Physica Medica 64 (2019) 89–97

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Model observers for Low Contrast Detectability evaluation in dynamic angiography: A feasible approach

Raffaele Villa[a,*], Nicoletta Paruccini[a], Antonia Baglivi[a,b], Michele Signoriello[a,b], Roberto Alejandro Montezuma Velasquez[a], Sabrina Morzenti[a], Elena De Ponti[a], Andrea Crespi[a]

[a] Medical Physics Department, ASST Monza, Italy
[b] Scuola di Specializzazione in Fisica Medica, University of Milan, Italy

ARTICLE INFO

Keywords:
Digital Angiography
Model Observer
Channelized Hotelling Observer
NPWE
Low Contrast Detectability
2AFC

ABSTRACT

Purpose: To evaluate the feasibility of spatio-temporal generalisation of mathematical methods for protocol optimisation in interventional radiology.

Materials and methods: Two model observers were considered:

- Channelized Hotelling Observer with spatio-temporal Gabor channels;
- Non-Prewhitening Model Observer with spatio-temporal contrast sensitivity function.

Furthermore, Low Contrast Detectability (LCD) was evaluated with a generalised statistical method by taking into account the noise integration capability of the human eye.

A series of two alternative force choices (2AFC) experiments performed by four observers were used to evaluate the reliability of the proposed models.

The evaluation of the mathematical methods was performed by comparing their results to the human observer performances in two steps:

1. Firstly, a series of simulated images were used to tune the models
2. In the second phase, tuned models were applied both to simulated images and actual images obtained with a commercial phantom to evaluate detectability scores.

Results: Evaluation with simulated images shows a good agreement with 2AFC results (RMSE < 10%). Phantom-based evaluations show a general decrease of such agreement, characterized by an RMSE lower than 16%.

Conclusions: The agreement with human observer experiments supports the feasibility of the proposed generalisations. Thus, they could be introduced in quality control programmes for a deeper protocol-characterisation or for clinical protocol-optimization when dynamic images are involved.

1. Introduction

Digital Angiography is a particular field of radiology, characterized by a wide range of exams, techniques, modalities and, therefore, acquisition parameters. Fluoroscopic pulse rates range from 30 images per second (coronary angiography imaging) to 1–4 images per second (cerebral studies), with dose-rate (Air Kerma at the entrance of the image detector) ranging from 10 *nGy/frame* to 5000 *nGy/frame*.

Generally, the image quality of fluoroscopic exams is poorer when compared to other radiological techniques; that calls for the relevance of acquisition protocols optimization in order to achieve the highest detectability for a given radiation dose to the patient.

Although the evaluation of detectability scores of radiographic protocols by means of human-observer studies is the gold standard, this method is high resource-demanding, heavily time-consuming and additionally suffers from variability even when the same observers are involved. Therefore, different mathematical methods were introduced to overcome these limitations, e.g. the task-based metrics. Among the major methods, there are the so-called Model Observer methods: out from 2 groups of images (insert-present and insert-absent), these methods are able to evaluate the detectability scores by reproducing the detection performances of human observers [1]. Another method is based on statistical hypotheses; starting from a single homogeneous image, it is able to predict the Low Contrast Thresholds (LC_{th}) for a

* Corresponding author.
E-mail address: raffaele.vill@gmail.com (R. Villa).

https://doi.org/10.1016/j.ejmp.2019.06.015
Received 14 July 2018; Received in revised form 29 May 2019; Accepted 29 June 2019
Available online 04 July 2019

fixed confidence level, e.g. 95%. Unfortunately, these mathematical approaches, such as Model Observer or statistical method, are designed to analyse one single image at once, that is detectability is evaluated using only the information taken from each single image, without considering any mutual information contained in contiguous images, thus making them unsuitable for the application to spatio-temporal images. Actually, the human visual system considers each run of images globally, therefore detectability is influenced by the content of contiguous frames and by the visualized frame-rate (*fps*). As a consequence, in ROC analysis, evaluating the Proportion Correct (PC) only by means of individual frames may lead to high underestimation of the detectability score, this effect being more relevant with increasing *fps* rate.

The proper analysis of cine-images requires a generalisation of the mathematical methods by taking into account how the human visual system merges the information to increase the global detectability.

The aim of this study is to investigate the ability of the above mentioned methods to predict the human observer performances with spatio-temporal images.

The investigated mathematical methods were:

- the Channelized Hotelling observer with Gabor channels [2];
- the Non-Prewhitening with Eye filter observer with two different approaches [3,4];
- the statistical method [5].

The reliability and consistency of the generalised methods were evaluated through a comparison with a series of human observer experiments, i.e. two alternative force choices (2AFC).

2. Materials and methods

In this section, after a brief introduction about model observer methods, a description of the two main classes of pseudo-anthropomorphic model observer is presented: Hotelling Observer (HO) and Non-Prewhitening Model Observer (NPW). These methods are described together with a series of degradation techniques aimed to reproduce the effects of the human vision system on Low Contrast Detectability: channels mechanism combined with Hotelling Observer and Contrast Sensitivity Function (CSF) combined with Non-Prewhitening Model. Moreover, since these methods are designed to deal only with two-dimensional images without considering the effects of the temporal dimension in the human visual system, a series of spatio-temporal generalisations of the methods are also presented.

The statistical method is described together with its generalization aimed to properly deal with spatio-temporal images by means of the introduction of a preliminary averaging step.

Simulated images were initially used to allow a better preliminary investigation of the models thanks to their lower variability and thereafter to tune the models themselves by comparison with the human response. Tuned models were then used to evaluate LCD both with simulated images and with actual images acquired with a commercial phantom.

2.1. Model observer

Considering binary classification tasks, images can be divided into two different classes: H_1 (signal absent) and H_2 (signal present). The observer classifies each image using a scalar test statistic λ, derived from the probability density of an image g to be taken from class k (k = 1,2). Constraining the test statistic to be a linear function of g, the mathematical observer evaluates the scalar test statistic λ applying a linear template to the image data vector *g*:

$$\lambda_k = w^t g_k$$

where w^t is the transpose vector of the template of the mathematical observer, a vector with dimension equal to that of g_k.

A first figure of merit related to the distribution of the test statistic is the detectability index (d'), defined as:

$$d' = \frac{\bar{\lambda}_1 - \bar{\lambda}_2}{\sqrt{\frac{(\sigma_{\lambda 1})^2}{2} + \frac{(\sigma_{\lambda 2})^2}{2}}}$$

where $\bar{\lambda}_k$ is the mean of the test statistic related to the class k and $\sigma_{\lambda k}$ is the corresponding standard deviation.

