying statistical analysis, the Wilcoxon signed rank test for matched-pair samples in the first case and Fleiss' kappa in the latter. Then, the average human observer performance was calculated for each

analysed condition as the mean PC value with its related uncertainty and compared with the model observers results. There are different approaches to deal with human observers variability. Some of them based on analysis of variance (ANOVA) which estimate the number of images and observers needed to obtain robust results[62]. Other methods are based on the analysis of clustered binary data[40,63]. The variance of the AUC human response can be estimated also when not all the observers analyse every case, using bootstrapping and Monte Carlo simulations[64].

At the moment, model observers in the literature can to some extent predict human performance for simple detection and discrimination tasks, even involving object search[1]. Despite this, they are not sufficiently developed to mimic human performance in complex diagnostic clinical tasks. The radiologists diagnostic performance is inherently subjective and influenced by internal and external factors, such as experience or fatigue, which can affect the final decision. Current model observers do not incorporate all these aspects, though they emulate to some stand human visual perception and the decision making processes for simple tasks[65].

2.2. Model observers used in CT

There is not a gold standard regarding the model observer that can be used as a surrogate for humans for any low contrast detectability task in CT. The most appropriate model in each case is highly dependent on the task and the reconstruction algorithm.

Multiple options are available to implement the model observers used in this thesis. Their implementation can be done in the frequency or in the spatial domain. For the NPWE, there are several published contrast sensitivity functions, which are the base of the eye filter. The one selected had been successfully applied in other medical imaging modalities and it showed a good correlation with humans[1,43,46]. Regarding the CHO model, Gabor channels were implemented with a given set of parameters, which had also been validated in different detection and discrimination studies in CT[40]. Different settings for the number of channels, orientations and phases are described in the literature. As an alternative for Gabor channels, Laguerre-Gauss and difference of Gaussian channels (DDOG) have also been successfully applied in combination with the Hotelling observer for detection tasks in mammographic backgrounds, SPECT and CT images[50,66-68].

For CT, different studies have investigated LCD in phantoms with uniform backgrounds and simple objects, such as disks, in the past few years. The findings of different studies applying model observers to LCD in CT phantom images reconstructed with iterative algorithms, similar to **[III],** are discussed in section **2.3** of Discussion.

Considering the implementation in the frequency domain, the NPW and the NPWE have been successfully applied to detection experiments with hybrid images (embedding simulated objects in real CT images in Boedeker et al), with a commercial phantom (Christianson et al, using the ACR CT phantom) or testing a custom water based phantom (Ott et al)[51,69,70]. In the first study, researchers studied the influence of different reconstruction algorithms in LCD with the NPW model in simulated images of a sphere of a given diameter and found that the detectability depended on the selected algorithm and it also was improved with increasing tube charge per rotation[51]. The other two studies

analysed the effect of iterative reconstruction in LCD with model observers and are mentioned in the next section[69,70].

There are different versions of the CHO model for CT images in the literature [40,53,67,71,72]. In this thesis, one of this versions, based on 60 Gabor channels was used **[IV]**, as it had been successfully used in detection tasks in uniform backgrounds with a custom water based phantom simulating the thorax shape (Yu et al)[40].

LCD assessment in CT is normally based on the analysis of the detectability of objects with different diameters and contrast levels, for a range of dose levels. To calibrate the internal noise used to tune the model to human performance, in general, an intermediate size and contrast object is used[40,53]. Internal noise is added to the model decision variables until the output scores the same AUC as the human observer for that condition. That internal noise value is used to recalculate the model detectability results. In **[IV]** it was discussed that with this approach, the model can overestimate or underestimate human performance for other object sizes and contrast levels. Eck et al discussed four different methods for the internal noise addition and CHO model observer, and obtained better correlation with human observers when internal noise was proportional to the channel output standard deviations[68].

Different papers have compared the performance of NPWE and CHO for the same detection tasks in CT and in simulated images, emphasizing that the selection of a model that can accurately predict human performance is highly dependent on the intended task [1,27,72-75]. For detection or discrimination studies of simple objects in CT images of uniform phantoms, the NPWE and the CHO models have shown a good correlation with humans. Applying these models to more complex tasks or to lesion detection in anatomical backgrounds is still under study [III,IV], [40,53,73].

2.3. Iterative reconstruction algorithms in CT: Effect on low contrast detectability

Noise in CT is correlated (also called filtered or coloured noise), as its power is frequency dependent and thus it is highly affected by the selected reconstruction filter and method (FBP or iterative reconstruction). The effect of using iterative reconstruction algorithms in low contrast detectability is still under investigation. The noise level is decreased and CNR increased with the current IR algorithms used in CT, but as stated in the introduction, these metrics are not totally suitable to measure the noise spectrum changes and how the objects detectability can be affected.

The level of dose reduction that might be achieved when selecting IR instead of FBP reconstruction algorithms is not going to be discussed, as in the published studies on this subject, the tasks, algorithms, dose levels and phantoms used can differ greatly and thus, results are not directly comparable. Only research based on phantom images is addressed, as real patient images assessment was out of the scope of the goals of this dissertation.

In the literature, studies showing that LCD can be either deteriorated or improved with iterative reconstruction in CT can be found, which are discussed next. On the first category, Schindera et al acquired images of a custom phantom that simulated the attenuation of the liver parenchyma with a uniform material, containing spheres

simulating hypo-attenuating liver metastases of different sizes and contrast selecting an abdominal protocol[25]. An additional ring was placed around the phantom to mimic the attenuation in an average male patient. Images were reconstructed with FBP and selecting the IR of the manufacturer. The tube current was modified in steps, starting by the default selected value in the abdominal protocol and 1/5 of it. Radiologists scored the location of the objects in the images and their degree of conspicuity in a scale. Lesion detection was significantly lower when comparing the images reconstructed based on the original protocol and FBP and the IR reconstructed set with the lowest dose level.

Goenka et al performed a multi-reader study, based on images of an anthropomorphic liver phantom containing spherical lesions or varying size and contrast acquired at different dose levels and reconstructed with FBP and iterative algorithms. Radiologists assessed if the lesion was present or absent together with a scale of confidence. Comparing the scorings for the original FBP set, acquired with the usual dose level for this protocol, with the IR sets for lower doses, diagnostic accuracy dropped for dose reductions ≥50%[75].

Similar findings were obtained in McCollough et al, based on images of the commercial ACR CT accreditation phantom comparing the visibility of objects for one of the available diameters derived from radiologists' performance[26]. Images were acquired in two CT manufacturers' scanners for several dose values, reconstructed selecting FBP and different levels of their respective iterative algorithms. LCD diminished as the dose level decreased for IR and FBP. For strong dose reductions ≥25%, LCD was much worse for IR than FBP at the original dose.

In the second category, different studies have found that there is leeway for dose reduction when using IR algorithms without low contrast resolution loss, at least in CT phantom images with model and human observers. Similar findings were obtained in the results presented in this PhD thesis[IV,55]. Research papers have been grouped together depending on if they used just human observers (**1**), NPWE (**2**) or CHO (**3**) model observers to support their conclusions.

