Numerical Observers for the Objective Quality Assessment of Medical Images

Lu Zhang

l'université nantes angers le mans
PÔLE DE RECHERCHE ET D'ENSEIGNEMENT SUPÉRIEUR

université angers

Thèse de Doctorat

Lu Zhang-Ge

Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université d'Angers
sous le label de l'Université de Nantes Angers Le Mans

Discipline : CNU 27 (Informatique) et CNU 61 (Génie informatique, automatique et traitement du signal)
Spécialité : Traitement des images et du signal
Laboratoire : Laboratoire d'Ingénierie des Systémes Automatisés (LISA);
et IVC - Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN)

Soutenue le 28 novembre 2012

École doctorale : 503 (STIM)
Thèse n° : 1274

Modèles Numériques pour l'Évaluation Objective de la Qualité d'Images Médicales

Numerical Observers for the Objective Quality Assessment of Medical Images

JURY

Rapporteurs :	**M^me^ Elizabeth KRUPINSKI**, Professeur, Université d'Arizona (USA) **M. Claude LABIT**, Directeur de recherches, IRISA / INRIA Rennes
Examinateurs :	**M^me^ Anne GUERIN-DUGUE**, Professeur, GIPSA-lab, Université de Grenoble **M. Pierre JANNIN**, Chargé de recherches 1 Inserm, LTSI, Université de Rennes 1 **M. Imants SVALBE**, Professeur, Monash University (Australie)
Directeur de thèse :	**M. Patrick LE CALLET**, Professeur, IRCCyN, Université de Nantes
Co-directrice de thèse :	**M^me^ Christine CAVARO-MÉNARD**, Maître de conférences, LISA, Université d'Angers

Acknowledgements

I would like to express my sincere gratitude towards my thesis advisors, Professor Patrick LE CALLET and Dr Christine CAVARO-MENARD for giving me the opportunity to work with them, following my research with great interest and productive advises, constantly supporting and encouraging me, reviewing my reports, papers and thesis and giving many comments and suggestions.

I am very thankful for Doctor Jean-Yves TANGUY (Hospital of Angers) for many interesting discussions on medical issues, providing medical image data sets and giving me feedback after each experiment. I also appreciate all the radiologists from the Hospital of Angers who participated in the subjective experiments, for generously giving their time and effort on performing the experiment and giving their feedback.

I am grateful to everyone in the research department TELIN (Telecommunications and Information Processing) of Gent University in Belgium, for being nice and showing their interests in my research during my stay. Special thanks go to Bart Goossens for his friendship, fruitful discussions on model observers and providing his source code of the SKS CHO for the detection of signal with varying orientation, and to Ljiljana Platisa for useful discussions, nice collaboration and making my stay enjoyable.

My appreciation also goes to all my colleagues from the laboratory LISA of University of Angers and from the research group IVC of Polytech Nantes, for the pleasure time.

I would also like to acknowledge all the people that I met during the conferences for sharing their knowledge, the nice discussions and correspondances.

This thesis would not have been possible without the financial funding from the “Région des Pays de La Loire, France” for supporting the EQuIMOSe project (Subjective et objective Evaluation of the Quality of Medical Images for an Optimal use of the display, archiving and transmission Systems).

Last but not least, I would like to thank my husband for his unceasing support, my parents for their encouragement during the past three years, and my sweetheart, my little baby Léo, whose

smile and love are worth it all.

List of Acronyms

AFAUC:	Area Under the AFROC Curve
AFROC:	Alternative Free-response ROC
AUC:	Area Under the ROC Curve
BIC:	Bayesian Information Criterion
BKE:	Background Known Exactly
BKS:	Background Known Statistically
CGB:	Correlated Gaussian Background
CHO:	Channelized Hotelling Observer
CJO:	Channelized Joint detection and estimation Observer
CLB:	Clustered Lumpy Background
CR:	Computed Radiography
CSF:	Contrast Sensitivity Function
CT:	Computed Tomography
D-DOG:	Dense DOG
DICOM:	Digital Imaging and Communications in Medicine
DOG:	Difference-Of-Gaussians
DR:	Digital Radiography
EM:	Expectation Maximization
EROC:	Estimation ROC

FAUC:	Area Under the empirical FROC curve
FFT:	Fast Fourier Transform
FLAIR:	FLuid Attenuated Inversion Recovery
FN:	False Negative
FOM:	Figure Of Merit
FP:	False Positive
FPF:	False Positive Fraction
FROC:	Free-response ROC
GMM:	Gaussian Mixture Model
GUI:	Graphical User Interface
HDR-VDP:	High Dynamic Range Visible Difference Predictor
HGG :	High-Grade Glioma
HO :	Hotelling Observer
HVS :	Human Visual System
IGMM:	Infinite Gaussian Mixture Model
IO:	Ideal Observer
JAFROC:	Jackknife Alternative Free-Response ROC
JDE:	Joint Detection and Estimation
JND:	Just Noticeable Difference
LAUC:	Area Under the LROC curve
LB:	Lumpy Background
LG:	Laguerre-Gaussian
LROC:	Localization ROC
LUT:	Look-Up Table

MAP: Maximum A Posteriori
MO: Model Observer
MRI: Magnetic Resonance Imaging
MS: Multiple Sclerosis
MSE: Mean Square Error
msCHO: multi-slice CHO
msPCJO: multi-slice PCJO

NPWMF: NonPreWhitening Matched Filter
NRMSE: Normalized Root-Mean-Square Error

PACS: Picture Archiving and Communication Systems
PC: Percentage of Correct decisions
PCJO: Perceptually relevant Channelized Joint Observer
PDF: Probability Density Function
PDM: Perceptual Difference Model
PET: Positron Emission Tomography
PSF: Point Spread Function
PSNR: Peak Signal-to-Noise Ratio

RMSE: Root-Mean-Square Error
ROC: Receiver Operating Chracteristic

S-DOG: Sparse DOG
SKE: Signal Known Exactly
SKEV: Signal Known Exactly but Variable
SKS: Signal Known Statistically
SNR: Signal-to-Noise Ratio
SPECT: Single-Photon Emission Computed Tomography
SSO: Spatial Standard Observer
ssCHO: single-slice CHO

TN:	True Negative
TP:	True Positive
TPF:	True Positive Fraction
VBI:	Variational Bayesian Inference
VCGC:	Visual Contrast Gain Control
vCHO:	volumetric CHO
VDM:	Visual Discrimination Model
VDP:	Visible Difference Predictor
WNB:	White Gaussian Background

Contents

List of Tables

List of Figures

1

Introduction

Over the past twenty years, computer technology has facilitated the phenomenal development of the digital medical imaging, which is now widespread for diagnosis and therapy. Different digital imaging modalities have revolutionized health care delivery around the world: Computed tomography (CT), Magnetic Resonance Imaging (MRI), ultrasound, Computed Radiography (CR)/Digital Radiography (DR), fluoroscopy, Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), etc.. The performance of a digital medical imaging system (including acquisition system, image post-processing system and visualization system) has obviously a strong influence on the diagnostic quality, the safety and efficacy of the interventional therapies performed by radiologists (or physicians).

"Facing a large number of digital medical imaging systems, how to choose the right systems to ensure the image quality and ultimately the patient care?" This is a ticklish question which any medical institute (or radiologist) must deal with. For example, numerous medical imaging display stations (part of imaging system) are proposed. They offer different contrast ratio, resolution and grey-level representation (8-bit or 12-bit). One question is then important: what are the best parameters for a certain application? Moreover, image compression is an effective means to archive medical images. When there are several tera bytes to archive per institute per year, are lossless compression algorithms with low compression ratio (3:1 to 4:1) sufficient compared to lossy compression algorithms with a high compression ratio (10:1 to 100:1)?

To answer the above type of questions, there is a need for good measurements of medical image quality and for measuring the impact of quality on diagnostic accuracy and efficacy.

Many studies have reported the use of simple numerical approach such as the Peak Signal-to-Noise Ratio (PSNR) to evaluate the medical image quality for a new acquisition protocol, a new image post-processing algorithm (compression, watermarking, etc.) or a new display station. However, this type of simple numerical approaches does not take into account the complex anatomical nature of the images and does not correlate well with perceived quality measurement [1].

A more reasonable measurement method is the so-called *task-based approach* [2], where image quality should be assessed in the context of a specific diagnostic task. The underlying paradigm is to quantify the quality of the image by its effectiveness with respect to its intended purpose.

In the frame of *task-based approach*, one or more diagnostic tasks can be performed by a panel of human observers (radiologists). During the assessment process, radiologists are asked to perform the same diagnostic task given different medical imaging systems, which could be acquisition devices, image post-processing algorithms or image visualization systems. A better

task performance means a better diagnostic accuracy. Thus the system that enables radiologists to gain a better task performance or to spend less time for interpretation with the same diagnostic reliability is said to be better. The radiologists are the end-users of medical images, thus the task performance of human observers is always the ultimate test of image quality. However, the assessment methods using human observers requires time-consuming and expensive efforts of radiologists, often highly trained and experienced experts [3], besides which variance exists between and within radiologists' responses [4].

Therefore, there is a growing demand for numerical observers (mathematical models). Once the numerical observer achieves the task performance of medical experts, it can be used to evaluate different medical imaging systems or other digital specialities with these same requirement such as telemedicine, by employing the same paradigm as the assessment method using human observers (mentioned above). And it is much easier, faster and cheaper to carry out, and has no variance within its responses. Then 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 imaging systems. The optimal system enables radiologists to make a reliable diagnosis in the shortest possible time. It has been argued that numerical observers are useful for medical image quality assessment in place of human observers [5]. Though we need to note that the task performance of human observers is still needed for the validation of a numerical observer [6].

The objective of this thesis is indeed to propose more clinically relevant [1] numerical observers that can approach the diagnostic task performance of radiologists for the objective quality assessment of medical images. The new numerical observers will be ultimately applied to quantify the performance of medical imaging systems.

1.1 Studied tasks, modality and pathology

1.1.1 Studied tasks

The key problem of designing a numerical observer lies in modeling the diagnostic process of radiologists. It is useful to firstly well characterize the diagnostic process in order to model it. Except for images obtained to rule out the presence of disease, one commonly accepted approach that we find in the literature [7, 8, 9] is to divide the diagnostic process into three tasks: detection, localization and characterization.

1. "More clinically relevant" here means the numerical observer can perform the diagnostic task as radiologists do in the clinical routine. For example, it cannot only detect but it can also localize multiple signals with unknown parameters (amplitude, orientation, size and location).

- The *detection task* simply requires the observer to find potential lesions while making as few false decisions as possible.
- The *localization task* consists in indicating the location of each lesion. The location is a clinically relevant variable, especially if it guides subsequent therapeutic interventions (e.g. biopsy or surgery).
- 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 (such as benign vs malignant, etc.), which is the most complicated task. Part of some characterization tasks is the *estimation task* [2], which means estimation of one or a few parameters of the lesion (e.g. intensity, size and orientation) to be detected. The estimation task is of direct relevance to the purpose for which the image was obtained. For example, for some pathologies, 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 joint estimation-detection task.

So far to our knowledge, existing numerical observers are limited in task range: most of them are dedicated to the detection task [10, 11, 12, 13, 14], much fewer are concerned with the localization task [15], and none with the entire characterization task.

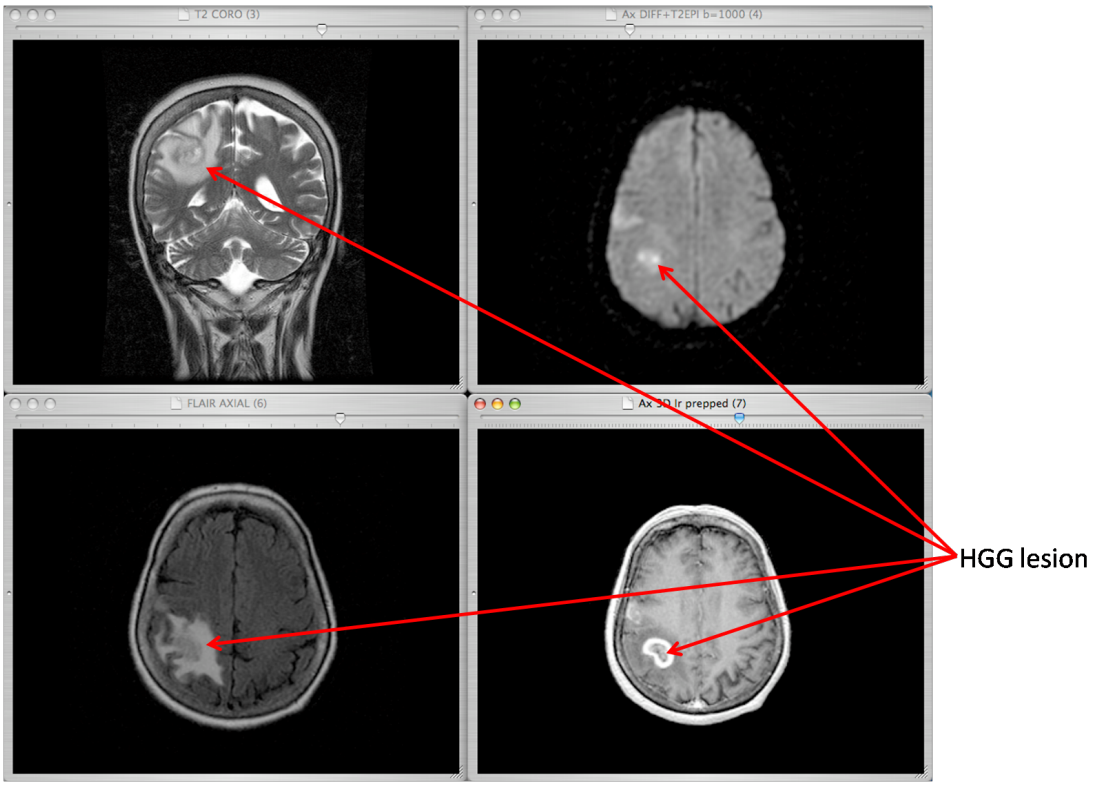
The scope of this thesis ranges from the detection task to the localization task, and the estimation task (only a part of the characterization task considering the complexity of this task).

1.1.2 Studied modality and pathology

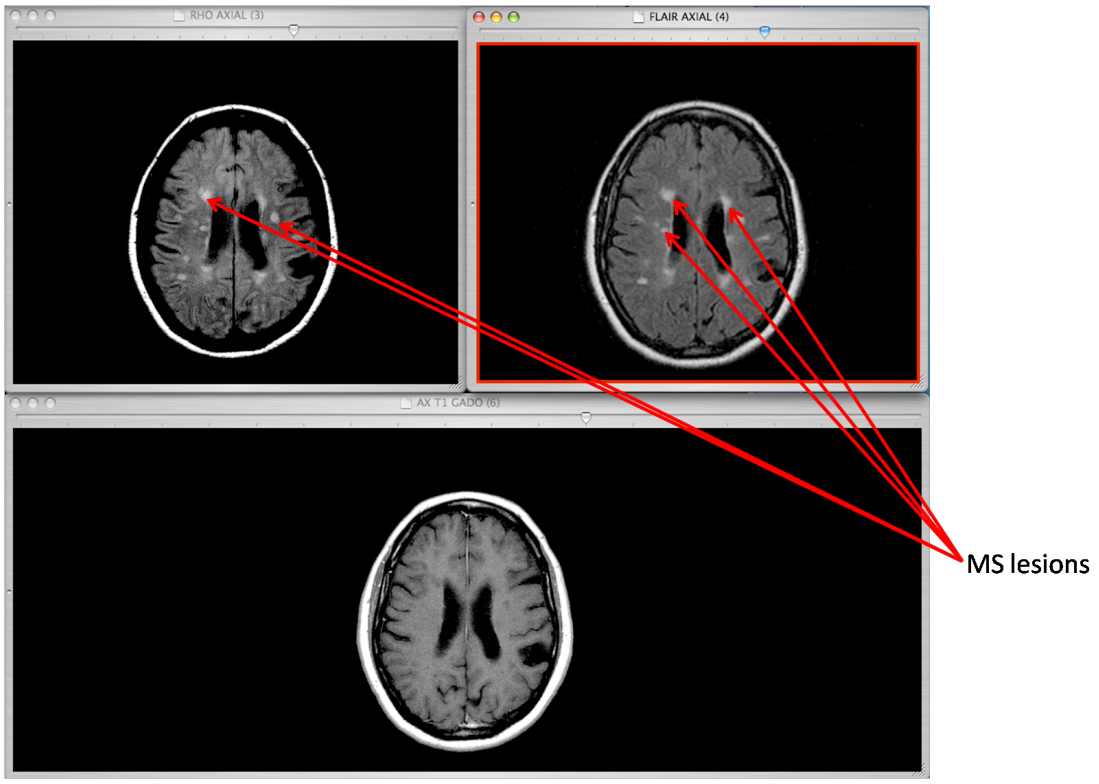
In this thesis, we restricted the scope to the MRI modality and the pathology of Multiple Sclerosis (MS) for all our studies.

Figure 1.1: MR images of two pathologies on different sequence (figures from [16]): Figure 1.1a High-Grade Glioma (HGG) and Figure 1.1b Multiple Sclerosis (MS).

(a)



(b)



The choice of the studied pathology is connected with the tasks that we studied. As far as the detection and localization tasks are concerned, the abnormalities must not be too conspicuous. Otherwise, observers (radiologists or numerical observers) could always detect and localize the abnormalities easily. In that way, different imaging systems can not be differentiated through different task performances of observers. Taking Magnetic Resonance Imaging (MRI) modality for example, Figure 1.1 shows MR images of two pathologies: High-Grade Glioma (HGG) and Multiple Sclerosis (MS). It is not hard to see from Figure 1.1a that the HGG lesion is so obvious on any one of the six sequences that even a person without any medical background can tell where it is. The evidence of the HGG lesion would not change even with moderate degradation of the imaging system for detection and localization tasks. In contrast, the MS lesions are much more subtle and difficult to be perceived, as seen from Figure 1.1b. Furthermore, an accurate detection and localization of MS lesions is the first and the most important step for the diagnosis and treatment of MS patients.

Moreover, for this particular pathology, orientation, size, position and number of lesions are all considered as diagnostic criteria [17]. MS pathology is then well adapted for our numerical observer study (with the aim of proposing new numerical observers which cannot only detect but also localize multiple signals with unknown amplitude, orientation, size and location). 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 [18].

Since the criteria for MS are largely based on brain MRI scan outcomes [19], we chose MRI as the studied modality.

1.2 Organization of this thesis

This thesis is organized as illustrated in Figure 1.2. We begin with a general introduction in this chapter and finally summarize the thesis in Chapter 10 while discussing some directions for future research. The major portion of the thesis consists of three parts:

- The first part (Chapters 2 to 4) deals with the task performance quantification methods and the existing numerical observers, which lays the foundation of this thesis.
- The second part (Chapters 5 and 6) presents the experimental pre-studies for investigating the suitability of the use of existing numerical observers in our application, which lead to our proposition of novel numerical observers.
- The third part (Chapters 7 to 9) is dedicated to our proposed novel numerical observers,

which try to overcome the limitations of existing numerical observers in our application. Their validation subjective studies are also detailed in this part.

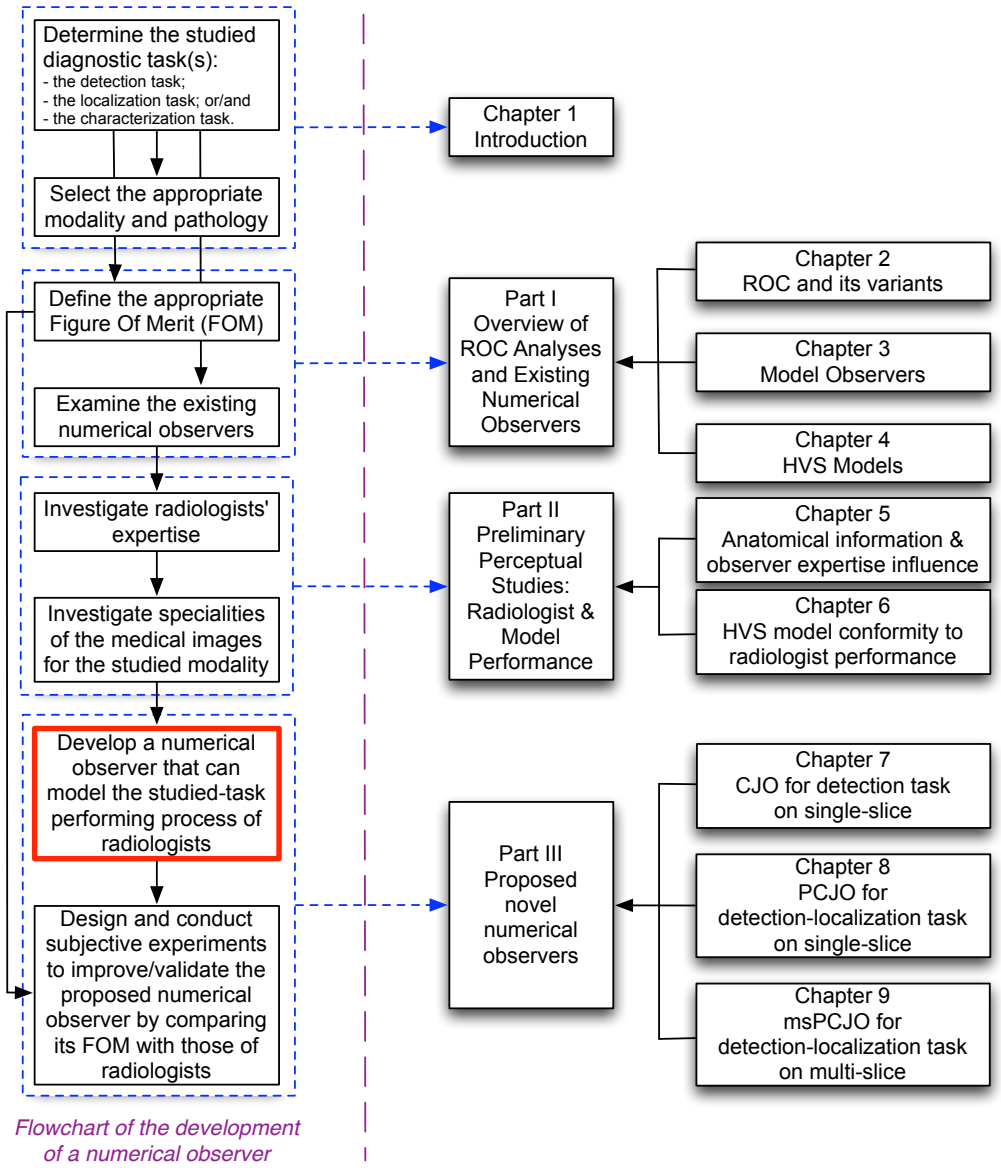


Figure 1.2: Organization of this thesis.

Part I

Overview of ROC Analyses and Existing Numerical Observers

Introduction of Part I

The first part of this thesis is dedicated to the state-of-the-art of the observer performance evaluation methods in the medical image quality assessment domain and the existing numerical observers.

Regarding medical image quality, the best medical image is not the one that looks most beautiful, but the one that enables observers to achieve the best *average* performance in a specific diagnostic task. Thus the first and important knowledge for the medical image quality assessment is how to quantify the task performance of observers, either human observers or numerical observers. Given the confidence rating responses of observers for a set of test images in the detection task, we have the classic Receiver Operating Characteristic (ROC) curve to depict the observer's performance. Given the confidence rating and location coordinates responses of observers for a set of test images in the joint detection-localization task, we have the variants of ROC curve. Since our studied tasks include both the detection task and the localization task, we will firstly introduce these increasingly popular methods in the medical domain.

With ROC analysis and its variants, medical imaging systems can certainly be assessed by empirical studies using human observers (as said in Chapter 1). But this was generally costly and time-consuming. Correspondingly, numerical observers have been proposed as a surrogate for human observers for medical image quality assessment. Numerous numerical observers have been proposed over last decades and they can be divided into two categories: model observer (MO) and human visual system (HVS) model.

One category of existing numerical observers - MO - considers the image quality as a statistical concept and uses statistical decision theory to mathematically characterize the image quality. The other category of existing numerical observers - HVS models - evaluates the image quality based on the characteristics of the HVS, considering that human observers are the ultimate receivers of the visual information contained in an image. Both of the categories have their advantages and limitations. Our proposed numerical observers are inspired from both of the families, and are designed to overcome some of their limitations. Thus the first part of this thesis presents a variety

of numerical observers of the two categories.

This part comprises three chapters.

In Chapter 2, we review the commonly used analysis method to summarize and quantify observer's different diagnostic task performance: the ROC analysis, the Localization ROC (LROC), the Free-response ROC (FROC) and the Alternative Free-response ROC (AFROC). We also describe some ways to establish the gold standard, which is indispensable for the ROC analyses.

In Chapter 3, we discuss a number of MOs in the order of their evolution, from SKE to SKS, from single-slice to multi-slice. We discuss their mathematical backgrounds, their distinguishing features, and their limitations.

In Chapter 4, we present basic characteristics of the HVS, as well as the frequently-used modeling of these characteristics. In particular, we devote attention to their applications in medical image quality assessment.

2

ROC and its variants

In this chapter, we will discuss ROC and its variants:

1. Receiver Operating Characteristic (ROC) analysis [20] for the detection task;
2. Localization ROC (LROC) analysis [21] for the detection-localization task of one signal on an image;
3. Free-response ROC (FROC) analysis [22] for the detection-localization task of multiple signals on an image;
4. Alternative Free-response ROC (AFROC) analysis [23] for the detection-localization task of multiple signals on an image.

The above series of analysis will be called "the ROC analyses" hereafter.

The diagnostic accuracy is the extent to which the results of a diagnosis agree, in some statistical sense, with patients' actual states of health or disease. Thus for the detection task, this means the ability to distinguish between two composite states of truth, "positive" versus "negative" with respect to a specified disease; for the localization task, this means the ability to localize the disease abnormality (e.g. lesion) position. The ROC analyses provide the most comprehensive description in the sense that it displays all of the combinations of sensitivity and specificity that a diagnostic test is able to provide as the observer's "decision criterion" is varied [24].

Different figures of merit (FOMs) have been developed based on the ROC analyses to characterize and quantify a certain task performance. FOMs can be the percentage of correct decisions (PC), the area under the ROC curve (AUC), the intercept of the ROC curve with the line at 90 degrees to the no-discrimination line (also called Youden's J statistic) [25], the area between the ROC curve and the no-discrimination line, the d' (d-prime, that is the distance between the mean of the distribution under lesion-absent hypothesis and that under lesion-present hypothesis, divided by their standard deviation, under the assumption that both these distributions are normal with the same standard deviation), partial ROC areas, the area under the LROC curve (LAUC), the area under the empirical FROC curve(FAUC), the augmented area under FROC curve (AFAUC), jackknife alternative free-response ROC (JAFROC1 and JAFROC2), etc.. The fundamental idea in the FOM design is to compute a FOM that rewards correct decisions and penalizes incorrect decisions.

The ROC analyses are essential for comparing or assessing different imaging systems, as well as for validating a numerical observer by comparing its FOM results with those gotten from subjective experiments.