Another figure of merit is the area under the ROC curve, denoted AUC. If the distribution of λ can be considered Gaussian, d' is related to the AUC by

$$AUC = \frac{1}{2}\left[1 + erf\left(\frac{d'}{2}\right)\right]$$

where $erf\left(\frac{d'}{2}\right)$ is the Gaussian error function [6].

2.1.1. Hotelling observer

The Hotelling Observer is a particular type of linear observer in which the template vector takes the form of:

$$w_{HO} = \left[\frac{1}{2}(K_1 + K_2)\right]^{-1} \Delta\bar{g}$$

where K_k is the covariance matrix related to classes $H_{1,2}$, and $\Delta\bar{g}$ is the difference between the average vectors under the two hypotheses. In Hotelling Observer studies, many authors suggest the use of independent datasets to train and to test the HO [7]. Therefore, $\Delta\bar{g}$ is calculated using the sample mean of the training set, while the test statistic λ_k is calculated using the test dataset.

2.1.1.1. Channels. To predict human-observer performance, the so-called "channel mechanism" was introduced, aiming to represent the selective sensitivity to different limited ranges of spatial frequencies that is thought to exist in the human visual system [8,9]. Its introduction can greatly reduce the dimensionality of the Hotelling Observer computation, that is another critical aspect of the model. The implementation of channels requires images to be multiplied by a series of channel template images. Considering a total of T channel vectors, the application of each channel vector to the image vector results in one scalar response:

$$v_i = U_i^t g$$

where U_i is the i^{th} channel and v_i is the i^{th} channel response. The final group of channel response is the channelized data vector *v*:

$$v = (v_1, v_2, \cdots, v_T)$$

which has a dimensionality of T * 1.

In the early 1980s a number of researchers suggested Gaussian-modulated sinusoids as models of the receptive fields of simple cells in the visual cortex [2,10]. The Spatio-temporal Gabor elementary functions were proposed and described to include the temporal aspects of vision [10,11]. From these general functions is possible to extract a set of channels $U = (U_1, U_2, \cdots, U_T)$ that can be used with the Hotelling Observer. This process requires a proper selection of the set of channels, characterised by different parameters (i.e. spatial (Δf) and temporal (Δw) bandwidth, orientations (θ) and phases).

2.1.2. Non-Prewhitening Model Observer

There is some evidence that humans cannot prewhiten correlated noise [12] suggesting the introduction of the so-called NPW. Firstly proposed by Wagner et al [13], the test statistic is defined as:

$$\lambda_{NPW} = (\bar{g}_2 - \bar{g}_1)^t g_k = w^t g_k$$

This model is characterized by a test statistic easy to compute and does not require any detailed knowledge of the noise statistics.

To account for differences with human observer studies, Burgess [4] proposed a modification to the NPW observer; this approach, denoted NPWE, combines the model observer with a convolution filter representing the effects of the CSF in the human visual system and is commonly referred to as Eye Filter. The eye filter is specified in the spatial-frequency domain by a radially symmetric transfer function [14].

It takes the form of:

$$E(f_r) = f_r^{\eta} exp(-cf_r^2)$$

where f_r is the spatial frequency and the parameter c was set so that the maximum value of $E(f_r)$ occurs at the peak of the human CSF at approximately 4.0 cycles per degree (cpd) visual angle [4]. Burgess found that $\eta = 1.3$ best fits observer data [4]. The template associated with this mathematical observer is given by:

$$w_{NPWE} = E^t E(\bar{g}_2 - \bar{g}_1),$$

where the matrix E implements the effects of the eye filter.

2.1.2.1. Spatio-temporal eye filter. The Eye filter model derives from the evaluation of the CSF. A series of studies, developed by Kelly et al, also accounts for the temporal domain of the vision, giving a spatio-temporal generalisation of the CSF [15–18].

A first implementation of this CSF, proposed by Burbeck et al., described it as the linear difference between the threshold response surfaces of two mechanisms characterized by a low pass shape [3]: the first, called excitatory mechanism, accounts for the peak sensitivities and the high frequencies fall-off, while the second, inhibitory mechanism, inhibits the first at lower frequencies. In this study, the CSF, as defined by Burbeck, was used as Eye Filter (E_K) with the NPWE, then renaming it $NPWE_K$.

Another attempt to define a spatio-temporal CSF derives from the generalisation of the Eye Filter as already implemented in the NPWE by Burgess et al. [14]. As the spatial and the temporal responses can be considered separable and independent, the temporal response can be defined as:

$$T(f_t) = f_t^{\eta} exp(-kf_t^2)$$

where the parameter k was set so that the maximum value of $T(f_t)$ occurs at approximately 5 Hz (literature report values between 5 and 8 Hz [16]). The final CSF is derived as:

$$E(f_r, f_t) = \|f_r, f\|_t^{1.3} exp(-cf_r^2 - kf_t^2)$$

and used as Eye Filter (E_B) with the NPWE, then renaming it $NPWE_B$. Figs. 1a and 1b shows the x-t values of the two Eye Filters.

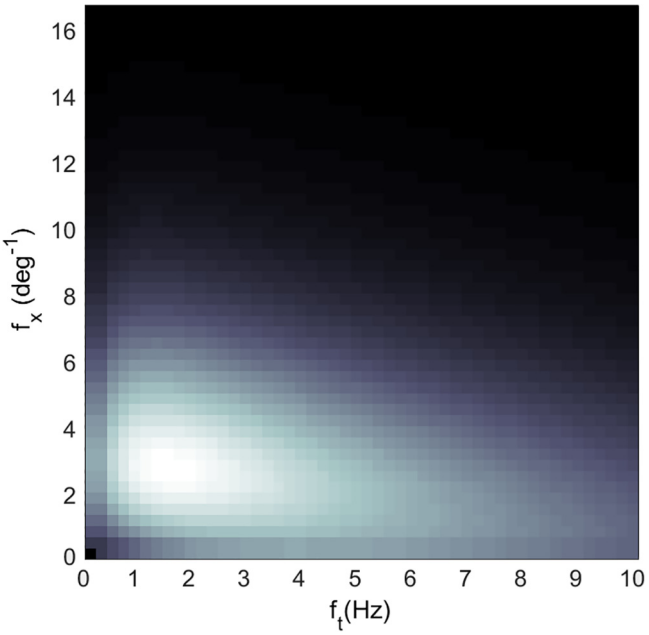

Fig. 1a. Spatio-temporal representation of the Eye Filter function as defined by Kelly's equations.

