A Perceptually relevant Channelized Joint Observer (PCJO) for the detection-localization of parametric signals

Lu Zhang, Christine Cavaro-Ménard, Patrick Le Callet, Jean Yves Tanguy

A Perceptually relevant Channelized Joint Observer (PCJO) for the detection-localization of parametric signals

Lu Zhang*, Christine Cavaro-Ménard, Patrick Le Callet, and Jean-Yves Tanguy

***Abstract*—Many numerical observers have been proposed in the framework of task-based approach for medical image quality assessment. However, the existing numerical observers are still limited in diagnostic tasks: the *detection* task has been largely studied, while the *localization* task concerning one signal has been little studied and the *localization* of multiple signals has not been studied yet. In addition, most existing numerical observers need *a priori* knowledge about all the parameters of the under-detection signals, while only a few of them need at least two signal parameters. In this paper, we propose a novel numerical observer called the Perceptually relevant Channelized Joint Observer (PCJO), which cannot only detect but also localize multiple signals with unknown amplitude, orientation, size and location. We validated the PCJO for predicting human observer task performance by conducting a clinically relevant free-response subjective experiment in which six radiologists (including two experts) had to detect and localize Multiple Sclerosis (MS) lesions on Magnetic Resonance (MR) images. By using the jackknife alternative free-response operating characteristic (JAFROC) as the figure of merit (FOM), the detection-localization task performance of the PCJO was evaluated and then compared to that of the radiologists and two other numerical observers - Channelized Hotelling Observer (CHO) and Goossenss CHO for detecting asymmetrical signals with random orientations. Overall, the results show that the PCJO performance was closer to that of the experts than to that of the other radiologists. The JAFROC1 FOMs of the PCJO (around 0.75) are not significantly different from those of the two experts (0.7672 and 0.7110), while the JAFROC1 FOMs of the numerical observers mentioned above (always over 0.84) outperform those of the experts. This indicates that the PCJO is a promising method for predicting radiologists' performance in the joint detection-localization task.**

***Index Terms*—Numerical observer, image quality assessment, Magnetic Resonance Imaging (MRI), Human Visual System (HVS), detection task, localization task,**

I. Introduction

MEDICAL image quality assessment is critical and indispensable for comparing and optimizing medical imaging systems such as acquisition systems, image post-processing systems and visualization systems [1]. Since the ultimate goal of medical images is to aid clinicians in rendering a diagnosis, it is widely accepted that in order to optimize diagnostic decisions image quality should be as good as possible. Additionally, image quality should be assessed in the context of a specific diagnostic task, the so-called *task-based approach* [2], [3]. During the assessment process, one or more diagnostic tasks could be performed by either human observers (radiologists) or numerical observers (mathematical models). Using human observers is generally difficult, time consuming and expensive [4]; besides, between and within their responses variance exists [5]. Conversely, numerical observers are much easier, faster and cheaper to carry out, and always yield the same results. Thus, it has been argued that numerical observers are useful for medical image quality assessment in place of human observers [2], [6]. However, in most cases, the task performance of human observers is the ultimate test of image quality and is needed for the validation of a numerical observer [3].

Once the numerical observer is validated, the number of quality evaluation tests without radiologists can be increased. The comparison of numerous systems with different parameters then allows to define the best adapted imaging systems. The optimal system enables radiologists to make a reliable diagnosis at a minimum cost and in the shortest possible time.

It may be useful to well characterize the diagnostic process in order to model it. One commonly accepted approach that we find in the literature [7]–[9] is to divide the diagnostic process into three tasks: detection, localization and characterization.

- The *detection task* simply requires a confidence rating concerning the presence of a signal, e.g. a lesion.
- The *localization task* consists in indicating the location of each lesion.
- The *characterization task* is related to the analysis of the different elements of the lesions (contour, texture, etc.) for the differential diagnosis, and normally involves a linguistic response (anatomical, tissular, etc.), which is the most complicated task. Part of the characterization task is sometimes called the *estimation task* [2], which means estimation of one or a few parameters of the lesion to be detected, and of direct relevance to the purpose for which the image was obtained. For example, a lesion may first be detected, then its size is estimated to determine the stage of the disease. When estimation and detection tasks are combined, the overall procedure is termed a

*L. Zhang and C. Cavaro-Ménard are with the Laboratory Lisa, University of Angers, 62 avenue Notre Dame Du Lac, 49000 Angers, France and the Medical Imaging Departement, Hospital of Angers, 4 Rue Larrey, 49933 Angers, France (e-mail: lu.ge@etud.univ-angers.fr).

P. Le Callet is with the Polytech Nantes/IRCCYN, University of Nantes, Rue Christian Pauc, 44300 Nantes, France.

J. Y. Tanguy is with the Radiology Department, Hospital of Angers, 4 Rue Larrey, 49933 Angers, France.

joint estimation-detection task.

So far to our knowledge, numerical observers are limited in task range: most of them are dedicated to the detection task [10]–[14], much fewer are concerned with signal localization in images that contain at most one signal [15], none is concerned with signal localization in images that contain more than one signal, and none for the characterization task. In this paper, we propose a new numerical observer for the signal detection-localization in images that contain multiple signals.

In Section II, we first present the ROC analyses that are needed to quantify, characterize and compare the task performance given observers' responses in a certain task, as well as a brief review of the existing numerical observers in the literature by addressing the problems to resolve. To overcome the limitations, we propose a novel numerical observer - Perceptually relevant Channelized Joint Observer (PCJO) - that is detailed in Section III and Section IV. The protocol of a free-response subjective experiment on MR images in order to evaluate the PCJO's detection-localization performance, and the performance evaluation method is illustrated in Section V. The performance results and discussions are reported in Section VI. Finally, a conclusion and a perspective are presented in Section VII.

II. ROC ANALYSES AND EXISTING NUMERICAL OBSERVERS

A. ROC and its variants

The Receiver Operating Characteristic (ROC) curve and its variants have been proposed to describe different diagnostic task performances, as summarized in Figure 1: ROC curve [16] for the detection task; Localization ROC (LROC) curve [17] for the detection-localization task of one signal on an image; Estimation ROC (EROC) curve [18], [19] for the joint detection-estimation task; Free-response ROC (FROC) curve [20] and Alternative Free-response ROC (AFROC) curve [21] for the detection-localization task of multiple signals on an image. Different figures of merit (FOMs) have been developed based on the ROC/LROC/EROC/FROC/AFROC analyses to characterize and quantify the diagnostic task performance. FOMs can be the area under the ROC curve (AUC), the area under the LROC curve (LAUC), the area under the empirical FROC curve, the augmented area under FROC curve, etc.

Since the numerical observer proposed in this paper detects and localizes multiple possible signals on an image, we considered using FROC or AFROC for the detection-localization task performance evaluation of our numerical observer in comparison with that of radiologists.

Within the FROC paradigm, the "area under the empirical FROC curve" is not a good FOM to summarize FROC curve, since a larger value of area under the empirical FROC curve can result either from an increase in True Positive (TP) responses with correct localization or an increase in the number of False Positive (FP) responses on each image [22]; likewise, the "augmented area under FROC curve" [20] should be avoided when the lesion density is really low (big image size with a small quantity of lesions) since the FOM would depend only on the degree of localization accuracy, but has nothing to do with the rating performance.

Within the AFROC paradigm, two jackknife alternative free-response operating characteristics JAFROC1 and JAFROC2 [23] have been proposed. They differ in how to count the number of highest rated FPs. While JAFROC2 counts only the highest rated FPs on actually negative images, JAFROC1 counts the number of highest rated FPs on actually negative and actually positive images. Because more cases are used to form the estimate of fraction of FPs (FPF) in JAFROC1, the values tend to be more stable and the statistical power is higher than JAFROC2 [24]. To our knowledge, JAFROC1 is the FOM that has the highest statistical power for the detection-localization task up to now. Note that JAFROC1 chance value is close to zero [25], since the chance of hitting a lesion approximates the lesion area divided by the clinical image area. For small lesions, it approximates 0%. It is not like ROC, where localization is not required and the chance value is 50%.

Therefore, in this paper,we chose JAFROC1 to evaluate the detection-localization task performance of our numerical observer versus that of radiologists. Note that although the FROC/AFROC paradigm has already been widely used to characterize the human observer performance, this is the first time that a FROC/AFROC method is used to evaluate the performance of a numerical observer (since no other numerical observers has been used before to process the signal localization in images that contain multiple signals).

B. Existing numerical observers

Current numerical observers fall into two families: human visual system (HVS) model and model observer (MO). The two families all need signal present and absent images. The difference between them is that MO backgrounds are not identical for signal present and absent cases.

The HVS models used in the evaluation of medical image quality [10]–[12], [26]–[28] belong to the *perceptual difference model* (PDM) [29]–[33] focusing on the spatial contrast detection. They are especially efficient for the detection of near-visibility-threshold distortion, suitable for the diagnostic task performance evaluation where the detection targets are not too conspicuous. All PDMs have two input images: a *reference image* and a *distorted image*. The two images are firstly transformed into the luminance domain (in cd/m^2). After being processed by several functions modeling various HVS properties, the two input images can be interpreted as two perceptual images. Their difference is thus a perceptual difference map. A spatial pooling function is applied on the map to produce a scalar overall score that reflects the detection confidence rating. The PDMs differ by their function definitions.

It is important to note that in psychology, human vision processing can be conventionally separated into two stages: *sensation* and *perception* [34]. Sensation is viewed as a function of low-level biochemical and neurological events. The perception is a dynamic cognitive process of interpreting

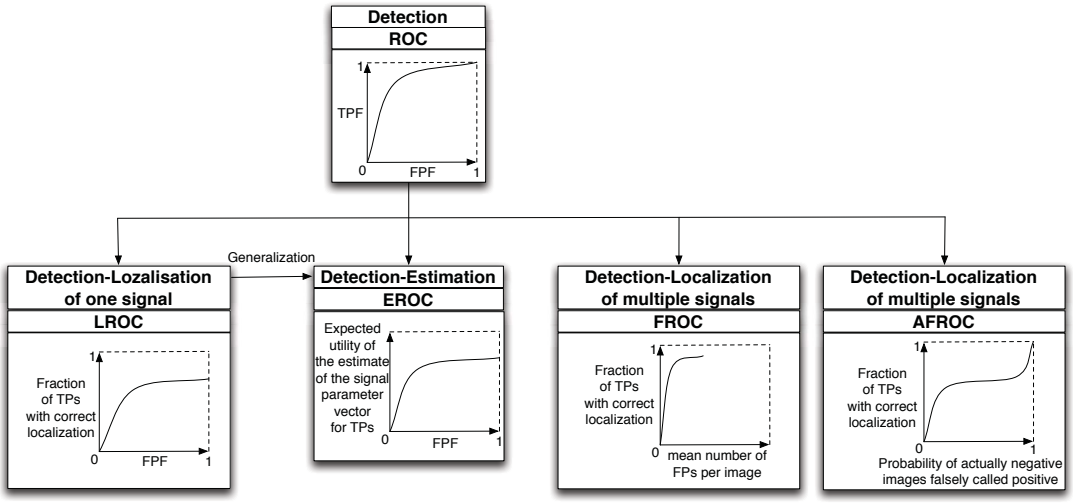

Fig. 1. Different figures of merit (FOMs) have been proposed for different diagnostic tasks to characterize the task performance.

information. Existing HVS models are mostly based on the sensation modeling, while the perception modeling is much more complex, as it depends on various factors and is tough to identify. For example, in the medical domain, the perception stage should integrate the radiologists' expertise, the given modality, the studied pathology, etc.