2.1 Gold standard

A fundamental concept in the ROC analyses is *gold standard* (ground truth), which is the true status of the images, that is to say the presence or absence of an abnormality for the detection task, and the true abnormality positions for the localization task.

It is difficult to establish the gold standard without external corroborating evidence such as pathology or other accepted tests, especially when the studied imaging modality itself is the gold standard. For example, the brain MRI exam result is normally used as the gold standard for the pathology of MS [19]. When the studied images are brain MR images of MS, what would be the gold standard?

Even when a method for establishing the gold standard exists, the method is more often a "bronze" standard rather than "gold" [26]. For example, biopsy proof is often used as the gold standard for mammogram interpretation, but biopsy needles may miss their mark and pathologists may make mistakes as well [2].

One alternative is to call upon a group of experts (truth-panel) who are then asked to establish, according to consensus, a gold standard on the images being studied. It is also possible to ask these experts to give their individual opinion on the image, and then keep only the images on which all physicians raise a similar diagnosis [27]. Another useful way is to use simulated images (adding simulated lesions on clinical images or simulated images), for which the gold standard is known perfectly because the status of images are totally controlled by the investigator.

2.2 ROC

2.2.1 ROC curve

A development of signal detection theory, the ROC analysis has been widely applied in the medical fields. To evaluate the observer's task performance, ROC analysis has been used for the detection task, where a binary positive/negative decision is made by comparing observer's confidence rating with a decision criterion. A positive decision is called a *true positive* (TP) when the ground truth of an image is lesion-present, or a *false positive* (FP) on the contrary; a negtive decision is called a *true negative* (TN) when the ground truth of an image is lesion-absent, or a *false negative* (FN) on the contrary, as illustrated in Table 2.1 and Figure 2.1.

Table 2.1: Four decision states of observer's decision: true positive(TP), false positive(FP), false negative (FN), true negative (TN).

		Gold Standard	
		lesion-present	lesion-absent
Observer's Decision	positive	true positive(TP)	false positive(FP)
	negative	false negative (FN)	true negative (TN)

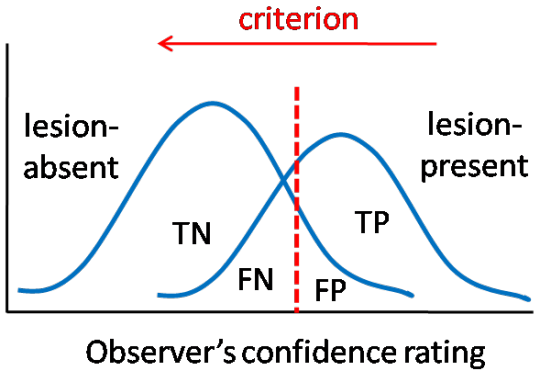


Figure 2.1: The probability density functions (PDFs) for the observer's response (confidence rating) under the two hypotheses (lesion-absent and lesion-present). And the varying decision criterion with the resulting four decision states: true positive(TP), false positive(FP), false negative (FN) and true negative (TN).

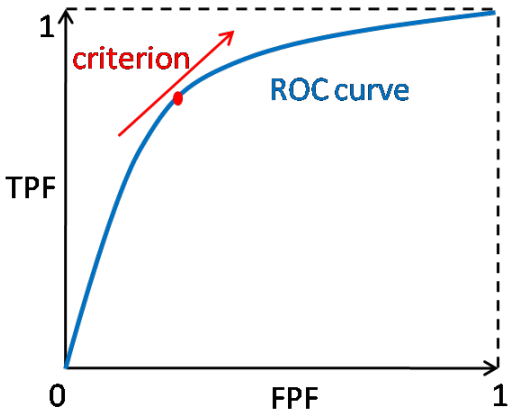


Figure 2.2: An example of ROC curve, where the vertical coordinate at each point on the ROC curve indicates the fraction of TPs out of the actual positives (TPF) that the observer can achieve at each value of the horizontal coordinate, the fraction of false positives out of the actual negatives (FPF). Different points on the curve are obtained by different settings of the observer's decision criterion, above which observer's responses (confidence ratings) are classified as positive and below which they are classified as negative.

A ROC curve (cf. Figure 2.2) is a graphical plot of the fraction of TPs out of the actual positives (TPF) vs. the fraction of false positives out of the actual negatives (FPF). It is a comparison of two operating characteristics (TPF & FPF) as the decision criterion changes. In medical domain, TPF is also known as *sensitivity*, and FPF is one minus the *specificity*. The sensitivity is the ability to correctly identify those with the disease, whereas the specificity is the ability to correctly identify those without the disease.

The best possible decision of an observer would yield a point in the coordinate $(0, 1)$ of the ROC space in Figure 2.2 (the upper left corner), representing 100% sensitivity (no false negatives) and 100% specificity (no false positives). A completely random guessing would give a point along the diagonal line (the so-called line of no-discrimination) from the coordinate $(0, 0)$ to the coordinate $(1, 1)$ (from the left bottom corner to the top right one). The diagonal divides the ROC space. Points above the diagonal represent good decisions (better than random), points below the line represent poor decisions (worse than random). Note that the output of consistently poor decisions (which means that the observer always makes an opposite diagnosis compared to the gold-standard) could simply be inverted to obtain good decisions.

2.2.2 Area under the ROC curve (AUC)

The AUC is a summary statistic of the ROC curve, used to reduce the ROC performance to a single scalar value representing the performance. Since the AUC is a portion of the area of the unit square, its value will always be between 0 and 1. The best possible decision of an observer yields an AUC of 1; and a completely random guessing produces an AUC of 0.5 (chance value).

One important statistical property of the AUC is: AUC is equal to the probability that an observer will rank a randomly chosen positive instance higher than a randomly chosen negative one (assuming 'positive' ranks higher than 'negative'). This is equivalent to the Mann-Whitney U-test [28], which tests whether positives are ranked higher than negatives. It is also equivalent to the Wilcoxon signed-rank test. In this way, it is possible to calculate the AUC by using an average of a number of trapezoidal approximations [28]. In addition, a variety of ROC analysis softwares can be downloaded from websites. A list is given on http://mips.ws/.

Supposing that one ROC curve (corresponding to observer A's performance) is above another one (corresponding to observer B's performance) everywhere, it suggests that:

1. for any given set of costs or benefits for the various correct and incorrect decisions and prevalence of the two diagnostic classes, the observer A always results in better task performance in terms of cost-benefit analysis;

2. the observer A always has a higher sensitivity for any given specificity;
3. the observer A always results in a greater number of correct decisions regardless of disease prevalence [29].

In this case, the higher the AUC is, the better the observer is. However, in cases where ROC curves cross each other, the AUC value only characterizes the average task performance.

Note that any attempt to summarize the ROC curve into a single number loses the particular discriminating procedure, even traditional ROC has been significantly improved over the years [24, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39].

2.3 Variants of ROC

The traditional ROC does not take location into account. Therefore, several extensions to the traditional ROC have been proposed to address the limitation, especially:

- Localization ROC (LROC) for the joint detection-localization of one lesion on an image;
- Free-response ROC (FROC) and Alternative Free-response ROC (AFROC) for the joint detection-localization of multiple signals on an image.

We then present the three ROC's variants.

It should be noted that the localization task concerns a closeness or proximity criterion in the presence of a lesion according to the gold standard. The proximity criterion could be an acceptance radius (for small lesions) or a percentage overlap of the outlined region and the true lesion (for big lesions). What constitutes the proximity criterion is a clinical decision that should be decided on the basis of the clinical application.

When the lesion is small, its position is often represented by its center. The LROC/FROC/AFROC paradigm classifies the marked coordinates, regardless of the marked rates, as a TP or FP by comparing the distance between the marked point and the actual center of the lesion to an predefined acceptance radius: TP when the distance is less than the acceptance radius; FP otherwise. The acceptance radius could be the diameter of the lesion. Then if the corresponding marked rate is greater than the decision criterion, the TP or FP mark is classified as "positive", and "negative" in reverse.

2.3.1 LROC

Consider the case where an observer has a correct decision by reporting an image as positive, but for a wrong reason since he may mistakenly identify a suspicious region on the background as

the lesion, while neglecting the actual true lesion [40]. Such mistakes may lead to serious clinical consequences. To avoid this ambiguity, LROC has been proposed [21]. In LROC paradigm, the observer is forced to detect and locate the lesion together with a confidence rating [41]. Obviously, a task that requires localization is harder than one that simply asks to observer to report "yes/no, something is in the image".

The ordinate of the LROC curve is the fraction of "positive" TP marks out of the number of actual positive images; while its abscissa is the fraction of "positive" FP marks out of the number of actual negative images, as shown in Figure 2.3. The LROC curve must contain the left bottom point $(0, 0)$, but does not necessarily include the top right point $(1, 1)$. In other words, when the fraction of "positive" FP marks out of the number of actual negative images is 1, the fraction of "positive" TP marks out of the number of actual positive images may be lower due to wrong localization results.

The area under the LROC curve (LAUC) is also a summary figure of merit of the LROC curve, which takes values in the range $[0, 1]$.

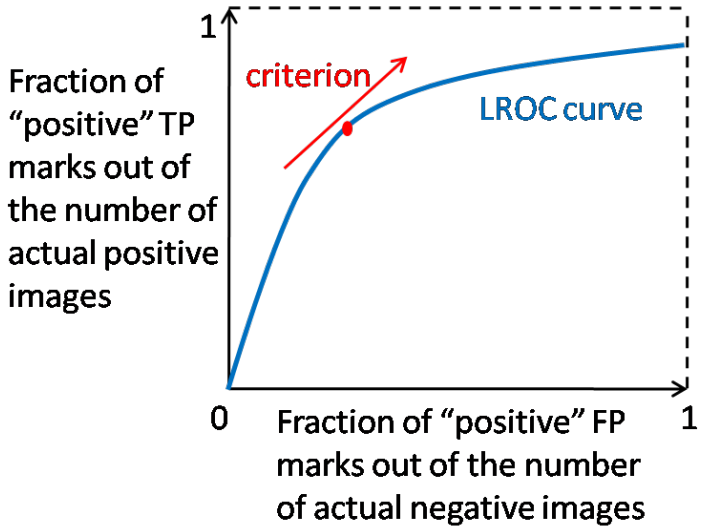


Figure 2.3: An example of LROC curve

Note that a generalization of the LROC curve to cases where the unknown lesion parameter other than location (e.g. the number of lesions, the amplitude, the orientation, etc.) is being estimated is the Estimation ROC (EROC) curve [42, 43]. In order to evaluate observers on more general combined detection and estimation task, the EROC analysis uses the "expected utility" of the estimate of the signal parameter for the TP cases as its ordinate and the same abscissa as the LROC.

2.3.2 FROC

Consider the case where an observer detects and localizes correctly one lesion on the image, but there are actually multiple lesions on the image and the observer misses other lesions. Such mistakes can not be considered in the LROC paradigm, but may introduce serious clinical consequences. Thus FROC has been proposed to allow the observer to specify an arbitrary number of lesions on an image in addition to providing a confidence rating for each lesion.

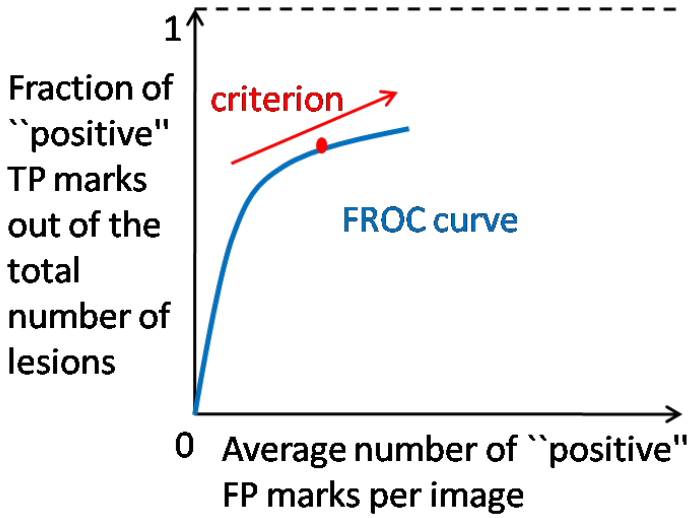


Figure 2.4: An example of empirical FROC curve

As shown in Figure 2.4, an empirical FROC curve is a graphical plot of the fraction of "positive" TP marks out of the total number of lesions (TPF) vs. the average number of "positive" FP marks per image (FPR). The empirical FROC is useful for knowing whether the observer is using all the rating scales. Ideally, the empirical FROC curve should approach a plateau towards the upper end; the top right endpoint of the FROC curve should not be too close to the point (1,1) in the FROC space. However, the area under the empirical FROC curve can no longer be a FOM to summarize FROC curve, since a larger value of area under the empirical FROC curve can result either from an increase in "positive" TP marks or an increase in the number of "positive" FP marks on each image [29].

One FOM that could be used to summarize FROC data is the "augmented area under FROC curve" [22], which can be interpreted geometrically as a measure of average performance superiority over an artificial *guessing* free-response process. This FOM rewards for higher ability to discriminate between FP and TP marks as well as for the higher detection fraction $\widehat{TPF}_{max}$, while penalizes for a higher $\widehat{FPR}_{max}$ and for a large effect of the acceptance radius, where

$(\widehat{FPR}_{max}, \widehat{TPF}_{max})$ are the TPF and FPR when the criterion is so low that all rated marks are considered "positive". However, the "augmented area under FROC curve" should be avoided when the lesion density is really low, for example, big image size with a small quantity of lesions. Since in this case, this FOM equals approximately $\widehat{TPF}_{max}$, which means that this FOM would depend only on the degree of localization accuracy, but has nothing to do with the rating performance.

2.3.3 AFROC

The AFROC paradigm provides an alternative way to analyze FROC data. The AFROC and the FROC curves have the same ordinates, but differ in the abscissas. For each criterion, no matter how many FPs are on one image, the AFROC only considers one FP mark, which is the one with the highest rating. Then it uses the number of highest rated FPs divided by the total number of cases as its abscissa, as shown in Figure 2.5. The AFROC curve include the point $(1, 1)$, the same as the ROC curve.

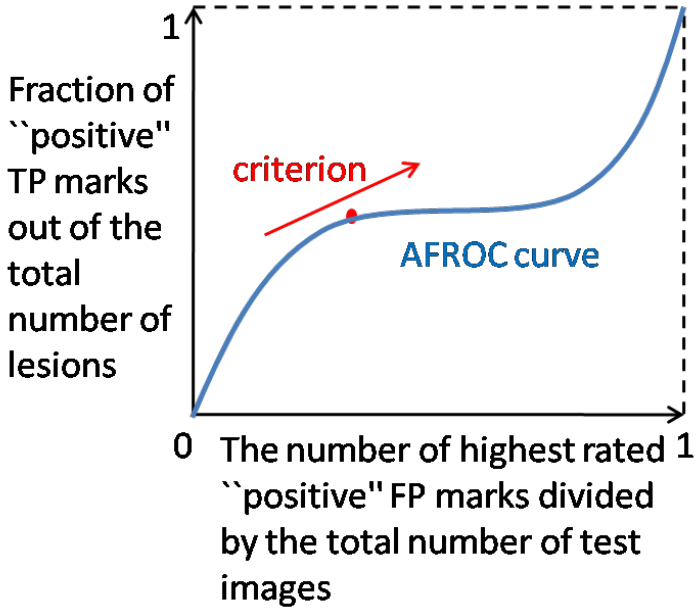


Figure 2.5: An example of AFROC curve

Two most common FOMs within the AFROC paradigm are: the area under the AFROC1 curve estimated using the jackknife procedure (JAFROC1) and the area under the AFROC2 curve estimated using the jackknife procedure (JAFROC2) [44]. 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 [45]. To the author's knowledge, JAFROC1 is the FOM that has the highest statistical power up to now for the detection-localization task.

Note that a completely random guessing would give a JAFROC1 FOM value of zero, or close to zero (chance value) [46], 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 the ROC analysis, where localization is not required and the chance value is 50%, as said in Section 2.2.2.

The JAFROC software can be downloaded for free from the website: http://www.devchakraborty.com/downloads.html.

2.4 Conclusion

In this chapter, we have discussed the ROC analyses that have been proposed to describe different diagnostic task performances, as summarized in Figure 2.6.

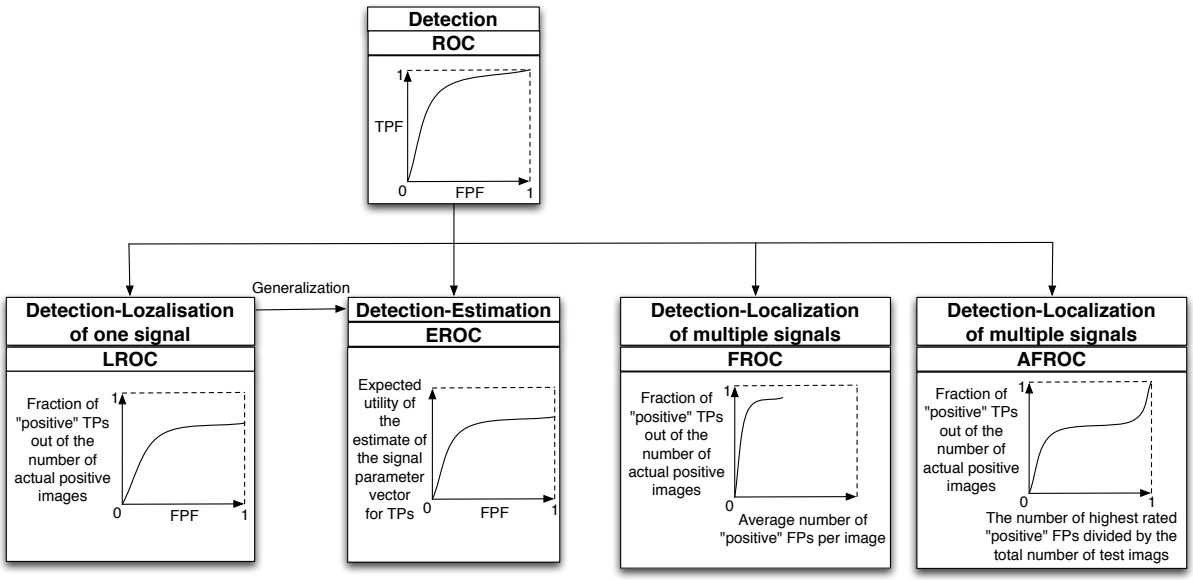


Figure 2.6: ROC analyses characterize the task performance for different diagnostic tasks.

These ROC analyses are also useful for the validation of numerical observers: the FOM can be calculated for both a numerical observer and radiologists; then the FOM of the numerical observer can be compared to those of radiologists; if there is not a big difference between them, the numerical observer can be said to approach well the radiologists' performance.

In Part III of this thesis, we will propose novel numerical observers for the detection task and the joint detection-localization task of multiple signals. Thus the ROC analysis and the AFROC analysis (JAFROC1) will be employed to validate our numerical observers.

3

Model Observers (MO)

This chapter is dedicated to a review of the existing MOs. We begin with the introduction of MO's mathematical background, including some of the most popular image background (White Gaussian Background, Correlated Gaussian Background, Lumpy Background and Clustered Lumpy Background) and the usually used signal model (Gaussian signal). Note that we will use the Correlated Gaussian Background and the Gaussian signal for our proposed numerical observers in Part III of this thesis.

Then in Section 3.2, we will present the different basic model observers:The ideal observer (IO) 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 in general are high-dimensional functions and difficult to compute except for some special cases. Its discriminant function, the likelihood-ratio, is usually a nonlinear function of the image data. Instead of PDFs, the Hotelling observer (HO) 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. The channelized Hotelling observer (CHO) emerged as required to reduce the dimentionality through *channelization*. The channelization is a technique that decomposes an image into different spatial frequency and orientation channels by a series of filters. And the choice of channel type determines the optimality of CHO: if the channels are designed to achieve the performance of HO, the CHO is still an optimal linear observer; if the channels are designed to model the intrinsic spatial frequency and orientation selectivity of the human visual system (HVS), then the CHO is an anthropomorphic observer. Note that the channelization technique of the CHO, which has several advantages, will also be employed in our proposed numerical observers (cf. Part III).

Section 3.3 reviews the Signal Known Statistically (SKS) MOs, which are more clinically relevant and may have a great potential to behave more closely to human observers than the Signal Known Exactly (SKE) MOs. Several SKS MOs have been proposed for different cases. In case of the detection of one signal with varying size and shape, different templates for different combinations of signal parameters could be established and the optimal template response in a maximum likelihood sense is output as the final test statistic. In the case of one signal with varying location, the *scanning* CHO could exhaustively scans the image and output the location giving the largest test statistic. For the detection of one signal with varying orientation, size or amplitude, a series of SKS CHOs based on the joint detection and estimation (JDE) theory were proposed recently. Note that our proposed numerical observers will also treat the SKS task (cf.

Part III).

Finally, multi-slice MOs proposed in the literature are presented, considering the trend to volumetric images in medical domain. Three categories of the existing multi-slice MOs are single-slice CHO (ssCHO), volumetric CHO (vCHO) and multi-slice CHO (msCHO). The ssCHO does not incorporate information in the z direction, and only considers the central slice; the vCHO may have more calculation burden when the image size increases; the msCHO firstly obtains a score for each slice in x-y plane then weights the score in the z direction to result in a scalar. Note that the existing multi-slice MOs only process the SKE task, and we will propose a novel multi-slice numerical observer for SKS task (cf. Chapter 9 in Part III).

3.1 General considerations

In general, model observer (MO) is designed to detect a signal on a noisy background. The problem can be seen as the validation of one of the two following exclusive hypotheses:

$$\begin{aligned} &\mathcal{H}_0 \text{ :the signal } \boldsymbol{x} \text{ is absent, only the noisy background } \boldsymbol{b} \text{ is observed;} \\ &\mathcal{H}_1 \text{ :the signal } \boldsymbol{x} \text{ is present on the noisy background } \boldsymbol{b}. \end{aligned} \tag{3.1}$$

Normally the amount of noise is assumed to be so small that it does not disturb the statistical properties of the background.

The two hypotheses can be formulated as follows:

$$\mathcal{H}_h : \ \boldsymbol{g} = h\boldsymbol{x} + \boldsymbol{b}, \ h = 0, 1 \tag{3.2}$$

where $\boldsymbol{g}$ is an $M \times 1$ column vector representing the digital image under consideration consisting of M pixels of the image; and the absence or presence of the signal is controlled by the binary variable h. Hereafter, all 2D data (image, signal or background) will be represented by a column vector through vertical concatenation.

3.1.1 Background models

In medical image quality assessment domain, the noisy background, denoted as $\boldsymbol{b}$ in Eq.(3.2), is the modeling of clinical image without the presence of signal. The background complexity thus depends on the considered application.

There are mainly four categories of background models used in the literature:

1. White Gaussian Background (WNB): a simple statistical model, since each pixel is an independent, identically distributed (i.i.d.) Gaussian. However this model is certainly not clinically realistic.
2. Correlated Gaussian Background (CGB): an extension to the WNB, generated by convolving white noise following the distribution $\mathcal{N}$ (0,1) with a 2D Gaussian kernel characterized by $\sigma_{\boldsymbol{b}}$. This model allows modeling the spatial correlation in clinical images while remains easy to be mathematically manipulated.
3. Lumpy Background (LB) [47]: a more complex model in which a random number of Gaussian functions (called lumps) are placed at random locations in the image. The number of lumps is selected using a Poisson probability distribution; while lump locations are selected using a uniform probability distribution. Note that the first- and second-order statistics of this model are similar to those of CGB.
4. Clustered Lumpy Background (CLB) [48]: the most complex model among these four background models, produced similarly as for LB. The main difference is that "clusters of lumps", instead of lumps, are placed at random locations in the image (locations are still selected using a uniform probability distribution). The number of clusters and the number of lumps in each cluster are also selected using a Poisson probability distribution. Moreover, the lump profile here is a generalized Gaussian function, instead of a Gaussian function. This spatially inhomogeneous model is widely used in mammography, since it is a good facsimile of real textures observed in mammograms.

Examples of four background model images are shown in Figure 3.1, as a comparison. Certain source codes of these background models (in Matlab) are available on the website http://www.radiology.arizona.edu/cgri/IQ/page2/page4/page4.html.

We say that the task is a Background Known Exactly (BKE) task if all the parameters in the background model $\boldsymbol{b}$ are known *a priori* by the observers. Otherwise, if at least one of the parameters in the background model $\boldsymbol{b}$ is not known in advance, but specified by a probability density function (PDF), the task corresponds to a Background Known Statistically (BKS) [49] task.

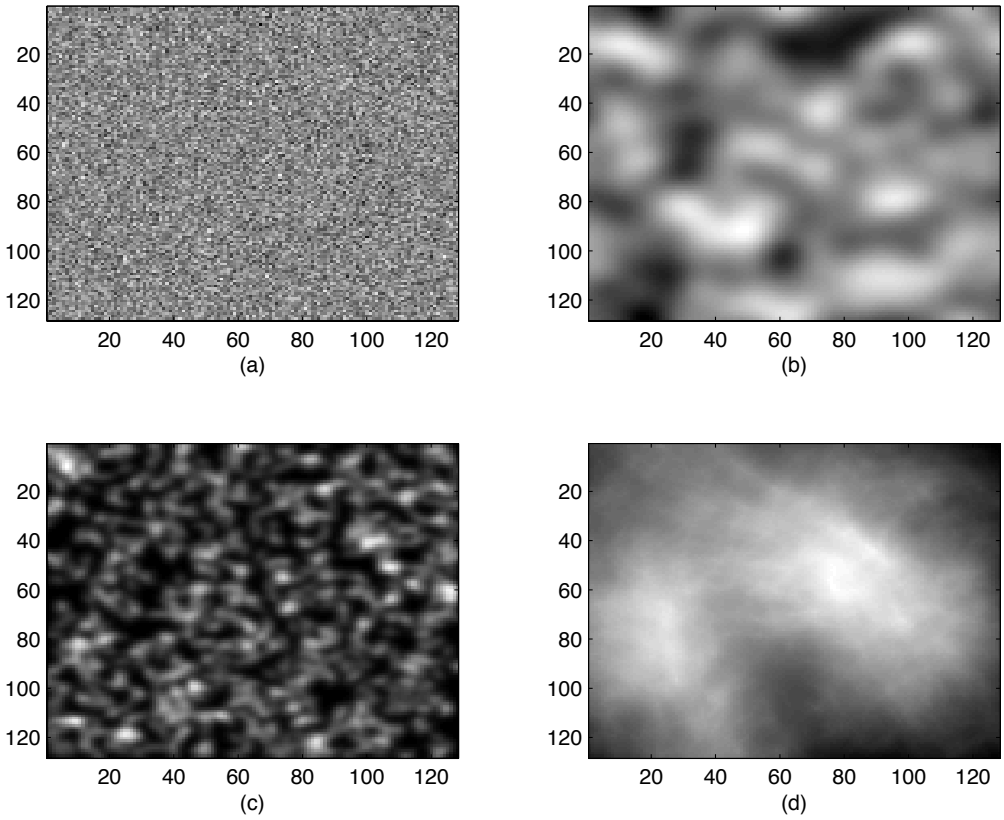


Figure 3.1: Examples of four background model simulations: (a) White Gaussian Background (WNB) ; (b) Correlated Gaussian Background (CGB); (c) Lumpy Background (LB); (d) Clustered Lumpy Background (CLB).