(1) Miéville et al studied the influence of the selected IR algorithm in LCD performing a 4-AFC human observer study using two paediatric phantoms containing objects of different sizes and contrast levels[76]. They found that LCD performance was maintained compared to FBP for one of the analysed reconstructions even at ultra-low doses.

(2) Regarding the use of the NPWE model observer, Ott et al implemented the NPW model to detect disks of different materials embedded in a custom made water phantom, finding an improvement in detectability with increasing dose, contrast and with iterative reconstruction compared to FBP[70]. The model observer results were not compared to human observer performance.

In Joemai et al, the NPWE model was applied together with an automated software to analyse LCD in the Catphan phantom[55]. For a range of mAs, the CT iterative algorithms rendered images with improved LCD, for *soft* and *sharp* kernels. The model LCD trends were not compared with human observers.

Christianson et al and Saiprasad et al implemented a version of the NPWE model in the frequency domain, with the filter proposed by Burgess and internal noise and analysed

images of the low contrast module of the ACR CT phantom acquired in scanners of three CT manufacturers[69,77]. They found a strong correlation between the model results and human observers for images reconstructed selecting IR or FBP algorithms for a range of dose levels. Another result was that CNR did not correlate with humans to the same stand and this metric overestimated the potential of dose reduction keeping the low contrast resolution level. For all the studied scanners, they found that low contrast resolution could be maintained using iterative algorithms enabling different dose reduction levels, depending on the manufacturer. Similar findings were published by Chen et al, applying the methodology of Christianson et al to images of a low contrast phantom containing cylindrical inserts of different contrasts[78,79]. The potential for dose reduction, when selecting IR, depended highly on the object size and contrast.

Solomon et al used a proprietary phantom to assess LCD with three model observers, implemented in the frequency domain, (Rose model, NPW and NPWE) and human observers in 2-AFC experiments[80]. Images were acquired selecting conditions that lead to noise values similar to those observed in typical CT clinical protocols, for a range of mAs and reconstructing the images applying FBP and different strengths for the manufacturer iterative algorithm. The NPWE model showed the highest correlation with human performance. LCD improved with increasing object size, contrast, dose level and IR strength in this study.

(3) There are several publications in which the CHO model was used for LCD tasks in CT. Yu et al implementation was based on the inclusion of 60 Gabor channels. In their study, LCD was assessed in a water phantom containing rods of different contrasts, which was imaged in a CT scanner selecting FBP and iterative algorithms for different dose levels[40]. The model showed an excellent correlation with human performance which was characterized in 2-AFC experiments. Regarding the detectability improvement with IR over FBP reconstruction, results were inconclusive. Leng et al applied the same model and phantom to assess LCD performance when the lesion location is unknown for an abdominal clinical protocol, FBP reconstruction, different dose levels and automatic exposure control[53]. The CHO model results proved that it was a good surrogate for human performance for the proposed detection task. The selection of parameters proposed by Yu et al for the CHO channels was used in paper **[IV]** in this thesis to assess the influence of kV in LCD in phantom images reconstructed with FBP[40].

The CHO model has also been applied in discrimination tasks by Zhang et al[71]. The CHO model proposed by Yu et al, with the same Gabor channels settings was modified with the addition of an edge mask. It was applied to discrimination tasks between hexagonal and disk shaped objects in a water phantom for CT images acquired for FBP and IR algorithms, showing a good correlation with human results[71]. Selecting IR improved the objects detectability indicating that images could be acquired at a lower dose, depending on the object size and contrast. The objects in this study had well defined edges and they were embedded in a uniform background, unlike real lesions in a patient.

Other types of channels have been also successfully used to implement CHO, like in Tseng et al, which developed two versions of the model one with Gabor and another with DDOG channels[43,67]. They analysed LCD in CT images of the MITA IQ LCD phantom (CCT183, The Phantom Laboratory, Salem, NY) for a wide range of dose levels, selecting protocols for head and body, reconstructing the images with FBP and IR algorithms. They

found that with the latter, an equivalent performance to that observed with FBP could be obtained for much lower doses[67].

Eck et al have developed a complete framework, similar to the one presented in this PhD thesis, to assess LCD in CT phantom images[68]. For the CHO model, five Laguerre-Gauss channels were implemented. Internal noise was selected as proportional to the channels output standard deviation. Two phantoms were used, the MITA IQ phantom (CCT183) and a virtual phantom, containing objects of different contrast and size, used in combination with a CT simulator. Images were acquired and simulated for a range of dose levels and reconstructed with FBP and the manufacturer iterative algorithm. Results were validated with a 4-AFC study in which human observers scored the detectability of the objects separately. The detectability of the objects increased with object contrast, size and selected dose for both, model and humans. LCD was substantially improved in the IR images compared to FBP and authors propose to use the model in dose reduction experiments leading to protocol optimization in CT[68].

Few papers in CT have used more realistic anatomical phantoms and model observers to assess LCD. For instance, Li et al investigated the influence of IR in LCD in an anthropomorphic phantom, which mimicked a paediatric patient thorax with the CHO model observer with Gabor channels proposed by Yu et al[40,81]. The detectability of objects of different sizes, which were inserted in a uniform background section in the phantom thorax increased with the iterative algorithm. The results were not validated with a human observer study[81].

All the studies reviewed in sections **2.2**, **2.3** of Discussion and those included in this PhD thesis **[I-IV]** were applied model observers to very simple detection or discrimination tasks in phantoms with uniform backgrounds. Model observers, especially NPWE and CHO, with different possible implementations, can predict human performance for this type of tasks to a high stand. They can be a useful tool to analyse LCD in an objective and fast way in CT but the results have to be taken cautiously. The use of iterative reconstruction algorithms is rapidly becoming common practice in CT and some manufacturers use model observers and phantom studies to validate the dose reductions that can be achieved with their algorithms compared to FBP. In 2011, one of the main CT manufacturers obtained the FDA clearance to use an IR algorithm based on a model observer[82]. The other major manufacturers have followed the same path and they have incorporated these models to their image quality analysis to endorse their performance and marketing claims[12]. The influence of patients' anatomy or tissue structures on LCD is still an open field for research and it was not investigated in the aforementioned studies. Model observers are not adapted by now to predict the radiologists' performance in more realistic clinical tasks in CT. In the US the FDA has made a recommendation regarding IR for all CT manufacturers, that is to include a disclaimer that might read as follows (quoting Vaishnaav et al): 'In clinical practice, the use of this algorithm may reduce CT patient dose depending on the clinical task, patient size, anatomical location, and clinical practice. A consultation with a radiologist and a physicist should be made to determine the appropriate dose to obtain diagnostic image quality for the particular clinical task'[12].

2.4. Phantoms for the assessment of low contrast detectability

The image quality phantoms that are normally used in LCD assessment in CT were designed to perform subjective studies in which the whole image is displayed and the observer scores the number of visible objects of a given contrast level[III]. These studies can be biased, because as it has been stated, the observer knows beforehand the objects distribution. This bias can be surpassed performing M-AFC studies, in which the visibility of each object is assessed individually.