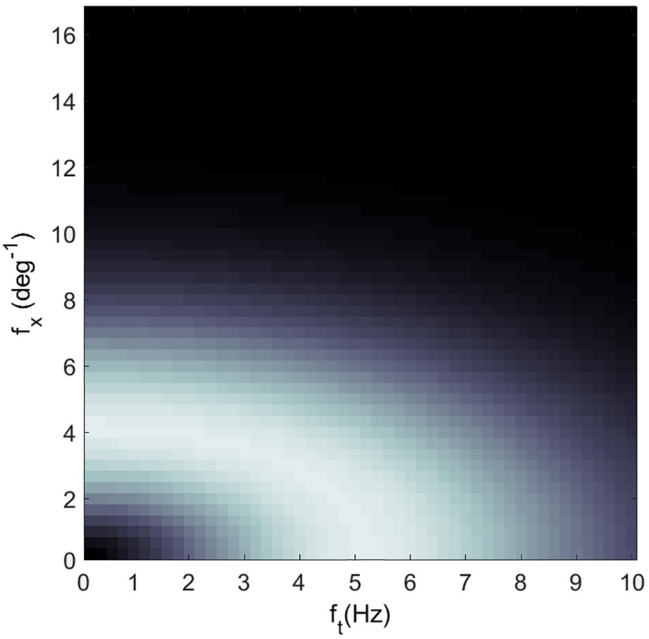

Fig. 1b. Spatio-temporal representation of the Eye Filter function as generalised from Burgess equation.

2.2. Statistical method

The statistical method applied in this paper is based on a work of Chao et al. [19] and already proposed for different applications [20,21]. It is generally acknowledged that Low Contrast Detectability (LCD) is influenced by both the amount of noise and the noise frequency spectrum, hence the standard deviation σ of the pixel values in a homogeneous image is not by itself useful in determining the detection threshold LC_{th}. However, considering several identical regions of interest (ROI) on a uniform background, each with mean pixel value μ, the standard deviation σ_μ is relevant when determining the LC_{th}, as the size of the subROIs sets a noise limit to the spatial frequencies. Since, according to the central limit theorem, σ_μ is normally distributed, a prediction on the detectability of an insert, identical in size to the considered ROI, can be made: the LC_{th} characterised by a PC of 95% is:

$$LC_{th} = F_s \sigma_\mu = 3.29\sigma_\mu$$

Thus, LC_{th} can be derived from standard deviation throughout a simple multiplicative factor (F_s) obtained from the analysis of the normal cumulative distribution function (Φ):

$$\mathrm{PC} = \Phi(x) = \frac{1}{2}\left[1 + erf\left(\frac{F_s}{2\sqrt{2}}\right)\right]$$

where *erf* is the Gaussian error function.

2.2.1. Generalisation introducing noise integration time

A large and growing body of literature has investigated the temporal characteristic of the human visual system; in particular, authors have emphasised the importance of the noise integration time of the human eye-brain system [5]. Overall, these studies fixed the time for noise integration as 200 *ms* [5].

In this approach, information derived from cine images is integrated

and merged over a fixed period of time; in this way, image noise is globally reduced (increasing detectability) even though at the cost of degradation in motion information. The noise integration time is then introduced in the statistical method via a preliminary step in the process. Instead of dealing with individual frames, the statistical method was applied on the average of a certain number of frames, the number depending on the visualised fps. In particular, the averaging process was performed on groups of frames that correspond to the noise integration time of 200 *ms*, e.g. 3 frames with 15 *fps*. The effect of averaging over a group of frames is a reduction of LC_{th}, inversely related to the number of merged frames.

2.3. Human observer experiments

Human observer experiments were conducted in the form of two alternative force choices experiments (2AFC).

A software with graphical user interface, developed in Matlab (Mathworks Inc., Natick, MA), was used to present to the observers a number of observation experiments each with thirty 2AFC trials, with known frame-rate and inserts dimension, but unknown contrast. Dynamic images were displayed side-by-side with approximately 1 cm gap between them (Fig. 2). The observer was request to select the side where the insert was present for each 2AFC trial. The percent correct (PC) for each observer was calculated as the number of correct responses divided by the number of trials.

Four medical physicists were recruited to perform the detection task; observers gained prior knowledge of the insert characteristics through a preliminary training phase, acquiring the required skills. Results of this preliminary phase were excluded from the subsequent observer study.

The experiments were conducted in a darkened room (<10*lux*) and images were displayed on a monitor (NEC MDview232, NEC Display Solutions LTD, Japan) calibrated with the GSDF Dicom Curve.

The display window was fixed at window level of 512 *Pv* and window width of 256 *Pv*, the observer-monitor distance was approximately 50*cm* and a maximum observation period of 1 h per session was imposed to avoid fatigue.

2.4. Acquisition setup

An Allura FD20 (Philips, The Netherlands) angiograph was employed in this study; it is characterized by a dynamic flat-panel, complying with the present standard technological level of angiographic equipment. Although clinical applications require to work with the X-ray tube positioned under the couch table and the detector above the patient for radioprotection optimisation, in this work an inverse configuration was selected; in that way, employed phantoms are easily placed over the detector. A 1.5 *mm* thick layer of Copper, placed directly on the X-ray source, was employed to simulate the X-ray attenuation due to the patient. Being not strictly required, patient table and scatter grid were removed from beam Field of View (FOV) before any acquisition.

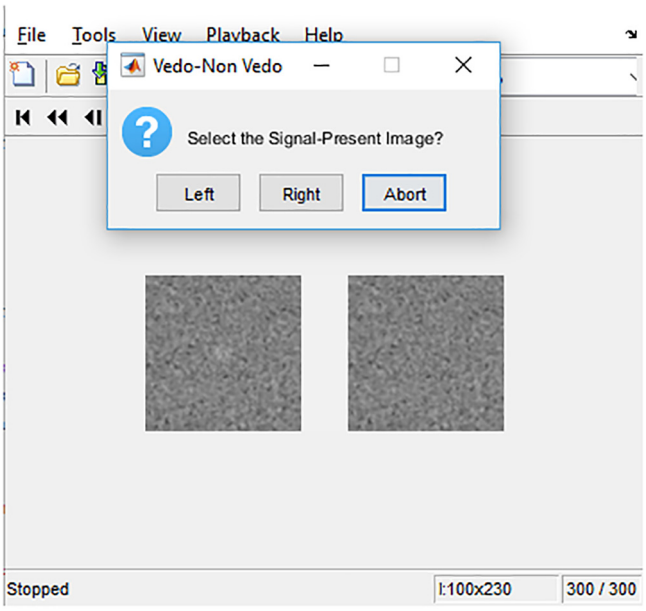

Fig. 2. Screenshot of the displayed images used for the 2AFC experiments.

Table 1
Acquisition parameters of Low Dose Fluoroscopy.

Parameters	Fluoro 1
KVp	72
mA	2
fps	15
Additional filtration	0.4 mmCu + 1mmAl
Dose-rate at detector surface (nGy/frame)	24

2.5. Analysis with simulated images

A 1 cm thick slab of PMMA was used to acquire homogeneous images selecting a low dose fluoroscopy protocol; the slab was placed over the detector and centred within a 37 cm-FOV, allowing the acquisition of the whole slab. Acquisition parameters are reported in Table 1.