3.1.2 Signal Models

To facilitate the mathematical manipulation while keeping the analysis general and representative, the signal is usually modeled with a two-dimensional (2D) elliptical Gaussian function:

$$[\boldsymbol{x}]_{\boldsymbol{p}} = a \, \exp\left(-\frac{1}{2} \left(\boldsymbol{p}-\boldsymbol{q}\right)^{\mathrm{t}} \mathbf{B}^{\mathrm{t}} \mathbf{D}^{-1} \mathbf{B} \left(\boldsymbol{p}-\boldsymbol{q}\right)\right) \tag{3.3}$$

where $[\boldsymbol{x}]_{\boldsymbol{p}}$ denotes the intensity value of the added signal at the 2D coordinate $\boldsymbol{p}$. The signal intensity attenuation is hereby modeled by a Gaussian function of peak amplitude a, centered at $\boldsymbol{q}$. The diagonal matrix $\mathbf{D}$ in Eq.(3.4) specifies the ellipse's scale σ and shape b ($b \geq 1$). Note that $\sqrt{b}\sigma$ and σ are one-half of the ellipse's major and minor axes respectively; when $b = 1$, the ellipse arises as a circle. The rotation matrix $\mathbf{B}$ in Eq.(3.5) rotates the signal by θ around the

signal center.

$$\mathbf{D} = \begin{bmatrix} b\sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \tag{3.4}$$

$$\mathbf{B} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3.5}$$

The signal parameters can be denoted as a vector:

$$\alpha = [a, \theta, b, \sigma, \boldsymbol{q}] \tag{3.6}$$

$\boldsymbol{x}_\alpha$ denotes then a particular realization. Note that other choices of signal models are also possible with our numerical observer. In general, it could be written as:

$$[\boldsymbol{x}]_{\boldsymbol{p}} = f(\alpha) \tag{3.7}$$

Similartly to the concepts of BKE and BKS, we say that the task is a Signal Known Exactly (SKE) task if all the signal parameters α are known *a priori* by the observers. Otherwise, if at least one of the signal parameters is not known exactly, but specified by a PDF, the task corresponds to a Signal Known Statistically (SKS) task.

Given the signal model in Eq.(3.3), the signal information will be coded via the *a priori* probability function of α, for which we assume that the different variables are independent:

$$P(\alpha) = P(a)P(\theta)P(b)P(\sigma^2)P(\boldsymbol{q}) \tag{3.8}$$

The signal is only known from its statistical characteristics.

Figure 3.2 illustrates examples of one signal added on the four noisy backgrounds presented in Figure 3.1 (the elliptical signal is at the centers). Note that with the presence of spatial correlation, the same signal seems more difficult to be detected by human eyes.

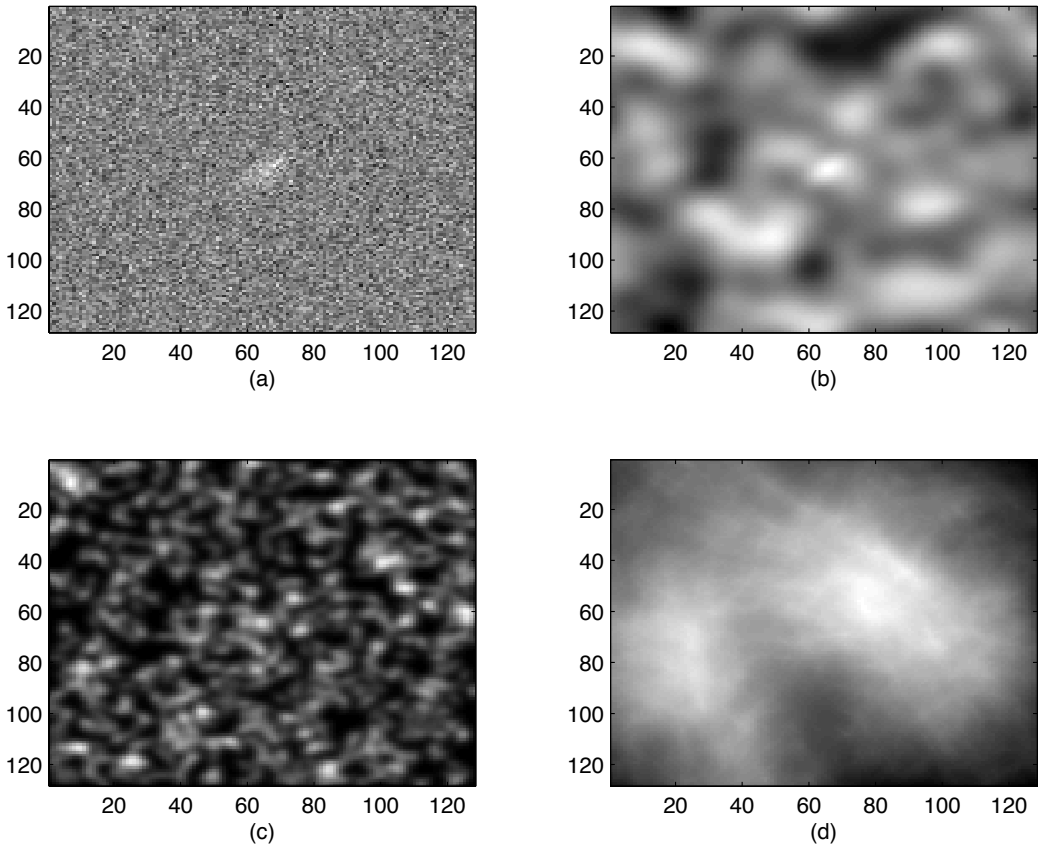


Figure 3.2: Examples of one elliptical signal (with $\sigma = 3$, $\sqrt{b} = 2$ and $\theta = \pi/4$) added at the center of the four noisy backgrounds presented in Figure 3.1.

3.2 Basics of MO

In case of the detection of a signal, every MO computes a scalar test statistic $\lambda(\boldsymbol{g})$ via a discriminant function of the image and they differ by their discriminant functions. Then a decision is made in favor of hypothesis $\mathcal{H}_1$ if the test statistic $\lambda(\boldsymbol{g})$ is greater than a decision criterion λ_c; otherwise $\mathcal{H}_0$ is selected. The decision rule can be represented as:

$$\lambda(\boldsymbol{g}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda_c \tag{3.9}$$

Then by changing the decision criterion λ_c, the MO's task performance can be characterized by a figure of merit (FOM), e.g. the area under the Receiver Operating Characteristic (ROC) curve (AUC) (cf. Chapter 2 for more information on the FOM).

3.2.1 Ideal Observer (IO)

Any observer that performs a likelihood-ratio test (in the following form) is referred to as an ideal observer [50]:

$$\lambda_{\text{IO}}(\boldsymbol{g}) = \frac{P(\boldsymbol{g} \mid \mathcal{H}_1)}{P(\boldsymbol{g} \mid \mathcal{H}_0)} \tag{3.10}$$

More conveniently, the log-likelihood-ratio $\ln \lambda_{\text{IO}}(\boldsymbol{g}) = \ln \frac{P(\boldsymbol{g} \mid \mathcal{H}_1)}{P(\boldsymbol{g} \mid \mathcal{H}_0)}$ is calculated instead of the likelihood-ratio.

The ideal observer (IO) achieves the maximum AUC attainable amongst all the MOs. It is "optimal" (or "ideal") in the sense of AUC and sets an upper bound on the signal-detection performance of any observer. Consequently, it is usually used for the design or optimization of data acquisition hardwares [2].

However, the calculation of the likelihood-ratio requires full knowledge of the PDFs of the image for each hypothesis, $P(\boldsymbol{g}|\mathcal{H}_h)$, including the descriptions of the objects to be classified and complete information regarding the measurement process and the noise statistics. In practice, these PDFs are high-dimensional functions and difficult, if not impossible, to compute for real clinical data sets. Moreover, its discriminant function, the likelihood-ratio, is usually a nonlinear function of the image data $\boldsymbol{g}$.

The IO is tractable only for some simple cases. For example, the analytical expression for the likelihood-ratio in Eq.(3.10) can be derived for the SKE/BKE task where the image data are assumed to be independent Gaussian or Poisson [2]. In more complex cases, the likelihood-ratio can be estimated using the more sophisticated and powerful approach of Markov-chain Monte Carlo (MRMC), used with LB and CLB background models in [49, 51].

3.2.2 Linear Observer

Linear observer is an observer who implements a linear discriminant which is a linear function of the image data $\boldsymbol{g}$. The general form of a linear discriminant (a test statistic) could be written as:

$$\lambda_l(\boldsymbol{g}) = \boldsymbol{w}^t \boldsymbol{g} \tag{3.11}$$

where $\boldsymbol{w}$ is an $M \times 1$ column vector, elements of $\boldsymbol{w}$ and $\boldsymbol{g}$ are assumed to be real, and $\boldsymbol{w}^t \boldsymbol{g}$ is the scalar product of $\boldsymbol{w}$ and $\boldsymbol{g}$.

3.2.2.1 Nonprewhitening Matched Filter (NPWMF)

One linear observer is Nonprewhitening Matched Filter (NPWMF):

$$\lambda_{\text{NPWMF}}(\boldsymbol{g}) = (\langle \boldsymbol{g} \,|\, \mathcal{H}_1 \rangle - \langle \boldsymbol{g} \,|\, \mathcal{H}_0 \rangle)^t \boldsymbol{g} \tag{3.12}$$

where the angle brackets $\langle ... \rangle$ denotes an expectation hereafter.

Many experiments have proved that when the correlated noise is introduced, the efficiency of the human observer is reduced [2]. That illustrates that the human observer can not perform the prewhitening operation. Therefore, the NPWMF is proposed to model the human observer.

3.2.2.2 Hotelling Observer (HO)

The Hotelling observer (HO) [2] which employs a linear discriminant is a practical alternative to the IO. HO does not need the PDFs and requires only the first- and second- order statistics of the image data for each hypothesis.

Given the *intraclass scatter matrix* $\mathbf{S}_2$ describing the average covariance matrix of the image data:

$$\mathbf{S}_2 = \frac{1}{2} \sum_{j=0}^{1} \mathbf{K}_j = \frac{1}{2} (\mathbf{K}_0 + \mathbf{K}_1) \tag{3.13}$$

where

$$\begin{aligned} \mathbf{K}_0 &= \langle (\boldsymbol{g} - \langle \boldsymbol{g} \,|\, \mathcal{H}_0 \rangle)(\boldsymbol{g} - \langle \boldsymbol{g} \,|\, \mathcal{H}_0 \rangle)^{\mathrm{t}} \,|\, \mathcal{H}_0 \rangle \\ \mathbf{K}_1 &= \langle (\boldsymbol{g} - \langle \boldsymbol{g} \,|\, \mathcal{H}_1 \rangle)(\boldsymbol{g} - \langle \boldsymbol{g} \,|\, \mathcal{H}_1 \rangle)^{\mathrm{t}} \,|\, \mathcal{H}_1 \rangle \end{aligned} \tag{3.14}$$

where $\mathbf{K}_0$ and $\mathbf{K}_1$ are the ensemble covariance matrice of the image data, without and with signal respectively.

The linear test statistic of the HO is hereby:

$$\lambda_{\text{HO}}(\mathbf{g}) = \mathbf{w}_{\text{HO}}^t \mathbf{g} \tag{3.15}$$

where

$$\mathbf{w}_{\text{HO}} = \mathbf{S}_2^{-1} (\langle \boldsymbol{g} \,|\, \mathcal{H}_1 \rangle - \langle \boldsymbol{g} \,|\, \mathcal{H}_0 \rangle) \tag{3.16}$$

The HO is also known as the “ideal linear observer” since it maximizes the *Signal-to-Noise*

Ratio (SNR), also called *detectability index*, among all linear observers. This is equivalent to maximizing AUC if the image data are Gaussian distributed.

Note that the definition of *SNR* here is different from the term used in electrical enginerring. In medical image quality assessment domain, *SNR* is defined as the distance between the test statistic mean divided by the square root of their average variance [52]:

$$SNR = \frac{\langle \lambda(\boldsymbol{g}) \,|\, \mathcal{H}_1 \rangle - \langle \lambda(\boldsymbol{g}) \,|\, \mathcal{H}_0 \rangle}{\sqrt{\frac{1}{2}\text{var}_0 + \frac{1}{2}\text{var}_1}} \tag{3.17}$$

where

$$\text{var}_0 = \langle (\lambda(\boldsymbol{g}) - \langle \lambda(\boldsymbol{g}) \,|\, \mathcal{H}_0 \rangle)(\lambda(\boldsymbol{g}) - \langle \lambda(\boldsymbol{g}) \,|\, \mathcal{H}_0 \rangle)^{\mathrm{t}} \,|\, \mathcal{H}_0 \rangle$$
$$\text{var}_1 = \langle (\lambda(\boldsymbol{g}) - \langle \lambda(\boldsymbol{g}) \,|\, \mathcal{H}_1 \rangle)(\lambda(\boldsymbol{g}) - \langle \lambda(\boldsymbol{g}) \,|\, \mathcal{H}_1 \rangle)^{\mathrm{t}} \,|\, \mathcal{H}_1 \rangle$$

The HO is “optimal” (or “ideal”) in the sense of SNR.

The HO has all the advantages that a linear observer has: it’s easy to compute; its performance is easy to summarize; far less information regarding the image statistics is needed, etc.. However, it needs the ensemble covariance matrix of the image and the ensemble covariance matrix of the image needs to be inverted. This becomes problematic in the face of modern image high-dimensionality. For example, if the image dimensions are 256×256, its covariance matrix would be 65536×65536, which is practically not feasible.

3.2.3 Channelized Hotelling Observer (CHO)

To encounter the problem of the HO, another interesting linear observer CHO was proposed.

3.2.3.1 Channelization

One approach to reduce the calculation burden (introduced by the high-dimensionality) of one image to a reasonable level is to pre-process the image data through a channelization stage.

The channelization, via a series of filters, is a technique that decomposes an image into different channels [53]. The formalization matrix is written as:

$$\boldsymbol{g}' = \mathbf{U}^{\mathrm{t}} \boldsymbol{g}, \tag{3.18}$$

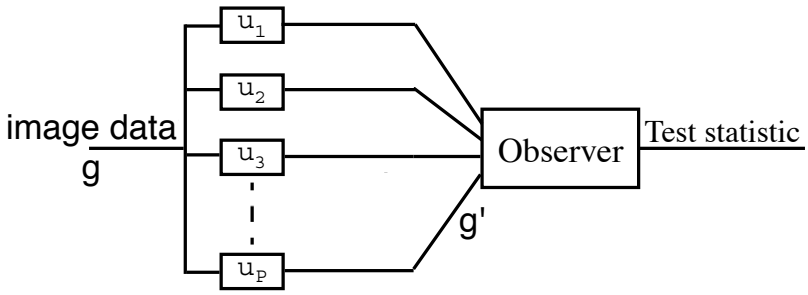


Figure 3.3: An example of a channelized Model Observer (MO), where the input image is decomposed by P filters.

where $\mathbf{U}$ (dimension $M \times P$) represents the linear filtering operator:

$$\mathbf{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_p, ..., \boldsymbol{u}_P] \tag{3.19}$$

where $p = 1, 2, 3, ..., P$. Note that the dimension of the channelized output $\boldsymbol{g}'$ is reduced to $P \times 1$, and normally $P \ll M$. Then all the calculations are executed on these channelized data, which is denoted by $(\cdot)'$ hereafter. The block diagram of a channelized MO is illustrated in Figure 3.3.

3.2.3.2 Implementation of CHO

The Channelized Hotelling Observer (CHO) [53] emerged as required to resolve the problem of HO (dimentionality) through channelization. It is well-known for its calculability and mathematical traceability among the channelized MOs. Similar to the HO, the CHO also employs a linear discriminant which is a linear function of the image $\boldsymbol{g}$.

The CHO paradigm consists of two stages: training and test, as shown in Figure 3.4.

We take the SKE CHO as an example to illustrate the practical implementation of the training and test stages, it is detailed as follows:

1. In the training stage, the inputs are two sets of images, one contains signal while the other does not. Their ground truth is known for the CHO. The output is a template:

$$\boldsymbol{w}_{\text{CHO}} = \widehat{\Sigma'}^{-1}\widehat{\boldsymbol{x}'} \tag{3.20}$$

where $\widehat{\boldsymbol{x}}$ is the estimated signal, which is the difference of the sample averages of the channelized two sets; and $\widehat{\Sigma}$ is the average of the covariance matrices of the channelized

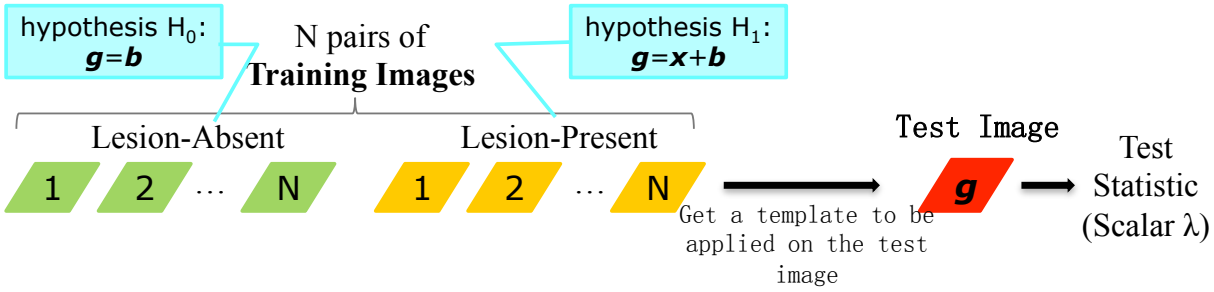


Figure 3.4: The CHO paradigm consists of two stages: training and test. In the training stage, the inputs are two sets of images, one contains signal while the other does not; the output is a template. In the test stage, the template is used to calculate a test statistic for each input test image.

two sets.

$$\widehat{\boldsymbol{x}'} = \langle \boldsymbol{g}' \,|\, \mathcal{H}_1 \rangle - \langle \boldsymbol{g}' \,|\, \mathcal{H}_0 \rangle \tag{3.21}$$

$$\begin{aligned} \widehat{\Sigma'} = \frac{1}{2} (&\langle (\boldsymbol{g}' - \langle \boldsymbol{g}' \,|\, \mathcal{H}_0 \rangle)(\boldsymbol{g}' - \langle \boldsymbol{g}' \,|\, \mathcal{H}_0 \rangle)^{\mathrm{t}} \,|\, \mathcal{H}_0 \rangle \\ &+ \langle (\boldsymbol{g}' - \langle \boldsymbol{g}' \,|\, \mathcal{H}_1 \rangle)(\boldsymbol{g}' - \langle \boldsymbol{g}' \,|\, \mathcal{H}_1 \rangle)^{\mathrm{t}} \,|\, \mathcal{H}_1 \rangle) \end{aligned} \tag{3.22}$$

2. In the test stage, the template $\boldsymbol{w}_{\mathrm{CHO}}$ is used to calculate a test statistic λ_{CHO} for each input test image $\boldsymbol{g}'$, the ground truth of which is however unknown to the CHO:

$$\lambda_{\mathrm{CHO}} = \boldsymbol{w}_{\mathrm{CHO}}^{\mathrm{t}} \boldsymbol{g}' \tag{3.23}$$

By performing the test processing on a pile of test images (with and without signal), the output test statistics are then used to calculate the area under the ROC (receiver operating characteristic) curve for characterizing the detection performance.

From Eq.(3.23), we deduce that the CHO is also a linear HO.

3.2.3.3 CHO case study

Note that the idea of the channelization to decompose the image into some spatial frequency and orientation channels (or sub-bands) also corresponds to the frequency and orientation selectivity of the HVS, one of the most important intrinsic characteristics of the HVS, cf. Section 4.2.6. Thus the choice of the channel profile determines the optimality of CHO:

– If the channels are designed to achieve the performance of HO [54], then the CHO is still an

"optimal" linear observer. By definition, the optimal MOs are the best possible in some sense and normally have a much higher task performance than human, far from our objective.

– To correlate better with human performance, one possibility is to integrate the intrinsic spatial frequency and orientation selectivity of the HVS into the channel profile by using the CHO paradigm [5]. In this way, the CHO is considered as "anthropomorphic" rather than "optimal".

One type of channels commonly used in an "optimal" [1] CHO [54] is the rotationally symmetric Laguerre-Gauss (LG) functions, which are the product of Laguerre polynomials and Gaussian functions. Let L_p denote Laguerre polynomials:

$$L_p(x) = \sum_{p'=0}^{p} (-1)^{p'} C_p^{p'} \frac{x^{p'}}{p'!} \tag{3.24}$$

The LG functions are defined by

$$u_p(r \mid a_u) = \frac{\sqrt{2}}{a_u} \exp\left(\frac{-\pi r^2}{a_u^2}\right) L_p\left(\frac{2\pi r^2}{a_u^2}\right) \tag{3.25}$$

where $r \in \mathbb{R}^2$, a_u is the spread parameter of the LG channel. These functions are then sampled to yield the linear $\boldsymbol{u}_p$. The exponential weighting $\exp\left(-\pi r^2/a_u^2\right)$ concentrates the weight of the polynomials around a Gaussian envelope with spread σ_u ($2\sigma_u^2 = a_u^2/\pi$).

As for the channels used in an "anthropomorphic" CHO [5], one example is the radially symmetric difference-of-Gaussians (DOG) functions that have long been used to model spatial-frequency selectivity in the HVS. The radial frequency profile of the pth channel is given by

$$u_p(\rho) = \exp\left[-\frac{1}{2}\left(\frac{\rho}{Q\sigma_p}\right)^2\right] - \left[-\frac{1}{2}\left(\frac{\rho}{\sigma_p}\right)^2\right] \tag{3.26}$$

where ρ is a 2D spatial-frequency variable (pixels^{-1}), the multiplicative factor $Q > 1$ defines the bandwidth of the channel, and σ_p is the standard deviation of each channel. From an initial σ_0, each of the σ_p values in Eq.(3.26) is defined by $\sigma_p = \sigma_0 a_u^p$ ($a_u > 1$). In [5], two types of DOG channels were mentionned: the sparse DOG (S-DOG) using three channels, and the dense DOG (D-DOG) using ten channels.

1. in the case of rotationally symmetric signal and background statistics

3.2.4 Comparison of MOs

We conclude the discussion on the basics of MO by providing a comparison table Table 3.1, which allows us to see more clearly how these basic MOs are evolved.

Table 3.1: Comparison of the Ideal Observer (IO), the Hotelling Observer (HO) and the Channelized Hotelling Observer (CHO).

	IO	HO	CHO
Linearity of the discriminant function	Nonlinear	Linear	Linear
Optimal or Anthropomorphic	Optimal (maximum AUC among all MOs)	Optimal (maximum SNR among all linear MOs)	Optimal or anthropomorphic (depending on channel profile)
Necessary data to calculate the template	Full knowledge of the PDFs of the image for each hypothesis $P(\boldsymbol{g} \mid \mathcal{H}_0)$ and $P(\boldsymbol{g} \mid \mathcal{H}_1)$	Only the first- and second-order statistics of the image for each hypothesis	Only the first- and second-order statistic of the channelized image for each hypothesis
Calculability	The PDFs are high-dimentional functions and difficult to compute except for a few simple cases	High-dimensionality: the ensemble covariance matrix of the image needs to be inverted.	Dimentionality reduction resulting in less calculation

3.3 Signal Known Statistically (SKS) MO

The goal of anthropomorphic channelized MOs (by integrating the frequency and orientation selectivity of the HVS into the channel profile) is to predict human performance, thus to allow us to evaluate the image quality from the standpoint of end-users (human observers).

The anthropomorphic CHOs have successfully predicted human performance in SKE experiments [2], where the signal attributes parameters (e.g. intensity amplitude, size, shape, orientation and location) are exactly known by observers and do not vary throughout the entire experiment.

However, SKE is not clinically realistic, because human observers (radiologists) do not know exactly 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 SKS experiments (where the signal attributes are known statistically) showed that the uncertainty and variability in signal parameters can significantly influence human performance [55, 15, 56]. It is thus reasonable to believe that the SKS MOs would be more clinically relevant and may have a great potential to behave closer to human observers than the SKE MOs. Since then, the SKS MOs have begun to draw more attention in the last decade.

In recent publications, several authors proposed different approaches for different SKS tasks.

- Eckstein *et al.* [57, 58, 59] proposed to establish different templates for different combinations of signal parameters and output the optimal template response (in a maximum likelihood sense) as the final test statistic. As the authors pointed out, the limitation of this method is that the required number of training images and the calculation amount increases rapidly with the number of possible signal parameters. But this type of method is efficient for the signal known exactly but variable (SKEV) [2] task [60], which could be considered as a simple case of SKS task. Lartizien *et al.* [61] extended this method to 3D (volumetric) images for signal location known exactly but variable.
- For the SKS task with signal location known statistically, Park *et al.* [15] proposed a *scanning* CHO which exhaustively scans the image. Then the location that gives the largest test statistic is chosen as the tentative location while that test statistic is the final test statistic. This scanning CHO has the same problem of calculation burden as Eckstein's method.
- Gifford *et al.* [62, 63, 64] proposed a visual-search (VS) model which firstly identifies some candidate blobs guided by features of the test image and then applies the scanning CHO on each candidate blob. With this pre-processing step, the visual-search (VS) model solves the problem that the calculation amount increases rapidly with the image size.
- Clarkson [42] proposed an ideal Estimation ROC (EROC) [42, 43] observer, whose EROC curve lies above those of all other observers for the given joint detection estimation task. Within his theoretical framework, all signal parameters could be estimated along with the calculation of the test statistic, but practically this has not been tackled yet.
- Whitaker *et al.* [65] proposed a scanning-linear estimator (SLE) for the pure estimation task. Here "scanning" means performing a global-extremum search to maximize a linear metric. He proved that the linear metric to be optimized can be derived as a special case

2. In the SKEV task, the signal varies from trial to trial but is known to the observer.

of maximum *a posteriori* (MAP) estimation when the likelihood is Gaussian and a slowly varying covariance approximation is made. But Whitaker et al. focused on the estimation task and did not investigate the detection task performance, they evaluated the SLE's estimation performance when the signal has varying amplitude, size and location.

– Later, based on the joint detection and estimation (JDE) theory [66], Goossens *et al.* [67, 68] proposed a series of corresponding SKS CHOs, each treats a signal with varying amplitude, orientation or size separately, which deal with the estimation and the detection task.

Our objective is indeed to develop a predictive MO of human detection task performance, for which the SKS task has a great potential. We remark that

1. SKS MOs have not been extensively studied yet;
2. the practicability needs to be improved and the range of variable signal parameters of the existing SKS MOs has to be widened;
3. furthermore, no numerical model has studied the localization task of multiple signals per image yet.

These inspire us to propose a novel numerical observer able to detect and localize multiple parametric signals with variable amplitude, orientation, scale and location, in order to represent more realistically the routine diagnostic process. This novel numerical observer, called Perceptually relevant Channelized Joint Observer (PCJO), will be presented in Chapter 8.