The objects in these phantoms are frequently packed close together, which makes it difficult to subtract samples to use in M-AFC studies or to be analysed with model observers. Besides, CT noise is radially dependent, and to investigate LCD accurately, both the samples with and without objects should be taken close in the image so the noise distributions are as similar as possible. These limitations are shared by most current commercial phantoms, such as Catphan (Phantom laboratories, Salem, NY), ACR CT phantom (Gammex, Middlleton, WI), QRM 2D-LC and QRM-2DMC phantoms or CIRS spiral/helical CT phantom (model 061).

Different research groups, phantom manufacturers and task forces are investigating new phantom designs for LCD assessment, which enable an easy extraction of image samples. In their design these phantoms have to offer several object sizes and contrast levels similar to low contrast lesions in patients. Different locations for the signal absent samples should be possible in the phantom too. Enough space around each object should be available, so the samples are big enough for the human observers to score. An additional feature is the inclusion of anthropomorphic shells to mimic the thorax or abdominal patient geometry and attenuation. Next, some of these new designs are described.

Popescu and Myers (FDA) proposed a framework to simulate phantom designs, testing the distribution of the objects to find the optimal settings before manufacturing. With their approach, different modules can be created with shapes, sizes and contrast levels chosen by the user. They conclude that it would be advisable to include in each module only objects of the same size and contrast and an extra module without any objects to subtract the lesion absent samples[83].

A proprietary phantom was presented in Solomon et al to assess LCD with both models and human observers in M-AFC experiments. It contained radial distributions of cylinders of five contrast levels and three diameters, placed at different distances from the phantom centre[79,80].

The COCIR CT manufacturers, HERCA and FDA are involved in the MITA CT image quality Task Force, which has led to the development of a phantom (MITA IQ LCD phantom, CCT183, The Phantom Laboratory, Salem, NY). It contains objects of different contrast levels and sizes and an additional module to subtract the lesion absent samples. Different external rings can be added to simulate the attenuation in patients of different sizes. This phantom has been used in several research papers and recently, most CT manufacturers have started to give commercial information regarding LCD with this particular phantom and model observers for certain protocols[67,68,84]. One of the voluntary commitments of COCIR was to develop methods that enable a fair benchmark between LCD manufacturers dose reduction claims, in particular, when iterative reconstruction algorithms are used[84].

Some phantoms for LCD assessment include spherical objects, instead of the more usual cylindrical rods. For example, the AAPM CT performance phantom (CIRS, model 610) has a module cast in a material equivalent to water with spheres of different diameters and three contrast levels. A custom phantom was developed by QRM to mimic the liver parenchyma, with spherical hypoattenuating hepatic metastases that was used in Schindera et al[25]. Another option includes water-based phantoms like that used by Yu et al, which consisted on a PMMA container shaped like a torso and contained rods with different attenuations[40].

2.5. Anthropomorphic phantoms for clinical image quality assessment

To perform and adequate CT protocol optimization, studies based on patients are crucial to assess image quality. Noise texture and magnitude is highly affected by the anatomical structures in the patient and also by the patient size and characteristics. This is especially critical for IR algorithms in CT as manufacturers support the possible dose reductions with image quality measured in uniform phantoms, as those discussed in section **2.4.** Thus, the potential dose reduction can be overestimated with these methods, as the influence of anatomy and patient habitus is overlooked. Anthropomorphic phantoms are an alternative, as they reproduce the patient anatomy and attenuation. These phantoms can be designed to make dose measurements, image quality assessment or both.

The degree of realism in the available commercial phantoms varies, ranging from simple phantoms that with a few materials reproduce the main structures in the thorax and its shape, to very realistic phantoms that reproduce organs attenuation and appearance in detail. There also exist phantoms of different sizes to mimic the spectrum of patient habitus, from paediatric to obese patients.

There are anthropomorphic phantoms that mimic the movement of the heart to assess the ability of CT scanners to cope with the possible related artifacts or that allow contrast injection.

Two examples of commercial anthropomorphic CT phantoms are shown in **Fig.18**. On the left, the chest phantom N1 'lungman' (Kyoto Kagaky Co., Tokyo) reproduces the pulmonary vascular trees covered by structures mimicking the thoracic cage and nodule-like objects can be inserted in it[85,86]. On the right the anthropomorphic cardio CT phantom in combination with a thorax phantom from QRM, which can be filled with water and allows the insertion of rods with similar attenuation and diameter as coronary artery calcifications which can be connected to a motor and reproduce different heart rates and cycles, including arrhythmias.

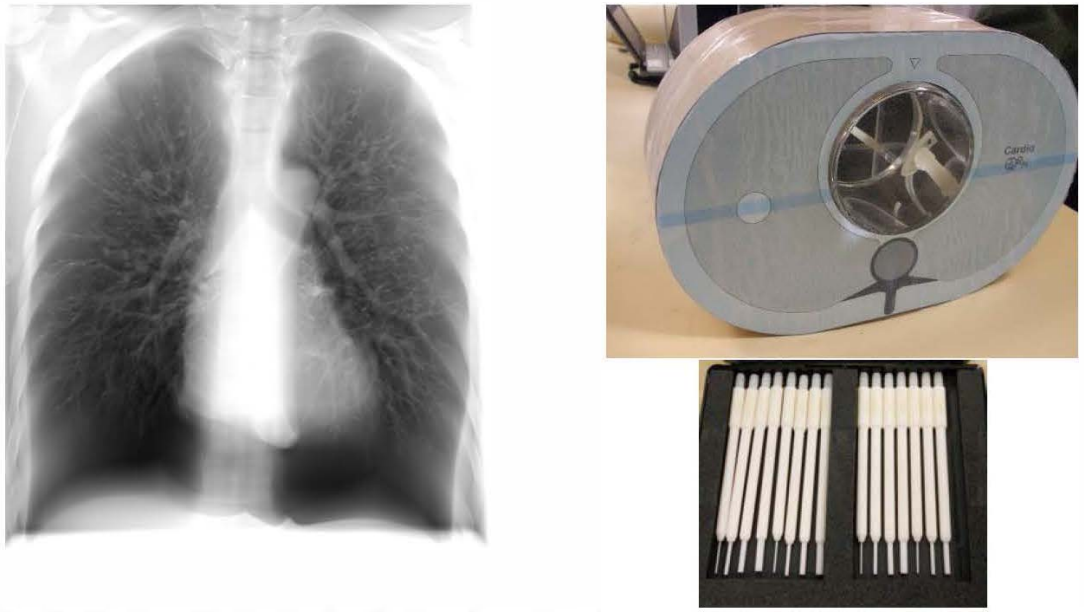


Fig. 18. Tomographic image of a thorax anthropomorphic phantom reproducing the pulmonary vascular trees in which an object mimicking a nodule can be seen (left). Cardiac CT phantom to assess the detectability of coronary artery calcifications and that can reproduce the heart dynamics (right).

CT image quality optimization has to be oriented to a given clinical task or a diagnostic indication. Clinical studies with patients have to be justified from an ethical point of view. Realistic anthropomorphic phantoms can play an important role as a surrogate.

Recently, 3D printing techniques, which are available since 1980 have started to be used in medical applications, such as creating printed models of certain parts of patients, based on medical images for teaching or informative purposes and creating orthopaedic prosthesis, for hands and parts of the leg for example. A new technique has been developed to obtain 3D printed segments of trachea that in combination with patient cells and collagen have been successfully used in paediatric patients with different malformations[87].