On the other hand, MOs [13], [14], [35]–[38] for the detection task compute a scalar test statistic via a discriminant function of the test image and they differ by their discriminant functions. The ideal observer (IO) [13] achieves the maximum AUC attainable amongst all the MOs, but requires full knowledge of the probability density functions (PDFs) of the image for each hypothesis (presence or absence of a signal) which are difficult to compute except for some special cases. The Hotelling observer (HO) [39] requires only the first- and second- order statistics of the image data for each hypothesis.It maximizes the Signal-to-Noise Ratio (SNR) among all linear observers. This is equivalent to maximizing AUC if the statistics are Gaussian. But the HO needs the inverted ensemble covariance matrix of the image which is hard to calculate in the face of modern image high-dimensionality [2]. The channelized Hotelling observer (CHO) [40] emerged as required to reduce the dimentionality through *channelization*, a technique that decomposes an image into different channels. To correlate better with human performance, one possibility is to integrate the intrinsic spatial frequency and orientation selectivity of HVS into the channel profile by using the CHO paradigm [6]. This type of CHOs has successfully predicted human performance in *signal-known-exactly* (SKE) experiments [2], where the signal parameters (e.g. intensity amplitude, size, shape, orientation and location) are exactly known to observers and do not vary throughout the entire experiment.

However, SKE is not clinically realistic because in clinical routine human observers do not exactly have the *a priori* knowledge about the signal parameters before the diagnostic analysis. Moreover, for pathologies with multiple lesions, the signal parameters differ for each lesion depending on the stage of the lesion, the background tissue characteristics, etc. Several *signal-known-statistically* (SKS) experiments (where the signal attributes are known statistically) showed that the uncertainty and variability in signal parameters can significantly influence human performance [15], [41], [42]. This suggests that the SKS MOs may have a great potential to be closer to human observers and they have begun to draw more attention in the last decade. In case of the detection of one signal with varying size and shape, Eckstein *et al.* [43]–[45] proposed to establish different templates for different combinations of signal parameters and to output the optimal template response (in a maximum likelihood sense) as the final test statistic. In the case of one signal with varying location (joint detection-localization task), Park *et al.* [15] proposed a *scanning* CHO which exhaustively scans the image, then the location giving the largest test statistic is chosen as the tentative location while that test statistic is the final one. Note that the required number of training images and the calculation amount increase rapidly with the number of possible signal parameters for Eckstein's method; they also increase rapidly with the image size for Park's scanning CHO. For the detection of one signal with varying orientation, size or amplitude, Goossens [46] proposed a series of SKS CHOs based on the joint detection and estimation (JDE) theory [47]. He evaluated the SKS CHO for signal with varying orientation only compared to IO performance [48], but not to human performance. For the joint detection and estimation task, Clarkson [18] proposed a theoretical framework of the ideal EROC observer, whose EROC curve lies above those of all other observers for the given joint detection estimation task. Theoretically, within his framework, all signal parameters could be estimated along with the calculation of the test statistic, but practically, this has not been tackled yet. For the pure estimation task, Whitaker et al. [49] proposed a scanning-linear estimator (SLE) which performs a global-extremum search to maximize a linear metric. They evaluated its estimation performance when the amplitude, size and location of the only one possible

signal are known statistically, but they did not investigate the detection task performance.

Considering the above overview and analysis about the existing numerical observers, we remark that:

- the SKS MOs have not been extensively studied yet;
- the range of variable signal parameters of the existing SKS MOs still has to be widened;
- no numerical model has been proposed for the localization task of multiple signals per image;
- the perception stage needs to be considered in the numerical observer.

In order to represent the routine diagnostic process more realistically, we propose a novel nonlinear numerical observer - PCJO for the detection-localization of multiple parametric signals with random amplitude, orientation, size and location. The only known signal parameter for the PCJO is the signal shape, bearing in mind that Castella [42] reported that uncertainty and variability in signal shape do not influence human performance significantly. The PCJO also involves the perception stage to some extent.

III. Channelized Joint detection and estimation Observer (CJO) for the detection task

Before including the localization task, let us first look at the detection task of one single signal with fixed location on a noisy background, for which we implemented an SKS MO in the presence of random signal amplitude, size and orientation. This is an extension of Goossens' SKS CHO [46], with several important modifications.

The detection problem can be seen as the validation of one of two exclusive hypotheses, $\mathcal{H}_0$ (signal is absent) and $\mathcal{H}_1$ (signal is present). The observed image data $\boldsymbol{g}$ is thus given by:

$$\mathcal{H}_h:\ \boldsymbol{g} = h\boldsymbol{x} + \boldsymbol{b},\ h = 0, 1 \tag{1}$$

where the signal is denoted as $\boldsymbol{x}$, the background as $\boldsymbol{b}$, and the absence or presence of the signal is controlled by the binary variable h. Here a 2D image (test image data, signal and background) is represented as column vectors through vertical concatenation. Let the observed image have M pixels, $\boldsymbol{g}$ is then an $M \times 1$ vector.

A. Background and signal models

The correlated Gaussian background model in Eq.(2) is adopted in the CJO for its simplicity and calculability:

$$\boldsymbol{b} \sim \mathcal{N}(\mu_{\boldsymbol{b}}, \Sigma_{\boldsymbol{b}}). \tag{2}$$

where $\mu_{\boldsymbol{b}}$ is the mean and $\Sigma_{\boldsymbol{b}}$ is the covariance matrix.

The signal is modeled by a 2D Gaussian function:

$$[\boldsymbol{x}]_{\boldsymbol{p}} = a\, \exp\left(-\frac{1}{2}\left(\boldsymbol{p}-\boldsymbol{q}\right)^{\mathrm{t}} \mathbf{A}^{\mathrm{t}}\mathbf{D}^{-1}\mathbf{A}\left(\boldsymbol{p}-\boldsymbol{q}\right)\right) \tag{3}$$

where

$$\mathbf{D} = \begin{bmatrix} b\sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \tag{4}$$

$$\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{5}$$

The signal parameters can be denoted as a vector:

$$\alpha = [a, \theta, b, \sigma, \boldsymbol{q}] \tag{6}$$

where a denotes the signal amplitude, θ the signal orientation, b the signal shape, σ the signal scale (size), $\boldsymbol{q}$ the signal position (signal center's coordinate), and $\boldsymbol{x}_\alpha$ is then a particular realization. SKS thus means that at least one of the parameters in α is not known exactly a priori, but specified by a probability density function (PDF).

B. The use of joint detection and estimation (JDE) theory

The principle of JDE is that α and $\mathcal{H}_h$ estimates should be chosen jointly to maximize the joint posterior probability $P(\alpha, \mathcal{H}_h \,|\, \boldsymbol{g})$:

$$\begin{aligned}\widehat{(\alpha, \mathcal{H}_h)} &= \arg\max_{\alpha, \mathcal{H}_h} P(\alpha, \mathcal{H}_h \,|\, \boldsymbol{g}) \\ &= \arg\max_{\alpha, \mathcal{H}_h} P(\boldsymbol{g} \,|\, \alpha, \mathcal{H}_h) P(\alpha) P(\mathcal{H}_h)\end{aligned} \tag{7}$$

where the statistical independence of α and $\mathcal{H}_h$ has been exploited. Under the assumption of a zero-mean correlated Gaussian background in Eq.(2), the conditional probability density function $P(\boldsymbol{g} \,|\, \alpha, \mathcal{H}_h)$ can be written as:

$$P(\boldsymbol{g} \,|\, \alpha, \mathcal{H}_h) = \frac{1}{\sqrt{(2\pi)^M \,|\, \Sigma_{\boldsymbol{b}} \,|}} \exp\left\{-\frac{1}{2}(\boldsymbol{g} - h\boldsymbol{x}_\alpha)^{\mathrm{t}} {\Sigma_b}^{-1} (\boldsymbol{g} - h\boldsymbol{x}_\alpha)\right\} \tag{8}$$

In view of the monotonic logarithmic function, the maximization in Eq.(7) is equivalent to:

$$\widehat{(\alpha, \mathcal{H}_h)} = \arg\max_{\alpha, \mathcal{H}_h} \left\{\ln P(\mathcal{H}_h) + \ln P(\alpha) + h\boldsymbol{x}_\alpha^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} (\boldsymbol{g} - \frac{1}{2} h\boldsymbol{x}_\alpha)\right\} \tag{9}$$

1) Estimation: The estimate of α is actually equivalent to maximizing separately $P(\alpha \,|\, \mathcal{H}_h, \boldsymbol{g})$ for $h = 0, 1$, then comparing the two results and choosing the bigger one. Note that the estimated parameters have no physical meaning under the hypothesis $\mathcal{H}_0$ (signal is absent).

Without loss of generality, two additional assumptions are taken in practice to further simplify the expression and facilitate the derivation, though an extension to a more generic model can be derived directly.

- Parameters are uniformly distributed over the admissible space ($P(\alpha) \propto 1$);
- Two hypotheses are equiprobable: $P(\mathcal{H}_0) = P(\mathcal{H}_1)$.

Thus from Eq.(9), the estimation problem becomes:

$$\widehat{\alpha} = \arg\max_{\alpha} \left\{\max\left\{\boldsymbol{x}_\alpha^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} (\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_\alpha), 0\right\}\right\} \tag{10}$$

where 0 corresponds to the case $h = 0$. Note that Goossens's estimation ignored the case $h = 0$ and is only correct when $\boldsymbol{x}_{\alpha_t}^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} (\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_{\alpha_t}) > 0$ [46].