3.4 Multi-slice (3D) MO

For numerous studied modalities, such as MRI, radiologists normally view all image slices of a stack (volumetric image) in their entirety by scrolling back and forth. Thus it becomes increasingly important to investigate multi-slice/three-dimensional (3D) MOs, which would be more clinically realistic than 2D MOs.

However, until now, the existing multi-slice MOs restrict the task of interest to the SKE/BKE detection task, where one isotropic signal (of the same size for all images) is known to be located in the center of the 3D stacks. The signal is usually modeled with a 3D circular Gaussian function. For this type of task, several authors proposed different multi-slice MOs [69, 70, 71, 72, 73, 62, 74] and a well-founded overview of these MOs can be found in [69].

As said in [69], the existing multi-slice MOs could be divided into three categories: single-slice CHO (ssCHO), volumetric CHO (vCHO) and multi-slice CHO (msCHO). They will be presented in Section 3.4.1, Section 3.4.2 and Section 3.4.3, respectively.

3.4.1 Single-slice CHO (ssCHO)

The ssCHO is the most direct and simplest approach, which uses a conventional 2D (planar) CHO and applies it on a single image slice of one 3D stack only [70]. As illustrated in Figure 3.5, within the paradigm of the ssCHO, the CHO template, $\boldsymbol{w}_{\text{CHO}}$ in Eq.(3.20), is calculated from the central slices of the 3D training stacks, and the CHO test statistic, λ_{CHO} in Eq.(3.23), is obtained by using the central slice of 3D test stack.

Note that the ssCHO does not incorporate information about signal contrast in the z direction or the spatial correlation of the background and signal in the adjacent slices. Its performance is expected to be poor.

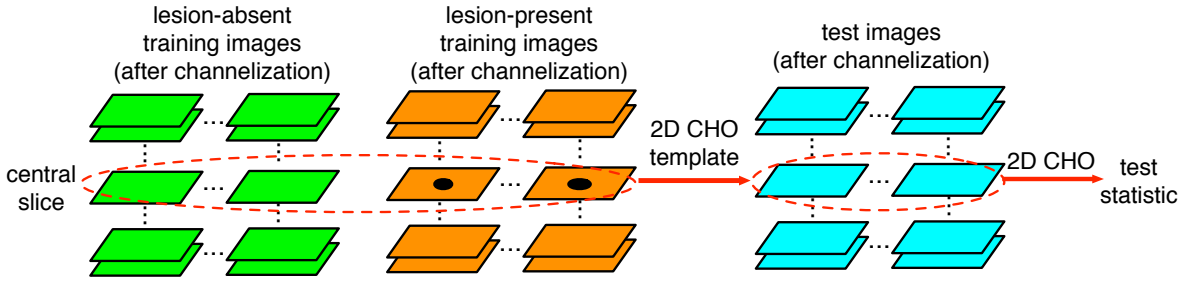


Figure 3.5: The ssCHO paradigm: the CHO template $\boldsymbol{w}_{\text{CHO}}$ is calculated from the central slices of the 3D training stacks, and the CHO test statistic λ_{CHO} is obtained by using the central slice of 3D test stack. Here each column represents one 3D stack of channelized image slices.

3.4.2 Volumetric CHO (vCHO)

The CHO is not limited by the dimensionality of the problem by definition. Therefore, a straightforward approach is to extend the 2D CHO to 3D CHO, using 3D images, 3D signals and 3D channels, so-called the vCHO [71, 72], as illustrated in Figure 3.6.

Note that the vCHO becomes "aware" of the contrast and correlation between the adjacent image slices. The vCHO is only useful when the 3D image data is not huge, if not (such as in MRI), the related calculations will be not manageable.

3.4.3 Multi-slice CHO (msCHO)

The multi-slice CHO (msCHO) assumes that humans interpret the multi-slice image in a two-stage process [73]:

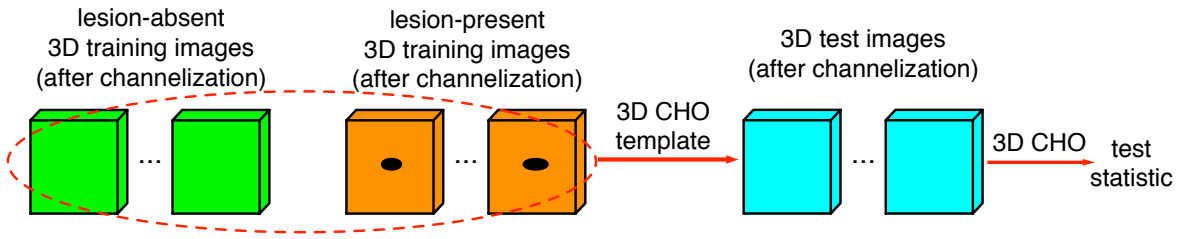


Figure 3.6: The vCHO paradigm: the 3D CHO template is calculated from the entire volumetric (3D) training images using 3D channels, and the CHO test statistic is obtained by using the entire volumetric (3D) test images. Here each cube represents one channelized 3D images.

1. the image is preprocessed in planar views (x-y plane), slice after slice, and the scores (forming a vector) obtained for each slice are buffered;
2. the score vector is processed in the z direction to result in the stack test statistic (a scalar).

In [69], the msCHO is divided into three types: msCHO$_a$, msCHO$_b$ and msCHO$_c$, as shown in Figure 3.7, Figure 3.8 and Figure 3.9 respectively. For all three msCHOs, the image data is first processed with a set of 2D channels. Then the three msCHOs process differently the channelized images:

1. Within the msCHO$_a$ paradigm, the 2D CHO test statistic of the nth slice is obtained by using a template calculated from the corresponding slice in the training images (that means all the nth slices of all the training images).
2. Within the msCHO$_b$ paradigm: the 2D CHO test statistic of the nth slice is obtained by using a single unique template calculated from the central slice in the training images.
3. Within the msCHO$_c$ paradigm, column vectors are obtained from the channelized training images, then the 2D HO template is applied directly on the column vectors.

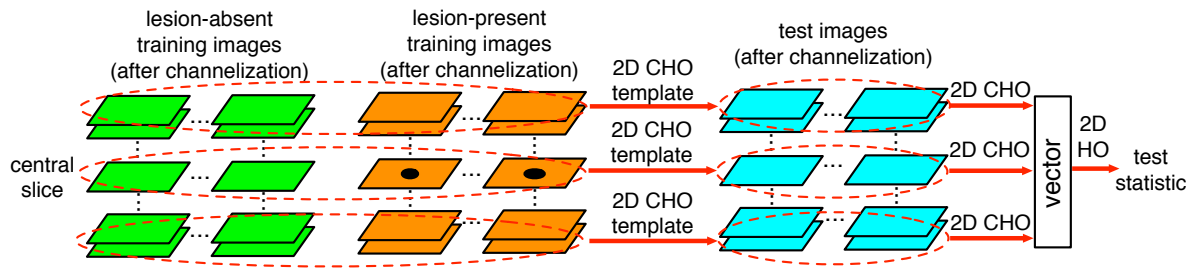


Figure 3.7: The msCHO$_a$ paradigm: the 2D CHO test statistic of the nth slice is obtained by using a template calculated from the corresponding slice in the training images (that means all the nth slices of all the training images). Here each column represents one 3D stack of channelized image slices.

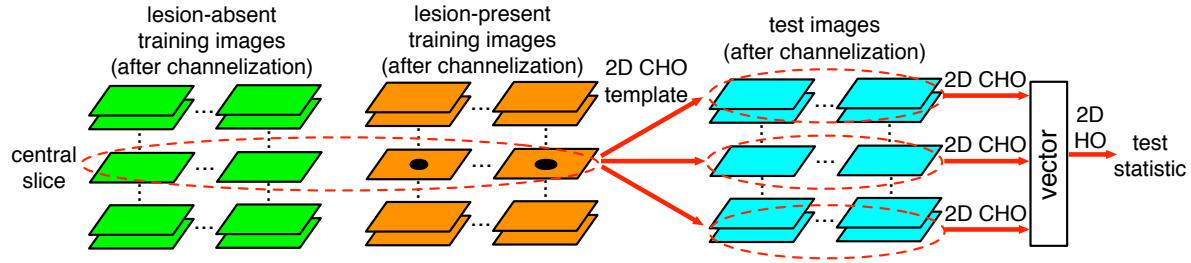


Figure 3.8: The msCHO$_b$ paradigm: the 2D CHO test statistic of the nth slice is obtained by using a single unique template calculated from the central slice in the training images. Here each column represents one 3D stack of channelized image slices.

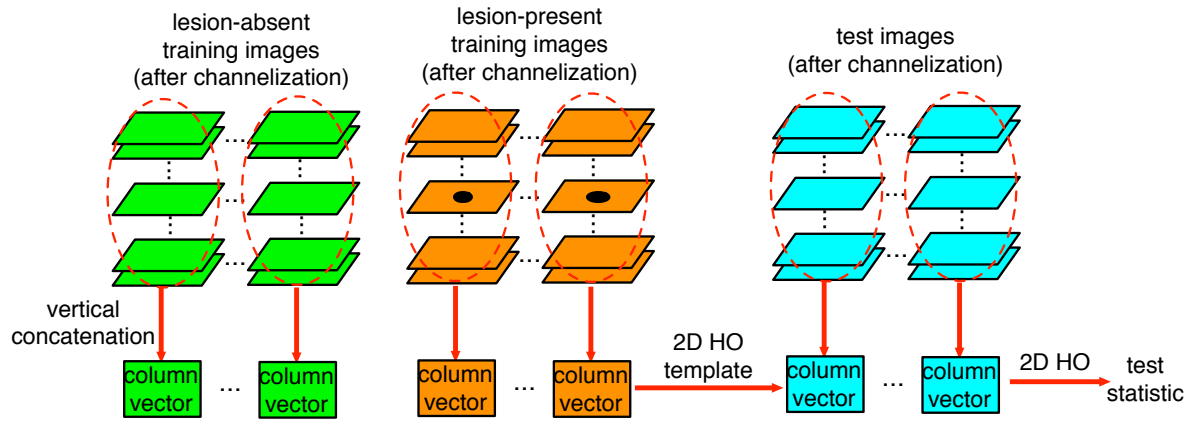


Figure 3.9: The msCHO$_c$ paradigm: column vectors are obtained from the channelized training images, then the 2D HO template is applied directly on the column vectors.

3.5 Conclusion

In this chapter, we firstly reviewed the most commonly used medical image background models and signal (lesion) models in the literature. Then we reexamined the historical development and applications of MOs:

1. from IO, HO to CHO where the mathematical tractability became more affordable;
2. from SKE task to SKS task, the latter considers the variability in signal parameters and ultimately the task of interest becomes more clinical relevant;
3. form single-slice MO to multi-slice MO, in consequence of more use of volumetric/multi-slice medical images in the clinical routine.

Complying with these trends of MOs, we will propose novel numerical observers for the SKS task, as well as multi-slice images in Part III of this thesis.

4

Human Visual System (HVS) models

Popular objective approaches in the medical image quality assessment domain, such as model observers (cf. Chapter 3), that have been specially proposed for medical images, mostly place emphasis on the image itself (e.g. its statistical characteristics) without considering the human observers' perceptual characteristics. An exception is the channelized Hotelling observer (CHO) that integrates the intrinsic spatial frequency and orientation selectivity of the human visual system (HVS) into its paradigm. However the selectivity is not the only one characteristic of the HVS. In order to correlate with the notion of "quality" as perceived by the HVS, we need to further simulate the complicated processes of the HVS.

Beside medical imaging community, e.g. in the context of "natural" image and video quality assessment, HVS models have been proposed and widely used for assessing the quality from the human's perspective by modeling the known response of the early vision system as measured by a variety of physiology and psychophysics experiments. This inspires researchers to investigate the utilization of the HVS models in the medical domain.

In this chapter, we will give an overview on the HVS models used in the medical domain, which concern the perceptual difference models (PDMs), including their general structure, common functions and applications in the medical image quality assessment domain. All PDMs have two input images: a *reference image* and a *distorted image*. The output is some sort of perceptual difference map between the two input images. Finally, a global note (scalar) can be obtained by combing the information on the perceptual difference map, usually in a simple way.

It is important to note that in psychology, human vision processing can be conventionally separated into two stages: *sensation* and *perception* [75]. The sensation stage is viewed as a function of low-level biochemical and neurological events, it reacts with the impinging of a stimulus upon the photoreceptors. The perception stage is a dynamic cognitive process of interpreting 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.

4.1 General structure

The HVS models used in the evaluation of medical image quality [76, 77, 10, 78, 11, 12] belong to the perceptual difference model (PDM) [79, 80, 81, 82, 83] focusing on 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.

To understand the general structure of PDMs, Figure 4.1, Figure 4.2, Figure 4.3 and Figure 4.4 give the structural block diagrams of four of the most well-known PDMs, respectively:

1. Visible Difference Predictor (VDP) [79];
2. High Dynamic Range Visible Difference Predictor (HDR-VDP) [81];
3. Visual Contrast Gain Control model (VCGC) [82];
4. Spatial Standard Observer (SSO) [83].

We can see that 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). Then they are passed through several functions, preprocessing functions and those modeling various HVS properties: display model, calibration, amplitude nonlinearity, contrast sensitivity function (CSF), decomposition channel function, contrast conversion, masking function, psychometric function, sub-band pooling, etc.. After being processed by these functions, 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.

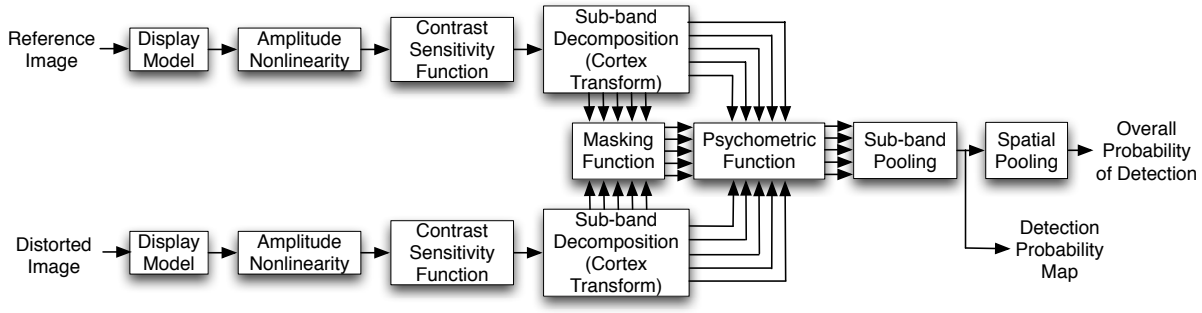


Figure 4.1: Block diagram of the Visible Difference Predictor (VDP) model

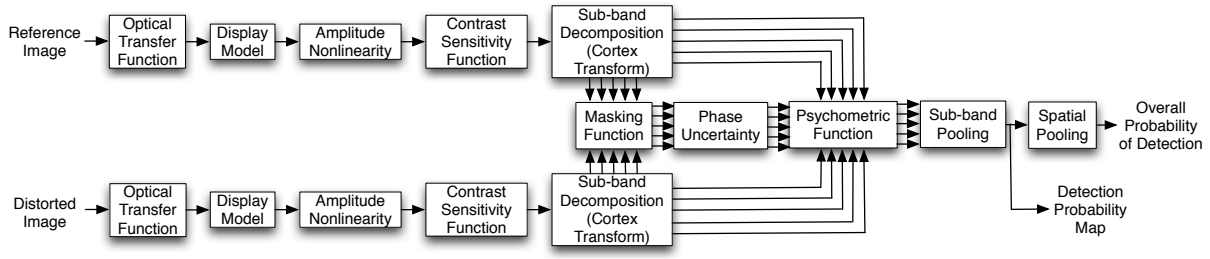


Figure 4.2: Block diagram of the High Dynamic Range Visible Difference Predictor (HDR-VDP) model

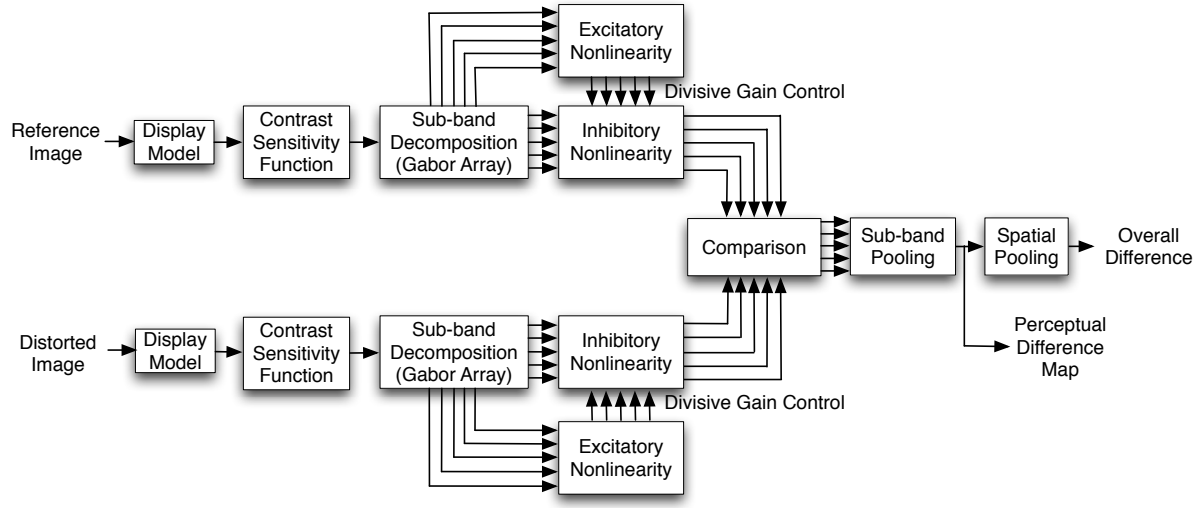


Figure 4.3: Block diagram of the Visual Contrast Gain Control model (VCGC) model

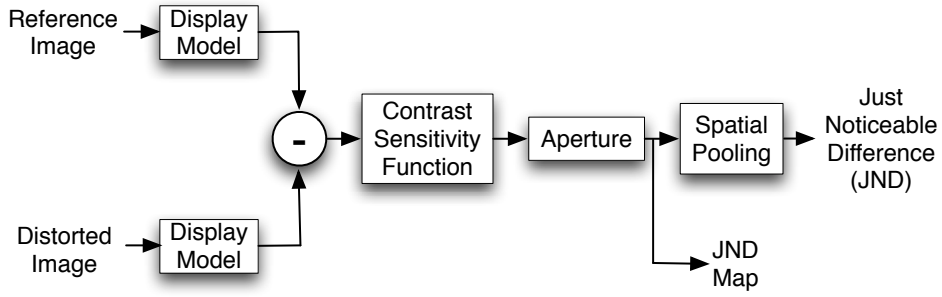


Figure 4.4: Block diagram of the Spatial Standard Observer (SSO) model

4.2 Common functions

The PDMs differ by their function definitions. In this section, we will concisely present the functions of HVS models, as well as the corresponding HVS characteristics that the functions try to model.

4.2.1 Display Model

The *Display Model* converts the digital code values of the input images into physical luminances, considering the relationship between the grey levels (same for the digital values) of the image and the luminances from the display instrument.

4.2.2 Calibration

During the *Calibration* stage, several input parameters (like viewing distance, physical pixel spacing, maximum light level and image size) are used to calculate the values needed in the subsequent stages, like visual frequency ρ in c/deg, orientation θ in deg, light adaptation level l in cd/m^2, image size i^2 in deg^2 and eccentricity e in deg, where the c, deg, cd and m are the abbreviations of $cycle(s)$, $degree(s)$, $Candela$ and $meter(s)$.

4.2.3 Amplitude Nonlinearity

4.2.3.1 Corresponding HVS characteristics

Naturally confronted with a highly dynamic light intensity, the HVS has several mechanisms which allow us to detect constantly objects in different environments despite large changes in the level of ambient illumination, for example:

1. The pupillary light reflex controls the diameter of the pupil (1.5 to 8 mm) in response to the intensity (luminance) of light that falls on the retina of the eye [84]. This reflex is maintained by the pretectum, a region of neurons between the thalamus and midbrain, which receives binocular sensory input from retinal ganglion cells of the eyes.
2. The photochemical concentration of the photoreceptor is relatively high in darkness and decreases in the light. Correspondingly, this modifies the sensibility of the photoreceptor.
3. The neuronal responses of all the retinal cell layers are adjusted when the environment changes. This adaptation is less effective than the previous one, but much faster [85].

4.2.3.2 Amplitude nonlinearity functions

Amplitude Nonlinearity functions are used to model the HVS adaptation to the luminance and the nonlinear response of retinal neurons.

One mostly used in the literature is Daly's amplitude nonlinearity function [79], which is a simplified version that is shift-invariant, invertible and implemented as simple point nonlinearities. This amplitude nonlinearity function assumes that the HVS observer can view the image from any arbitrarily close distance thus the HVS adapts to each pixel of an image. However, physiological experiments have proved that the HVS can not adapt to infinitely small areas of an image, not to mention one pixel. Thus Daly's amplitude nonlinearity function remains to be a coarse simple approximation of the luminance adaptation in the retina.

An alternative way is to use the optics *point spread function* (PSF) [86], since the blurring effect from convolving the PSF with the input images could lead to a better approximation. But introducing the PSF will cause a shift-variant nonlinearity and the accompanying problem of noninvertibility. The invertible process (of the amplitude nonlinearity function) is usually preferred for signal processing.

4.2.4 Contrast Conversion

4.2.4.1 Corresponding HVS Characteristics

The HVS is more sensitive to contrast than absolute luminance. This conversion from the luminance to the contrast makes an object (or its representation in an image) distinguishable from other objects and the background.

In the HVS, each *retinal ganglion cell* bears a *receptive field*, the organization of which provides a way of detecting contrast. We can find detailed explanation in [87].

4.2.4.2 Contrast functions

There are many possible definitions of contrast, various definitions of contrast are used in different situations. Here, only monochrome and static images are concerned, so we will use the luminance contrast, which can be generalized as:

$$\text{Contrast} = \frac{\text{Luminance Difference}}{\text{Average Luminance}} \tag{4.1}$$

which is simply denoted as DI/I. The rationale behind this is that a small difference is negligible if the average luminance is high, while the same small difference matters if the average luminance

is low (according to Weber-Fechner law).

The contrast functions can be divided into two types:

– when the average luminance intensity is calculated over the whole input image, it is called *global contrast*, like the contrast function defined in [79];
– when the average luminance intensity is calculated over the area of adaptation in the input image, it is called *local contrast*, like the contrast function defined in [88].

4.2.5 Contrast Sensitivity Function (CSF)

4.2.5.1 Corresponding HVS Characteristics

Contrast sensitivity is a measure of the ability to discriminate between luminances of different levels in an image. The minimum contrast required for an observer to detect a change in luminance is known as the *contrast threshold*. Results of many psychophysical experiments showed that the contrast threshold is affected by several factors such as the spatial frequency, the orientation, the temporal frequency, the target size, the background luminance, the viewing duration, the observer's age, etc. The contrast threshold is the reciprocal of the contrast sensitivity, therefore the lower the contrast threshold is the higher the contrast sensitivity and visual performance are [89].

The HVS is more sensitive to contrast at certain spatial frequencies. This is firstly illustrated by F.W. Campbell and J.G. Robson [90] (cf.Figure 4.5). They showed that the HVS is more sensitive to middle spatial frequencies. The pixel luminance is depicted as a sine-wave grating along the horizontal axis. The spatial frequency increases exponentially from left to right, and the contrast increases logarithmically from 100% at the bottom to 0.5% at the top. At the top, the contrast is too low to see the grating (for 0.5% contrast, only homogeneous grey is seen). Very wide (low spatial frequency) and very thin (high spatial frequency) gratings are harder to see than the middle bars, even with high contrast [90].

This tells us that the HVS is most sensitive in detecting contrast differences occurring at intermediate spatial frequencies (4-13 cycles per degree), i.e. at these spatial frequencies humans can detect lower contrast differences than at any other spatial frequencies. The high-frequency cut-off represents the limitation of HVS's ability to resolve details: frequencies higher than 50 cycles per degree are undetectable even at maximum contrast. The high-frequency cut-off is related to the packing density of the retinal photoreceptor cells: a finer matrix can resolve finer gratings [91]. The low-frequency drop-off is due to lateral inhibition within the retinal ganglion cells. A typical retinal ganglion cell presents a central disk with either excitation or inhibition in the *center* and the *surround* (a concentric ring) responding oppositely. The bright bands fall on the

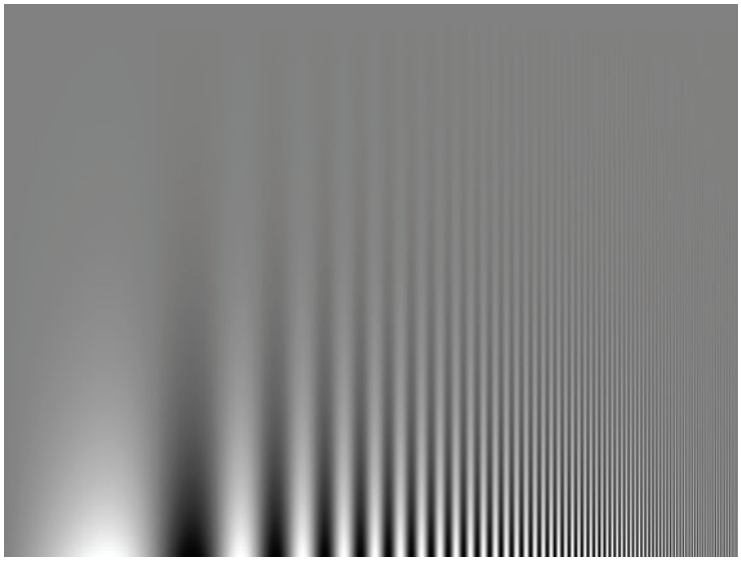


Figure 4.5: Illustration of the contrast sensibility (vertical axis) as a function of the spatial frequency (horizontal axis). Figure originally produced in [90].

inhibitory as well as the excitatory region of the retinal ganglion cell resulting in lateral inhibition and account for the low-frequency drop-off of the human contrast sensitivity function [92].

4.2.5.2 CSF

With the purpose of modeling the variance of the HVS response according to different spatial frequencies, Contrast Sensitivity Function (CSF) reflects the HVS sensitivity of a contrasted stimulus, normally as a function of spatial frequency.

As a complicated model, the CSF can be an anisotropic function of radial spatial frequency ρ in cy/deg, orientation θ in deg, light adaptation level l in cd/m^2, image size i^2 in deg^2, eccentricity e in deg and lens accommodation due to the distance d in m, e.g. the CSF used in [79]; as a simple model, the CSF can be an isotropic function of only radial spatial frequency ρ in cy/deg and eccentricity e in deg, e.g. the CSF used in [93].

Note that most CSFs are used in the frequency domain, and frequency domain analysis can easily use certain well-understood mathematical computations (e.g. FFT and FFT^{-1}). If the CSF stage is put before the decomposition channel function stage (cf. Section 4.2.6), thus the nonlinear feature of CSF may introduce some distortions into the signal in the frequency domain which would consequently interfere the application of the frequency selectivity. As indicated in [94], the ordering of the stages is an important feature of the masking model and it is better to place the linear stages before the nonlinear stages.