One of the possible applications of 3D printing is to create low-cost custom made phantoms to use in image quality assessment. In particular, in CT different papers have been published showing phantom prototypes for lung and liver indications.

Solomon et al were the first to design 3D printed models of the tissue structure of the lung, including vessels and 3D printed lesions with different attenuations[88]. They also designed a liver phantom. They analysed the influence of the presence of the anatomical structures in noise texture with FBP and iterative reconstruction algorithms. Solomon and Samei, have recently proposed a method to simulate lesion-like objects that could stand for lung, liver and renal pathologies after being 3D printed[89].

Leng et al built a 3D printed liver phantom, based on CT patient images, which were segmented into different materials, including vessels filled with iodine to simulate contrast-enhanced CT protocols[90]. Two versions of the phantom were created, one with lesions and another without, and the detectability of the objects was assessed with a CHO model observer comparing images reconstructed with FBP and different levels of iterative reconstruction for a range of doses.

In the final months of this PhD thesis, as part of a project called CLUES (Clinical Image Quality Assessment) funded by STW, a 3D printed lung phantom has been developed together with nodules surrogates[91]. It was presented at the Medical Imaging Perception Society (MIPS) XVI conference held in Ghent in June of 2015[92]. A view of some details in the phantom and a CT scan is shown in **Fig.19**.

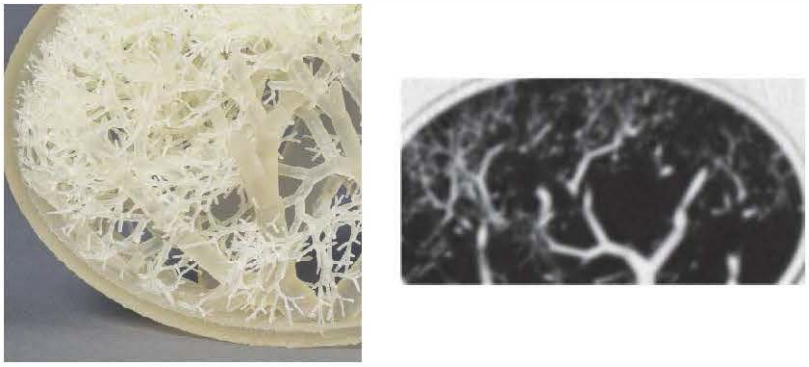


Fig. 19. Example of a 3D printed phantom reproducing the structure of the lung vessels (left). On the right, a selection of a CT image of the same phantom.

2.6. Other applications for model observers in medical imaging

Model observers can be applied to analyse image quality in detection tasks in different imaging modalities. Their use has increased in the past few years, especially in mammography[93,94].

With the tools developed during this thesis, a model observer (NPW) was used in Garayoa et al to analyse the detectability of low contrast objects in a phantom (TOR MAM) for image quality in digital breast tomosynthesis (DBT)[95]. The use of DBT has been approved for breast cancer screening in the USA, consisting on the acquisition of the 2D mammoghaphy image together with a DBT sweep[96-98]. This procedure almost doubles the dose received by the patient in conventional mammography. Some manufacturers propose to use the projections acquired during the DBT to generate a synthetic image that might eventually substitute the 2D acquisition. Images of the TOR MAM phantom were acquired selecting the 2D mammography and the DBT default protocols. An improvement in LCD was shown in the synthetic images compared to conventional mammography. These results are promising but the use of more realistic phantoms, mimicking breast tissue attenuation and structure, is necessary to determine if DBT can be used for breast cancer screening and the conventional 2D be unnecessary for certain groups of patients[96,97,99].

Conclusions

This PhD thesis presents a framework for CT image quality assessment in CT phantom images using model observers. The use of these mathematical algorithms has grown in the past few years. They are an objective and fast alternative to human observer studies as they can predict human performance in simple detection and discrimination tasks. They are especially interesting in medical imaging modalities, such as CT, because a wide range of parameters affect image quality. The use of CT has expanded in the past decades and it will continue so, as it has recently been approved for screening in certain indications in the US (lung cancer) and others are under study (colorectal cancer). Different organizations, like FDA, recommend the use of model observers to assess certain aspects of image quality in CT in an objective way. CT manufacturers have voluntarily compromised to develop and apply a common method in this direction to asses low contrast detectability, together with a scientific task force[84].

Regarding the milestones of this thesis:

1. A software was developed to automatically extract samples from phantom images to be analysed with model observers. It was created for a particular phantom, Catphan, widely used in quality control in CT. This software can be easily adapted to other phantoms.

2. A model observer (non-prewhitening matched filter with an eye filter, NPWE) was implemented in the software to assess LCD in CT phantom images. The performance of the model was compared successfully with results in the literature regarding detection experiments in simulated Gaussian white noise backgrounds.

3. Images of the phantom were acquired varying different acquisition and reconstruction parameters. Their influence in LCD was assessed with the model observer and the trends validated with a human observer study. The detectability of the objects increased with object diameter and contrast. It was also improved with increasing tube current, kV and for *soft* reconstruction kernels.

4. A software was developed to perform 2-alternative forced choice experiments with human observers. This software can be easily adapted to other M-AFC experiments if needed. It was used to compare model observers and humans, performing the same detection tasks, analysing each object of interest individually.

5. The channelized Hotelling observer (CHO) was implemented in our framework. The performance of NPWE and CHO was investigated to assess the influence of kVp in LCD. The NPWE model results showed a higher correlation with human observers.

6. Images of the Catphan phantom were acquired and reconstructed with filtered back projection and iterative (IR) algorithms for a range of dose levels. The NPWE model showed an improvement in LCD for the IR images, especially for low doses and low contrast objects.

Model observers have been successfully applied to predict human performance in simple detection and discrimination tasks in CT phantom images. The developed methodology is a fast and objective tool to assess low contrast detectability in CT. It can be used in protocol optimization or to compare different CT manufacturers in terms of image quality.

Model observers are not intended to be a substitute of clinical validation of systems based on patient images assessed by radiologists. Further research is needed to investigate the correlation between humans and model observers in more complex tasks and in anatomical backgrounds. Anthropomorphic phantom images, including 3D printed phantoms, can be a good thread to follow for these goals.

Future work

The selected model observers, NPWE and CHO with a particular eye filter and set of Gabor channels predicted human performance for simple detection tasks. Their performance in more complex tasks or backgrounds still needs to be investigated. Also other eye filters or types of channels can be easily tested using our framework, as they are input settings.

The software that was developed to subtract object present and absent samples from phantom images will be adapted to automatically analyse images of other phantoms, acquired either in CT scanners or using other medical imaging modalities.

The most common approach to implement model observers, based on the detection of simple bidimensional objects in the images, has to be adapted to detect 3D objects, as patient lesions are tridimensional. Few studies have been performed in this field, and none in CT images so far[55,66].

Model observers need to be adapted to include more aspects of the radiologist performance, such as scrolling speed and find theroretical models to reproduce how the information of contiguous CT slices is integrated[66,100].