Consequently, the estimation algorithm can be written as:

1) Calculate α_t:

$$\alpha_t = \arg\max_{\alpha} \left\{ \boldsymbol{x}_{\alpha}^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} (\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_{\alpha}) \right\}; \quad (11)$$

2) Determine $\widehat{\alpha}$ according to the following rules:

$$\widehat{\alpha} = \begin{cases} \alpha_t, & \text{if } \boldsymbol{x}_{\alpha_t}^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} (\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_{\alpha_t}) > 0 \\ \text{any admissible value}, & \text{else} \end{cases} \quad (12)$$

There is no analytical solution for Eq.(11) because of the nonlinearity and non-concavity of the function. One alternative is to solve the maximization problem by an iterative method. The iterative algorithm is then run over the parameter space of α, according to some updating rule which often involves matrix multiplications for each iteration.

2) Detection: Let us now consider the posterior probabilities of the two hypotheses given $\widehat{\alpha}$ and a particular observation $\boldsymbol{g}$:

$$\begin{aligned} P(\mathcal{H}_h \,|\, \widehat{\alpha}, \boldsymbol{g}) &= \frac{P(\mathcal{H}_h) P(\widehat{\alpha}) P(\boldsymbol{g} \,|\, \mathcal{H}_h, \widehat{\alpha})}{P(\boldsymbol{g} \,|\, \widehat{\alpha}) P(\widehat{\alpha})} \\ &\propto P(\mathcal{H}_h) P(\boldsymbol{g} \,|\, \mathcal{H}_h, \widehat{\alpha}) \quad h = 0, 1 \end{aligned} \quad (13)$$

The classical decision approach, based on signal detection theory, is to choose $\mathcal{H}_1$ when $P(\mathcal{H}_1 \,|\, \widehat{\alpha}, \boldsymbol{g}) > P(\mathcal{H}_0 \,|\, \widehat{\alpha}, \boldsymbol{g})$ and $\mathcal{H}_0$ in the reverse case, which leads to the decision rule below:

$$\lambda = \boldsymbol{x}_{\widehat{\alpha}}^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} \left(\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_{\widehat{\alpha}} \right) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \ln \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} \quad (14)$$

The test statistic λ is linear in $\boldsymbol{g}$, given the optimal estimated signal parameters $\widehat{\alpha}$, the background covariance matrix $\Sigma_{\boldsymbol{b}}$ and the prior probabilities of two hypotheses $P(\mathcal{H}_h)$.

Note that Goossens omitted the minuend in Eq.(14) [46]. We argue that in the case of SKE, $\boldsymbol{x}_{\widehat{\alpha}}^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{x}_{\widehat{\alpha}}$ may be ignored, since $\boldsymbol{x}_{\widehat{\alpha}}$ could be the same for all test images. However, in the case of SKS, $\boldsymbol{x}_{\widehat{\alpha}}$ could vary for different test images, thus this influential and critical part must be included in the test statistic.

We can note that the test statistic is actually realized jointly with the estimation. In other words, the comparison $\boldsymbol{x}_{\alpha_t}^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} (\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_{\alpha_t}) > 0$ determines both the estimate of the signal parameters in Eq.(12) and the validation of $\mathcal{H}_h$ (detection decision) in Eq.(14), hence the name JDE. The JDE approach has a nonlinear estimation part (cf. Eq.(11)) and a linear detection part (cf. Eq.(14)).

C. The use of channelization

Despite the linearity of the test statistic, its calculation is often compute-intensive. In fact, the inversion of the covariance matrix $\Sigma_{\boldsymbol{b}}^{-1}$ costs $\mathcal{O}(M^3)$, knowing that the dimension of $\Sigma_{\boldsymbol{b}}$ is $M \times M$ when the number of pixels in the image is M. The channelization is the most commonly used preprocessing method to reduce the dimentionality which leads to less calculation.

The estimation based on the channelized image $\boldsymbol{g}'$ is then formulated as follows [46]:

$$\widehat{\alpha} = \arg\max_{\alpha} \frac{\boldsymbol{x}_{\alpha}^{\mathrm{t}}}{\|\mathbf{U}_{\alpha}\|_F^2} \left(\mathbf{U}_{\alpha} \left(\Sigma_{\boldsymbol{b}}' \right)^{-1} \mathbf{U}_{\alpha}^{\mathrm{t}} \right) \left(\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_{\alpha} \right) \quad (15)$$

where $\|\mathbf{U}_{\alpha}\|_F^2$ is a channel matrix energy normalization factor and $\|\cdot\|_F$ denotes the Frobenius norm of the matrix. Goossens pointed that the covariance matrix $\Sigma_{\boldsymbol{b}}'$ and the channel matrix $\mathbf{U}_{\alpha}$ should be calibrated properly, i.e. $\Sigma_{\boldsymbol{b}}^{-1} \approx \frac{1}{\|\mathbf{U}_{\alpha}\|_F^2} \mathbf{U}_{\alpha} \left(\Sigma_{\boldsymbol{b}}' \right)^{-1} \mathbf{U}_{\alpha}^{\mathrm{t}}$ [46].

An important contribution of Goossens is to solve the hardly realizable maximization problem in Eq.(15) by searching the optimal parameters in the channel domain, instead of the spatial domain, without loss of accuracy. This requires that a transform on the signal in the spatial domain can be expressed as an equivalent transform on the signal in the channel domain, thus he added an extra requirement for designing the channel matrix:

$$\mathbf{U}_{\alpha} = \mathbf{A}_{\alpha}^{\mathrm{t}} \mathbf{U} = \mathbf{U} \left(\mathbf{A}_{\alpha}' \right)^{\mathrm{t}} \quad (16)$$

where $\mathbf{U}$ is a fixed channel matrix which does not depend on α and leads to reduce the data dimensionality; $\mathbf{A}_{\alpha}$ serves to map the parametric signal $\boldsymbol{x}_{\alpha}$ onto a reference signal $\boldsymbol{x}_0$ (with known parameters): $\mathbf{A}_{\alpha}\boldsymbol{x}_{\alpha} = \boldsymbol{x}_0$; $\mathbf{A}_{\alpha}'$ map the channelized parametric signal $\boldsymbol{x}_{\alpha}'$ onto the channelized reference signal $\boldsymbol{x}_0' : \mathbf{A}_{\alpha}'\boldsymbol{x}_{\alpha}' = \boldsymbol{x}_0'$. The size of $\mathbf{A}_{\alpha}'$ is normally much smaller than that of $\mathbf{A}_{\alpha}$.

Then our estimates of the signal parameter and the test statistic, Eq.(11) and Eq.(14), could be rewritten as follows:

$$\alpha_t = \arg\max_{\alpha} \frac{1}{\left\| \mathbf{U} \left(\mathbf{A}_{\alpha}' \right)^{\mathrm{t}} \right\|_F^2} (\boldsymbol{x}_0')^{\mathrm{t}} \left(\Sigma_{\boldsymbol{b}}' \right)^{-1} \left(\mathbf{A}_{\alpha}' \boldsymbol{g}' - \frac{1}{2}\boldsymbol{x}_0' \right) \quad (17)$$

$$\lambda = \frac{1}{\left\| \mathbf{U} \left(\mathbf{A}_{\widehat{\alpha}}' \right)^{\mathrm{t}} \right\|_F^2} (\boldsymbol{x}_0')^{\mathrm{t}} \left(\Sigma_{\boldsymbol{b}}' \right)^{-1} \left(\mathbf{A}_{\widehat{\alpha}}' \boldsymbol{g}' - \frac{1}{2}\boldsymbol{x}_0' \right) \quad (18)$$

From now on, a numerical observer performing Eq.(17) and Eq.(18) is referred to as a *Channelized Joint detection and estimation Observer (CJO)*.

D. Channel design

Goossens has also derived suitable channels according to different specific detection task. The simplest task is the amplitude-unknown case for which any channel could be adopted since it suffices to choose

$$\mathbf{A}_a' = a^{-1}\mathbf{I} \quad (19)$$

For two other more complicated tasks, a set of *steerable* channels is adopted for the orientation-unknown case while a set of *scale-shiftable* channels is adopted for the scale-unknown case. In polar-frequency coordinates, let ω be the radial frequency and φ be the angular orientation. The steerable channels and the scale-shiftable channels are given in

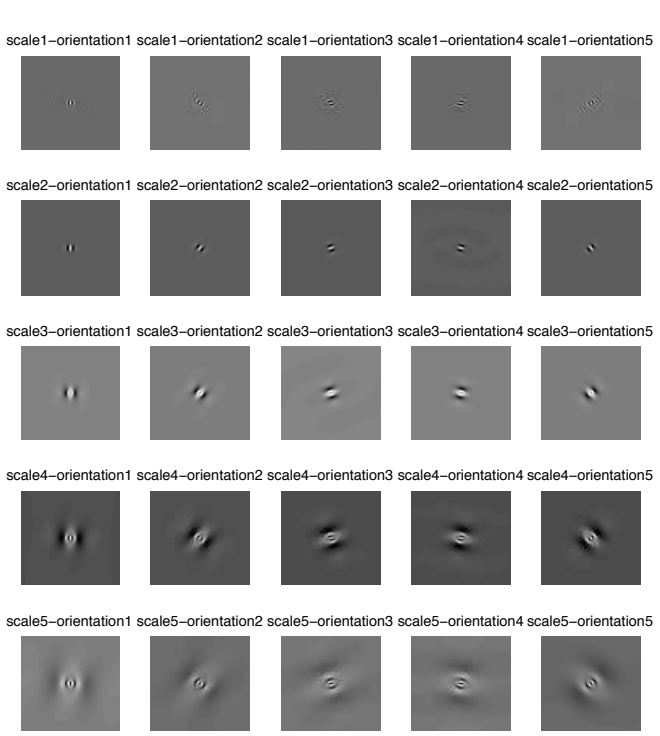

Fig. 2. An example of the filters for the amplitude-orientation-scale-unknown case, when the number of steerable channels $K = 5$ and the number of scale-shiftable channels $J = 5$

Eq.(20) and Eq.(21) respectively.

$$f_{\theta_k}(\varphi) = \frac{(K-1)!2^{K-1}}{\sqrt{K(2K-2)!}} \left(\cos(\varphi - \theta_k)\right)^{K-1} \qquad (20)$$

where $\theta_k = \frac{(k-1)\pi}{K}$ for $k = 1, \ldots, K$ are evenly spaced basic angles and K is the number of steerable channels.

$$f_{\sigma_j}(\omega) = \mathrm{sinc}\left(\log_2(\frac{|\,\omega\,|}{\pi} + \epsilon) + \log_2\sigma_{\mathrm{j}}\right), \; |\,\omega\,| < \pi; \quad (21)$$

where $\sigma_j = 2^{j-1}$ for $j = 1, \ldots, J$ are basic scales and J is the number of scale-shiftable channels; ϵ is a small positive number to make sure the result is defined for $\omega = 0$ (e.g. $\epsilon = 2^{-52}$).