4.2.6 Sub-band Decomposition

4.2.6.1 Corresponding HVS Characteristics

The results of many neurophysiological studies have shown that the specific selectivity based on spatial frequency and orientation (and color, etc.) is a basic property of neurons in visual cortical areas. The *visual cortex* refers to the *primary visual cortex* (also known as striate cortex or V1) and extrastriate visual cortical areas such as V2, V3, V4, and V5.

It is widely believed that early responses of V1 neurons consists of sets of selective spatiotemporal filters. In the spatial domain, the mechanism of V1 can be regarded as the complex Fourier transform or more approximately, Gabor transform [95]. Theoretically, these filters together can carry out neuronal processing of spatial frequency, orientation, motion, direction, speed (thus temporal frequency), and many other spatiotemporal features.

Besides V1, we should also notice V2 and V4. Orientation, spatial frequency, and color tuning properties of simple cells in V2 are similar to those in V1. The responses of many V2 neurons are also modulated by complex properties, such as the orientation of illusory contours and whether the stimulus is part of the figure or the ground [96]. It is also known that V4 has similar tuning properties for orientation, spatial frequency, and color. The difference with V1 is that V4 is tuned for object features of intermediate complexity.

4.2.6.2 Sub-band decomposition function

The sub-band decomposition functions are usually a hierarchy of filters, used to model the frequency selectivity and orientation selectivity of the HVS.

One example is the cortex filters [97, 79] which decompose data into some overlapping "dom" filters dedicated to modeling the radial frequency selectivity, some overlapping "fan" filters dedicated to modeling the orientation selectivity, and one separate base filter built at the lowest-frequency without regard to orientation. Gabor array [82] and wavelet transform are also often used to transform the input image into certain bands or channels, because their responses approximate as well the functioning of the primary visual cortex V1.

4.2.7 Masking Function

4.2.7.1 Corresponding HVS Characteristics

Visual masking is a well-known characteristic of the HVS, which is used commonly to refer to the visibility reduction of one stimulus, called the *signal*, by the presentation of a second stimulus,

called the *mask*. One masking effect is caused by the time interval on the order of the presentation of the two stimuluses. For still images, we only consider the *simultaneous visual masking* which indicates the effect of one stimulus on the detectability of another where the stimuli are coincident in space and simultaneous in time.

More specifically, Legge and Foley [94] proposed the concept of *contrast masking*, which is used to study the spatial frequency selectivity of pattern vision by measuring signal thresholds at one spatial frequency in the presence of masking patterns of other spatial frequencies.

There are actually two phenomena in the contrast masking: the masking effect and the facilitation effect. The masking effect, illustrated in Figure 4.6, describes that with the raise of the mask contrast, firstly (when $MC < MC_0$) the threshold of visibility (T) is almost a constant (T_0), but after a certain value of the mask contrast (when $MC > MC_0$), the threshold of visibility will increase, that means the signal becomes harder to distinguish when the contrast *mask* is high. The facilitation effect, illustrated in Figure 4.7, describes a contrary effect of the *mask* when $MC < MC_0$, then the threshold of visibility decreases with the augmentation of the mask contrast, rather than remains a constant.

In addition, for high contrast masks and signals at medium and high spatial frequencies, many studies have found that signal threshold elevation is maximal when the mask and signal have the same frequency. Threshold elevation decreases regularly as the masking frequency departs from the signal frequency [94]. The functions relating signal threshold elevation to masker frequency may be termed spread of masking functions or masking tuning functions.

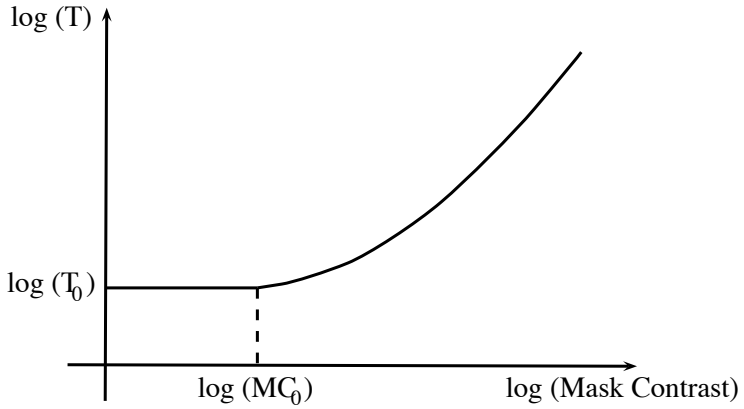


Figure 4.6: Masking Effect

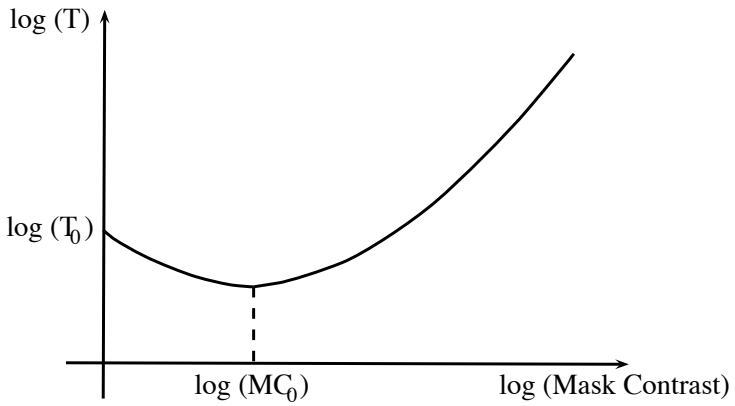


Figure 4.7: Facilitation Effect

4.2.7.2 Masking Functions

Masking functions could be divided into single-channel masking and multi-channel masking, depending on the number of channels used in the masking functions. Among multi-channel masking functions, there are two types: *inter-channel masking* and *intra-channel masking*.

The *inter-channel masking* functions represent the masking and facilitation properties mentioned above Section 4.2.7.1. A computationally simple modeling is to reproduce the masking effect by building the relationship between the *threshold elevation* and the mask contrast [79, 98], which however does not account for the facilitation effect.

The *intra-channel masking* functions [99, 100, 82] models the lateral inhibition and lateral excitation behaviors of the neurons in the receptive field (cf. Section 4.2.5.1 for this HVS characteristic). While the functions in [100, 82] are image-driven (that is to say it can be applied on an arbitrary pair of stimuli), the function in [99] can be applied only on the particular stimuli employed.

4.2.8 Psychometric Function

The psychometric function relates the observer's probability of detection/seen to the strength of a stimulus, with the abscissa being the strength of a stimulus (the input of the function) and the ordinate measuring the observer's probability of detection/seen (the output of the function) [101]. The strength of a stimulus could be signal contrast, luminance, etc. Summary measures, such as a detection/visibility threshold or a slope, may be also derived from the psychometric function.

The psychometric function is usually estimated based on experimental studies where observers have to give one of two possible responses (e.g., yes or no in the yes/no experimental paradigm;

or correct/incorrect in the two-alternative forced choice (2AFC) experimental paradigm) [102].

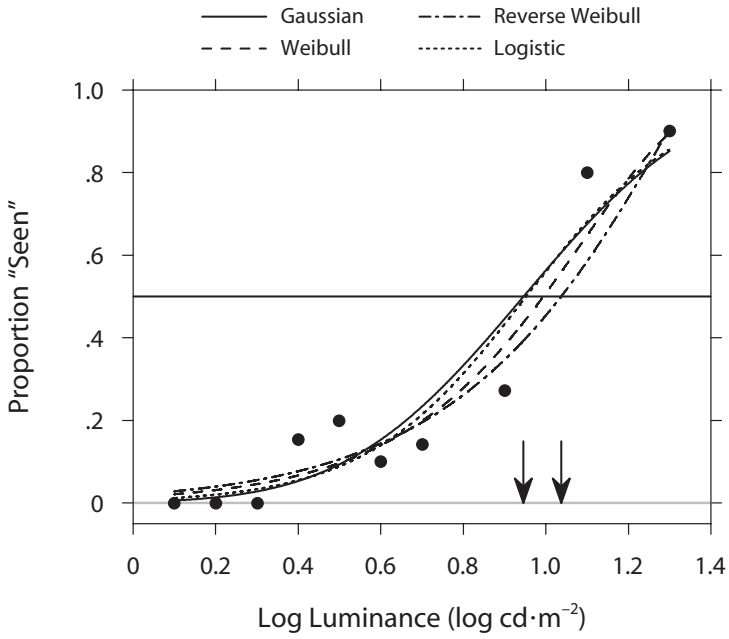


Figure 4.8: Figure from [103]: four parametric fitted psychometric function curves for frequency of seeing. The proportion of responses "seen" is plotted against log stimulus luminance. Filled circles are data from [104] and are based on 3-20 trials at each stimulus level. The horizontal solid line indicates the criterion level for a 50% threshold, and the arrows mark the extreme estimated thresholds, which were for the Gaussian and reverse Weibull cumulative distribution functions.

One common way of estimating the psychometric function is to fit a parametric model to the experimental data and the parameters of the model are estimated by maximizing the likelihood [105, 106, 107, 108]. However, it is difficult to specify the best parametric model in practice when there is insufficient knowledge about the underlying processes. This is illustrated graphically by Figure 4.8 [103] which represents four plausible models: Gaussian, Weibull, reverse Weibull, and logistic cumulative distribution functions. The four functions have different shapes, but all of them are acceptable on the basis of the goodness of their individual fits given the number of data points. However, threshold estimates (corresponding to a criterion level of 50% seen), indicated by the solid horizontal line in Figure 4.8 varied from 0.946 log cd·m^{-2} to 1.037 log cd·m^{-2}. Whether these differences in threshold estimates are important depending on the application, but it is clear that the choice of the model may lead to bias in the parametric curve.

Recently, a nonparametric approach based on local linear fitting was proposed in [103], where no assumption is made about the true model underlying the data, except its smoothness. This method removes the burden of deciding about the parametric model when there is insufficient

knowledge about the underlying processes. That is why we chose this model-free method in our experimental study in Section 5.2.1.

4.2.9 Error Pooling

Error pooling function includes sub-band pooling and spatial pooling, which are applied to combine together the error information on the sub-bands and on the perceptual difference map, respectively.

This last human visual processing stage is delicate, so far we do not know how this mechanism occurred. Thus in the literature we normally employ two simple *error pooling* methods: the probability summation (with product series) [79] and the Minkowski summation [98].

4.3 Applications in medical image quality assessment

In the literature, several researchers tried HVS models for the detection task on medical images. Jackson *et al.* [76] adopted the Sarnoff model [80] for the detection of a simulated tumor added on a region of a mammographic image without tumor. They compared the model's detection task performance with that of three human observers using the detectability index [2], and showed that the detectability for the human observers and the one predicted by HVS model are close. However they did not conduct any hypothesis test to demonstrate whether there is a significant detectability difference. Sarnoff model was also employed by Krupinski *et al.* [77] to compare the CRT and LCD viewing for mammographic images. The results suggested that observer performance with LCD displays is superior to CRT viewing. The correlation between the just noticeable difference (JND) outputs of Sarnoff model and the human observers is high. Huo *et al.* [78] and Jiang *et al.* [11] developed a Perceptual Difference Model (PDM) as a quantitative image quality evaluation tool to compare different reconstruction methods in the context of parallel magnetic resonance imaging (MRI). The results were obtained in the form of signal contrast sensitivity, while the detectability was fixed. The PDM and human observer showed the same comparison results and after scaling, the numerical differences between the reconstruction algorithms could be reasonably predicted from a perceptual point of view. Johnson *et al.* [12] recently presented new results investigating the utility of a Visual Discrimination Model (VDM) developed at Siemens Corporate Research for predicting JPEG 2000 bit rates corresponding to visually lossless compression of virtual slides for breast biopsy specimens. The results indicated that for test images compressed to their visually lossless thresholds, JND metrics computed by the VDM were nearly constant at the 95th percentile level or higher.

Note that the two input images of PDM vary for different applications: for the evaluation of image post-processing algorithm, the reference image is the original non-processed image and the distorted images are the images post-processed by different algorithms, see [78] for example; while for the detection performance evaluation, the reference image is a healthy image without abnormality and the distorted image is that with simulated abnormality, see [76] for example. We remark that PDMs haven't been applied on the diagnostic localization task to date.

4.4 Conclusion

In this chapter, we firstly analyzed and summed up the general structure of HVS models: they process the two input images (a reference image and a distorted image) with several functions modeling HVS properties and they output a scalar that quantify the perceptual difference between the two input images. Then we described their common functions by explaining the corresponding HVS characteristics. We ended with a brief overview of their applications in medical image quality assessment domain.

It should be remembered that existing HVS models only consider the sensation stage of the human vision processing, while the more complex perception stage presents substantial interests to be integrated. Consequently, we allow for part of the perception stage in the framework of our proposed numerical observers, as detailed in Part III of this thesis.

Conclusion of Part I

This first part was dedicated to the ROC analyses and existing MOs and HVS models. As said before, medical image quality must be defined in terms of the *average* performance of some observers for some diagnostic tasks.

To characterize and quantify the task performance, ROC analysis and its variants have been proposed for different diagnostic tasks. In this part, we introduced the ROC for the detection task; the LROC for the detection-localization task of one signal on an image; the FROC and the AFROC for the detection-localization task of multiple signals on an image. All these ROC analyses need the *gold standard* which in reality is difficult to obtain. Thus the consensus of experts is usually taken as the *gold standard*, or the *gold standard* can be "controlled" by using synthetic images and/or signals.

We remark that: although the JAFROC1 is showed to be the FOM that has the highest statistical power for the detection-localization task up to now and has already been widely used to characterize the human observer performance, it hasn't been used to evaluate the performance of a numerical observer (mainly because no other numerical observers has been used before to process the signal localization in images that contain multiple signals).

Regarding the image quality assessment using observers, one possibility is performing empirical studies with human observers. But this is costly and time-consuming. Thus the alternative is to use numerical observers, as substitutes for human observers, which have been widely used in today's medical image quality assessment community. In this part, we detailed two categories of numerical observers: MO and HVS model.

In the frame of MO, the detection task (defined in Section 1.1.1) is a binary classification task with two underlying hypotheses (presence or absence of a signal in the image). The IO was firstly proposed for the optimization of the binary classification problem, which chooses the hypothesis resulting in the greatest probability or likelihood of the image data given that hypothesis. It sets an

upper bound on the detection task performance of any observer, but is tractable only for simplified settings (e.g. when the image data is Gaussian). The HO and then the CHO emerged for solving the mathematically intractable problems of the IO. The CHO paradigm allows integrating the spatial frequency and orientation selectivity of the HVS. Today's trend in MO is towards SKS and multi-slice, considering the variability in signal parameters and more and more use of volumetric medical images, respectively.

We remark that:

– the SKS MOs have not been extensively studied yet;
– there is a great demand for widening the range of variable signal parameters of the existing SKS MOs.

In the frame of HVS model, the scalar output is some sort of perceptual difference between the two input images: a reference image and a distorted image. The input images are processed by a series of functions modeling different characteristics of the HVS: amplitude nonlinearity, CSF, sub-band decomposition, contrast conversion, masking function, psychometric function, error pooling, etc..

We remark that:

– existing HVS models are only sensation models;
– it is necessary to integrate the perception stage of the human vision processing in the numerical observers, since the perception plays an important role in medical image perception.

We also remark that: no numerical model (neither MOs nor HVS models) has been proposed for the localization task of multiple signals on single-slice images.

Part II

Preliminary Perceptual Studies: Radiologist & Model Performance

Introduction of Part II

The second part of this thesis is dedicated to the experimental pre-studies conducted during the preliminary stage of my PhD work. Given the studied modality MRI and pathology MS, the first question that we asked ourselves regarding whether we can employ the existing numerical observers directly for the signal detection on single-slice MR image. Thus we designed some subjective experiments to investigate the suitability of the use of existing numerical observers in our application.

As underlined in Chapter 3, MOs were normally applied to homogeneous background, such as the four types of background introduced in Section 3.1.1. Homogeneous background may be realistic in some cases (e.g. mammography), but is not realistic for brain MR image where radiologists normally view images in their entirety, not a small homogeneous region only (e.g. an extracted white matter region). It is thus worthwhile to investigate the importance of the supplemental anatomical information in the entire single-slice image. In addition, as underlined in Chapter 4, existing HVS models are not obtained from clinically relevant experiments with radiologists. The end users of medical images are radiologists, not naive observers as for natural images. Radiologists' expertise certainly influences their task performance, especially in the cognitive process to interpret perceived information. Thus we firstly explored the above aspects and provided further insight into human visual system processing for this specific task, which laid the groundwork for our numerical observer modeling in MRI image quality assessment by addressing the perspective of radiologist diagnostic performance.

The ultimate goal of medical images is to help radiologists gain a high diagnostic accuracy. Thus evaluating medical image quality from radiologists' perspective is a useful alternative compared to optimal observer approach. While several HVS models exist in the literature (Chapter 4), few works have been conducted to evaluate the suitability of these models w.r.t. the diagnostic task performance, and none has investigated their utilization for the perception task. Thus it is worth comparing the suitability of HVS models to radiologists' sensation-perception behavioral performances in the context of medical images. Although several HVS models have

been compared for natural images, their comparison in the context of medical images is not a straightforward exercise, since the characteristics of medical images require totally different investigation method.

This part comprises two chapters.

In Chapter 5, we present the experimental studies using human subjects, which investigated the influences of MR image anatomical information and observer expertise on the human visual system processing (including the sensation stage and the perception stage) for specific tasks.

In Chapter 6, we propose to use the AUC and its variance estimates as the figure of merit to study the suitability of four of the most well-known HVS models: : VDP, HDR-VDP, VCGC, SSO.

5

Anatomical information & observer expertise influence

In this chapter, the study on the effects of anatomical information and observer expertise on the detection task will be presented and analyzed in detail.

For brain MRI, radiologists view images in their entirety, but the MO studies are usually carried on a relatively homogeneous region [109, 110]. Besides, HVS models are composed of functions based on the results of experiments with naive observers (observers without medical knowledge), rather than radiologists who are accustomed to reading the medical images. However the expertise influence of radiologists cannot be neglected in the image perception. For example, it was showed in [111] that experience combined with training provides the basis for generating efficient visual search strategies and developing distinctive conceptual criteria for perceptual differentiation and interpretation of true breast masses from image artifacts and structured noise that mimics breast abnormalities; and it is demonstrated in [112] that the development of expertise improved pattern recognition (taking more information in during the initial Gestalt or gist view) as well as improved allocation of attention and visual processing resources.

We therefore designed experimental observer studies with the objective of:

1. comparing the ability of human observers to detect a lesion in the presence of an entire brain slice (with more anatomical information than a white matter region considered as homogeneous);
2. comparing the detection ability of radiologists to that of naive observers.

The studies are detailed in this chapter. This work has resulted in the conference paper [113].

5.1 Method

5.1.1 Experiment protocol

The experiments were conducted in the same viewing conditions, e.g. a simulated radiology reading environment in the Hospital of Angers, France. One expert in brain MRI with 21 years' experience (expert), three radiologists (radiologist 1, radiologist 2, radiologist 3) and eight Naive observers (Naive 1, Naive 2, Naive 3, Naive 4, Naive 5, Naive 6, Naive 7, Naive 8) participated in the study. The decision time was unlimited.

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).

Brain MR images of a 25-year-old healthy volunteer (who gave his approval for using the images), with simulated multiple sclerosis (MS) lesions of different degrees of intensity, were

used in this study. The subject identifier was stripped to anonymize it. A single case was adopted here to eliminate background as a variable.

The studied image is one axial slice of the T2 FLAIR sequence, and in this type of sequence, the MS lesions appear as hyper-signal ("bright spot").

Before the experiment, we have conducted a quick pre-experiment, in which the adjustment method was used to get the estimated visibility threshold of the radiologists. In the procedure of the adjustment method, the stimulus intensity was adjusted according to the radiologist's response until it is just detectable. Then fourteen intensities around the average estimated visibility threshold (with a step of 10% of the average estimated visibility threshold) were chosen to make the lesions not too conspicuous.

Finally a model-free estimation method [103] was used to generate the psychometric curves (a single-variable function for each subject and each experiment, where y-axis represents the proportion of "seen" and x-axis the lesion intensity), and to get the thresholds at the criterion level of 50% of "seen".

5.1.2 Lesion simulation

The synthetic MS lesions here are modeled by the two-dimensional elliptical Gaussian function defined in Eq.(3.3). The scale parameter $\sigma = 7$ and the shape parameter $b = \sqrt{2}$ considering the entire slice size is 996×1274.

The experiment only included lesions within the white matter cortical tissue as MS is characterized by lesions within the white matter. The proprieties of lesion shape, size, position and orientation were verified and confirmed by a consultant radiologist who specializes in MS. During the experiments, six non-overlapped lesions of different intensities are simulated for each trial, while their size, position and orientation were fixed for each trial.

This simulation protocol is adopted to avoid the problem of obtaining a healthy subject slice from a lesion present slice (the only data available in the hospital database).

5.1.3 Study experiments

For the first objective, two experiments were conducted with each participant on two separate days (one month between the two days) to reduce the memory effect:

1. Experiment one (Exp 1) consisted of six white matter blocks being selected within the white matter. Lesions were placed at the centre of each block, as shown in Figure 5.1. Six blocks were selected from the same slice used in Experiment two (Exp 2) to ensure

their characteristics remain the same and the results could be compared. Participants were requested to report whether they could see an abnormality in each of the six blocks. Participants reported findings by selecting either; a) "Yes, I see an abnormality in this block" or b) "No, I can't see any abnormality in this block".

2. Experiment two (Exp 2) comprised of one axial, cerebral slice, with six lesions located in separated areas of the slice, as shown in Figure 5.2 (the slice contains the same white matter blocks as in Exp 1). As in the previous experiment, participants were asked to report whether they were able to see an abnormality in each of the selected six areas as denoted and labeled for reference in Figure 5.2.

The two experiments differ in the amount of anatomical information available to each participant. The results of Exp 1 and Exp 2 allow the comparison of observer detection when anatomical information is present and withheld.

For the second objective, we asked different questions to different observers in Exp 1 and Exp 2:

- the three radiologists and eight Naive observers are required to note if they were aware of the presence of hyper-signals (the abnormality here is a hyper-signal);
- the expert was asked to note if clinically significant signs were present or not, that means that he thinks that hyper-signal is not just a noise or a background structure, but a clinical abnormality candidate which deserves further check (the abnormality here is a clinically significant sign).

All observers were trained using 20 training images. Radiologists and Naive observers were trained for the hyper-signal detection, and the expert was trained for the clinically significant sign detection.

The results of the two experiments then also yield a comparison of inter-rater performance, e.g. between expert (specialized reader of MR image analysis with MS lesions), general radiologist (specialized in medical image analysis only) and Naive observer (not specialized in image analysis or medicine).

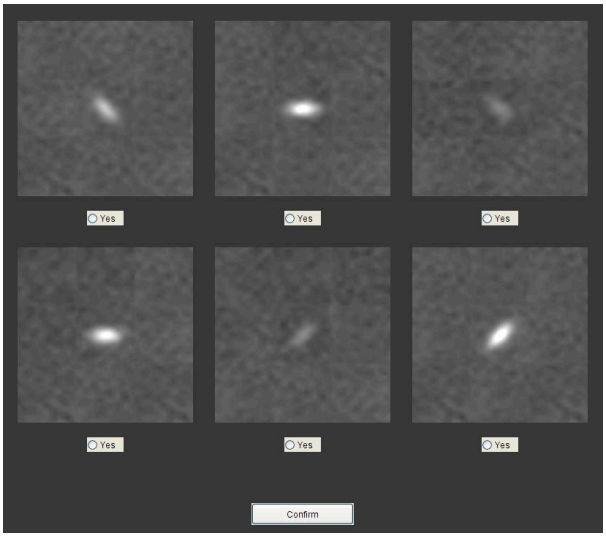


Figure 5.1: An example of one trial of experiment one (Exp 1) with six synthetic lesions.

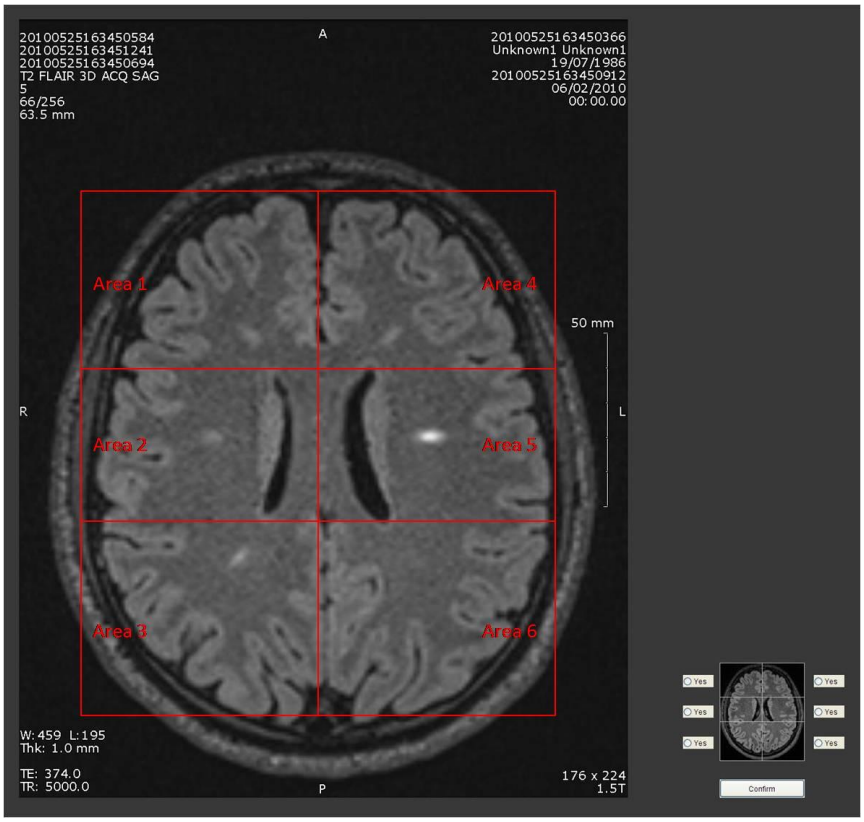


Figure 5.2: An example of one trial of experiment two (Exp 2) with six synthetic lesions (Red labels were applied only for reader demonstration).

5.2 Results

5.2.1 Psychometric curves

The model-free estimation method [103] was applied to the data to generate psychometric curves for each lesion location, participant, and experiment. The perception thresholds were gotten at the criterion level of 50% of "seen". The twelve psychometric curves for all observers for Exp 1, Area 6, and for Exp 2, Area 6, are presented in Figure 5.3a and Figure 5.3b respectively.

Since the following statistical analyses are based on the thresholds derived from the model-free estimation method, it's important to know how accurately this method predicts study outcomes. Thus, the coefficient of determination, R^2, was calculated to measure the goodness of fit. The results are shown in Table 5.1, note that an R^2 of 1.0 indicates perfect curve fitting data, and many of our R^2 values are close to 1. However, two Naive observer scores (Naive 3 and Naive 4) were outside of the normal distribution for this study: their R^2 values were less than 0.9. As a consequence the results of these observers are eliminated in the following statistical analyses. Finally, the remaining observer results (including one expert, three radiologists and six Naive observers) were included in subsequent analyses.