Signal-absent images were obtained from a region located in the middle of homogeneous images; signal-present images were then produced by adding a proper template to the homogeneous ones. The template was created starting from a circular detail of the desired size on which unsharpness was introduced by means of a filter to reproduce the spatial resolution properties of the selected acquisition protocol. The modulation function was then evaluated at different spatial frequencies with Drodge Method [22] applied to the circular detail. Images were then multiplied by a scalar factor to obtain different contrast levels; in Fig. 3(a) and (b) is illustrated a visual representation of the process used to create the signal-present images from the mathematical circular detail (a) to the final template (b) after the filtration used to reproduce the spatial resolution properties of the selected protocol.

Signal-present and signal-absent images were saved in Dicom format: final images are characterized by a size of 100x100 px^2, with a pixel size of 0.25 *mm*, and a total extension of 10800 frames, for a total time of 720 *s* (at 15 frames per second).

Due to the high complexity of 2AFC experiments, evaluation and comparison were performed only for three detail dimensions (i.e. 1.25, 2 and 3 *mm*), each characterized by 25 different contrast levels, ranging from 1 to 40 *Pv*. In addition, the study was performed with different frame-rates, in particular 5, 10, 15 (standard) and 30 *fps*, for a total amount of 300 different conditions.

2.5.1. Model observer analysis

A Matlab Code (Mathworks Inc., Natick, MA) was developed to perform the model observer evaluations. Both signal-present and signal-absent images are automatically divided in a series of packets of the duration of 1 sec and processed by the model observer as single images. As previously mentioned, signal blur and contrast are completely known, therefore, model observer templates, usually requiring a subset of images as training, can be derived directly without reducing the number of images used for testing. While both eye filter approaches (E_K and E_B) employed in the NPWE model observer are completely defined from their equations, Gabor channels applied to the Hotelling observer have to be extracted from the general Gabor equation by selecting a

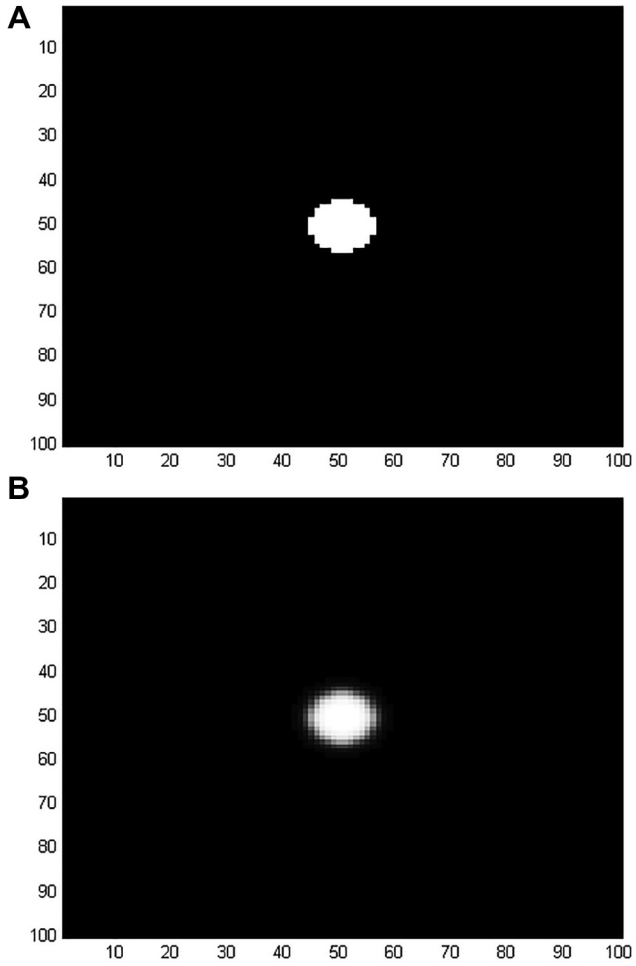

Fig. 3. Visual representation of the process used to create the signal present-images from the mathematical circular detail (a) to the final template (b) after the filtration used to reproduce the spatial resolution properties of the selected protocol.

proper set of parameters. The selected parameters are:

- 4 spatial frequencies bandpass;
- 2 spatial angles orientations;
- 3 temporal frequencies bandpass;
- 1 phase.

with a total of 24 channels.

Detectability scores were evaluated considering an observer viewing distance of 50 cm; this is obtained by scaling the spatial dimensions of the images, in terms of cycles per degree, according to the observer viewing distance. Uncertainty related to the model observer evaluations was estimated by dividing the total amount of images in 9 different sets. Results were then averaged to find the mean detectability scores, with the standard deviation used as uncertainty.

2.5.1.1. Models tuning. Mathematical models have to be tuned to match human observer results. The process of model tuning (i.e. parameter tuning) consists in finding the values for the model parameters that produce the best agreement with human observer detectability scores, in terms of root mean square error (RMSE) and Pearson correlation coefficient. This process required a series of preliminary sets of 2AFC experiments to evaluate the model parameters; therefore, because their uncertainty affects the final calibration reliability, experiments were performed for 3 different conditions with a high number of trials. Table 2 shows the number of experiments and the considered conditions.

Table 2
2AFC experiments used during the tuning phase.

Insert size (mm)	3.25	2	1.25
Insert thickness (mm of air gap)	7	8	20
Total number of 2AFC trials	800	800	800
Fps	15	15	15

2.5.1.1.1. Channelized Hotelling Observer. Due to the high number of degrees of freedom, Gabor Channels were tuned in two different phases.

Firstly, the parameters of the channels, in terms of bandwidth, orientations, and phases, were investigated to achieve the maximum agreement with the human observer performance. These parameters were iteratively changed over fixed ranges to select the best ones in terms of RMSE.

Secondly, the Channel set selected was further tuned with the internal eye noise to achieve the best agreement with the human observer results; the internal eye noise is defined as a multiplicative factor applied directly to the detectability index d':

$$d_c^{'} = \eta d^{'}$$

where η is the internal eye noise and $d_c^{'}$ is the degraded detectability index. Tuning was performed minimizing the RMSE and maximizing Pearson Correlation coeffcient with human observer experiments in terms of internal-eye noise.

2.5.1.1.2. Non-Prewhitening with eye filter model observer. As the CSF is completely defined without any pre-set parameter, the NPWE approaches are tuned simply with the introduction of the internal eye noise η, previously defined for the Channelized Hotelling Observer. Tuning was performed minimizing the RMSE and maximizing Pearson Correlation coeffcient with human observer experiments in terms of internal-eye noise.