The product of the two functions then defines the channel functions for the orientation-scale-unknown case:

$$f_{\theta_k,\sigma_j}(\omega, \varphi) = f_{\theta_k}(\varphi) f_{\sigma_j}(\omega), \; k = 1 \ldots K; j = 1 \ldots J \quad (22)$$

An example of channels ($K = 5$ and $J = 5$) is shown in Figure 2.

Since any channel could be adopted for the amplitude-unknown case, the channel function for the amplitude-orientation-scale-unknown case could be the same as in Eq.(22).

By storing each sampled channel function given by Eq.(22) in a column, we construct the channel matrix ($M^2 \times JK$) for the amplitude-orientation-scale-unknown case:

$$[\mathbf{U}]_{m,n} \qquad m = 1, \ldots, M^2; \; n = (j-1)K + k \qquad (23)$$

where $[\cdot]_{m,n}$ denotes the matrix element of the m-th line and n-th column.

Two corresponding transform matrices for the orientation-unknown case and scale-unknown case are given by Eq.(24) and Eq.(25) respectively:

$$[\mathbf{A}'_\theta]_{m,n} = \frac{1}{K} \frac{\sin(\pi(m-n) - \theta K)}{\sin(\pi(m-n)/K - \theta)}, \; m, n = 1 \ldots K \quad (24)$$

$$[\mathbf{A}'_\sigma]_{m,n} = \mathrm{sinc}\left(((\mathrm{m} - \mathrm{n}) + \log_2\sigma)\right), \; \mathrm{m}, \mathrm{n} = 1 \ldots \mathrm{J} \quad (25)$$

where θ is the signal orientation and σ is the signal scale; and the sinc function is:

$$\mathrm{sinc(t)} = \begin{cases} 1 & t = 0 \\ \frac{\sin(\pi t)}{\pi t} & t \neq 0 \end{cases} \qquad (26)$$

We then design the transform matrix for the orientation-scale-unknown case which depends on the sample storage ordering during the construction of channel matrix $\mathbf{U}$. For the ordering defined in (23), we have:

$$\mathbf{A}'_\alpha = \mathbf{A}'_{a,\theta,\sigma} = \frac{1}{a} \mathbf{A}'_\sigma \otimes \mathbf{A}'_\theta \qquad (27)$$

where $\otimes$ denotes a Kronecker product of two matrices.

For an odd number K, the matrix $\mathbf{A}'_\theta$ has the interesting unitary property: $(\mathbf{A}'_\theta)^{\mathrm{t}} \mathbf{A}'_\theta = \mathbf{I}$. We find that this is useful for reducing the calculation burden of the normalization factor $\left\| \mathbf{U} \left(\mathbf{A}'_\alpha\right)^{\mathrm{t}} \right\|_F^2 = \frac{1}{a^2} \left\| \mathbf{U} \left(\mathbf{A}'_\sigma \otimes \mathbf{A}'_\theta\right)^{\mathrm{t}} \right\|_F^2$, which costs $\mathcal{O}(M^2 \cdot (JK)^2)$ by direct evaluation. We provide a simplified numerical method hereafter.

From the definition of the Frobenius norm, we have:

$$\begin{aligned} \left\| \mathbf{U} \left(\mathbf{A}'_\sigma \otimes \mathbf{A}'_\theta\right)^{\mathrm{t}} \right\|_F^2 &= \mathrm{tr}\left(\left(\mathbf{A}'_\sigma \otimes \mathbf{A}'_\theta\right) \mathbf{U}^{\mathrm{t}} \mathbf{U} \left(\mathbf{A}'_\sigma \otimes \mathbf{A}'_\theta\right)^{\mathrm{t}} \right) \\ &= \mathrm{tr}\left(\mathbf{U}^{\mathrm{t}} \mathbf{U} \left[\left((\mathbf{A}'_\sigma)^{\mathrm{t}} \mathbf{A}'_\sigma \right) \otimes \mathbf{I} \right] \right) \end{aligned} \qquad (28)$$

Let $(\mathbf{A}'_\sigma)^{\mathrm{t}} \mathbf{A}'_\sigma = S = [s_{ij}]$ $(i, j = 1, \ldots, J)$, then divide the matrix $\mathbf{U}^{\mathrm{t}}\mathbf{U}$ (dimension $JK \times JK$) into $J \times J$ submatrices T_{ij} $(i, j = 1, \ldots, J)$, each of dimension $K \times K$, the right hand can be written as follows:

$$\begin{aligned} \sum_{i=1}^{J} \mathrm{tr}\left(\sum_{j=1}^{J} s_{ji} T_{ij} \right) &= \sum_{i=1}^{J} \mathrm{tr}\left(\sum_{j=1}^{J} s_{ij} T_{ij} \right) \\ &= \sum_{i=1}^{J} \sum_{j=1}^{J} s_{ij} \mathrm{tr}\left(T_{ij}\right) \end{aligned} \qquad (29)$$

Therefore it costs only $\mathcal{O}(J^2)$ to calculate the normalization factor given an odd value of K. We also note that T_{ij} $(i, j = 1, \ldots, J)$ is known a priori, independent of image data. Therefore, $\mathrm{tr}\left(T_{ij}\right)$ can be pre-calculated.

E. Implementation of the CJO

Like for SKE CHO, we could also implement the CJO for the amplitude-orientation-scale-unknown task in two stages: training and test. In the training stage, the channelized reference signal $\boldsymbol{x}'_0$ is estimated, and a template $\boldsymbol{w}$ is calculated, which is the multiplication of the inversion of the estimated background covariance matrix and the estimated channelized reference signal. In the test stage, the test statistic λ for each input test image is calculated.

IV. Perceptually relevant CJO (PCJO) for the joint detection-localization task

Let us now consider a more clinically realistic but more complex situation: a test image may have multiple signals, and observers do not exactly know the position, amplitude, orientation and scale of each signal, nor the number of signals on each test image.

In fact, if there is only one signal on the image and the signal position is the only one unknown parameter, then the CJO approach can be essentially applied by using position-shiftable channels for this particular localization task. But the required number of channels increases rapidly if a fine spatial resolution is needed, e.g. 65^2 channels are needed for a grid of 65×65 spatially selective channels, as explained in [46]. Moreover position shiftability requires the channels to have limited bandwidth (regions of support) in the frequency domain while scale shiftability requires them to have compact regions of support in the spatial domain. Thus, it is impossible to design channels that are shiftable in position and in scale at the same time. An alternative is the scanning CHO [15], but this is only efficient for spatially stationary images; otherwise, a huge amount of training images is required and the computation burden is very high. The localization task of multiple signals on an image is more challenging.

To solve this problem, we propose a novel Perceptually relevant CJO (PCJO) by taking advantage of the state-of-the-art in HVS modeling.

A. Structure diagram of the PCJO

The structure diagram of the PCJO is shown in Figure 3. The PCJO consists of two parts: (1) Candidate selection (outlined by a red dashed line in Figure 3) which generates several test blocks for each image under test, and outputs the center position of each test block on the image (performing the localization task); and (2) Application of the CJO on candidates which calculates a test statistic for each test block (performing the detection task).

The underlying paradigm consists in dividing the diagnostic process into two steps:

- the first step is a global search to locate the abnormality candidates deserving further check, based on the human vision functioning used to detect abnormalities from the perceived image under test (the *sensation* stage of human vision processing);
- the second step is a cognitive analysis and interpretation of each candidate, mainly depending on radiologists' semiological knowledge and past experience (the *perception* stage of human vision processing), which leads to their final decision.

B. Candidate selection

We remark that most numerical observers have ignored the intermediate result of the PDMs - the perceptual difference map, which actually can provide extremely useful information for the localization task. Our last study [50] showed that the Visible Differences Predictor (VDP) [29] presents a good approximation of radiologists' *sensation* performance in the context of multiple sclerosis (MS) lesion detection on MR image (which is the same scenario as in our performance evaluation study in Section V). Thus the VDP is adopted in this paper to model the radiologists' global search and to yield a candidate map.

The VDP schema is shown in Figure 4. Basically, the VDP is used to produce a map of the probability of detecting the differences between two input images. On the assumption that radiologists know what a healthy brain image should be like and refer to it in their mind during the diagnostic process, we take an image acquired from a healthy person as *Reference image* and the same image with simulated signals (modeled by Eq.(3)) as *Distorted image*, as Jackson did in [26]. Thus, the VDP actually predicts the probability that the signal is visible in radiologists' perceptual domain. In VDP, a display model is involved to relate pixel values to displayed luminance levels; an amplitude nonlinearity function is used to model the nonlinear response to luminance of the photoreceptors in the retina; one CSF is adopted in VDP as a function of radial spatial frequency, orientation, light adaptation level, image size in deg^2, eccentricity and lens accommodation due to the viewing distance; the cortex transform is applied to decompose the image into several sub-bands (spacial frequency- and orientation- selectivity); an intra-channel masking function quantifies the visibility decrease due to the presence of a supra-threshold background; a psychometric function characterizes the increase in the probability of detection as the signal contrast increases; finally a sub-band pooling combines the detection probability of each sub-band into a single map that describes the overall detection probability p_{ij} for every pixel (i, j) in the image. More details about the above functions, including the parameter values, can be found in [29].

Since the VDP behaves well in terms of approximate sensation performance, we just use a simple method to get the *signal candidate map* for the PCJO's preliminary version. This is generated from the VDP output *detection probability map* in five steps:

1) Creating a binary image by setting a probability threshold T_p on the detection probability map:

$$b_{ij} = \begin{cases} 1 & \text{if } p_{ij} \geq T_p \\ 0 & \text{else} \end{cases} \tag{30}$$

   where b_{ij} denotes the pixel value at the position (i, j) on the binary image, T_p is chosen empirically.
2) Finding and labeling 8-connected objects in the binary image.
3) Counting the number of pixels for each labeled object and eliminating the objects smaller than the smallest possible signal.
4) Relabeling 8-connected objects in the processed binary image.
5) Calculating the gravity centers of the labeled objects.