Figure 5.3: Psychometric curves for Exp 1, Area 6 (Figure 5.3a) and for Exp 2, Area 6 (Figure 5.3b). The red lines highlight the radiologist results, while the black lines highlight Naive observer results.

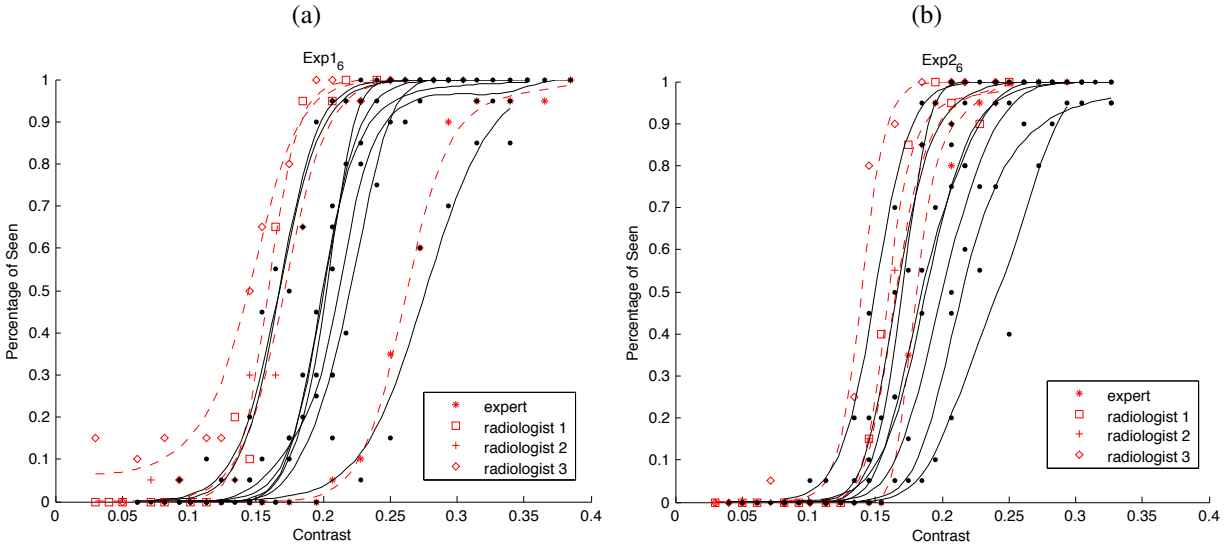


Table 5.1: The coefficient of determination, R^2, of the psychometric curve for each lesion location, each participant and each experiment.

		Expert	Radiologists			Naive observers							
			1	2	3	1	2	3	4	5	6	7	8
Exp 1	Area 1	0.997	0.997	0.997	0.997	0.993	0.983	0.998	0.958	0.996	0.995	0.986	0.995
	Area 2	0.992	0.989	0.984	0.993	0.995	0.949	0.997	0.976	0.966	0.990	0.971	0.994
	Area 3	0.985	0.970	0.986	0.990	0.988	0.977	0.893	0.995	0.974	0.986	0.994	0.993
	Area 4	0.989	0.967	0.955	0.923	0.991	0.961	0.989	0.984	0.993	0.992	0.995	0.990
	Area 5	0.997	0.973	0.993	0.989	0.995	0.967	0.982	0.995	0.993	0.990	0.975	0.985
	Area 6	0.997	0.988	0.986	0.978	0.998	0.964	0.978	0.989	0.989	0.995	0.993	0.994
Exp 2	Area 1	0.972	0.998	0.992	0.994	0.996	0.974	0.976	0.810	0.989	0.974	0.997	0.993
	Area 2	0.995	0.996	0.994	0.920	0.976	0.990	0.991	0.991	0.993	0.990	0.987	0.963
	Area 3	0.987	0.628	0.978	0.990	0.995	0.949	0.993	0.964	0.964	0.985	0.983	0.954
	Area 4	0.994	0.978	0.992	0.993	0.983	0.923	0.976	0.942	0.991	0.914	0.979	0.979
	Area 5	0.995	0.996	0.988	0.997	0.997	0.988	0.994	0.987	0.989	0.999	0.994	0.967
	Area 6	0.989	0.995	0.995	0.988	0.997	0.988	0.996	0.960	0.999	0.962	0.998	0.991

5.2.2 Differences between Observer Group

The observers were divided into three groups: expert, radiologists and Naives. As previously discussed, each experiment and observer had different research questions (distinguish the presence of hyper-signals for radiologists and Naives, or the presence of clinically significant signs for the expert). The analysis was performed taking into account the question (hyper-signal detection vs clinically significant sign detection) and the expertise (radiologist vs Naive). As a result, observer threshold defined here is threshold of simple visibility for the three radiologists and six Naive observers; while the threshold of the expert is threshold of clinically significant sign's detection.

Two non-parametric comparisons were conducted, since our samples were not normally distributed. The following studies include results of Exp 1 and Exp 2. The first comparison was conducted to examine simple visibility differences between radiologist group and naive group. The second comparison examined between expert group and radiologist group differences for two levels of detection task. The significance level was set to $p < 0.05$. The p-values of each test are shown in Table 5.2 and Table 5.3, respectively. The box plots of the corresponding visibility thresholds are shown in Figure 5.4 and Figure 5.5.

From Table 5.2, it is evident that, the p-value is much smaller than 0.05, denoting that there is a significant difference between the visibility threshold of radiologists and Naive readers. Furthermore, from Figure 5.4, we can see that the visibility thresholds of radiologists are much lower than those of Naives, except for Area 3.

From Table 5.3, we can see that for all areas, the p-value is much smaller than 0.05, indicating that there is significant difference between expert detection thresholds and radiology visibility thresholds. And Figure 5.5 highlights that expert detection thresholds are higher than radiologist visibility thresholds. One reason may be that the expert needs more information (noisy background, location...) to interpret cognitively a hyper-signal as a clinically significant sign.

Table 5.2: The p-values of Mann-Whitney test between radiologists and Naive observers.

	Area 1	Area 2	Area 3	Area 4	Area 5	Area 6
p-value	<0.001	<0.001	0.506	0.019	0.016	<0.001

Table 5.3: The p-values of Mann-Whitney test between expert and radiologists.

	Area 1	Area 2	Area 3	Area 4	Area 5	Area 6
p-value	<0.001	<0.001	0.010	<0.001	0.001	<0.001

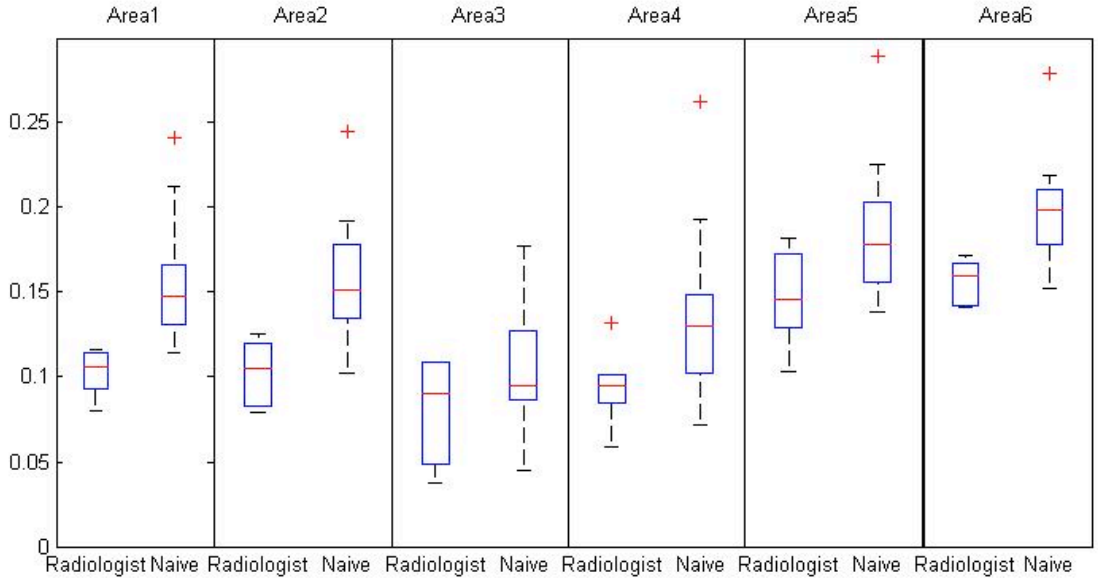


Figure 5.4: Comparison of thresholds between radiologists and naive observers.

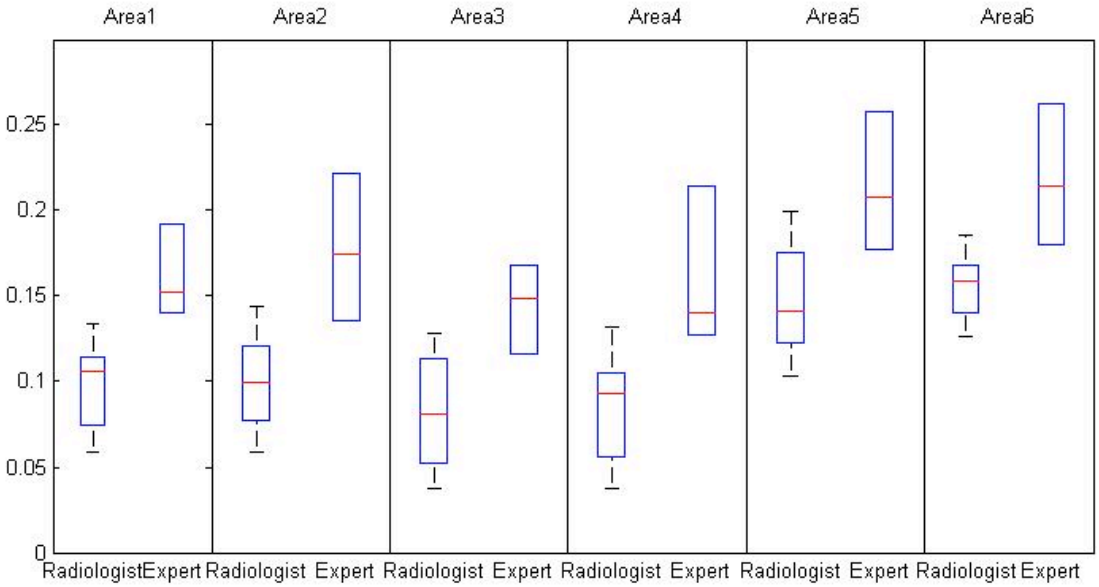


Figure 5.5: Comparison of thresholds between radiologists and expert.

5.2.3 Differences between Experiment

For this analysis, Mann-Whitney was applied to the two experiments (Exp 1 and Exp 2), and the significance level remained at $p < 0.05$ for both studies. Firstly, the simple visibility thresholds of the three radiologists and six Naives were compared, for six areas, between both experiments and the results are shown in Table 5.4. It is apparent that there was no significant difference between the visibility thresholds of Exp 1 and those of Exp 2, indicating that there might be no contribution of anatomical information in the simple visibility task.

Secondly, the clinically significant detection thresholds of expert observer were compared between Exp 1 and Exp 2. The p-value is 0.009 ($p < 0.05$) for the 6 areas, and from Figure 5.7 it is apparent that the detection threshold of Exp 2 is lower than that of Exp 1. Then there is difference between the clinical significant detection threshold of Exp 1 and that of Exp 2; this indicates that the anatomical information contributes to a reduction of the threshold in the clinical significant detection task.

Table 5.4: The p-values of Mann-Whitney test between Exp 1 and Exp 2 for radiologists and Naives.

	Area 1	Area 2	Area 3	Area 4	Area 5	Area 6
p-value	0.749	0.881	0.328	0.026	0.815	0.374

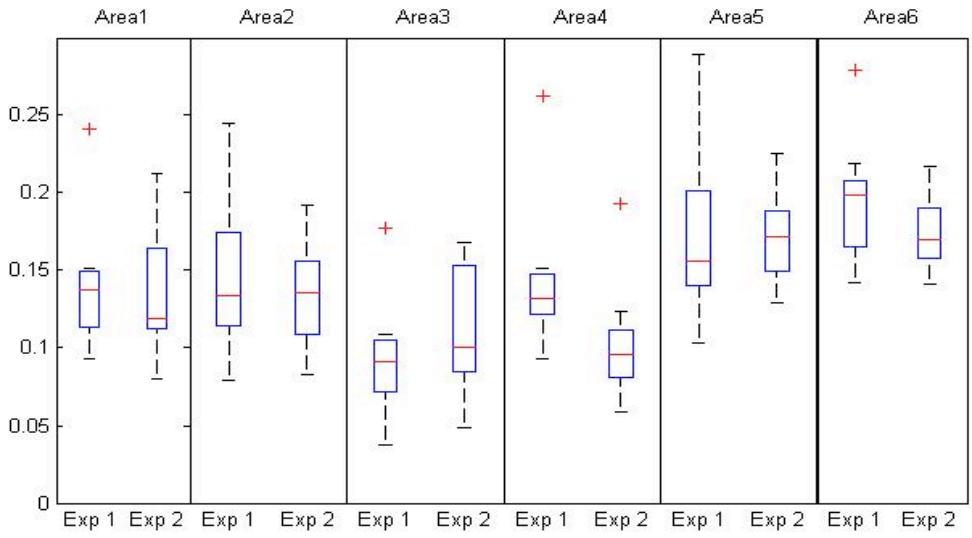


Figure 5.6: Comparison of thresholds between Exp 1 and Exp 2 for radiologists and naives, highlighting that there is no contribution of anatomical information in the simple hyper-signal visibility task.

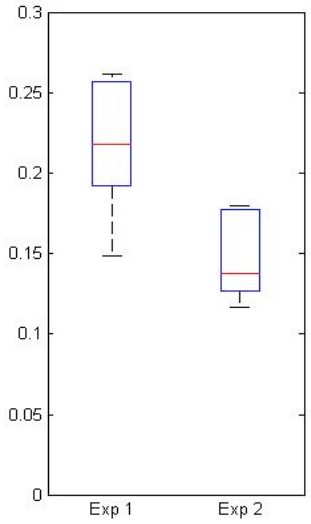


Figure 5.7: Comparison of thresholds between Exp 1 and Exp 2 for the expert.

5.3 Discussion

As said in Chapter 4, most previous researches have favored the opinion that visual perception processing includes two stages: sensation and perception. In the present study, when radiologists and Naives perform the same task (as Naives do not have relevant clinical knowledge, they can only simply detect the hyper-signal), they are actually performing the sensation, low-level stage of visual processing. However, when radiologists consider the clinical implications of a signal, they are interpreting the findings cognitively, which may be considered as a high-level of perceptual processing here.

Altogether, the results suggest that at the sensation stage, radiologists have better detectability of simple hyper-signals than Naive observers. This reason may exist as radiologists receive much more training in their daily life and are far more experienced in image assessment. To go further in the visual processing, it would be interesting to explore more on the perception stage where expertise appears to influence the interpretation of the hyper-signals with an elevation of the detection threshold.

It is well known that in daily practice, radiologists always read an entire slice, rather than a small homogeneous region with limited anatomical information. However, the results suggest that with regard to the visibility threshold of sensation at the low-level stage of visual perception processing, no difference exists between assessing the synthesized lesion region on the entire slice or a small homogeneous area with the lesion implanted within it. However, if the primary aim is to evaluate the threshold of perception, anatomical information must be taken into consideration

as it facilitates radiological performance.

5.4 Conclusion

Anatomical information coupled with the experience of the radiology reader are non-negligible factors, which must be assessed when evaluating diagnostic performance, however, these factors have not yet been accounted for in numeric observer models or HVS models. This study aimed at assessing a small piece of the puzzle – how much information do radiologists need of the original image to detect a synthesized lesion in terms of a) its actual presence and b) whether the lesion would be considered of clinical importance.

This study demonstrates different detection abilities between radiology and naive observers, confirming that radiologists have lower simple detection thresholds than naïves, due to their medical education background, clinical experience and expertise; it also demonstrates little impact of anatomical information at the low-level stage of perception processing, these simple visibility thresholds can be acquired by psychometric experiments on either the entire slice or on the small homogeneous region.

We also notice the impacts at the high-level stage of perception processing. In this small cross-sectional study, anatomical information plays an important role and facilitates radiology detection and assessment.

This study intended to elaborate on questions surrounding the prediction of human observer performance, to facilitate understanding of visual perception, particularly at different levels of abnormality visual processing in the specific domain of diagnostic interpretation.

6

HVS model conformity to radiologist performance

In this chapter, the conformities of HVS models to radiologists' sensation-perception behavioral performances will be studied.

Beside medical imaging community, e.g. in the context of natural image and video quality assessment, the widely used HVS models have been proposed for assessing the quality from the human's perspective by modeling the known response of the early vision system as measured by a variety of psychophysical experiments. However, these psychophysical experiments have been performed by naive observers with no specialized medical education background. Radiologists, as human observers, share the common HVS characteristics. Nevertheless, our previous study (Chapter 5) demonstrated that radiologists have lower simple detection thresholds than naive observers in this task (perhaps more top-down processes that are better tuned to the task due to better training etc.). Besides, compared to natural images, medical image possesses specific features (for example, images are monochrome in MRI and CT, the noise and the contrast depend on the modality...). This inspires us to investigate the utilization of HVS models in the medical image domain.

In the literature, several researchers tried HVS models for the sensation task on medical images. Normally HVS models underlined the perceptual difference between a non-degraded image and a degraded image [78, 11, 12], although some have considered the abnormality detection [76, 77] and used HVS models within the framework of task-based approach, but without separating the sensation and perception tasks. Note that while it has been poorly addressed by "natural" image quality assessment community, task-based approaches are fundamental in the context of medical imaging.

From literature overview, those HVS models have been adopted in different ways (for example, the two input images of PDM vary for different applications: for the evaluation of image post-processing algorithm, the reference image is the original non-processed image and the distorted images are the images post-processed by different algorithms; while for the detection performance evaluation, the reference image is a healthy image without abnormality and the distorted image is that with simulated abnormality), applied for different modalities, and evaluated with different figures of merit (detectability, AUC, signal contrast sensitivity, etc.), so that they can not be compared directly. Consequently, identifying which HVS model correlates the best with radiologists' task performance is still an open issue.

In this chapter, we present a pilot study in response to the following two questions in a specific context (MRI and MS):

1. As a task-based approach, could the generic HVS models be used to approach radiologists' detection task performance? And if yes, which HVS model has a closest performance to

that of radiologists?

2. To what extent the HVS models could approximate the radiologists' performance in the perception stage? Which is the optimal model in this sense?

To our knowledge, no attempt has been made before to compare HVS models in the medical image quality assessment domain. Although several HVS models have been compared for natural images, the characteristics of medical images require different investigation method. This work has resulted in the conference paper [114] and the Student Scholar Award in [115].

6.1 Method

6.1.1 Studied HVS models and their setup

Four of the most well-known quantitative HVS models are selected for this study:

1. Visible Difference Predictor (VDP) [79];
2. High Dynamic Range Visible Difference Predictor (HDR-VDP) [81];
3. Visual Contrast Gain Control model (VCGC) [82];
4. Spatial Standard Observer (SSO) [83].

They are chosen for three reasons:

1. Quantitative HVS models, based on the spatial contrast detection, is efficient for the detection of near-visibility-threshold distortion. In the case of the MS lesion detection, their contrast levels are typically near the human visibility threshold.
2. Though the quantitative HVS models used in the literature [80, 78, 11, 12, 116, 76, 77] showed convincing results, they are kept private to the corporate. We would like to develop a free software, and these four models provide details for the implementation.
3. We did not choose qualitative HVS models, e.g. structural similarity index, since it has been empirically and analytically proved to be directly related to the conventional mean squared error (MSE) which often produces misleading values that do not correlate well with perceived quality [117].

The structural block diagrams of these four HVS models are shown in Section 4.1. Their different HVS function implementations are compared in Table 6.1.

The four HVS models need two input images (reference image and distorted image); and their output is a prediction of the perceptual differences between the two images. Default parameters of these four HVS models were used for the programming.

Table 6.1: Differences among the four HVS models: Visible Difference Predictor (VDP), High Dynamic Range Visible Difference Predictor (HDR-VDP), Visual Contrast Gain Control model (VCGC) and Spatial Standard Observer (SSO) ("—" means the function part is not included in the corresponding HVS model).

Functions	VDP	HDR-VDP	VCGC	SSO
Amplitude Nonlinearity	arbitrary units	just noticeable difference (JND) units	—	—
Contrast Sensitivity Function (CSF)	one CSF filter	a set of CSF filters for different adaptation luminance	one CSF filter [118]	one CSF filter [119]
Decomposition into Multiple Channels	cortex filters	cortex filters	Gabor array	—
Sub-band Pooling	probability summation	probability summation	Minkowski summation	—
Spatial Pooling	probability summation	probability summation	Minkowski summation	Minkowski summation
Masking Function	intra-channel masking: threshold elevation function	same as VDP	inter-channel masking	—

After years of medical education and clinical experience, radiologists diagnose without any reference image. On the assumption that the radiologist has a reference healthy image [1] in mind [116, 76], we take a brain MR slice of a healthy volunteer as the reference image, on which there is no lesion; distorted images are generated adding simulated multiple sclerosis (MS) lesions of different degrees of intensity to the reference image. The MS lesion has been modeled in the same way as in Section 5.1.2.

6.1.2 Subjective experiments

Three radiologists participated on four separate days to neglect the fatigue influence. All experiments were conducted in a simulated radiology reading environment in the Hospital of Angers, France. The decision time was unlimited. The viewing distance was 40cm (the distance observed in clinical routine in the Hospital of Angers), and the visual angle is 42 pixels per degree.

The medical display used in this study was KEOSYS Positoscope, which was calibrated to the DICOM GSDF. We established a look-up table (LUT), which is the most accurate specification of the display response function, as the display model (relating pixel values to displayed luminance levels) involved in the HVS models. The LUT was derived from the measurements of a luminance meter (included in X-Rite i1XTreme).

In order to study the HVS model' conformity to radiologists' sensation and perception behavior, four subjective experiments were conducted. 84 distorted images were generated using 14 contrast levels and 6 locations in white matter region (since MS is a white matter disease) on the slice. In each trail, one distorted image was shown to observers, which leads to 84 trials per experiment.

1. Experiment 1 (sensation task on entire slice): In each trial, a healthy cerebral MR slice with one superimposed simulated lesion was shown to the radiologist. We asked radiologist if he could detect a hyper-signal on the slice.
2. Experiment 2 (sensation task on homogeous region): In each trial, a homogeneous white matter region, without other anatomical cues, selected from the same slice as in the experiment 1 (to ensure its characteristics remain the same, cf. Figure 6.1) with one superimposed simulated lesion was shown to the radiologist. We asked radiologist if he could detect a hyper-signal on the region.
3. Experiment 3 (perception task on entire slice): In each trial, the same whole cerebral slice as in experiment 1 was shown to the radiologist in each trial. Radiologists were asked to

1. There are lots of ways for radiologists to obtain this reference image: for example, some radiologists will refer to normal atlases and other sources when examining an image; in MS there are longitudinal sets of images and the previous image is a reference in terms of looking for change (as in many diseases).

report if a clinical sign existed on the slice.

4. Experiment 4 (perception task on homogeous region): In each trial, the same homogeneous white matter region as in experiment 2 was shown to the radiologist. The condition differs from experiment 2 as radiologists were asked if a clinical sign existed on the region.

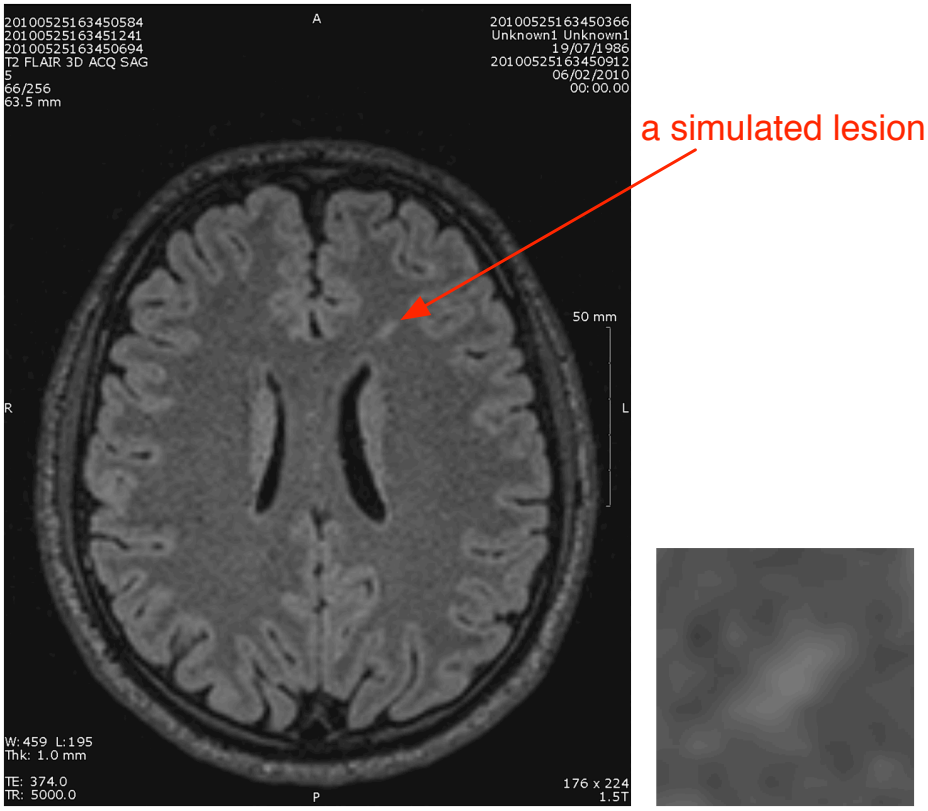


Figure 6.1: The left image is a healthy cerebral MRI slice with a simulated lesion on the top-right of the slice; the right image is the zoomed white matter region containing the simulated lesion.

6.1.3 Performance evaluation method

One difficulty of model comparison is that the four models' outputs are in different ranges: both the VDP and the HDR-VDP provide a detection probability ranging from 0 to 1; the VCGC gives arbitrary units of response difference (greater than 0); the output of SSO is in units of just-noticeable difference (JND) greater than 0. Thus there is no direct way to compare their outputs quantitatively.

The usual solution in the medical domain to evaluate an objective metric's performance for the detection task is to use the figure of merit AUC, which is adopted here to evaluate the HVS models in a quantitative manner.

The input images are first divided into two classes: signal absent and signal present, according to the gold standard. The output of each HVS model is then considered as a test statistic, to be compared with a decision threshold t_c resulting in signal present if $t > t_c$, and signal absent otherwise. According to the decision of a HVS model, four states (TP, FP, TN, FN) exist for drawing the ROC curves and calculating the AUC values. In finite dataset experiments, t_c values are chosen to explore the whole range of the model output. The output range of an HVS model does not affect the binary classification task performance as long as the decision rule remains unchanged.