2.5.2. Statistical method analysis

Noise integration time is the only parameter of the statistical method that is not completely defined in literature and which has to be properly tuned; the tuning process was performed by estimating the required noise integration time, and consequently the number of frames, that produced the best agreement with the human observer results in terms of RMSE and Pearson correlation coefficient.

Since the model does not take into account the effect of Spatial Contrast Sensitivity Function and Modulation Transfer Function of the system over detectability, results related to different detail dimensions were compared. Since the statistical method resulted to overestimate human observer performances depending on the specific size of the considered detail, it was further tuned by introducing a correction factor related to the specific insert size.

2.5.3. Final comparison with human observer

Tuning parameters of the models and eye internal noise, as evaluated from the previously described analysis, were investigated, in terms of RMSE and Pearson correlation coefficient, in a final comparison to determine further differences with the human observer experiments for all the considered conditions. To obtain high statistical significance, approximately 40,000 decision task experiments were performed. Images and 2AFC experiments employed to tune the mathematical methods were removed from this comparison in order not to affect results.

2.6. Analysis with phantom images

Phantom images were acquired by means of a CDRAD phantom (Artinis Medical Systems, The Netherlands) (Fig. 4); it is designed for testing both the physical properties of radiology systems and the observer's perception. It consists of a square grid with inserts of different

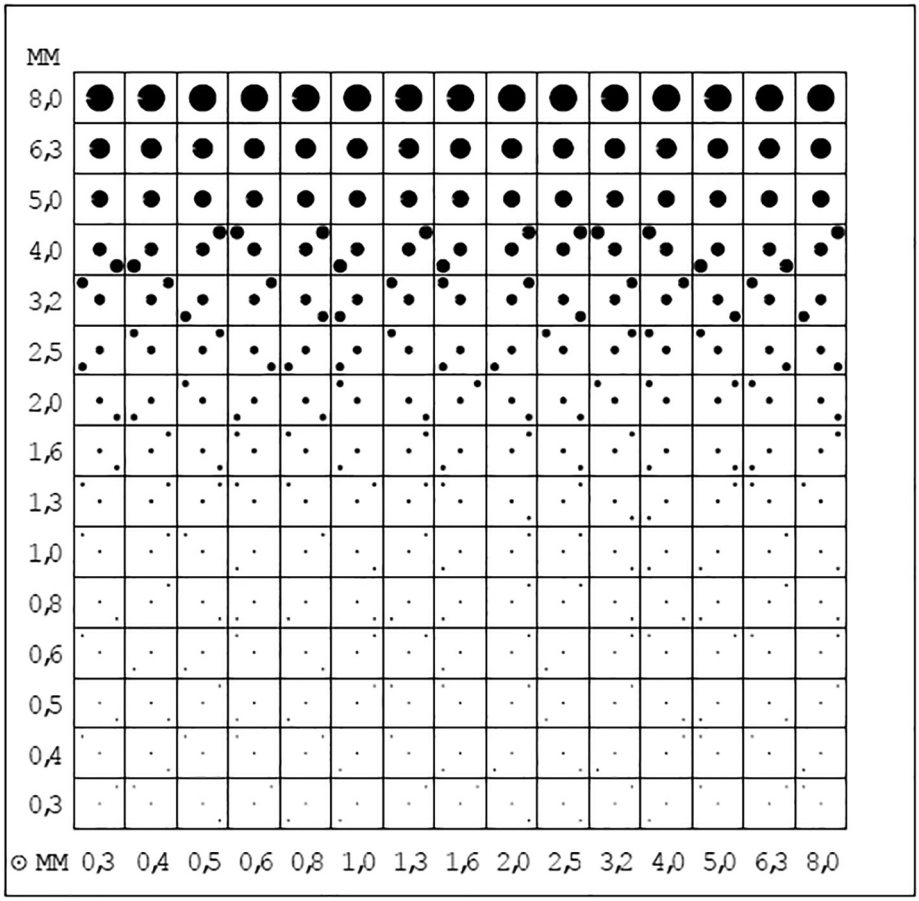

Fig. 4. Artinis CDRAD phantom.

thicknesses, from 0.3 to 8 mm of air gap, simulating different contrasts, and different diameters, from 0.3 to 8 mm.

The phantom was placed directly over the detector with a field of view of 37 cm, allowing the acquisition of the whole phantom. Furthermore, a series of homogeneous images were acquired using a 1 cm PMMA slab, with the same dimensions of the CDRAD phantom. Proper segmentation was applied to the acquired images to obtain signal-present images and thereafter the same segmentation was applied to the PMMA slab images to obtain signal-absent images.

Segmented images are characterized by a size dimension of $28x28\ px^2$ with a pixel size of $0.25\ mm$, and a total time of $160s$; the number of frames is related to acquisition frame-rate, e.g. 2400 and 4800 frames for 15 *fps* and 30 *fps*, respectively. The previously tuned mathematical methods were then applied to the images acquired considering four protocols, characterized by different image quality, fps, and dose:

- 3 different fluoroscopy modes (low, mid and high dose);
- one coronary angiography protocol.

Acquisition parameters are reported in Table 3.

Evaluations and comparison with human observer were performed for three insert dimensions (1.3, 2 and 3.2 *mm*) characterized by 15 different thicknesses (from 0.3 to 8 *mm*). In conclusion, 4 different protocols, each with its specific frame rate (Table 3), 3 different insert dimensions and 15 different thicknesses were investigated, for a total amount of 180 different conditions.

2.6.1. Model observer analysis

The same Matlab code as before was employed to perform the model observer analysis. Both signal-present and signal-absent images (160 s of total time) were automatically divided in a series of packets of the duration of 1 s and processed by the model observer as single images. Since the signal blur and contrast are not completely known, template evaluation required a subset of images as training: images were randomly divided into two separate groups for training and test (in our analysis 50% of the total for each group). This process was repeated 9 times to estimate the uncertainty related to the evaluation. Model parameters, such as eye internal noise η estimated with simulated images, was then applied to obtain the final Proportion Correct.

2.6.2. Statistical method evaluation

A correction factor related to insert size, obtained from the analysis

Table 3
Acquisition parameters of the different protocols employed in the analysis with Phantom Images.

Parameters	Fluoro 1	Fluoro 2	Fluoro 3	Coro
kVp	72	61	63	63
mA	2	6	6	235
fps	15	15	30	15
Additional filtration	0.4 mmCu + 1mmAl	0.1 mmCu + 1mmAl	0.1 mmCu + 1mmAl	–
Dose-rate at detector surface (nGy/frame)	24	30	20	93

of the simulated images, was applied.

2.6.3. Final comparison with human observer

The different mathematical methods were compared with the human observer results of the 2AFC experiment. To obtain high statistical significance, approximately 21,000 decision task experiments were performed. The RMSE and Pearson correlation coefficient was investigated to evaluate the reliability of the characterization obtained with simulated images.