These labeled objects are considered as signal candidates. Then a test block of predetermined size $\sqrt{M} \times \sqrt{M}$ centered around each object center is extracted from the distorted image. These test blocks are the inputs for the test stage of

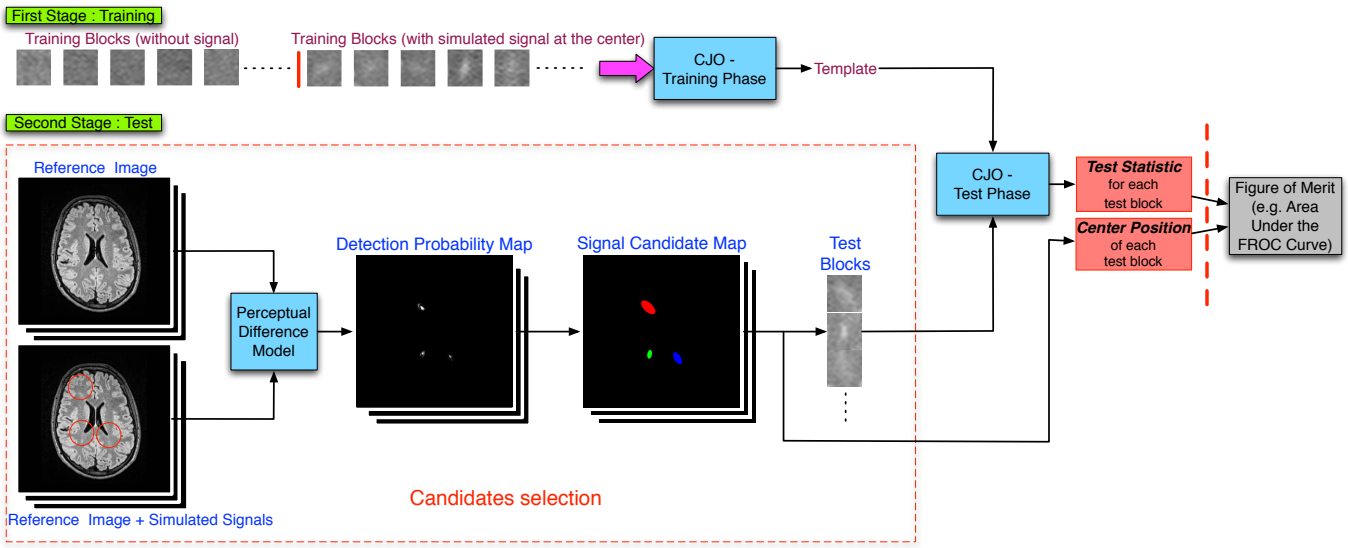

Fig. 3. The structure diagram of the proposed Perceptually relevant Channelized Joint Observer (PCJO).

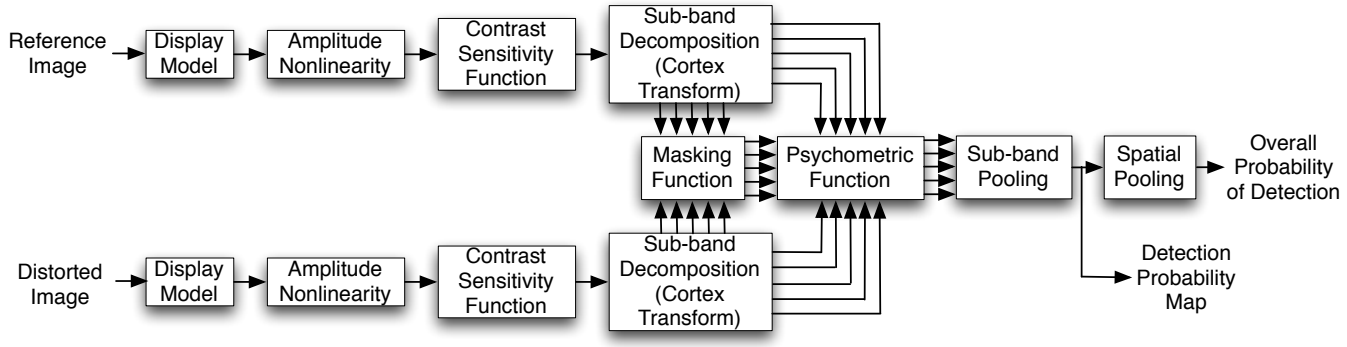

Fig. 4. The VDP schema: in which a map of the probability of detecting the perceptual difference between the *Reference image* and the *Distorted image* is produced as a function of their location in the input images.

TABLE I
COMPARISON OF OUR PCJO WITH EXISTING MODEL OBSERVERS.

	Concerned diagnostic task	Knowledge about signal	Allowable number of signals on one image	Human visual system (HVS) features included
CHO [35]	Detection task	Full knowledge about all the signal parameters	one	Frequency and orientation selectivity of HVS
Eckstein's model [43]–[45]	Detection task	Signal size and shape are unknown, while other signal parameters are known exactly	one	Frequency and orientation selectivity of HVS
Goossens's SKS CHO [48]	Detection task	Signal orientation is unknown, while other signal parameters are known exactly	one	Frequency and orientation selectivity of HVS
Park's model [15]	Localization task	Signal position is unknown, while other signal parameters are known exactly	one	Frequency and orientation selectivity of HVS
PCJO	Detection and localization task	Signal amplitude, orientation, scale, position are unknown, while signal shape is known exactly	multiple	Many characteristics of HVS: Nonlinear response of photoreceptors to luminance; Variations in visual sensitivity as a function of spatial frequency; Spatial frequency and orientation selectivity; Visual masking properties; Visual psychometric function.

the CJO, and the center positions are saved as the localization response of the PCJO.

C. Application of the CJO on candidates

Our last study [50] also showed that there is still a gap between the VDP's performance and radiologists' *perception*

performance. One important reason may be that radiologists can "learn" from experience. Thus we also "train" our perceptually relevant numeric observer, using empirical data, to make it behave more like radiologists. This is realized by using the training stage of the CJO. Then we apply the CJO technique on every $\sqrt{M} \times \sqrt{M}$ test block and save its test statistic as the detection response of the PCJO.

As outputs of the PCJO, the test statistics and the center positions of the test blocks can then be used to characterize its joint detection-localization task performance by an FROC or AFROC FOM.

D. Comparisons of PCJO with other numerical observers

Table I gives a comparison of our PCJO with several existing numerical observers.

V. Performance Evaluation Study - Methods and Materials

A. Studied pathology and modality

To study the PCJO task performance, we selected MS as the studied pathology. The main reason is that for this particular pathology, orientation, size, position and number of lesions are all considered as diagnostic criteria [51], which is well adapted for our PCJO study. Note that MS patients generally have multiple lesions and the number of lesions is considered as a diagnostic criterion for MS, where the minimum number of MS lesions is considered to be 13 in the diagnostic criteria used in clinical studies [52]. Besides, MS lesions are subtle and difficult to be perceived. This also makes MS a suitable pathology for the joint detection-localization task, for which the abnormality of the pathology should not be too conspicuous.

Since the criteria for MS are largely based on brain Magnetic Resonance Imaging (MRI) scan outcomes [53], and under certain circumstances a diagnosis of MS can be confirmed based on a single MR exam [54], we chose MRI as the studied modality.

B. Experimental images

As preparation, we acquired brain MR axial images from 20 healthy volunteers using the T2 FLAIR sequence which is the most efficient sequence for the MS lesion detection. In this sequence, MS lesions appear as hyper-signals, namely the amplitude is higher than its surrounding white matter. The lesions were modeled by Eq.(3), which had been verified to mimic well the MS lesions by experts in our first study [51]. From a retrospective database of the University Hospital of Angers, we collected 20 healthy subjects' MR brain images (3D axial stacks) of the T2 FLAIR sequence which is the most efficient sequence for the MS lesion detection. We then chose 90 independent non contiguous slices among these stacks. According to the radiologists' feedback, none of them could tell whether some slices were from same subjects. Thus we considered these images as independent.

The slices were originally acquired at 256×256 and 16-bit. Then we asked a neuroradiologist (who did not participate in the experiment) to use the digital imaging processing tools (zoom/interpolation, windowing) to modify each slice as radiologists are accustomed to do in clinical routine. The resulting 1024×1024 and 8-bit images were optimized for MS diagnosis and could not be modified during the experiment.

On T2 FLAIR sequence, MS lesions appear as small hyper intense signals of elliptical shape, inferior to 10 millimeters. Their amplitude is higher than the surrounding white matter [55]. We used Eq.(3) to model the MS lesions on this sequence, since our precedent study [56] had verified that experts found it well mimics MS lesions. According to this study, when we added lesions with a luminance contrast between 0.1 and 0.35, we obtained an expected detection rate between 60% and 80%. Overall 145 lesions (with random luminance contrasts between 0.1 and 0.35) were simulated. All the 90 images had 1 to 4 lesions.

C. Subjective experiment

We conducted a free-response subjective experiment [57] in order to assess human performances.

1) Participants: Six radiologists participated in the experiment in a simulated radiology reading environment at the University Hospital of Angers, France. Radiologists 1 and 2 are MS experts, with respectively 21 and 10 years' experience; Radiologists 3-6 are not MS experts, with respectively 6, 3, 8 and 5 years' experience.

2) Experimental protocol: According to [58], it is important to verify that the ambient lighting in the diagnostic reading room is below 15 lux. The ambient lighting level in our experiment room was set to 11 lux (same as the radiology reading environment at the University Hospital, Angers, France). The medical display used in this study was KEOSYS Positoscope, which was calibrated to the Digital Imaging and Communications in Medicine (DICOM) Grayscale Standard Display Function (GSDF). The viewing distance was 40cm, and the visual angle was 42 pixels per degree.

Each radiologist had to read 90 images, a 10-image training was also offered beforehand. The display was set to be gray during the presentation of two slices for 0.5s. The decision time was unlimited. Radiologists performed the detection-localization task without knowing exactly the lesion characteristics (intensity amplitude, orientation, size, position and the number of lesions on each image). The experiment was conducted in a controlled environment, thus the readers could not modify the images.

3) Experiment's graphical user interface (GUI): For this experiment, we developed a special GUI (shown in Figure 5).