Designing and validating anthropomorphic 3D printed phantoms for medical imaging is a quite recent research line. We aim to continue with the work started at the end of this PhD thesis, investigating different materials to reproduce the attenuation of lesions and human tissue. Other design improvements, to include patient habitus in the phantom will be investigated. Finally, human and model observers will be used to investigate the visibility of lesion like objects in the 3D printed phantoms.

References

1. Samei E, Krupinski E, editors. The handbook of medical image perception and techniques, 1st ed. New York: Cambridge University Press; 2010.

2. Bushberg JT, Seibert JA, Leidholdt Jr, EM, Boone JM. The essential physics of medical imaging, 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2012.

3. Hsieh J. Computed tomography: Principles, design, artifacts, and recent advances, 2nd ed. Bellingham: John Wiley & Sons Inc.; 2009.

4. Kalender WA. Computed tomography. Fundamentals, system technology, image quality, applications, 3rd ed. Erlangen: Publicis Kommunikationsag; 2011.

5. Buzug TM. Computed tomography: from photon statistics to modern cone-beam CT, 1st ed.: Springer-Verlag Berlin Heidelberg; 2010.

6. Wang G, Kalra M, Marugan V, Xi Y, Gjesteby L, et al. Vision 20/20: Simultaneous CT-MRI–Next chapter of multimodality imaging. Med Phys 2015;42:5879-89.

7. Brenner DJ, Hall EJ. Computed tomography–An increasing source of radiation exposure. N Engl J Med 2007;357:2277–2284.

8. Berrington de González A, Mahesh M, Kim KP, Bhargavan M, Lewis R, Mettler F, et al. Projected cancer risks from computed tomography scans performed in the United States in 2007. Arch Intern Med 2009;169:2071–2077.

9. 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;259:311–316.

10. Tack D, Kalra MK, Gevenois PA, editors. Radiation dose from multidetector CT, 2nd ed. Springer-Verlag Berlin Heidelberg; 2012.

11. ICRP Publication 103. The 2007 recommendations of the International Commission on Radiological Protection. Ann ICRP 2007;37.

12. Vaishnav JY, Jung WC, Popescu LM, Zeng R, Myerz KJ. Objective assessment of image quality and dose reduction in CT iterative reconstruction. Med Phys 2014;41:071904.

13. Thibault JB, 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.

14. Beister M, Kolditz D, Kalender WA. Iterative reconstruction methods in X-ray CT. Phys Medica 2012;28:94-108.

15. Computed tomography dose check (NEMA Standards Publication XR 25-2010, October 2010. Available at: http://www.nema.org/stds/sr25.cfm).

16. HERCA position paper: The process of CT dose optimization through education and training and role of CT manufacturers (2012). Available at: http://www.herca.org/uploaditems/documents/HERCA%20Position%20paper%20Education%20and%20Training%20in%20CT_website.pdf .

17. American Association of Physicists in Medicine (AAPM). Alliance for Quality Computed Tomography Working Group. CT scan protocols. Available at: http://www.aapm.org/pubs/CTProtocols/ .

18. COCIR: CT Manufacturer's Voluntary Commitment Regarding CT Dose (2013). Available at: http://www.cocir.org/fileadmin/5_Initiatives/COCIR_CT_MANUFACTURER_List_of_Dose_ManagementFeatures_05_July_2013.pdf .

19. Sodickson A, Warden GI, Farkas CE, Ikuta I, Prevedello LM, Andriole KP, et al. Exposing exposure: automated anatomy-specific CT radiation exposure extraction for quality assurance and radiation monitoring. Radiology 2012;264:397–405.

20. Matsubara K, Sugai M, Toyoda A, Koshida H, Sakuta K, Takata T, et al. Assessment of an organ-based tube current modulation in thoracic computed tomography. J Appl Clin Med Phys 2012;13:148–158.

21. Winklehner A, Goetti R, Baumueller S, Karlo C, Schmidt B, Raupach R, et al. Automated attenuation-based tube potential selection for thoracoabdominal computed tomography angiography: Improved dose effectiveness. Invest Radiol 2011;46:767–773.

22. International Comission on Radiation Units and Measurements. Receiver operating characteristic analysis in medical imaging. ICRU Report No. 79. Bethesda, MD: International Comission on Radiation Units and Measurements; 2008.

23. International Comission on Radiation Units and Measurements. Radiation dose and image-quality asssessment in Computed Tomography. ICRU Report No. 87. Oxford University Press; 2012.

24. Richard S, Husarik DB, Yadava G, et al. Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms. Med Phys 2012;39:4115-4122.

25. Schindera ST, Odedra D, Raza SA, Kim TK, Jang HJ, Szucs-Farkas Z, et al. Iterative reconstruction algorithm for CT: Can radiation dose be decreased while low-contrast detectability is preserved? Radiology 2013;269(2):511-518.

26. McCollough CH, Yu L, Kofler JM, Leng S, Zhang Y, Li Z, et al. Degradation of CT low-contrast spatial resolution due to the use of iterative reconstruction and reduced dose levels. Radiology 2015;276(2):499-506.

27. Burgess AE. Visual perception studies and observer models in medical imaging. Semin Nucl Med 2011;41:419-436.

28. Beutel J, Kundel JL, Van Metter RL, editors. Handbook of medical imaging. Physics and psychophysics, Vol. 1. SPIE publications; 2000.

29. Green DM, Swets JA. Signal detection theory and psyhophysics. Peninsula Pub; 1989.

30. MacMillan NA, Creelman CD. Detection theory: a user's guide. 2nd edition, Psychology Press; 2004.

31. He X, Park S. Model observers in medical imaging researh. Theraostics 2013;3(10):774-786.

32. 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.

33. Chao EH, Toth TL, Bromberg NB, Williams EC, Fox SH, Carleton DA. A statistical method of defining low contrast detectability. Radiology 2000;217:162.

34. Torgensen 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(5):603-611.

35. Obuchowski NA, Schoenhagen P, Modic MT, Meziane M, Budd GT. Incidence of advanced symptomatic disease as primary endpoint in screening and prevention trials. AJR 2007;189:19-23.

36. International Comission on Radiation Units and Measurements. Medical imaging–The assessment of image quality. ICRU Report No. 54. Bethesda, MD: International Comission on Radiation Units and Measurements; 1996.

37. Barrett HH, Myers KJ. Foundations image science. John Wiley & Sons; 2004.

38. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. Proc Natl Acad Sci U.S.A. 1997;90:9758-65.

39. Burgess AE, Colborne B: Visual detection. IV. Observer inconsistency. J Opt Soc Am A 1998;5:617-27.

40. 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.

41. Tanner WP, Birdsall TG. Definitions of d' and η as psychophysical measures. J Acoust Soc Am 1958;30(10):922-928.

42. Pelli DG, Bex P. Measuring contrast sensitivity. Vision Research 2013;90:10-14.

43. Abbey CK, Barrett HH. Human and model-observer performance in ramp-spectrum noise: effects of regularization and object variability. J Opt Soc Am A 2001;18:473-487.

44. Bouwman RW, van Engen RE, Dance DR, Young KC, Veldkamp WJH. Evaluation of human contrast sensitivity functions used in the nonprewhitening model observer with eye filter. LNCS 2014;8539:715-722.