3. Results

3.1. Simulated images

3.1.1. Models tuning

The tuning process of Channelized Hotelling Observer in terms of channels parameters allows to select the optimal parameter set; from our analysis, the selected parameters resulted:

- 4 spatial frequencies bandpass ([1/8, 1/4], [1/4, 1/2], [1/2, 1] and [1, 2] cycles/degree);
- 2 spatial angles orientations (0 and 1/2π);
- 3 temporal frequencies bandpass ([1/8, 1/4], [1/4, 1/2], [1/2, 1] and [1, 2] cycles/sec);
- 1 phase (0°).

with a total of 24 spatio-temporal Gabor channels.

Table 4 shows the results of the analyses for each model observer with the eye internal noise η values corresponding to the best agreement in terms of RMSE: Channelized Hotelling observer required a lower correction compared to both $NPWE_B$ and $NPWE_K$, $NPWE_K$ being the approach characterized by higher correction. The good agreement is also confirmed by the Pearson correlation coefficient, reported in Table 4. The uncertainty related to the Proportion Correct results 5% and 7% for the CHO and both NPWE approaches, respectively.

Results of the tuning process of the Statistical method are reported in Table 5. Optimal noise integration time underlines a clear dependency from the detail dimension: the optimal noise integration time related to small details is significantly smaller than that related to larger inserts. Analysis seems to indicate the presence of an effect related to the detail dimension, caused by both the Modulation transfer function of the equipment and the spatial contrast sensitivity function.

After this evaluation, the optimal noise integration time for all the detail dimensions was selected as the one estimated for the larger detail (i.e. 200 ms); the agreement with human observer results has been obtained introducing a further tuning factor related to the detail size applied directly on the LC_{th}. Results of the tuning process, in terms of RMSE and Pearson correlation coefficient, are reported in Table 5.

3.1.2. Evaluation and final comparison with human observer

Fig. 5(a) and (b) show the results for a visualised frame-rate of 15 *fps*. Plots underline a good agreement between the different methods.

Quantitative analyses of the comparison between model observer methods and human 2AFC experiments were reported in Fig. 6(a) and (b) in terms of RMSE and Pearson Correlation Coefficient, respectively. Results highlight a good agreement for every model observer considered. More in detail, Channelized Hotelling Observer shows the lowest differences with the human experiments, compared to both NPWE approaches. Nevertheless, the mean difference results to be lower than 10% in every considered condition. Results of the analysis with the Pearson Correlation Coefficient (Fig. 6(b)) confirmed the good agreement.

Similar considerations can be proposed for the Statistical method;

Table 4
Model Observer results of the tuning process.

	CHO	$NPWE_B$	$NPWE_K$
Internal eye noise (η)	1	1.32	1.79
RMSE	0,78%	3,81%	0,90%
Pearson correlation coefficient	0,999	0,536	0,970

Table 5
Statistical method results of the tuning process. Correction factors related to different insert dimensions were introduced to consider the effect of the spatial contrast sensitivity function.

		Optimal Noise Integration Time (ms)	Selected Noise Integration Time (ms)	Correction Factor
Insert Size	3.25 mm	200	200	1
	2 mm	156		1.13
	1.25 mm	90		1.45
RMSE		0,34%		
Pearson correlation coefficient		0,999		

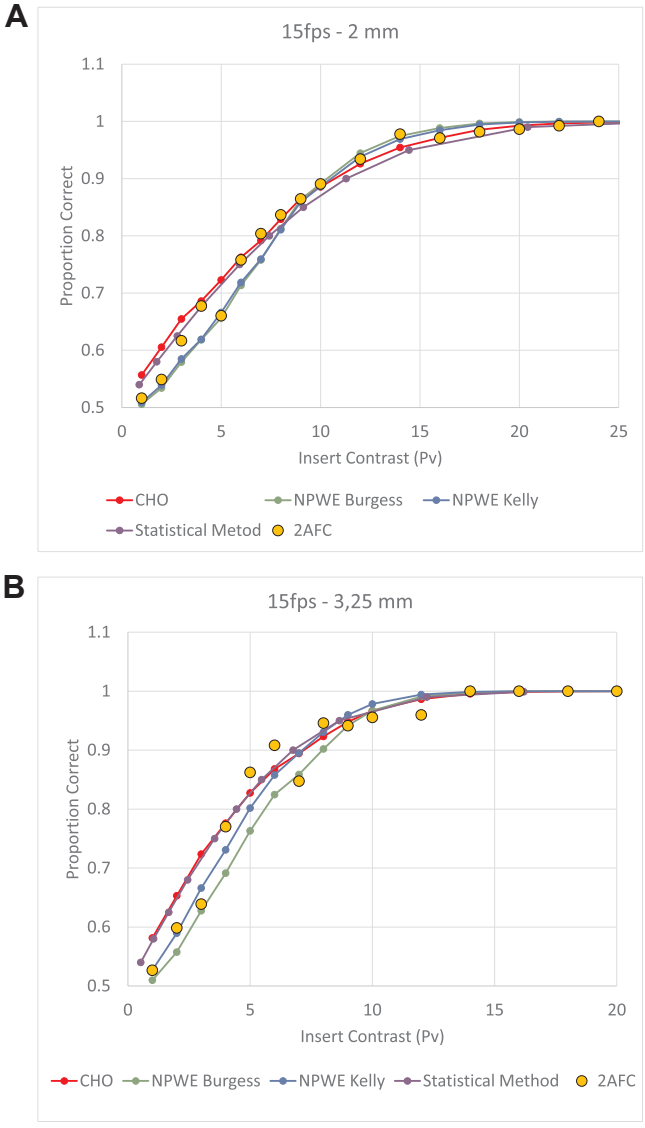

Fig. 5. Comparison of the different approaches with a frame-rate of 15 fps related to an insert of 2 mm (a) and 3.25 mm (b).