It is widely accepted that the gold standard is difficult to define; it is more often a bronze standard rather than gold (cf. Section 2.1). In our simulation case, the truth status is signal present for all images since a simulated signal is added on the reference input image, and consequently FP and TN are nonexistent. An alternative is to take the consensus of radiologists as the gold standard, for which some images of the lower intensity stimulus are considered as signal absent. The AUC performance thus measures the distance between a particular HVS model and the radiologists' diagnostic results, which is in line with our initial goal - to evaluate medical image quality from the radiologist's point of view. In this study, the consensus was established for each trial using the majority decision of radiologists. Thus $N0$ signal absent images and $N1$ signal present images ($N0 + N1 = 84$) are obtained according to this consensus.

6.1.4 One-shot estimate

For each HVS model, a metric should be derived to evaluate its performance given the test statistics on the set of $N0 + N1$ images. As a figure of merit, the exact value of AUC is only computable given the probability density functions (PDF) of the test statistics, for which the four HVS models in this study do not yield analytical expressions due to their complexity. With finite experiment datasets, we used Gallas's one-shot method [33] to calculate an unbiased AUC estimate together with the estimate of its variance, since it is proved to be unique, uniformly minimum variance and unbiased; and it is very easy to calculate.

In our application, each HVS model plays the role of a *single non-random reader* (term definition cf. [33]) due to its deterministic algorithm behavior (lack of randomness). Thus the corresponding AUC estimate and the variance estimate are given by:

$$\widehat{\text{AUC}} = \frac{1}{N_0 N_1} \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} \text{step}(t_{1j} - t_{0i}) \tag{6.1}$$

$$\widehat{\text{var}} = c_1 \widehat{M_1} + c_2 \widehat{M_2} + c_3 \widehat{M_3} + (c_4 - 1) \widehat{M_4} \tag{6.2}$$

where t_{1j} denotes the test statistic output of an HVS model for the j^{th} signal-present image and t_{0i} for the i^{th} signal-absent image. The step function step$(\cdot)$ in Eq.(6.1) is 1 when the argument is positive, 0 when the argument is negative, and 0.5 when the argument is zero. The linear combination coefficients and the cross moments are also given in [33]:

$$c_1 = \frac{1}{N_0 N_1}$$

$$c_2 = \frac{N_0 - 1}{N_0 N_1}$$

$$c_3 = \frac{N_1 - 1}{N_0 N_1}$$

$$c_4 = \frac{(N_0 - 1)(N_1 - 1)}{N_0 N_1}$$

$$\widehat{M_1} = \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} \frac{[\text{step}(t_{1j} - t_{0i})]^2}{N_0 N_1}$$

$$\widehat{M_2} = \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} \sum_{i' \neq i}^{N_0} \frac{\text{step}(t_{1j} - t_{0i}) \text{step}(t_{1j} - t_{0i'})}{N_0 N_1 (N_0 - 1)}$$

$$\widehat{M_3} = \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} \sum_{j' \neq j}^{N_1} \frac{\text{step}(t_{1j} - t_{0i}) \text{step}(t_{1j'} - t_{0i})}{N_0 N_1 (N_1 - 1)}$$

$$\widehat{M_4} = \sum_{i=1}^{N_0}\sum_{j=1}^{N_1}\sum_{i'\neq i}^{N_0}\sum_{j'\neq j}^{N_1} \frac{\text{step}(t_{1j} - t_{0i})\text{step}(t_{1j'} - t_{0i'})}{N_0 N_1 (N_0 - 1)(N_1 - 1)}$$

It is noteworthy that for each HVS model, the AUC estimate and its variance estimate expressions only involve the ordering of test statistic values through the use of the step function. Thus two HVS model yield the same AUC estimate values if and only if the number of instances of $t_{1j} - t_{0i} > 0$, $i = 1, ...N_0$, $j = 1, ...N_1$ is equal [33].

6.2 Results

6.2.1 ROC curves of HVS models

The ROC curves of the four studied HVS models using the consensuses obtained from the four experiments are presented in Figure 6.2, Figure 6.3, Figure 6.4 and Figure 6.5. These graphical representations show that in all of the experiments, VDP has the largest AUC value and SSO has the smallest AUC value.

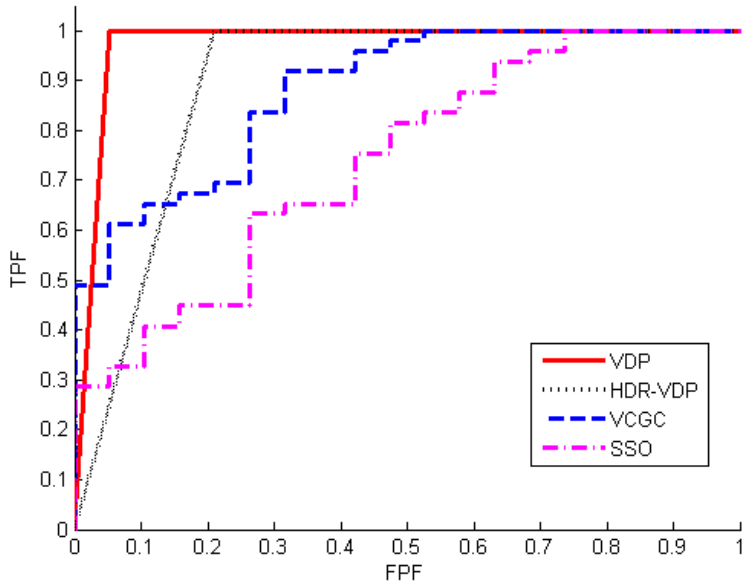


Figure 6.2: Experiment 1 (sensation task on entire slice): the ROC curves for the four HVS models.

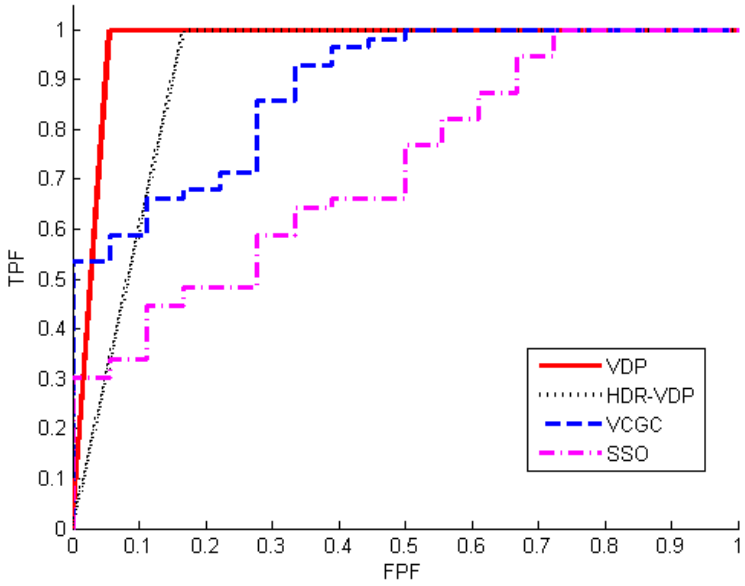


Figure 6.3: Experiment 2 (sensation task on homogenous region): the ROC curves for the four HVS models.

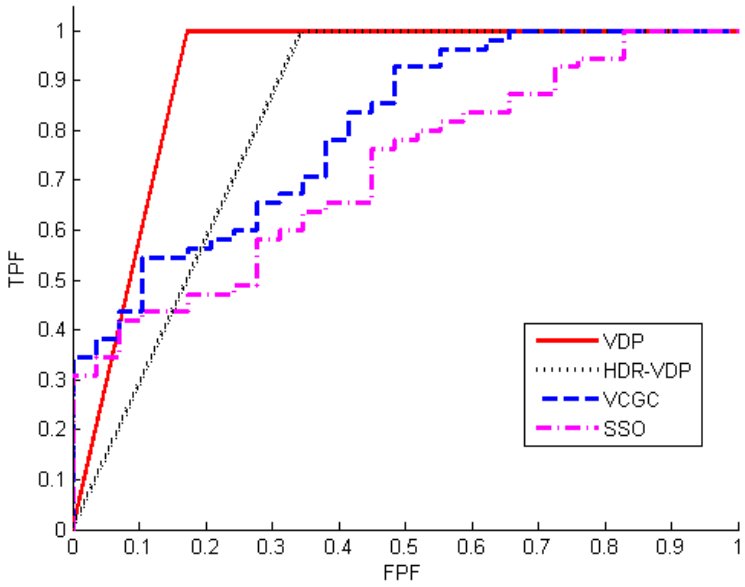


Figure 6.4: Experiment 3 (perception task on entire slice): the ROC curves for the four HVS models.

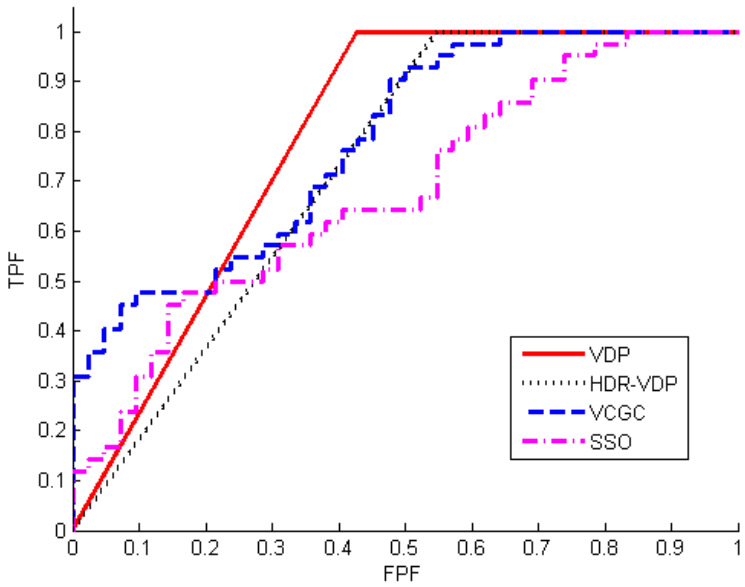


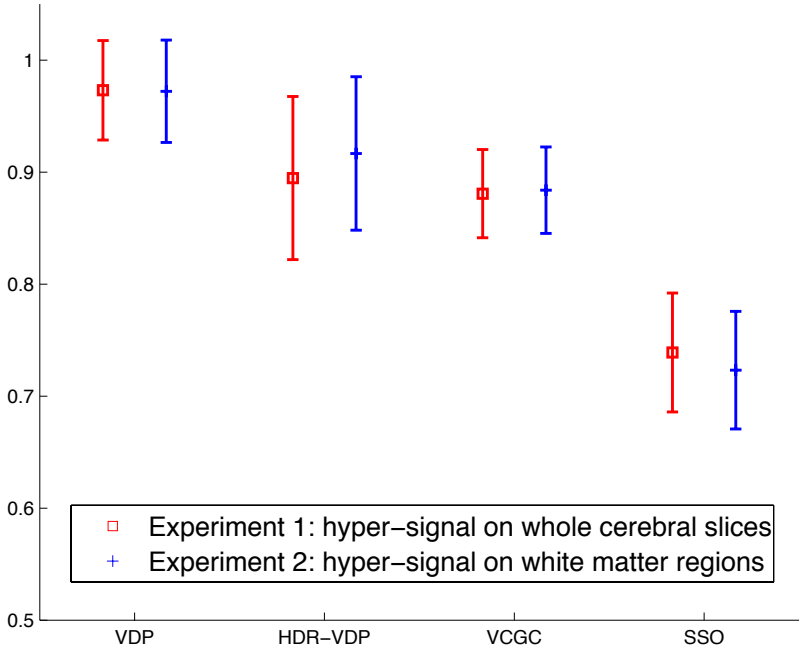
Figure 6.5: Experiment 4 (perception task on homogenous region): the ROC curves for the four HVS models.

6.2.2 AUCs for the sensation and the perception tasks

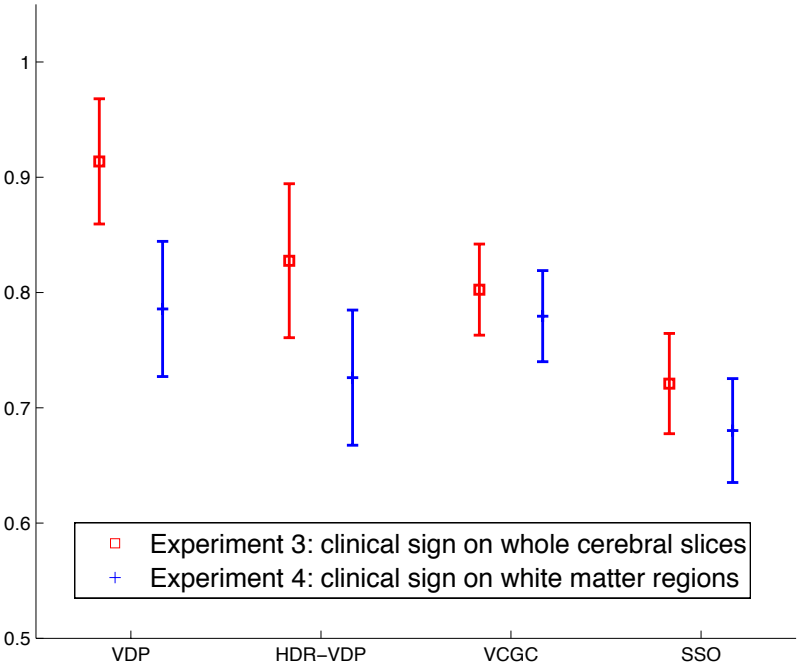
The AUC comparisons between Experiment 1 and Experiment 2, Experiment 3 and Experiment 4 are demonstrated in Figure 6.6a and Figure 6.6b, respectively. They do not only show the estimates of the AUCs' standard deviations and the ordering of the four models' AUC values in each stage of the HVS (sensation and perception), but also reveal the influences of anatomical information in the two different stages of the HVS. Discussions about these results will be found in Section 6.3.

Figure 6.6: The Area Under the receiver operating characteristic Curve (AUC) comparison between Experiments 1 and 2 (Sensation task), and between Experiments 3 and 4 (Perception task).

(a)



(b)



6.2.3 Pair wise comparisons of the AUCs

We performed the t-test for the pair wise comparisons of the AUCs with 95% confidence interval. The statistical hypothesis tests clarify whether the differences between HVS models (cf. Table 2) and between Experiments (cf. Table 3) are importantly significant or not (in our case, when the p-value is less than 0.05, it means that the two AUCs are significantly different.). Discussions about these results will be found in the next section.

Table 6.2: AUC comparison results among the four HVS models of Experiment 1, 2, 3 and 4, the four symbols in the table cells denote the four experiments in order. "○" indicates p-value < 0.05 (there is a significant difference between the two models) and "-" indicates p-value $>= 0.05$ (there is no significant difference between the two models). For example, in the cell of VDP vs. VCGC, "○○○-" means there is a significant difference between VDP and VCGC in Experiment 1, 2, 3, but there is no significant difference between VDP and VCGC in Experiment 4.

	VDP	HDR-VDP	VCGC	SSO
VDP				
HDR-VDP	○○○○			
VCGC	○○○-	○○○○		
SSO	○○○○	○○○○	○○○○	

Table 6.3: AUC comparison results among the four Experiments of HVS models, the four symbols in the table cells denote the four HVS models (VDP, HDR-VDP, VCGC, SSO) in order. "○" indicates p-value < 0.05 (there is a significant difference between the two models) and "-" indicates p-value $>= 0.05$ (there is no significant difference between the two models). For example, in the cell of Experiment 2 vs. Experiment 3, "○○○-" means there is a significant difference between Experiment 2 and 3 for VDP, HDR-VDP and VCGC, but there is no significant difference between Experiment 2 and 3 for SSO.

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Experiment 1				
Experiment 2	- - - -			
Experiment 3	○○○○	○○○-		
Experiment 4	○○○○	○○○○	○○○○	

6.3 Discussion

From Figure 6.6a and Table 6.2, it can be observed that the VDP model presents a significantly higher AUC score estimate (> 0.95) with a small variance for the sensation task. The estimates of

AUC differ significantly between the VDP model and any other HVS model. In that sense, the VDP model performance is the closest to the radiologists' sensation task performance and it is significantly better than the other HVS models chosen for comparison.

Figure 6.6b and Table 6.3 confirm the same trend for the perception task: the VDP model remains the closest to the radiologists' perception task performance and it is significantly better than other models. In addition, it is showed that the HVS models perform closer to radiologists when more anatomical information is available, since the AUCs of all HVS models are much lower in Experiment 4 (clinical sign on white matter region) than in Experiment 3 (clinical sign on whole cerebral slice).

Table 6.3 indicates that there is a significant difference of AUC values between the sensation task and the perception task (Experiment 1 vs. 3, and Experiment 2 vs. 4) for all of the four HVS models. The results presented Figure 6.6a and Figure 6.6b pointed out that even if the VDP approximates almost perfectly the radiologists in the sensation task, the distance between them in the perception task is still important. Besides, Table 6.3 shows that for the sensation task, HVS models' performances remain nearly the same with or without anatomical information (Experiment 1 and 2); however, for the perception task, the performances are significantly different with or without anatomical information (Experiment 3 and 4).

Finally we note that in both the sensation and the perception experiments, the SSO model is always the worst, i.e. the most distant model to the radiologists' detection consensus using the AUC metric. Note that the SSO is a simplified model by eliminating several sensation features, such as the spatial frequency and the orientation selectivity, the visual masking effect based on these selectivity sub-bands, and the sub-band pooling. These simplifications lead to lower complexity and faster computation. Although it has been demonstrated to be an effective tool for the application of display metrology [83], our results suggest that the spatial frequency and orientation selectivity modeling are important for the medical image inspection.

6.4 Conclusion

In this chapter, four HVS models are compared for their potential use as a perceptual model observer for the detection task on MR images. The results suggest that the VDP model outperforms the other three HVS models and can be used to correctly model the radiologists' sensation task; nevertheless, there is a considerable distance between the HVS models' performance and radiologists' perception performance. Moreover, in the perception task it is important to include the anatomical information which helps the HVS models gain a better and closer performance. Our

proposed assessment methodology remains valid for comparing and choosing the most suitable model observer that best fits the radiologists' consensus results. This type of study offers an insight into modeling the diagnostic process and highlights the characteristics to be considered for developing a suitable model for medical image quality assessment.

Conclusion of Part II

This second part was dedicated to the experimental pre-studies for exploring the application of existing numerical observers on the MS lesion detection on MR images. In our experiments, MS lesions were simulated as the target stimuli for detection.

The purpose of the first study was to analyze the influences of MR image anatomical information and radiologist expertise on the detection task. Two different image backgrounds were used in the following experiments: a) homogeneous region of white matter tissue, and b) one slice of a healthy brain MR image. One expert radiologist, three radiologists and eight Naive observers (without any prior medical knowledge) performed these experiments, during which different questions dependent were asked depending upon the level of experience; the three radiologists and eight Naive observers were asked if they were aware of any hyper-signal, likely to represent an MS lesion, while the most experienced expert was asked if a clinically significant sign was present. With the percentages of responses "yes" displayed on the y-axis and the lesion intensity contrasts on the x-axis, psychometric function was generated from the observer' responses. Results of psychometric functions and calculated thresholds indicated that radiologists had better hyper-signal detection ability than Naive observers, which was intuitively shown by the lower simple visibility thresholds of radiologists. Moreover, the study indicated that for the radiologists, the simple visibility thresholds remained the same with and without the anatomical information. But anatomical information reduced the threshold for the clinically significant sign detection task.

In the second study concerning the comparison of HVS models, healthy cerebral MR slice of T2 FLAIR sequence was taken as reference image; that with one simulated MS lesion was taken as distorted image. 84 differently distorted images are generated given fourteen intensity levels and six positions. The model performance is evaluated against gold standard derived from three radiologists' consensus. Thus the AUC measures the distance between the model and the radiologist performance. Four experiments have been specially designed to identify radiologists' sensation and perception behavior. The AUC estimates differ significantly between one HVS model - Visible Difference Predictor (VDP) and other HVS models. VDP presents a significantly

higher AUC score estimate (>0.95) with a small variance in the four cases, although it is still far away from the human performance in the perception task.

Results in Chapter 5 and Chapter 6 indicate that there is a great demand for:

- a novel numerical observer which can treat not only small homogeneous regions, but also entire images with anatomical structures;
- a novel numerical observer which can take into account the radiologists' expertise and the perception stage of visual processing, knowing that the sensation models (existing HVS models) cannot really approach radiologist performance in the perception task.

Part III

Proposed novel numerical observers

Introduction of Part III

Recall the technical barriers for existing numerical models in case of both single-slice and multi-slice images, as said in Part I and Part II:

1. the range of variable signal parameters of the existing SKS MOs urgently needs to be widened;
2. the perception of the HVS processing needs to be considered, knowing that the existing HVS models are only sensation models;
3. more complex diagnostic tasks (e.g. the localization and estimation tasks, task on multi-slice images) have to be treated;
4. in numerous modalities (e.g. MRI, CT scan...) the entire image with anatomical information has to be processed and not only a small homogeneous regions, because of the importance of anatomical information in the diagnostic decision;
5. radiologists' expertise needs to be taken into account.

In the last part of this thesis, we propose several novel numerical observers trying to surmount these barriers, for both single-slice and multi-slice medical images.

This part comprises three chapters.

In Chapter 7, we propose a novel SKS MO, named as Channelized Joint detection and estimation Observer (CJO), for the detection of one parametric signal with random amplitude, orientation and size on single-slice images.

In Chapter 8, we propose a nonlinear numerical observer, named as Perceptually relevant Channelized Joint Observer (PCJO) for the detection-localization of multiple parametric signals with random amplitude, orientation, size and location on single-slice images, while the number of signals on the image is also unknown.

In Chapter 9, we propose a multi-slice numerical observer, named as multi-slice PCJO (msPCJO) for the detection-localization of multiple parametric signals with random amplitude, orientation, size and location on multi-slice images, while the number of signals in the volumetric image is also unknown.

7

CJO for detection task on single-slice

In this chapter, we deal with the simplest task in the diagnostic process: the signal-detection on single-slice images.

As said in Section 3.3, the experiments on SKE MOs showed that the task performances of these models are much higher than that of human observer according to the AUC FOM. The main reason lies in the unrealistic hypothesis of the SKE MO concerning the exact knowledge of the signal of interest on the image, while the human observer only has a vague idea on the signal parameters (location, amplitude, shape, orientation and size). Contrarily, the SKS MO constitutes a model closer to human observation, by including the random characteristics of the signal. The SKS MO searches the presence of a certain category of signals, not the signal known *a priori*. Thus we expect that this flexibility can lead to a MO with a task performance closer to that of human observer.

Toward this goal, we deepen the research in the SKS MO and extend the SKS CHO presented in [120] to the case in which signal amplitude, orientation and size are unknown at the same time. We named our extended numerical observer the Channelized Joint detection and estimation Observer (CJO).

7.1 Joint Detection and Estimation (JDE)

Recall that for the detection task, the two hypotheses can be formulated as follows:

$$\mathcal{H}_h : \ \boldsymbol{g} = h\boldsymbol{x} + \boldsymbol{b}, \ h = 0, 1 \tag{7.1}$$

where $\boldsymbol{g}$ is an $M \times 1$ column vector representing the digital image with M pixels, and the binary variable h controls the absence or presence of the signal $\boldsymbol{x}$. $\boldsymbol{x}_\alpha$ denotes a particular signal with parameters vector α:

$$\alpha = [a, \theta, b, \sigma, \boldsymbol{q}] \tag{7.2}$$

where a represents the signal amplitude (intensity), θ represents the signal orientation, b represents the signal shape, σ represents the signal scale, and $\boldsymbol{q}$ represents the signal location (the coordinates of the signal center).

One possible solution for SKS tasks [120] is an approach based on the joint detection and estimation (JDE) theory [66], in which the estimates of α in Eq.(7.2) and $\mathcal{H}_h$ in Eq.(7.1) are chosen jointly to maximize the joint posterior probability $P(\alpha, \mathcal{H}_h \,|\, \boldsymbol{g})$. Thus the validation of a hypothesis in Eq.(7.1) is accompanied by the maximum a posteriori probability (MAP) estimation

of unknown signal parameters:

$$\begin{aligned}\widehat{(\alpha, \mathcal{H}_h)} &= \arg\max_{\alpha,\mathcal{H}_h} P(\alpha, \mathcal{H}_h \,|\, \boldsymbol{g}) \\ &= \arg\max_{\alpha,\mathcal{H}_h} \frac{P(\boldsymbol{g} \,|\, \alpha, \mathcal{H}_h)P(\alpha)P(\mathcal{H}_h)}{P(\boldsymbol{g})} \\ &= \arg\max_{\alpha,\mathcal{H}_h} P(\boldsymbol{g} \,|\, \alpha, \mathcal{H}_h)P(\alpha)P(\mathcal{H}_h) \end{aligned} \tag{7.3}$$

where the statistical independence of α and $\mathcal{H}_h$ has been exploited. Under the assumption of a zero-mean Correlated Gaussian Background (CGB, cf. Section 3.1.1), the conditional probability density function $P(\boldsymbol{g} \,|\, \alpha, \mathcal{H}_h)$ can be written as:

$$P(\boldsymbol{g} \,|\, \alpha, \mathcal{H}_h) = \frac{1}{\sqrt{(2\pi)^M \,|\, \Sigma_{\boldsymbol{b}} \,|}} \exp\left\{-\frac{1}{2}(\boldsymbol{g} - h\boldsymbol{x}_\alpha)^{\mathrm{t}} {\Sigma_b}^{-1} (\boldsymbol{g} - h\boldsymbol{x}_\alpha)\right\} \tag{7.4}$$

In view of the monotonic logarithmic function, the maximization of the Eq.(7.3) is then equivalent to:

$$\widehat{(\alpha, \mathcal{H}_h)} = \arg\max_{\alpha,\mathcal{H}_k} \left\{\ln P(\mathcal{H}_h) + \ln P(\alpha) - \frac{1}{2}(\boldsymbol{g} - h\boldsymbol{x}_\alpha)^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} (\boldsymbol{g} - h\boldsymbol{x}_\alpha)\right\} \tag{7.5}$$

$$\begin{aligned} &= \arg\max_{\alpha,\mathcal{H}_h} \{\ln P(\mathcal{H}_h) + \ln P(\alpha) \\ &\quad - \frac{1}{2}(\boldsymbol{g}^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{g} - h\boldsymbol{x}_\alpha^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{g} - h\boldsymbol{g}^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{x}_\alpha + h^2 \boldsymbol{x}_\alpha^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{x}_\alpha)\right\} \end{aligned} \tag{7.6}$$

$$\begin{aligned} &= \arg\max_{\alpha,\mathcal{H}_h} \{\ln P(\mathcal{H}_h) + \ln P(\alpha) \\ &\quad - -\frac{1}{2}(-h\boldsymbol{x}_\alpha^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{g} - h\boldsymbol{x}_\alpha^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{g} + h^2 \boldsymbol{x}_\alpha^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{x}_\alpha)\right\} \end{aligned} \tag{7.7}$$

$$= \arg\max_{\alpha,\mathcal{H}_h} \left\{\ln P(\mathcal{H}_h) + \ln P(\alpha) + h\boldsymbol{x}_\alpha^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} (\boldsymbol{g} - \frac{1}{2} h\boldsymbol{x}_\alpha)\right\} \tag{7.8}$$

where the term $\boldsymbol{g}^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{g}$ is ignored in Eq.(7.7) since it is not a function of α and $\mathcal{H}_h$. Notice that the transpose of a scalar term (e.g. $h\boldsymbol{g}^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{x}_\alpha$) is equal to itself.