45. Burgess AE, Jacobson FL, Judy PF. Human observer detection experiments with mammograms and power-law noise. Med Phys 2001;28:419-37.

46. Reiser I, Nishikawa RM. Identification of simulated microcalcifications in white noise and mammographic backgrounds. Med Phys 2006;33:2905-11.

47. Daugman JG: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J Opt Soc Am A 1987;2 (7):1160-1169, 1985.

48. Myers KJ, Barrett HH. Addition of a channel mechanism to the ideal-observer model. J Opt Soc Am A 1987;4:2447-2457.

49. Wunderlich A, Noo F. Image covariance and lesion detectability in direct fan-beam X-ray computed tomography. Phys Med Biol 2008;53:2471-93.

50. Gifford HC, King MA, de Bries DJ, Soares EJ. Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging. J Nucl Med 2000;41:514-21.

51. Boedeker KL, McNitt-Gray MF. Application of the noise power spectrum in modern diagnostic MDCT: part II. Noise power spectra and signal to noise. Phys Med Biol 2007;52:4047-4061.

52. Richard S, Siewerdsen JH. Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images. Med Phys 2008;35:5043-5053.

53. 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.

54. Tapiovaara MJ. Efficiency of low contrast detail detectability in fluoroscopic imaging. Med Phys 1997;24:655-664.

55. Joemai RM, Veldkamp WJ, Kroft LJ, Hernandez-Giron I, Geleijns J. Adaptive iterative dose reduction 3D versus filtered back projection in CT: evaluation of image quality. AJR Am J Roengenol 2013;201:1291-1297.

56. Woolson RF. Statistical methods for the analysis of biomedical data. 1st ed. New York. John Wiley & Sons, Inc; 1987.

57. Schindera ST, Diedrichsen L, Müller HC, Rusch O, Marin D, Schmidt B, et al. Iterative reconstruction algorithm for abdominal multidetector CT at different tube voltages: assessment of diagnostic accuracy, image quality, and radiation dose in a phantom. Radiology 2011;260:454-462.

58. Noël PB, Köhler T, Fingerle AA, Brown KM, Zabic S, Münzel D, et al. Evaluation of an iterative model-based reconstruction algorithm for low-tube-voltage (80 kVp) computed tomography angiography. J Med Imaging 2014;1(3):033501.

59. Mathieu KB, Turner AC, Khatonabadi M, McNitt-Gray MF, Cagnon CH, Cody DD. Varying kVp as a means of reducing CT breast dose to pediatric patients. Phys Med Biol 2013;58:4455-4469.

60. Marin D, Nelson RC, Barnhart H, Schindera SR, Ho LM, Jaffe TA, et al. Detection of pancreatic tumors, image quality, and radiation dose during pancreatic parenchymal phase: effect of a low-tube-voltage, high-tube-current CT technique – Preliminar results. Radiology 2010;256:450-459.

61. Lee KH, Lee JM, Moon SK, Baek JH, Park JH, Flohr TG, et al. Attenuation-based automatic tube voltage selection and tube current modulation for dose reduction at contrast-enhanced liver CT. Radiology 2012;265:437-447.

62. Klein Zeggelink WFA, Hart AAM, Gilhuijs KGA. Assessment of analysis-of-variance-based methods to quantify the random variations of observers in medical imaging measurements: guidelines to the investigator. Med Phys 2004;31:1996-2007.

63. Rao JNK, Scott AJ. A simple method for the analysis of clustered binary data. Biometrics 1992;48:577-585.

64. Gallas BD. One-shot estimate of MRMC variance: AUC. Acad Radiol 2006;13:353-362.

65. Eckstein MP, Ahumada AJ Jr, Watson AB. Visual signal detection in structured backgrounds. II. Effects of contrast gain control, background variations, and white noise. J Opt Soc Am A 1997;14(9):2406-2418.

66. Platiša L, Goossens B, Vansteenkiste E, Park S, Gallas BD, Badano A, Philips W. Channelized Hotelling observers for the assessment of volumetric imaging data sets. J Opt Soc Am A 2011;28(6):1145-1163.

67. Tseng H, 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.

68. Eck BL, Fahmi RF, Brown KM, Zabic S, Raihani N, Miao J, et al. Computational and human observer image quality evaluation of low dose, knowledge-based CT iterative reconstruction. Med Phys 2015;42:6098.

69. Christianson O, Chen JJS, Yang Z, Saiprasad G, Dima A, Filliben JJ, 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.

70. Ott JG, Becce F, Monnin P, Schmidt S, Bochud FO, Verdun FR. Update on the non-prewhitening model observer in computed tomography forthe assessment of the adaptive statistical and model-based iterative reconstruction algorithms. Phys Med Biol 2014;59:4047-4064.

71. 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;59:3389-3404.

72. Zhang Y, Pham BT, Eckstein MP. Task-based model/human observer evaluation of SPIHT wavelet compression with human visual system-based quantization. Acad Radiol 2005;12:324-336.

73. Solomon J, Samei E. What observer models best reflect low-contrast detectability in CT? Proc of SPIE Medical Imaging 2015. Vol 9416:9460I.

74. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. Proc Natl Acad Sci USA 1997;90:9758-9765.

75. Goenka AH, Herts BR, Obuchowski NA, Primak AN, Dong F, Karim W, et al. Effect of reduced radiation exposure and iterative reconstruction on detection of low-contrast low-attenuation lesions in an anthropomorphic liver phantom: An 18-reader study. Radiology 2014;272(1):154-163.

76. 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 Medica 2013;29:99-110.

77. Saiprasad G, Filliben J, Peskin A, Siegel E, Chen J, Trimble C, et al. Evaluation of low-contrast detectability of iterative reconstruction across multiple institutions, CT scanner manufacturers, and radiation exposure levels. Radiology 2015;277(1):124-133.

78. Chen B, Ramirez Giraldo JC, Solomon J, Samei E. Evaluating iterative reconstruction performance in computed tomography. Med Phys 2014;41:121913.

79. Wilson JM, Christianson OI, Richard S, Samei E. A methodology for image quality evaluation of advanced CT systems. Med Phys 2013;40:031908.

80. Solomon J, Mileto A, Ramirez-Giraldo JC, Samei E. 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 2015;275(3):735-745.

81. Li K, Garrett J, Ge Y, Chen GH. Statistical model based iterative reconstruction (MBIR) in clinical CT systems. Part II. Experimental assessment of spatial resolution performance. Med Phys 2014;41(7):071911.

82. Siemens Medical Systems, Inc., 510(k) Summary for the Somatom Definition Flash, FDA 510(k) Premarket Notification Database, FDA 510(k) (Siemens Medical Systems, Inc., Malvern, PA, 2011) (available at http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/dfpmn/pmn.cfm?ID=K113342).

83. Popescu LM, Myers KJ. CT image assessment by low contrast signal detectability evaluation with unknown signal location. Med Phys 2013;40:111908.

84. COCIR: CT Manufacturer's Voluntary Commitment Regarding CT Dose (2015). Available at: http://www.cocir.org/fileadmin/5_Initiatives/2015_COCIR_CT_Manufacturers_-_Optimization_Report_2015.pdf.