First, a radiologist decided whether there was one lesion on one image. Upon clicking on the "YES" button, a horizontal line and a vertical one appeared on the image and their point of intersection followed the cursor to indicate the positions on the image. By a left click on a position, the radiologist could define the center of a detected lesion, and a cross mark appeared at the position while the two lines disappeared. The radiologist could cancel a previous center definition by using the "CANCEL" button, which deleted the cross mark and made the horizontal and vertical lines appear again. Then

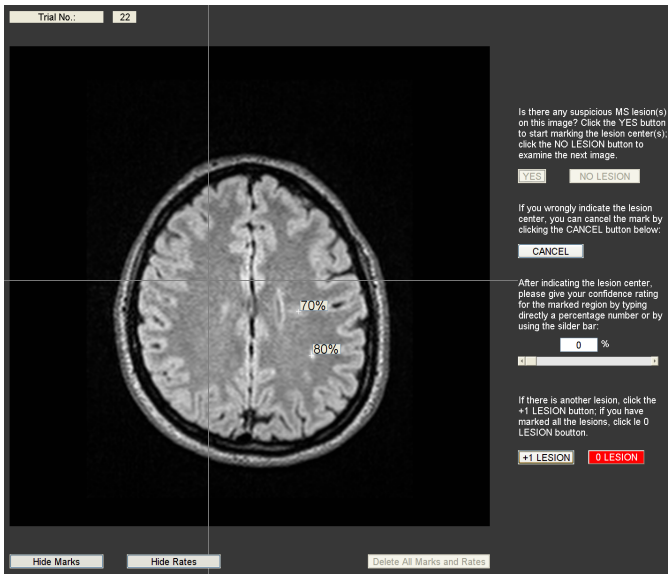

Fig. 5. GUI of the free-response subjective experiment.

the radiologist expressed his confidence rating for the marked region by typing directly a percentage number or using the slider bar. In case of visual interference, clicking the "Hide Marks" and the "Hide Rates" button toggled off either the marks or the rates, and changed them to "Show Marks" and "Show Rates", that allowed undoing the previous operation. If another lesion was detected on this image, the above process could be repeated by clicking the "+1 LESION" button.

Clicking the red "0 LESION" button terminated the diagnosis process on the current image. A dialog box then popped up showing all the marks and the rates and the next image was displayed upon validation of the results; otherwise, the diagnosis on the current image restarted.

D. PCJO's setup

The reference images for the VDP were from the stack of healthy axial slices; while the distorted images for the VDP were obtained by randomly adding simulated lesions on the healthy images. Note that the same set of 90 distorted images was used in our free-response subjective experiment (cf. Section V-C) as the images read by radiologists.

The possible positions of simulated lesions ($\boldsymbol{q}$) were the coordinates of the pixels within the white matter. The default parameter values in VDP were adopted, and the probability threshold T_p in Eq.(30) was set to 0.9.

The training blocks for the CJO were constructed by extracting white matter regions from the stack of healthy axial slices, but different from those regions used in the candidate selection and the subjective experiment. This ensured that the training blocks and the test blocks of the CJO were different. In this paper, 1000 different white matter blocks were extracted, 500 blocks without signal and 500 blocks with a simulated signal were used for the training step of the CJO. The block size was predetermined as $\sqrt{M} \times \sqrt{M} = 65 \times 65$. The scale range of simulated lesions was $[\sigma_{\min},\ \sigma_{\max}] = [1,\ 12]$, and the lesion shape was fixed ($\sqrt{b} = 2$ in this paper). The possible signal amplitude was $[a_{\min},\ a_{\max}] = [1,\ 255]$ (pixel intensity value). The signal orientation range was $[\theta_{\min},\ \theta_{\max}] = [0,\ \pi]$.

E. Performance evaluation method

Remember that the JAFROC1 [23] (cf. Section II-A) was selected to evaluate the detection-localization task performance of PCJO compared with that of radiologists.

The JAFROC1 analysis was conducted using the software *JAFROC 4.1* (downloaded from http://www.devchakraborty.com/). The *acceptance radius* R was set to 12 (one-half of the lesion's major axis) in our study. In addition, the DBM-MRMC significance testing integrated in the software was conducted to examine JAFROC1 FOM differences between observers.

VI. Performance Evaluation Study - Results and Discussion

A. PCJO performance evaluation with respect to radiologists' performances

The JAFROC1 FOMs of the six radiologists are shown in Table II. The first DBM-MRMC significance testing is conducted to examine JAFROC1 FOM differences between each pair of radiologists. The significance level is set to $p < 0.05$. Table II and p-value results illustrate that the experts have a significantly higher FOM than the other four radiologists, while no significant difference exists among the FOMs of the two experts, nor among the FOMs of the other four radiologists (cf. Section VI-C for more discussions).

TABLE II
THE JAFROC1 FOMS OF THE SIX RADIOLOGISTS. THE JAFROC1 FOMS OF THE EXPERTS ARE HIGHLIGHTED IN RED.

	JAFROC1 FOM	Standard Error
Radiologist 1 (expert 1)	0.7672	0.0326
Radiologist 2 (expert 2)	0.7110	0.0329
Radiologist 3	0.4736	0.0378
Radiologist 4	0.4278	0.0399
Radiologist 5	0.4742	0.0419
Radiologist 6	0.4728	0.0369

We then calculate the JAFROC1 FOMs for the PCJO with different combinations of the number of *steerable* channels K and the number of *scale-shiftable* channels J. The p-values for examining JAFROC1 FOM differences between the PCJO and the experts are also calculated. Part of the results are illustrated in Figure 6. The results show that the number of *steerable* channels K hardly influences the PCJO FOM, while the number of *scale-shiftable* channels J does. When J is equal or less than four, there is no significant difference between the PCJO FOMs and those of the experts ($p > 0.05$). The PCJO with $J = 4$ has the closest FOM to that of expert 1 and the PCJO with $J = 3$ has the closest FOM to that of expert 2. This also illustrates that even using a small number of channels (e.g. $K = 3$ and $J = 4$, $3 \times 4 = 12$ channels) can result in a good FOM, close to those of the experts; this allows to reduce the calculation burden in the PCJO.

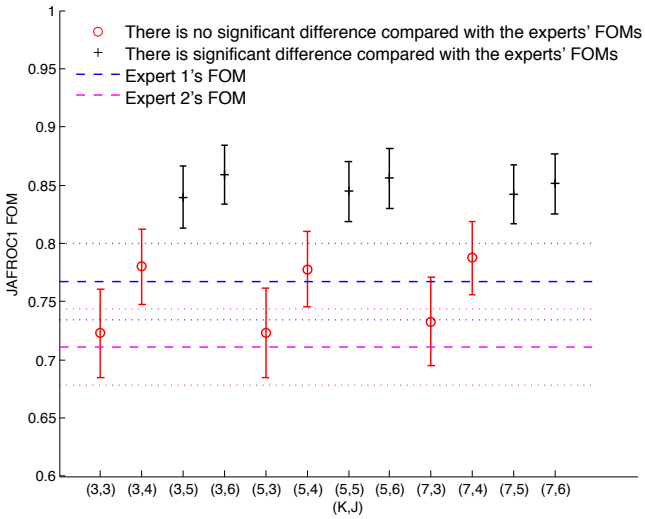

Fig. 6. JAFROC1 FOMs of PCJO for different combinations of K and J. The p-values for examining JAFROC1 FOM differences between the PCJO and the experts are also calculated, and those greater than 0.05 are highlighted in red, which means that the corresponding JAFROC1 FOMs of the PCJO are not significantly different from those of the experts.

B. PCJO performance evaluation with respect to other numerical observers' performances

Compared to other numerical observers, one distinctive point of the PCJO is that it does not need to have *a priori* knowledge about the signal amplitude, orientation and scale in the detection-localization task, thanks to the CJO part (cf. Section III) which can estimate the signal parameters. This obviously yields more clinical relevance (more clinically realistic).

To further verify the usefulness of the signal parameter estimation, we compare the PCJO with two other numerical observers selected from Table I: the CHO [35] and Goossens's SKS CHO [48]. The two numerical observers are selected since their source codes are available. In order to use the same FOM, it is necessary to perform the localization task as well, which is however not included in these two numerical observers. Therefore, to conduct the comparison, the two numerical observers are not directly used, but are used to replace the CJO part in the PCJO, while the candidate selection step in Section IV-B remains the same. This also allows us to compare them with the radiologists' detection-localization task performances.

Note that the signal amplitude, orientation and scale should be known exactly in the CHO and the signal amplitude and scale should be known exactly in Goossens's SKS CHO. Thus, these need-to-know parameters should be fixed in advance. We test their performances with different settings of signal parameters. The JAFROC1 FOMs of the CHO combined with the candidate selection step and those of Goossens's SKS CHO combined with the candidate selection step are presented in Table III and Table IV, respectively.

We observe that with any setting of signal parameters and number of channels, the CHO and Goossens's SKS CHO combined with the candidate selection step always have similar JAFROC1 FOMs. Furthermore, they are always significantly higher than the JAFROC1 FOMs of the experts, according to the DBM-MRMC significance testing results.

TABLE III
THE JAFROC1 FOMS OF THE CHO COMBINED WITH THE CANDIDATE SELECTION STEP, FOR DIFFERENT FIXED VALUES OF SIGNAL AMPLITUDE, ORIENTATION, SCALE, AND DIFFERENT NUMBER OF LAGUERRE-GAUSSIAN (LG) CHANNELS USED IN CHO.

amplitude	orientation	scale	# LG channels	JAFROC1 FOM	Standard Error
11	0	2	3	0.8663	0.0250
11	0	2	21	0.8484	0.0273
11	$\pi/4$	6	9	0.8644	0.0247
11	$\pi/4$	6	21	0.8621	0.0250
11	$\pi/2$	10	3	0.8648	0.0245
11	$\pi/2$	10	9	0.8533	0.0259
51	0	6	3	0.8675	0.0242
51	0	6	9	0.8644	0.0248
51	$\pi/2$	10	9	0.8536	0.0260
51	$\pi/2$	10	21	0.8485	0.0264
131	$\pi/4$	2	9	0.8559	0.0269
131	$\pi/4$	6	9	0.8646	0.0247
131	$\pi/2$	10	3	0.8656	0.0244
131	$\pi/2$	10	21	0.8488	0.0264

TABLE IV
THE JAFROC1 FOMS OF GOOSSENS'S SKS CHO COMBINED WITH THE CANDIDATE SELECTION STEP, FOR DIFFERENT FIXED VALUES OF SIGNAL AMPLITUDE, SCALE, AND DIFFERENT NUMBER OF STEERABLE CHANNELS USED IN GOOSSENS'S SKS CHO.

amplitude	scale	# steerable channels	JAFROC1 FOM	Standard Error
11	2	3	0.8559	0.0266
11	2	9	0.8502	0.0264
11	10	9	0.8606	0.0251
11	10	21	0.8587	0.0248
51	2	9	0.8585	0.0254
51	6	3	0.8661	0.0248
51	6	21	0.8567	0.0251
51	10	9	0.8643	0.0248
131	2	9	0.8499	0.0264
131	2	21	0.8211	0.0271
131	6	3	0.8669	0.0246
131	6	9	0.8650	0.0247

C. Discussion

It will be interesting to further investigate the radiologists' results (cf. Table II). Remember that the JAFROC1 chance value is near zero. The results mean that Radiologists 3-6 are not necessarily "bad", just not as good as the experts. In Table V, we present the number of TP marks, the number of FP marks for the laxest criterion (confidence rating > 0) on all 90 images, as well as the detection rate (number of TP marks divided by the total number of simulated lesions, that is 145). From Table V, we can see that all the radiologists have acceptable detection rates which are greater than 0.5; and five of six radiologists have expected detection rates (between 60% and 80%) since lesions' luminance contrasts were adjusted between 0.1 and 0.35, which is in coherence with our previous study [56]. The big difference between the two experts and Radiologists 3-6 is the number of FP marks: the experts have really few FP marks while the Radiologist 3-6 have a number of FP marks almost the same as the number

of TP marks. That is why the experts have a significantly higher performance than the other 4 radiologists. Actually, by observing the subjective experiment and talking to the radiologists after the experiment, we note that Radiologists 3-6 could not really tell the difference between the MS lesions and the cerebral cortex that also appears as hyper-signals, without referring to the information in the adjacent slices, in some difficult cases. They tended to mark more lesions rather than to miss one lesion when they had doubts. However the experts were far more experienced in MS image assessment and they could even imagine where the cerebral cortex locates without the adjacent slices in the difficult cases. This also reveals that the experience/training on a certain pathology can dramatically influence the diagnostic task performance on this pathology.