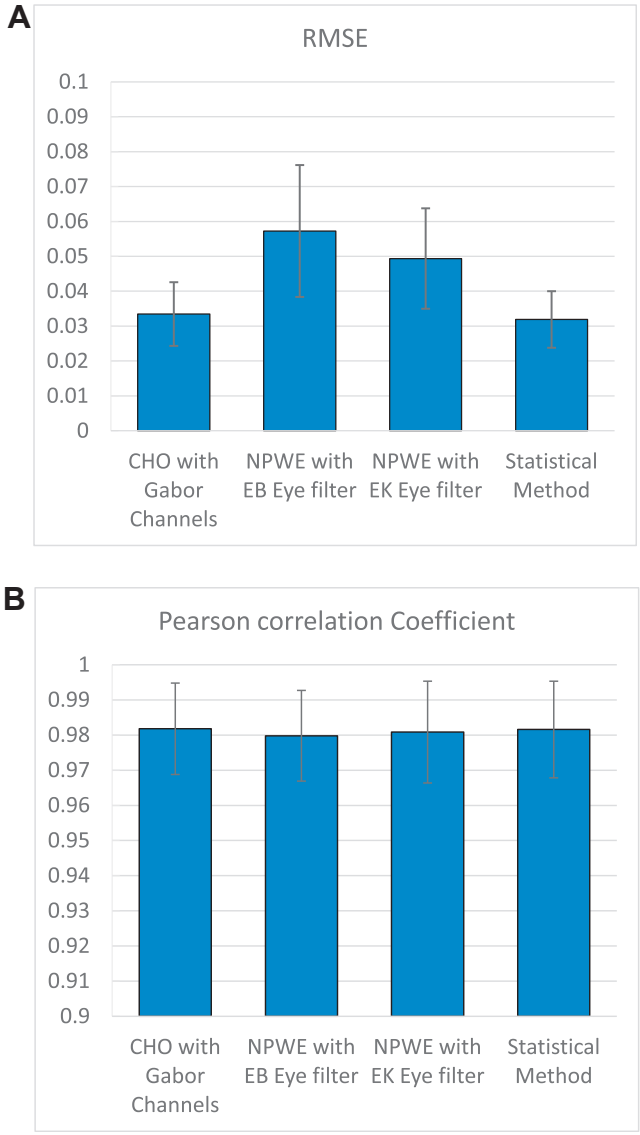

Fig. 6. Comparison between 2AFC experiments and mathematical approaches with simulated images: (a) and (b) show RMSE and Pearson Correlation Coefficients, respectively. Histograms shows the mean values, while error bars represent the range of variability between the minimum and maximum values.

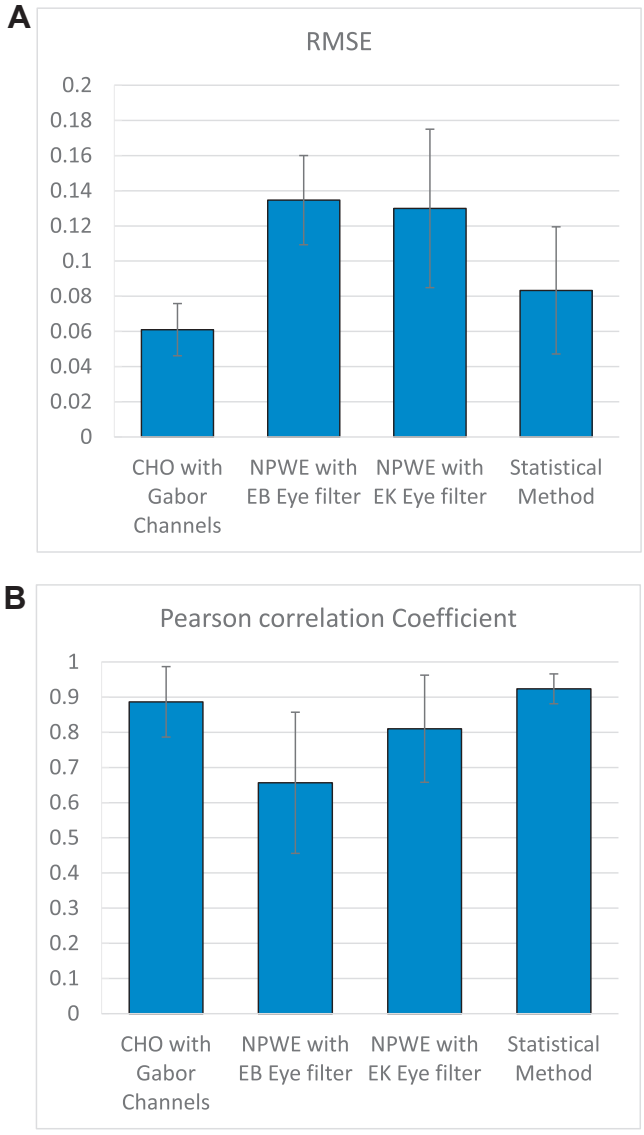

Fig. 7. Comparison between 2AFC experiments and mathematical approaches with phantom images: (a) and (b) show RMSE and Pearson Correlation Coefficients, respectively. Histograms shows the mean values, while error bars represent the range of variability between the minimum and maximum values.

comparison between LC_{th} corrected with the factor previously evaluated and human 2AFC experiments are summarized in Fig. 6(a) and (b). As already noticed, while the agreement is in some way related to the detail dimension, the RMSE results to be less than 6% in every considered condition.

3.2. Phantom images

Fig. 7(a) and (b) report a quantitative analysis of the comparison between model observer methods and 2AFC experiments in terms of RMSE and Pearson correlation coefficients, respectively. Although the differences are generally higher when compared to the analyses with simulated images (Fig. 6), results highlight a good degree of agreement. More in detail, CHO shows the lower difference with the human experiments, compared to other models, (lower than 8%). Nevertheless, the mean difference results lower than 17% in every considered condition. The uncertainty related to the Proportion Correct is 8% and 10% for the CHO and NPWE, respectively.

In Fig. 7(a) and (b) are also reported the results related to the Statistical method; LC_{th} was corrected with a tuning factor related to the insert size derived from the ones previously evaluated for simulated images. Histograms show RMSE and Pearson correlation coefficients between the statistical method and 2AFC experiments. Analyses show different degree of agreement: smaller inserts are characterized by a higher agreement compared to larger ones; nevertheless, differences result lower than 12% for every considered protocol and insert size.

4. Discussion

The characterization of the proposed mathematical methods highlights two main aspects.

Firstly, mathematical methods show a general overestimation of the detectability score as compared to the human observer experiments, therefore a tuning process is required to achieve a good degree of agreement. Secondly, there is high variability among the human observer 2AFC experiments. Since the degradation factor for Model Observer methods was evaluated from a direct comparison of the detectability, it obviously suffers from the intrinsic variability of human

2AFC experiments. In these analyses, the average value was selected to represent a good estimation of the correction factor. Nevertheless, other degradation-factor models could be introduced to compensate for the differences with the human observers.

Literature proposes different methods, characterized by different degrees of complexity [23,24]. These methods allow a better estimation of the effects of internal eye noise, but at the cost of a reduction of their robustness: compared to the proposed method, they can suffer from human-observer experiment variability, thus making the entire tuning phase highly unstable. A single factor applied directly on the detectability score is probably a good compromise for a first evaluation of the proposed generalisation of Model Observer. Furthermore, Channelized Hotelling Observer is characterized by a series of degrees of freedom (i.e. channel parameters) requiring laborious tuning to match the human observer performances. Since the channels do not form a set of orthogonal bases, the design of the channel bank is quite arbitrary and dependent on the particular properties of the images used in the study. Without a prior definition of some constraints (e.g. the number of channels), different combinations of the channel parameters, though characterised by the same RMSE, can be selected during the tuning phase [25].