85. Shim SS, Oh Y-W, Kong KA, Ryu YJ, Kim Y, Jang DH. Pulmonary nodule size evaluation with chest tomosynthesis and CT: a phantom study. Br J Radiol 2015;88(1047):20140040.

86. Gomi T, Nakajima M, Fujiwara H, Takeda T, Saito K, Umeda T, et al. Comparison between chest digital tomosynthesis and CT as a screening method to detect artificial pulmonary nodules: a phantom study. Br J Radiol 2012;85:e622-e629.

87. Morrison RJ, Hollister SJ, Niedner MF, Ghadimi Mahani M, Park AH, Mehta DK, et al. Mitigation of tracheobronchomalacia with 3D-printed personalized medical devices in pediatric patients. Science Translational Medicine 2015;7(285): 285RA64.

88. Solomon J, Bochud F, Samei E. Design of anthropomorphic textured phantoms for CT performance evaluation. Proc of SPIE Medical Imaging 2014. Vol 9033:90331U.

89. Solomon J, Samei E. A generic framework to simulate realistic lung, liver and renal pathologies in CT imaging. Phys Med Biol 2014;59:6637-6657.

90. Leng S, Yu L, Vrieze T, Kuhlmann J, Chen B, McCollough CH. Construction of realistic liver phantoms from patient images using 3D printer and its application in CT image quality assessment. Proc of SPIE Medical Imaging 2015. Vol 9412:84124E.

91. STW CLUES project website: http://www.stw.nl/nl/content/clinical-image-quality-assessment-bridging-gap-between-physical-measurements-and-clinical.

92. Den Harder JM, Hernandez-Giron I, Calzado A, Geleijns J, Veldkamp WJH. 3D printed lung phantom for clinical image quality assessment. Presented at the Medical Imaging Perception Society (MIPS) XVI conference held in Ghent (2015). Book of abstracts available at: http://mips.rogulski.com/documents/MIPS2015_BookOfAbstracts.pdf.

93. 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;56:4221-4238.

94. Das M, Gifford HC. Comparison of model-observer and human-observer performance for breast tomosynthesis: Effect of reconstruction and acquisition parameters. Proc of SPIE Physics of Medical Imaging 2011. Vol 7961:796118.

95. Garayoa J, Hernandez-Giron I, Castillo M, Valverde J, Chevalier M. Digital breast tomosynthesis: Image quality and dose saving of the synthesized image. IWDM 2014. LNCS 2014;8539:150-157.

96. Sechopoulos I. A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications. Med Phys 2013;40:014302.

97. Reiser I, Nishikawa RM. Task-based assessment of breast tomosynthesis: Effect of acquition parameters and quantum noise. Med Phys 2010;37:1591.

98. Hernandez-Giron I, Chevalier M, Castillo M, Valverde Moran J, Garayoa J. Model observer performance in detection tasks in mammography: 2D vs reconstructed planes in tomosynthesis.

Presented at the Medical Imaging Perception Society (MIPS) XVI conference held in Ghent (2015). Book of abstracts available at:
http://mips.rogulski.com/documents/MIPS2015_BookOfAbstracts.pdf .

99. Sechopoulos I, Sabol JM, Berglund J, Bolch WE, Brateman L, Christodoulou E, et al. Report of AAPM Tomosynthesis Subcommittee Task Group 223. Radiation dosimetry in digital breast tomosynthesis. Med Phys 2014:41(9):091501.

100. Ba A, Racine D, Ott JG, Verdun FR, Kobbe-Schmidt S, Eckstein MP, et al. Low contrast detectability in CT for human and model observers in multi-slice data sets. Proc of SPIE Image Perception, Observer Performance, and Technology Assessment 2015. Vol 9416:94160F.

APPENDIX: Other publications

The following peer-reviewed publications were completed during this PhD thesis:

Image quality in Radiology

1. Joemai RMS, Veldkamp WJH, Kroft LJ, Hernández-Girón I, Geleijns J. Adaptive iterative dose reconstruction 3D versus filtered back projection in CT: evaluation of image quality. AJR 2013;201:1291-7.

2. Garayoa J, Hernández-Girón I, Castillo M, Valverde J, Chevalier M. Digital breast tomosynthesis: Image quality and dose saving of the synthesized image. H.Fujita, T. Hara and C. Muramatsu (Eds): IWDM 2014. LNCS 2014;8539:150-7.

3. Racine DS, Ott JG, Tapiovaara MJ, Toroi P, Bochud FO, Veldkamp WJH, Schegerer A, Bouwman RW, Hernandez-Giron I, Marshall NW, Edyvean S. Image quality in CT: from physical measurements to model observers. Phys Medica 2015 http://dx.doi.org/10.1016/j.ejmp.2015.08.007 (accepted, corrected proofs available online).

CT and cone-beam CT dosimetry

4. Morant JJ, Salvadó M, Casanovas R, Hernández-Girón I, Velasco E, Calzado A. Validation of a Monte Carlo based code for dose assessment in dental cone beam CT examinations. Phys Medica 2011;28:200-209.

5. Morant JJ, Salvadó M, Hernández-Girón I, Casanovas R, Ortega R, Calzado A. Dosimetry of a cone beam computed tomography device for oral and maxilofacial radiology using Monte Carlo techniques and ICRP adult reference computational phantoms. Dentomaxillofacial Radiology 2013;42:92555893.

6. Calzado Cantera A, Hernández-Girón I, Salvadó Artells M, Rodríguez González R. State of the art and future trends in technology for computed tomography dose reduction. Radiologia 2013;55(S2):9-16.

7. Geleijns J, Joemai RMS, Cros M, Hernández-Girón I, Calzado A, Dewey M, Salvadó M. A Monte Carlo simulation for the estimation of patient dose in rest and stress cardiac Computed Tomography with a 320-detector row CT scanner. Phys Medica 2015 http://dx.doi.org/10.1016/j.ejmp.2015.08.008 (accepted, corrected proofs available online).

In other fields:

8. Casanovas R, Morant JJ, López M, Hernández-Girón I, Batalla E, Salvadó M. Performance of data acceptance criteria over 50 months from an automatic real-time environmental radiation surveillance network. Journal of Environmental Radioactivity 2011;102:742-748.

Book chapters and editions

During this thesis, the following book chapters were written:

J.J. Morant Echevarne, M. Alcaraz Baños, M. Salvadó Artells, I. Hernández-Girón. *"Producción de rayos X y su aplicación en radiodiagnóstico"*. In the book: *Formación básica en protección radiológica. Curso de protección radiológica dirigido al personal técnico de las empresas de venta y asistencia técnica de equipos de rayos X dentales.* Published by Edit.um (2013), pp 69-126, ISBN: 978-84-15463-89-4.

M. López, F. Borrull, M. Salvadó, C. Aguilar, R. Casanovas, M. Cros, I. Hernández-Girón, J.J. Morant, A. Nieto, S. Peñalver. Edition of the proceedings of the conference *VII Jornadas sobre calidad en el control de radioactividad ambiental*, Tarragona (Spain), 2012.