TABLE V
THE DETECTION RATE, NUMBER OF TP MARKS, NUMBER OF FP MARKS FOR THE LAXEST CRITERION (CONFIDENCE RATING > 0) ON ALL 90 IMAGES OF THE SIX RADIOLOGISTS.

	Detection Rate	# TP marks	# FP marks
Radiologist 1 (expert 1)	0.7241	105	22
Radiologist 2 (expert 2)	0.7241	105	36
Radiologist 3	0.7448	108	112
Radiologist 4	0.6690	97	107
Radiologist 5	0.6897	100	110
Radiologist 6	0.5172	75	76

What we expect from a numerical observer is, of course, to obtain an approximate performance compared to that of the experts. The results of JAFROC1 FOM (which has a high statistical power) demonstrate that our proposed single-slice numerical observer - PCJO - can correctly approach the experts' detection-localization task performances, with a relative small number of channels; while other existing numerical observers always outperform the experts' performances. This suggests that the PCJO paradigm is a promising method with a good clinical relevance.

VII. Conclusion and Future Work

We have introduced a novel numerical observer called the Perceptually relevant Channelized Joint Observer (PCJO) in the interest of the evaluation of medical image quality. As far as we know, we are the first ones to propose such a numerical observer that can perform the joint detection-localization task when multiple signals with random amplitude, orientation, size and location are possibly present on one image, and the number of lesions is unknown to the observer.

Compared to other numerical observers in the literature,

- we greatly extended the range of variable signal parameters by investigating further the SKS MOs;
- we considered the perception stage of the HVS by using a training process;
- we addressed the difficulty of localizing multiple signals on an image by using a PDM to select candidate blocks.

We also conducted a free-response experiment close to the clinical paradigm to get the detection-localization task performance of radiologists with which we evaluated that of the PCJO by using JAFROC1 FOM. We found that for the studied images (axial MRI slices of the T2 FLAIR sequence), the PCJO can correctly predict human performance. We also compared the PCJO with other mathematical observers using JAFROC1 FOM by adding the candidate selection part to these mathematical observers which cannot perform localization task originally. The results showed that other mathematical observers have a significantly higher performance than that of radiologists.

In the proposed PCJO, we used a perceptual difference model (PDM) to select candidate blocks for signal detection by a channelized model observer. The big advantage of this approach is that it is more computationally feasible for modeling combined detection-localization task performance. The main limitation of this approach is that the PDM requires two images differing only in the presence of one or more signals to be detected. This is addressed effectively with simulation methods that introduce lesion signals into healthy images, a simulation which is also required to generate signal-present images for MOs. But this model is well adapted for comparing and optimizing medical image systems since we can do it with simulated lesions.

On all accounts, this initial study suggests that the PCJO paradigm (a candidate search process plus a decision making process) is promising for accurately predicting radiologists' performance in the joint detection-localization task. Of course, it is necessary to further investigate and validate the PCJO with more datasets from different sequences in MRI, different modalities and with a greater number of experts.

In our further work, the PCJO will be used to quantify the medical imaging system performance and will be applied to compare different image processing algorithms or different medical displays, as shown in Figure 7. When the display system is modeled by the display model in the VDP part, with regard to different image processing algorithms, the PCJO will give different outcomes, based on which the different performances of these algorithms can be evaluated. Similarly, when the image processing algorithm is given, we can compare the performances of different displays by building their corresponding display models. In both cases, these different outcomes of the PCJO will also be compared with subjective evaluation scores to further validate the PCJO.

In addition, we will conduct another subjective experiment in which the adjacent slices will be shown to the radiologists and we will consider how to extend the proposed methodology for a more clinically realistic numerical observer - a 3D (multi-slice) numerical observer. For numerous studied modalities, such as MRI, radiologists normally view all images/slices of stacks in their entirety by scrolling back and forth. The radiologists would very probably change their behavior and performance in this more realistic condition. In the mean time, we will consider future extensions that incorporate factors in temporal perception and the characterization of signals (in terms of benign-malignant classes or diagnostic grading scales). These would help us to gradually reach the radiologists' performance in the most realistic possible clinical conditions.

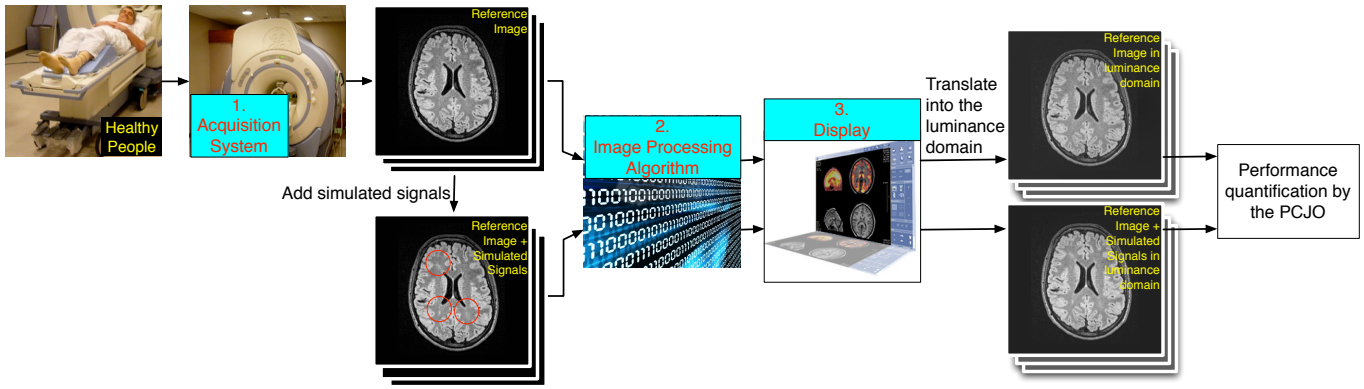

Fig. 7. Using the PCJO to quantify the medical imaging system performance: the reference images for the VDP and the training images for the CJO can be obtained from healthy subjects, then the distorted images for the VDP can easily be produced using a signal model.

ACKNOWLEDGMENT

The authors would like to thank all the radiologists from the University Hospital of Angers who participated in this study, for generously giving their time and effort on performing the experiment. We also thank Bart Goossens for helpful discussions and providing his source code; and Dev Chakraborty for help in using the JAFROC1 metric and interpreting the results. The authors are also grateful to anonymous reviewers for insightful and helpful comments about the manuscript.

This work was supported by the Pays de La Loire region in France. The authors are grateful for the funding of EQuIMOSe (Subjective et objective Evaluation of the Quality of Medical Images for an Optimal use of the display, archiving and transmission Systems).

REFERENCES

[1] A. Nai-Ali and C. Cavaro-Menard, *Compression of Biomedical Images and Signals*. Hoboken, NJ07030, USA: John Wiley & Sons, Inc., 2007.

[2] H. Barrett and K. Myerrs, *Foundations of Image Science*. New Jersey, USA: John Wiley and Sons, Inc., Hoboken, 2004.

[3] J. Brankov, Y. Yang, L. Wei, I. El Naqa, and M. Wernick, "Learning a channelized observer for image quality assessment," *IEEE Transactions on Medical Imaging*, vol. 28, pp. 991–999, 2009.

[4] M. Kupinski and E. Clarkson, "Extending the channelized hotelling observer to account for signal uncertainty and estimation tasks," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 5749, 2005, pp. 183–190.

[5] D. Lowe and A. Ginige, "Image quality assessment using an image activity weighting and the HVS response," *Image and Vision Computing NZ*, vol. 93, pp. 169–176, 1993.

[6] C. K. Abbey and H. H. Barrett, "Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," *Journal of the Optical Society of America A*, vol. 18, pp. 473–488, 2001.

[7] L. Dickinson, H. U. Ahmed, C. Allen, J. O. Barentsz, B. Carey, J. J. Futterer, S. W. Heijmink, P. J. Hoskin, A. Kirkham, A. R. Padhani, R. Persad, P. Puech, S. Punwani, A. S. Sohaib, B. Tombal, A. Villers, J. van der Meulen, and M. Emberton, "Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: Recommendations from a european consensus meeting," *European Urology*, vol. 59, no. 4, pp. 477 – 494, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0302283810011875

[8] C. Cavaro-Ménard, L. Zhang, and P. L. Callet, "Diagnostic quality assessment of medical images: Challenges and trends," in *in Visual Information Processing (EUVIP), 2010 2nd European Workshop on* , July 2010, pp. 277–284.

[9] R. Bar-Shalom, N. Yefremov, L. Guralnik, D. Gaitini, A. Frenkel, A. Kuten, H. Altman, Z. Keidar, and O. Israel, "Clinical performance of PET/CT in evaluation of cancer: additional value for diagnostic imaging and patient management," *Journal of nuclear medicine*, vol. 44, pp. 1200–1209, 2003.

[10] T. Kimpe and Y. Sneyders, "Impact of defective pixels in AMLCDs on the perception of medical images," in *SPIE Medical Imaging*, vol. 6146, 2006.

[11] Y. Jiang, D. Huo, and D. L. Wilson, "Methods for quantitative image quality evaluation of MRI parallel reconstructions: detection and perceptual difference model," *Magnetic Resonance Imaging*, vol. 25, no. 5, pp. 712 – 721, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0730725X06003675

[12] J. P. Johnson, E. A. Krupinski, M. Yan, H. Roehrig, A. R. Graham, and R. S. Weinstein, "Using a visual discrimination model for the detection of compression artifacts in virtual pathology images." *IEEE Transactions on Medical Imaging*, vol. 30, no. 2, pp. 306–314, 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20875970

[13] F. Shen and E. Clarkson, "Using fisher information to approximate ideal-observer performance on detection tasks for lumpy-background images," *Journal of the Optical Society of America A*, vol. 23, no. 10, pp. 2406–2414, 2006.