Similar considerations can be applied to the Statistical Method; the evaluated number of averaged frames as a function of visualized frame-rate (*fps*) highlights a dependency on the insert dimension. The higher correction factor needed with smaller inserts can be seen as an effect of the spatio-temporal contrast sensitivity function. The different correction for different frame-rates suggests the coexistence of two separable effects to be considered: the first effect, related to temporal processing characteristics of the human eye-brain system, was taken into account throughout the preliminary averaging step, the latter is due to the spatial contrast sensitivity function; in fact, the relative sensitivity at different spatial frequencies affects the Low Contrast Detectability scores for different insert dimensions. Thus, as not already implemented in the model, the relative sensitivity at different spatial frequencies must be considered to obtain an unbiased evaluation of the LC_{th}.

Different approaches can be implemented to deal with the relative sensitivity at different spatial frequencies, i.e. introducing a further preliminary step: single frames, before the averaging process (used to simulate the effects of the temporal noise-integration), can be filtered with the spatial contrast sensitivity function. In this work, correction factors were investigated for 3 insert dimensions only, however, results suggest the possibility of drawing a curve from which these factors could be derived for any insert dimension. Further analyses based on other insert dimensions are required to investigate the behaviour of this curve. The Statistical Method is characterized by a simple approach that requires just a single acquisition to assess LC_{th} and is far less time-consuming compared to the proposed Model Observers. Thus, these properties make it an interesting and feasible approach to assess LCD in a periodical constancy check program.

Another important consideration is related to the human observer 2AFC experiments.

As noted previously, this approach is highly time-consuming and characterized by intrinsic variability. Any additional condition to be investigated (as further insert dimension or visualized frame-rate) requires dealing with the increase of the total number of trials, or, on the other hand, the reduction of their numerosity in every considered condition; thus, a preliminary definition of a limited but meaningful set of conditions is mandatory in order to make the human observer experiment feasible in terms of total 2AFC trials performed with an acceptable degree of variability. Alternatively, an increase of the number of observers employed in the experiments, would be necessary to improve the robustness of the evaluation.

5. Conclusion

The proposed generalisations of mathematical methods for Low Contrast Detectability evaluation in dynamic angiographic images show good results. The agreement with human observer experiments proves the feasibility of the proposed generalisations, aimed to reproduce the effects of visualised frame-rates on human detection performances. In conclusion, they could be introduced for a deeper modalities-characterisation, in terms of image quality, or of clinical protocol optimization. After a proper preliminary tuning phase with a series of human observer experiments, the proposed methods can be employed to compare the image quality of equipment even from different vendors. In addition, these mathematical methods can be applied not only on angiographic equipment, but also on any modality characterized by spatio-temporal images.

References

[1] Noferini L, Taddeucci A, Bartolini M, Bruschi A, Menchi I. CT image quality assessment by a Channelized Hotelling Observer (CHO): application to protocol optimization. Phys Med 2016;32(12):1717–23.
[2] Marcelja S. Mathematical description of the responses of simple cortical cells. J Opt Soc Am 1980;70(11):1297–300.
[3] Burbeck AA, Kelly DH. Spatiotemporal characteristics of visual mechanisms: excitatory-inhibitory model. J Opt Soc Am 1980;70(9):1121–6.
[4] Burgess AE. Statistically defined backgrounds: performance of a modified nonprewhitening matched filter model. J Opt Soc Am 1994;11(4):1237–42.
[5] Burgess AE. The Rose model revisited. J Opt Soc Am 1999;16(3):633–46.
[6] Swets J, Pickett RM. Evaluation of diagnostic systems: methods from signal detection theory. New York: Academic; 1982.
[7] He X, Park S. Model observers in medical imaging research. Theranostics 2013;3(10):774–86.
[8] Myers KJ, Barrett HH. Addition of a channel mechanism to the ideal-observer model. J Opt Soc Am 1987;4(12):2447–57.
[9] Sachs MB, Nachmias J, Robson JG. Spatial-frequency channels in human vision. J Opt Soc Am 1971;61(9):1176–86.
[10] Daugman JG. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. IEEE Trans Acoustics Speech Signal Process 1988.
[11] MacLennan B. Gabor representation of spatiotemporal visual images. Technical Report CS-91-144 Computer Science Department, University of Tennessee Knoxville, TN, 1991.
[12] Burgess AE, Li X, Abbey CK. Visual signal detectability with two noise components: anomalous masking effects. J Opt Soc Am 1997;14(9):2420–42.
[13] Wagner RF, Brown DG, Pastel MS. Application of information theory to the assessment of computed tomography. Med Phys 1979;6(2):83–94.
[14] Barten PGJ. The SQRI method: a new method for the evaluation of visible resolution on a display. Proc Soc Inf Disp 1987;28(3):253–62.
[15] Kelly DH. Motion and vision I. Stabilized images of stationary gratings. J Opt Soc Am 1979;69(9):1266–74.
[16] Kelly DH. Motion and vision II. Stabilized spatio-temporal threshold surface. J Opt Soc Am 1979;69(10):1340–9.
[17] Kelly DH. Motion and vision III. Stabilized pattern adaptation. J Opt Soc Am 1980;70(11):1283–9.
[18] Kelly DH. Motion and vision IV. Isotropic and anisotropic spatial responses. J Opt Soc Am 1982;72(4):432–9.
[19] Chao EH, Toth TL, Bromberg NB, Williams EC, Fox SH, Carleton DA. A statistical method of defining low contrast detectability. Radiology 2000;217:162.
[20] Spadavecchia C, Villa R, Pasquali C, Paruccini N, Oberhofer N, Crespi A. A statistical method for low contrast detectability assessment in digital mammography. International workshop on breast imaging. Springer International Publishing; 2016. p. 532–9.
[21] Paruccini N, Villa R, Pasquali C, Spadavecchia C, Baglivi A, Crespi A. Evaluation of a commercial Model Based Iterative reconstruction algorithm in computed tomography. Phys Med 2017;41:58–70.
[22] Droege RT, Morin RL. A practical method to measure the MTF of CT scanners. Med Phys 1982;9(5):758–60.
[23] Zhang Y, Pham BT, Eckstein MP. Evaluation of internal noise methods for Hotelling observer models. Med Phys 2007;34:3312–22.
[24] Pereira NF, Gifford HC, King MA. A noise mechanism for model observers. IEEE nuclear science symposium conference record. 2008. p. 4433–9.
[25] Favazza CP, Fetterly KA, Hangiandreou NJ, Leng S, Schueler BA. Implementation of a channelized Hotelling observer model to assess image quality of X-ray angiography systems. J Med Imag 2015;2(1). 015503-1-12.