Available at:
https://www.csn.es/images/stories/publicaciones/otras_publicaciones/coediciones/vii_jornades_de_calidad_web.pdf

Abstracts in conferences and proceedings

2015

1. Authors: I. Hernandez-Giron, M. Chevalier, M. Castillo, J. Valverde Morán, J. Garayoa.

Title: Model observer performance in detection tasks in mammography: 2D vs reconstructed planes in tomosynthesis. (Oral presentation, presenter).

MEDICAL IMAGE PERCEPTION SOCIETY CONFERENCE XVI

http://mips.rogulski.com/documents/MIPS2015_BookOfAbstracts.pdf .

Ghent (Belgium). June 2015. Awarded with the MIPS Scholarship to attend.

2. Authors: J.M. Den Harder, I. Hernández-Girón, A. Calzado, J. Geleijns, WJH Veldkamp.

Title: 3D printed lung phantom for clinical image quality assessment (Oral presentation).

Medical Image Perception Society Conference XVI.

http://mips.rogulski.com/documents/MIPS2015_BookOfAbstracts.pdf .

Guent (Belgium). June 2015. Awarded with the MIPS Scholarship to attend.

3. Authors: J.E.M. Mourik, M. L. Overvelde, I. Hernández-Girón, W. J. H. Veldkamp, D. Zweers, K. Geleijns.

Title: Multicentre comparison of image quality for low contrast objects and micro-catheter tips in X-ray guided treatment of arteriovenous malformation in the brain. (Poster).

Fourth Malmö Conference on Medical Imaging: 'Optimization in X-ray and Molecular Imaging 2015'. Gothenburg (Sweden). May 2015.

4. Authors: J. M. den Harder, I. Hernández-Girón, A. Calzado, J. Geleijns, W. J. H. Veldkamp.

Title: 3D-printed lung phantom for clinical image quality assessment. (Poster).

29th Nederlandse vereniging voor Klinische Fysica. Woudschoten (The Netherlands).

March 2015.

2014

5. Authors: I. Hernandez-Giron, A. Calzado, J. Geleijns, R. M. S. Joemai, W. J. H. Veldkamp

Title: Model observers low contrast detectability performance at different kV levels in CT phantom images.

(Oral presentation, presenter).

8th European Conference in Medical Physics (ECMP).

http://www.efomp-2014.gr/images/ECMP_2014_Abstract_Book.pdf .

Athens (Greece). September 2014.

6. Authors: M. Cros, J. Geleijns, R.M.S. Joemai, I. Hernandez-Giron, A. Calzado, M. Dewey, M. Salvado.

Title: Patient dose estimation in cardiac Computed Tomography with a 320 detector row scanner based on Monte Carlo simulation (Oral presentation).

8th European Conference in Medical Physics (ECMP).

http://www.efomp-2014.gr/images/ECMP_2014_Abstract_Book.pdf .

Athens (Greece). September 2014.

7. Authors: I. Hernandez-Giron, A. Calzado, J. Geleijns, R. M. S. Joemai, W. J. H. Veldkamp

Title: Human and model observers performance in low contrast detection tasks with CT phantom images acquired at different dose levels (Poster).

Proceedings of the Third CT meeting.

http://www.ucair.med.utah.edu/CTmeeting/ProceedingsCTMeeting2014.pdf .

Third CT meeting. Salt Lake City (USA). June 2014.

8. Authors: J. Garayoa, I. Hernandez-Giron, M. Castillo, J. Valverde, M. Chevalier

Title: Digital breast tomosynthesis: image quality and dose saving of the synthetic image (Oral presentation, presenter).

Proceedings of the 12th International Workshop on Breast Imaging: IWDM 2014.

12th International Workshop on Breast Imaging. Nagoya (Japan). July 2014.

2013

9. Authors: I. Hernández-Girón, J. Geleijns, A. Calzado, W. J. H. Veldkamp

Title: Detectabilidad de bajo contraste en TC. Comparación entre observadores humanos y un modelo de observador.

(Oral presentation, presenter).

III Congreso conjunto SEFM-SEPR (XIX Congreso nacional de la Sociedad Española de Física Médica (SEFM), XIV Congreso nacional de la Sociedad Española de Protección Radiológica (SEPR). Cáceres (Spain). June 2013.

10. Authors: I. Hernández-Girón, J. Geleijns, A. Calzado, R. Joemai, W. J. H. Veldkamp

Title: Comparison between human and model observer performance in low contrast detection tasks in CT images reconstructed with iterative methods (Oral presentation, presenter).

Medical Image Perception Society Conference XV.

BOOK OF ABSTRACTS

http://home.comcast.net/~eakmips/documents/MIPS-XV-2013Abstracts.pdf .

Washington DC (USA). August 2013. Awarded with the MIPS Scholarship to attend.

2011

11. Authors: I. Hernández-Girón, J. Geleijns, A. Calzado, W. J. H. Veldkamp

Título: Evaluación automática de la detectabilidad de bajo contraste en equipos de tomografía computarizada.

(Oral presentation, presenter).

II Congreso conjunto SEFM-SEPR (XVIII Congreso nacional de la Sociedad Española de Física Médica (SEFM), XIII Congreso nacional de la Sociedad Española de Protección Radiológica (SEPR). Sevilla (Spain). May 2011.

12. Authors: M. Salvadó, I. Hernández-Girón, J. J. Morant, R.Casanovas, M. López, A. Calzado

Title: Cálculos de dosis en radiodiagnóstico sobre los maniquíes voxelizados ICRP 110 mediante el método de Monte Carlo (Poster).

II Congreso conjunto SEFM-SEPR (XVIII Congreso nacional de la Sociedad Española de Física Médica (SEFM), XIII Congreso nacional de la Sociedad Española de Protección Radiológica (SEPR). Sevilla (Spain). May 2011.

13. Authors: I. Hernández-Girón, J. Geleijns, A. Calzado, M. Salvadó, R. Joemai, W. J. H. Veldkamp

Title: Automated assessment of low contrast sensitivity for Computed Tomography (CT) using a model observer: influence of the reconstruction filter (Oral presentation, presenter).

Oral presentation (presenter).

Medical Image Perception Society Conference XIV. Dublin (Ireland). August 2011. Awarded with the MIPS Scholarship to attend.

14. Authors: I. Hernández-Girón, J. Geleijns, A. Calzado, M. Salvadó, R. M. S. Joemai, W. J. H. Veldkamp

Title: Objective assessment of low contrast detectability for real CT phantom and in simulated images using a model observer (Poster).

2011 IEEE Nuclear Science Symposium and Medical Imaging Conference. Valencia (Spain). October 2011.

15. Authors: I. Hernández-Girón, J. Geleijns, A. Calzado, M. Salvadó Artells, R. M. S. Joemai, W. J. H. Veldkamp

Title: Dose efficiency of a 320-detector row CT scanner by objective assessment of low contrast detectability (Oral presentation, presenter).

RSNA 2011 (RADIOLOGICAL SOCIETY OF NORTH AMERICA). Chicago (USA).

2010

16. Authors: W. J. H. Veldkamp, I. Hernández-Girón, A.Calzado, J. Geleijns

Title: Automated assessment of low contrast sensitivity for Computed Tomography (CT) (Poster).

At The first international meeting on image formation in X-ray computed tomography. Salt Lake City (Utah). June 2010.