[14] S. Kulkarni, P. Khurd, I. Hsiao, L. Zhou, and G. Gindi, "A channelized hotelling observer study of lesion detection in SPECT MAP reconstruction using," *Physics in Medicine and Biology*, 2007.

[15] S. Park, E. Clarkson, M. A. Kupinski, and H. H. Barrett, "Efficiency of the human observer detecting random signals in random backgrounds," *Journal of the Optical Society of America A*, vol. 22, no. 1, pp. 3–16, 2005.

[16] J. A. Swets, *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 1996.

[17] H. C. Gifford, M. A. King, R. G. Wells, W. G. Hawkins, M. V. Narayanan, and P. H. Pretorius, "LROC analysis of detector-response compensation in SPECT," *IEEE Transactions on Medical Imaging*, vol. 19, no. 5, pp. 463–473, 2000.

[18] E. Clarkson, "Estimation receiver operating characteristic curve and ideal observers for combined detection/estimation tasks," *JOSA A*, vol. 24, no. 12, pp. B91–B98, 2007.

[19] Clarkson, E., "Estimation ROC curves and their corresponding ideal observers," in *Proc. SPIE*, vol. 6515, 2007.

[20] A. I. Bandos, H. E. Rockette, T. Song, and D. Gur, "Area under the free-response ROC curve (FROC) and a related summary index," *Biometrics*, vol. 65, pp. 247–256, 2009.

[21] D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: modeling, analysis, and validation," *Medical Physics*, vol. 31, no. 8, pp. 2313–2330, 2004.

[22] X. He and E. Frey, "ROC, LROC, FROC, AFROC: an alphabet soup." *Journal of the American College of Radiology JACR*, vol. 6, no. 9, pp. 652–655, 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19720362

[23] D. P. Chakraborty, "New developments in observer performance methodology in medical imaging," *Seminars in Nuclear Medicine*, vol. 41, pp. 401–418, 2011.

[24] D. P. and Chakraborty, "Validation and statistical power comparison of methods for analyzing free-response observer performance studies," *Academic Radiology*, vol. 15, no. 12, pp. 1554 – 1566, 2008.

[25] D. P. Chakraborty, *The Handbook of Medical Image Perception and Techniques*. Cambridge University Press, 2010, no. 16, ch. Recent developments in free-response methodology, pp. 216–239.

[26] W. B. Jackson, M. R. Said, D. A. Jared, J. O. Larimer, J. L. Gille, and J. Lubin, "Evaluation of human vision models for predicting human-observer performance," in *SPIE Medical Imaging*, vol. 3036, 1997, p. 6473.

[27] E. Krupinski, J. Johnson, H. Roehrig, J. Nafziger, J. Fan, and J. Lubin, "Use of a human visual system model to predict observer performance with CRT vs LCD display of images," *Journal of Digital Imaging*, vol. 17, pp. 258–263, 2004, 10.1007/s10278-004-1016-4. [Online]. Available: http://dx.doi.org/10.1007/s10278-004-1016-4

[28] D. Huo, D. Xu, Z.-P. Liang, and D. Wilson, "Application of perceptual difference model on regularization techniques of parallel MR imaging," *Magnetic Resonance Imaging*, vol. 24, no. 2, pp. 123 – 132, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0730725X05003231

[29] S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity," in *Digital Image and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993.

[30] R. Mantiuk, S. Daly, K. Myszkowski, and H. Seidel, "Predicting visible differences in high dynamic range images - model and its calibration," in *Proc. of Human Vision and Electronic Imaging X, IS&T/SPIE's 17th Annual Symposium on Electronic Imaging*, 2005, pp. 204–214.

[31] A. Watson and J. Solomon, "Model of visual contrast gain control and pattern masking," in *Journal of the Optical Society of America A*, vol. 14, no. 9, 1997, pp. 2379–2391.

[32] A. Watson, "The spatial standard observer: A new tool for display metrology," in *Information Display*, vol. 23, 2007, pp. 12–15.

[33] J. P. Johnson, E. A. Krupinski, H. Roehrig, J. Lubin, and J. Nafziger, "Human visual system modeling for selecting the optimal display for digital radiography," in *CARS*, 2004, pp. 335–340.

[34] G. Mather, *Foundations of Sensation and Perception*. Hove and New York, USA: Psychology Press Ltd., Talor & Francis Inc., 2009.

[35] H. H. Barrett, J. Yao, J. P. Rolland, and K. J. Myers, "Model observers for the assessment of image quality," in *Proceedings of the National Academy of Sciences*, D. C. Wilson and J. N. Wilson, Eds., vol. 90. SPIE, 1993, pp. 9758–9765.

[36] S. Park, M. A. Kupinski, E. Clarkson, and H. H. Barrett, "Ideal-observer performance under signal and background uncertainty," in *Information processing in medical imaging proceedings of the conference*, vol. 18, 2003, pp. 342–353.

[37] X. He, B. S. Caffo, and E. C. Frey, "Toward realistic and practical ideal observer (IO) estimation for the optimization of medical imaging systems," *IEEE Transactions on Medical Imaging*, vol. 27, no. 10, pp. 1535–1543, 2008.

[38] C. Abbey and J. Boone, "An ideal observer for a model of x-ray imaging in breast parenchymal tissue," *Digital Mammography*, pp. 393–400, 2010.

[39] R. F. W. Harrison H. Barrett, J. L. Denny and K. J. Myers, "Objective assessment of image quality. II. fisher information, fourier crosstalk, and figures of merit for task performance," *Journal of the Optical Society of America A*, vol. 12, pp. 834–852, 1995.

[40] B. D. Gallas and H. H. Barrett, "Validating the use of channels to estimate the ideal linear observer." *Journal of the Optical Society of America A*, vol. 20, no. 9, pp. 1725–1738, 2003.

[41] R. M. Manjeshwar and D. L. Wilson, "Effect of inherent location uncertainty on detection of stationary targets in noisy image sequences," *Journal of the Optical Society of America A*, vol. 18, no. 1, pp. 78–85, 2001.

[42] C. Castella, M. P. Eckstein, C. K. Abbey, K. Kinkel, F. R. Verdun, R. S. Saunders, E. Samei, and F. O. Bochud, "Mass detection on mammograms: influence of signal shape uncertainty on human and model observers," *Journal of the Optical Society of America A*, vol. 26, no. 2, pp. 425–436, 2009.

[43] M. P. Eckstein and C. K. Abbey, "Model observers for signal-known-statistically tasks (SKS)," in *Proceedings of SPIE*, vol. 4324. SPIE, 2001, pp. 91–102.

[44] M. P. Eckstein, B. Pham, and C. K. Abbey, "Effect of image compression for model and human observers in signal-known-statistically tasks," in *Proceedings of SPIE*, vol. 4686. SPIE, 2002, pp. 91–102.

[45] M. P. Eckstein, Y. Zhang, B. Pham, and C. K. Abbey, "Optimization of model observer performance for signal known exactly but variable tasks leads to optimized performance in signal known statistically tasks," in *Proceedings of SPIE*, vol. 5034. SPIE, 2003, pp. 91–102.

[46] B. Goossens, "Multiresolution image models and estimation techniques," Ph.D. dissertation, Ghent University, Ghent, Belgium, 2010.

[47] G. Olmo, E. Magli, and L. L. Presti, "Joint statistical signal detection and estimation. part I: theoretical aspects of the problem," *Signal Processing*, vol. 80, pp. 57–73, 2000.

[48] E. V. Bart Goossens, Ljiljana Platia and W. Philips, "The use of steerable channels for detecting asymmetrical signals with random orientations," in *Proceedings of SPIE*, vol. 7627. SPIE, 2010.

[49] M. Whitaker, E. Clarkson, and H. Barrett, "Estimating random signal parameters from noisy images with nuisance parameters: linear and scanning-linear methods," *Optics express*, vol. 16, no. 11, p. 8150, 2008.

[50] L. Zhang, C. Cavaro-Ménard, and P. L. Callet, "Using AUC to study perceptual difference model suitability for the detection task on MR image," in *MIPS XIV Conference*, August 2011, student scholar award.

[51] K. Lövblad, N. Anzalone, A. Dörfler, M. Essig, B. Hurwitz, L. Kappos, S. Lee, and M. Filippi, "MR imaging in multiple sclerosis: review and recommendations for current practice," *American Journal of Neuroradiology*, vol. 31, no. 6, pp. 983–989, 2010.

[52] C. Polman, S. Reingold, G. Edan, M. Filippi, H. Hartung, L. Kappos, F. Lublin, L. Metz, H. McFarland, P. O'Connor *et al.*, "Diagnostic criteria for multiple sclerosis: 2005 revisions to the McDonald Criteria," *Annals of neurology*, vol. 58, no. 6, pp. 840–846, 2005.

[53] C. Polman, S. Reingold, B. Banwell, M. Clanet, J. Cohen, M. Filippi, K. Fujihara, E. Havrdova, M. Hutchinson, L. Kappos *et al.*, "Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria," *Annals of neurology*, vol. 69, no. 2, pp. 292–302, 2011.

[54] X. Montalban, M. Tintoré, J. Swanton, F. Barkhof, F. Fazekas, M. Filippi, J. Frederiksen, L. Kappos, J. Palace, C. Polman *et al.*, "MRI criteria for MS in patients with clinically isolated syndromes," *Neurology*, vol. 74, no. 5, pp. 427–434, 2010.

[55] M. Sahraian and A. Eshaghi, "Role of MRI in diagnosis and treatment of multiple sclerosis," *Clinical neurology and neurosurgery*, vol. 112, no. 7, pp. 609–615, 2010.

[56] L. Zhang, C. Cavaro-Ménard, P. L. Callet, and L. H. K. Cooper, "The effects of anatomical information and observer expertise on abnormality detection task," in *in Proc. SPIE Medical Imaging*, vol. 7966, February 2011.

[57] D. P. Chakraborty, "Recent advances in observer performance methodology: jackknife free-response ROC (JAFROC)," *Radiat Prot Dosimetry.*, vol. 114, pp. 26–31, 2005.

[58] E. Samei, A. Badano, D. Chakraborty, K. Compton, C. Cornelius, K. Corrigan, M. Flynn, B. Hemminger, N. Hangiandreou, J. Johnson *et al.*, "Assessment of display performance for medical imaging systems: executive summary of aapm tg18 report," *Medical physics*, vol. 32, p. 1205, 2005.