7.1.1 Estimation

The estimation problem here is actually to find the parameters maximizing the probability in the maximum a posteriori probability (MAP) sense:

$$\begin{aligned}\widehat{\alpha} &= \arg\max_{\alpha} P(\alpha \,|\, \boldsymbol{g}) \\ &= \arg\max_{\alpha} \ \max\left\{P(\boldsymbol{g} \,|\, \boldsymbol{x}_\alpha, \mathcal{H}_1)P(\alpha)P(\mathcal{H}_1), P(\boldsymbol{g} \,|\, \boldsymbol{x}_\alpha, \mathcal{H}_0)P(\alpha)P(\mathcal{H}_0)\right\} \\ &= \arg\max_{\alpha} \ \max\left\{P(\boldsymbol{g} \,|\, \boldsymbol{x}_\alpha, \mathcal{H}_1)P(\alpha)P(\mathcal{H}_1), P(\boldsymbol{g} \,|\, \mathcal{H}_0)P(\alpha)P(\mathcal{H}_0)\right\}\end{aligned} \tag{7.9}$$

which is equivalent to maximizing separately $P(\alpha \,|\, \mathcal{H}_k, \boldsymbol{g})$ for $k = 0, 1$, then comparing the two results and choosing the bigger one. Notice that the estimated parameters have no physical meaning in the case of hypothesis $\mathcal{H}_0$ (signal is absent).

Without loss of generality, two additional assumptions are often taken in practice to further simplify the expression:

- the parameters are uniformly distributed over the admissible space ($P(\alpha) \propto 1$);
- the two hypotheses $k = 0, 1$ are equiprobable: $P(\mathcal{H}_0) = P(\mathcal{H}_1)$.

This simplification is adopted in this work to facilitate the validation of the model observer for SKS tasks. Thus from Eq.(7.8) and Eq.(7.9), the estimation problem becomes:

$$\widehat{\alpha} = \arg\max_{\alpha} \left\{\max\left\{\boldsymbol{x}_\alpha^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1}(\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_\alpha), 0\right\}\right\} \tag{7.10}$$

where 0 corresponds to the case $k = 0$. It follows that:

- if for all α, $\boldsymbol{x}_\alpha^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1}(\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_\alpha) \leq 0$, then the maximization in Eq.(7.10) is meaningless since the maximum is always zero;
- otherwise $\widehat{\alpha} = \arg\max_\alpha \left\{\boldsymbol{x}_\alpha^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1}(\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_\alpha)\right\}$.

Consequently the estimation algorithm can be written as:

1. Calculate α_t:

$$\alpha_t = \arg\max_{\alpha} \left\{\boldsymbol{x}_\alpha^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1}(\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_\alpha)\right\}; \tag{7.11}$$

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

$$\widehat{\alpha} = \begin{cases} \alpha_t & \text{if } \boldsymbol{x}_{\alpha_t}^{\mathrm{t}} \Sigma_{\boldsymbol{b}}^{-1}(\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_{\alpha_t}) > 0 \\ \text{any value in the space of } \alpha & \text{else} \end{cases} \tag{7.12}$$

It should be noted that in general there is no analytical solution for Eq.(7.11) because of the nonlinearity and non-concavity of the function. An 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.

7.1.2 Detection

Now we consider the posterior probabilities of the two hypotheses given the estimated signal parameters $\widehat{\alpha}$ and a particular observation $\boldsymbol{g}$:

$$P(\mathcal{H}_k \,|\, \widehat{\alpha}, \boldsymbol{g}) = \frac{P(\mathcal{H}_k)P(\widehat{\alpha})P(\boldsymbol{g} \,|\, \mathcal{H}_k, \widehat{\alpha})}{P(\boldsymbol{g} \,|\, \widehat{\alpha})P(\widehat{\alpha})} \propto P(\mathcal{H}_k)P(\boldsymbol{g} \,|\, \mathcal{H}_k, \widehat{\alpha}) \quad k = 0, 1 \tag{7.13}$$

where $P(\boldsymbol{g} \,|\, \widehat{\alpha})$ in the denominator is omitted since it is not a function of $\mathcal{H}_k$.

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 as below:

$$\begin{aligned} &\ln P(\mathcal{H}_1) + \ln P(\boldsymbol{g} \,|\, \mathcal{H}_1, \widehat{\alpha}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \ln P(\mathcal{H}_0) + \ln P(\boldsymbol{g} \,|\, \mathcal{H}_0, \widehat{\alpha}) \\ \Leftrightarrow\ &-\frac{1}{2}(\boldsymbol{g} - \boldsymbol{x}_{\widehat{\alpha}})^{\mathrm{t}}\Sigma_{\boldsymbol{b}}^{-1}(\boldsymbol{g} - \boldsymbol{x}_{\widehat{\alpha}}) + \frac{1}{2}\boldsymbol{g}^{\mathrm{t}}\Sigma_{\boldsymbol{b}}^{-1}\boldsymbol{g} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \ln \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} \\ &\lambda = \boldsymbol{x}_{\widehat{\alpha}}^{\mathrm{t}}\Sigma_{\boldsymbol{b}}^{-1}\left(\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_{\widehat{\alpha}}\right) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \ln \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} \end{aligned} \tag{7.14}$$

It's easy to see that the test statistic λ is linear in $\boldsymbol{g}$, given the optimal estimated signal parameters, the background covariance matrix $\Sigma_{\boldsymbol{b}}$ and the prior probabilities of two hypotheses $P(\mathcal{H}_k)$ (often assumed to be equal without special knowledge of the data set).

Note that the test statistic is actually realized jointly with the estimation step. 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 Eq.(7.12) and the validation of $\mathcal{H}_k$ (detection decision) Eq.(7.14), where the name JDE comes from.

7.2 CJO

The calculation of the test statistic in Eq.(7.14) has the same compute-intensive problem as HO, in the face of modern images high dimensionality. 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. So the CHO's channelization method Eq.(3.18) is employed in CJO to reduce the dimentionality.

The maximization in Eq.(7.11) based on the channelized image $\boldsymbol{g}'$ is then formulated as follows:

$$\alpha_t = \arg\max_{\alpha} \frac{\boldsymbol{x}_\alpha^{\mathrm{t}}}{\|\mathbf{U}_\alpha\|_F^2} \left(\mathbf{U}_\alpha \left(\Sigma_{\boldsymbol{b}}'\right)^{-1} \mathbf{U}_\alpha^{\mathrm{t}}\right) \left(\boldsymbol{g} - \frac{1}{2}\boldsymbol{x}_\alpha\right) \tag{7.15}$$

where $\|\mathbf{U}_\alpha\|_F^2$ is a channel matrix energy normalization factor and $\|\cdot\|_F$ denotes the Frobenius norm of the matrix [120]. Ideally, we want the function to be optimized in Eq.(7.15) is close enough to that in Eq.(7.11), so that the impact of the dimentionality reduction is minimal. Intuitively, this implies that the covariance matrix $\Sigma_{\boldsymbol{b}}'$ and the channel matrix $\mathbf{U}_\alpha$ should be calibrated properly, namely $\Sigma_{\boldsymbol{b}}^{-1} \approx \frac{1}{\|\mathbf{U}_\alpha\|_F^2} \mathbf{U}_\alpha \left(\Sigma_{\boldsymbol{b}}'\right)^{-1} \mathbf{U}_\alpha^{\mathrm{t}}$. This is a sufficient condition but not necessary for that the α_t obtained by Eq.(7.15) is approximate to that given by Eq.(7.10).

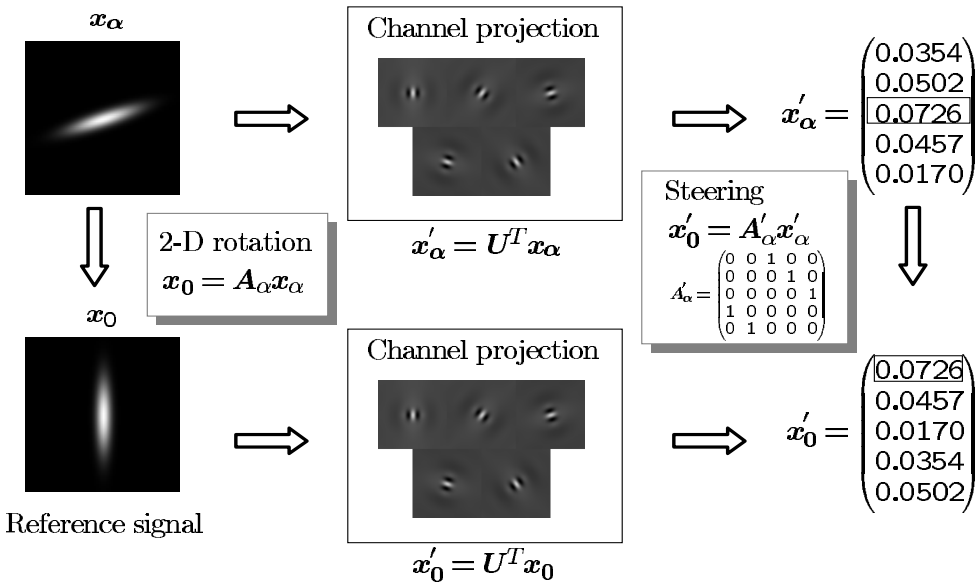


Figure 7.1: An illustration of the channel design condition (from Goossens' thesis [120]): a rotation in the image domain corresponds to a matrix multiplication in the channel domain. The channel matrix $\mathbf{U}$ and the matrices $\mathbf{A}_\alpha$, $\mathbf{A}'_\alpha$ need to be chosen suitably, such that in the figure the end result for transiting from $\boldsymbol{x}_\alpha$ to $\boldsymbol{x}'_0$ is the same, independent of the path being followed.

The maximization problem in Eq.(7.15) is hardly realizable since practically it needs a whole set of channel matrices $\mathbf{U}_\alpha$ with varying α (the search space of α is huge). An important

contribution of Goossens is to solve this problem 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:

$$\mathbf{U}_\alpha = \mathbf{A}_\alpha^{\mathrm{t}}\mathbf{U} = \mathbf{U}\left(\mathbf{A}'_\alpha\right)^{\mathrm{t}} \tag{7.16}$$

where $\mathbf{U}$ is a fixed channel matrix which does not depend on α and conduce to reduce the data dimensionality; $\mathbf{A}_\alpha$ serve to map the parametric signal $\boldsymbol{x}_\alpha$ onto a reference signal $\boldsymbol{x}_0$ (with a known orientation/scale): $\mathbf{A}_\alpha\boldsymbol{x}_\alpha = \boldsymbol{x}_0$; $\mathbf{A}'_\alpha$ map the channelized parametric signal $\boldsymbol{x}'_\alpha$ onto the channelized reference signal $\boldsymbol{x}'_0$: $\mathbf{A}'_\alpha\boldsymbol{x}'_\alpha = \boldsymbol{x}'_0$, and the size of $\mathbf{A}'_\alpha$ is normally much smaller than that of $\mathbf{A}_\alpha$. A clear illustration (taking the orientation-unknown detection task as an example) is shown in Figure 7.1 (from Goossens' thesis [120]).

Then the estimation and the detection (as in Eq.(7.11) and Eq.(7.14) respectively) can be rewritten as:

$$\alpha_t = \arg\max_\alpha \frac{1}{\left\|\mathbf{U}\left(\mathbf{A}'_\alpha\right)^{\mathrm{t}}\right\|_F^2}\left(\boldsymbol{x}'_0\right)^{\mathrm{t}}\left(\Sigma'_{\boldsymbol{b}}\right)^{-1}\left(\mathbf{A}'_\alpha\boldsymbol{g}' - \frac{1}{2}\boldsymbol{x}'_0\right) \tag{7.17}$$

$$\widehat{\alpha} = f(\alpha_t) \quad \text{as defined in (7.12)}$$

$$\lambda = \frac{1}{\left\|\mathbf{U}\left(\mathbf{A}'_{\widehat{\alpha}}\right)^{\mathrm{t}}\right\|_F^2}\left(\boldsymbol{x}'_0\right)^{\mathrm{t}}\left(\Sigma'_{\boldsymbol{b}}\right)^{-1}\left(\mathbf{A}'_{\widehat{\alpha}}\boldsymbol{g}' - \frac{1}{2}\boldsymbol{x}'_0\right) \tag{7.18}$$

7.2.1 Amplitude-unknown, orientation-unknown and scale-unknown task

In [120], suitable channels according to the different specific tasks of detection (amplitude-unknown, orientation-unknown or scale-unknown) have been derived.

The simplest task is the amplitude-unknown case for which any channel could be adopted since in order to satisfy Eq.(7.16) it suffices to choose

$$\mathbf{A}'_a = a^{-1}\mathbf{I} \tag{7.19}$$

For the other two more complicated tasks (orientation-unknown and scale-unknown), a set of *steerable* channels is adopted for the orientation-unknown case while a set of *shiftable* channels is adopted for the scale-unknown case. An important feature of these *steerable*/*shiftable* basis

channels is that another channel at any orientation/scale may be computed as a linear combination of the basis channels [121, 122].

In polar-frequency coordinates, let ω be the radial frequency and φ be the angular orientation. The steerable channels are given in Eq.(7.20):

$$f_{\theta_k}(\varphi) = \frac{(K-1)!2^{K-1}}{\sqrt{K(2K-2)!}} \left(\cos(\varphi - \theta_k)\right)^{K-1} \tag{7.20}$$

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

Next, the scale-shiftable channels are defined as follows:

$$f_{\sigma_j}(\omega) = \text{sinc}\left(\log_2(\frac{|\,\omega\,|}{\pi} + \epsilon) + \log_2 \sigma_{\text{j}}\right) \qquad |\,\omega\,| < \pi; \tag{7.21}$$

where $\sigma_j = 2^{j-1}$ for $j = 1, \ldots, J$ 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 defines then the channel functions for the amplitude-orientation-scale-unknown case:

$$f_{\theta_k,\sigma_j}(\omega, \varphi) = f_{\theta_k}(\varphi) f_{\sigma_j}(\omega) \qquad k = 1, \ldots, K; j = 1, \ldots, J \tag{7.22}$$

An example of channels ($K = 3$ and $J = 4$) is shown in Figure 7.2.

By storing each sampled channel function Eq.(7.22) in a column, we construct the channel matrix $\mathbf{U}_{M^2 \times JK}$ for the amplitude-orientation-scale-unknown case, for which the $(j-1)K + k$ -th column is the vectorized channel $f_{\theta_k,\sigma_j}(\omega, \varphi)$.

Two corresponding transform matrices for the orientation-unknown case and scale-unknown case are given by Eq.(7.23) and Eq.(7.24) respectively [120]:

$$[\mathbf{A}'_\theta]_{m,n} = \frac{1}{K} \frac{\sin(\pi(m-n) - \theta K)}{\sin(\pi(m-n)/K - \theta)} \qquad m, n = 1, \ldots, K; \tag{7.23}$$

$$[\mathbf{A}'_\sigma]_{m,n} = \text{sinc}\left(((\text{m} - \text{n}) - (-\log_2 \sigma))\right) \qquad \text{m, n} = 1, \ldots, \text{J}; \tag{7.24}$$

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

$$\text{sinc}(t) = \begin{cases} 1 & t = 0 \\ \frac{\sin(\pi t)}{\pi t} & t \neq 0 \end{cases} \tag{7.25}$$

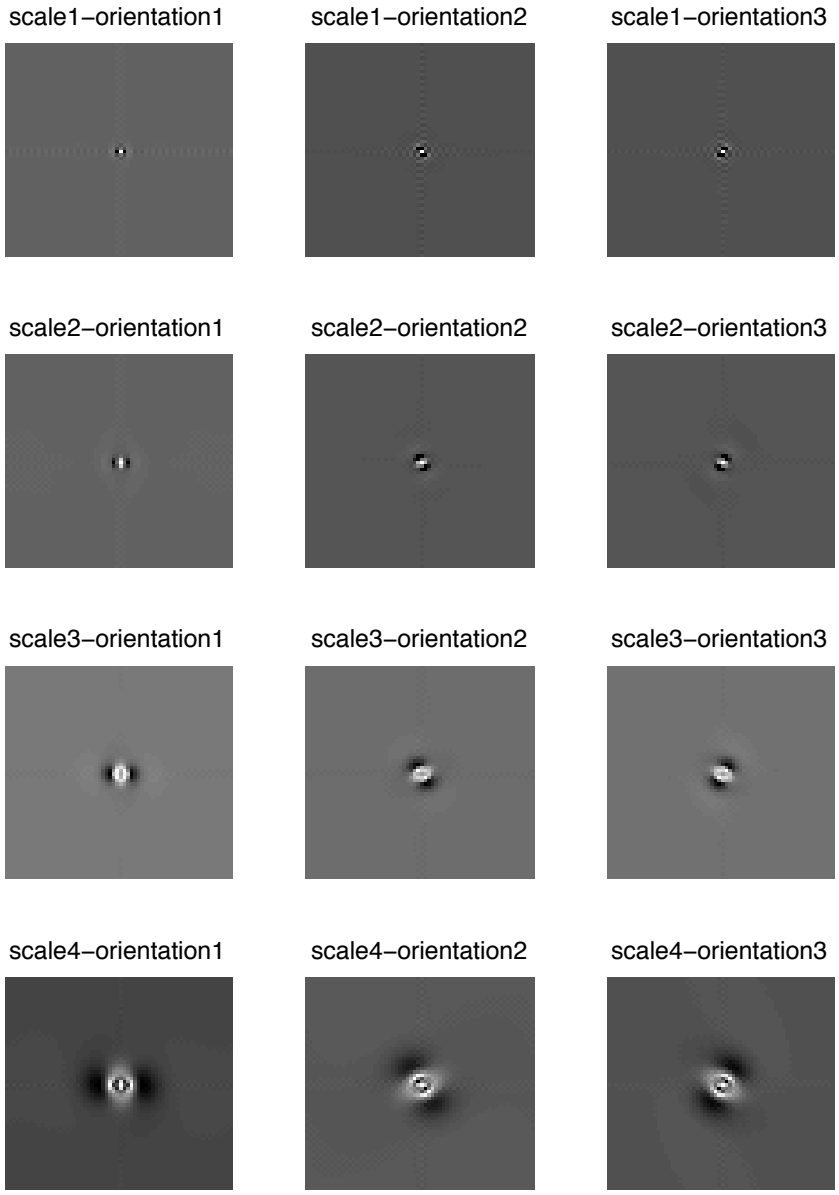


Figure 7.2: An example of the filters for the amplitude-orientation-scale-unknown case, when the number of steerable channels $K = 3$ and the number of scale-shiftable channels $J = 4$

7.2.2 CJO for the amplitude-orientation-scale-unknown task

The design of the transform matrix for the orientation-scale-unknown case depends on the sample storage ordering during the construction of channel matrix $\mathbf{U}$, we have in the present configuration:

$$\mathbf{A}'_{\alpha} = \mathbf{A}'_{a,\theta,\sigma} = \frac{1}{a}\mathbf{A}'_{\sigma} \otimes \mathbf{A}'_{\theta} \tag{7.26}$$

where $\otimes$ denotes the Kronecker product.

Notice that for an odd number K, the matrix $\mathbf{A}'_{\theta}$ has the interesting unitary property: $(\mathbf{A}'_{\theta})^{\mathrm{t}}\mathbf{A}'_{\theta} = \mathbf{I}$. 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. A simplified numerical method is provided here.

From the definition of the Frobenius norm, we have:

$$\begin{aligned}
\left\|\mathbf{U}\left(\mathbf{A}'_{\sigma} \otimes \mathbf{A}'_{\theta}\right)^{\mathrm{t}}\right\|_F^2 &= \mathrm{trace}\left(\left(\mathbf{A}'_{\sigma} \otimes \mathbf{A}'_{\theta}\right)\mathbf{U}^{\mathrm{t}}\mathbf{U}\left(\mathbf{A}'_{\sigma} \otimes \mathbf{A}'_{\theta}\right)^{\mathrm{t}}\right) \\
&= \mathrm{trace}\left(\mathbf{U}^{\mathrm{t}}\mathbf{U}\left(\mathbf{A}'_{\sigma} \otimes \mathbf{A}'_{\theta}\right)^{\mathrm{t}}\left(\mathbf{A}'_{\sigma} \otimes \mathbf{A}'_{\theta}\right)\right) \\
&= \mathrm{trace}\left(\mathbf{U}^{\mathrm{t}}\mathbf{U}\left(\left(\mathbf{A}'_{\sigma}\right)^{\mathrm{t}} \otimes \left(\mathbf{A}'_{\theta}\right)^{\mathrm{t}}\right)\left(\mathbf{A}'_{\sigma} \otimes \mathbf{A}'_{\theta}\right)\right) \\
&= \mathrm{trace}\left(\mathbf{U}^{\mathrm{t}}\mathbf{U}\left[\left(\left(\mathbf{A}'_{\sigma}\right)^{\mathrm{t}}\mathbf{A}'_{\sigma}\right) \otimes \left(\left(\mathbf{A}'_{\theta}\right)^{\mathrm{t}}\mathbf{A}'_{\theta}\right)\right]\right) \\
&= \mathrm{trace}\left(\mathbf{U}^{\mathrm{t}}\mathbf{U}\left[\left(\left(\mathbf{A}'_{\sigma}\right)^{\mathrm{t}}\mathbf{A}'_{\sigma}\right) \otimes \mathbf{I}\right]\right)
\end{aligned} \tag{7.27}$$

Let $(\mathbf{A}'_{\sigma})^{\mathrm{t}}\mathbf{A}'_{\sigma} = \mathbf{S} = [s_{ij}]$ $(i, j = 1, \ldots, J)$, and divide the matrix $\mathbf{U}^{\mathrm{t}}\mathbf{U}$ (of dimension $JK \times JK$) into $J \times J$ submatrices T_{ij} $(i, j = 1, \ldots, J)$, each of dimension $K \times K$, then Eq.(7.27) can be written as:

$$\begin{aligned}
\sum_{i=1}^{J}\mathrm{trace}\left(\sum_{j=1}^{J}s_{ji}T_{ij}\right) &= \sum_{i=1}^{J}\mathrm{trace}\left(\sum_{j=1}^{J}s_{ij}T_{ij}\right) \\
&= \sum_{i=1}^{J}\sum_{j=1}^{J}\mathrm{trace}\left(s_{ij}T_{ij}\right) \\
&= \sum_{i=1}^{J}\sum_{j=1}^{J}s_{ij}\mathrm{trace}\left(T_{ij}\right)
\end{aligned} \tag{7.28}$$

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, thus $\mathrm{trace}\left(T_{ij}\right)$ can be pre-calculated.

7.2.3 Practical implementation of the CJO for the amplitude-orientation-scale-unknown task

As SKE CHO's practical implementation (cf. Section 3.2.3.2), we can realize the CJO for the amplitude-orientation-scale-unknown task by training and test stages.

7.2.3.1 Stage 1: Training

In the training stage, the inputs are two sets of images (with-signal and without-signal) whose ground truth is known, including all the signal parameters α (so that $\mathbf{A}'_{a,\theta,\sigma}$ can be calculated). Let $\boldsymbol{g}$ denote a particular image, $\boldsymbol{x}_\alpha$ a particular signal.

The outputs of this stage are the estimated channelized reference signal $\boldsymbol{x}'_0$ and a template $\boldsymbol{w}$, unlike the SKE CHO in Section 3.2.3.2.

$$\widehat{\boldsymbol{x}'_0} = \langle \mathbf{A}'_{a,\theta,\sigma} \boldsymbol{x}'_\alpha \rangle = \langle \mathbf{A}'_{a,\theta,\sigma} (\mathbf{U}^{\mathrm{t}} \boldsymbol{x}_\alpha) \rangle \tag{7.29}$$

$$\boldsymbol{w} = \left(\widehat{\Sigma'_{\boldsymbol{b}}} \right)^{-1} \widehat{\boldsymbol{x}'_0} \tag{7.30}$$

where $\langle \cdot \rangle$ denotes the sample average. $\widehat{\Sigma'_{\boldsymbol{b}}}$ is the estimated background covariance matrix:

$$\begin{aligned} \widehat{\Sigma'_{\boldsymbol{b}}} =& \frac{1}{2} \langle (\boldsymbol{g}' - \langle \boldsymbol{g}' \mid H_0 \rangle) (\boldsymbol{g}' - \langle \boldsymbol{g}' \mid H_0 \rangle)^{\mathrm{t}} \mid H_0 \rangle + \\ & \frac{1}{2} \langle [(\boldsymbol{g}' - \boldsymbol{x}'_\alpha) - \langle (\boldsymbol{g}' - \boldsymbol{x}'_\alpha) \mid H_1 \rangle] [(\boldsymbol{g}' - \boldsymbol{x}'_\alpha) - \langle (\boldsymbol{g}' - \boldsymbol{x}'_\alpha) \mid H_1 \rangle]^{\mathrm{t}} \mid H_1 \rangle \end{aligned} \tag{7.31}$$

where $\boldsymbol{g}'$ is the channelized image with the M^2-by-JK channel matrix $\mathbf{U}$ constructed from Eq.(7.22):

$$\boldsymbol{g}' = \mathbf{U}^{\mathrm{t}} \boldsymbol{g} \tag{7.32}$$

7.2.3.2 Stage 2: Test

In the test stage, $\boldsymbol{w}$ and $\boldsymbol{x}'_0$ are used to estimate the signal parameters, as well as to calculate the test statistic λ for each input test image $\boldsymbol{g}$, the ground truth is however unknown to the CJO.

To find the maximum in Eq.(7.17), we used a trivial but very general problem-solving technique: brute-force search for a certain number (denoted by N) of equally spaced signal parameters. Then the maximum of the test statistics yielded by all the possible combinations of chosen param-

eters is chosen as the final test statistic:

$$\lambda = \max_{a,\theta,\sigma}(\lambda_{a,\theta,\sigma}) = \max_{a,\theta,\sigma}\left(\frac{\boldsymbol{w}^{\mathrm{t}}}{\left\|\mathbf{U}\left(\mathbf{A}'_{a,\theta,\sigma}\right)^{\mathrm{t}}\right\|_F^2}\left(\mathbf{A}'_{a,\theta,\sigma}\boldsymbol{g}' - \frac{1}{2}\widehat{\boldsymbol{x}'_0}\right)\right) \tag{7.33}$$

By comparing Section 3.2.3.2 and Section 7.2.3, we can find the influence introduced by the unknown variable.

7.3 Performance evaluation of the CJO

In this section, we will evaluate the CJO's estimation and detection performance using both the synthesized correlated Gaussian backgrounds and the real clinical backgrounds (white matter regions extracted from brain MR axial images of T2 FLAIR sequences).

7.3.1 Parameter setup for the CJO

For both synthesized images and real clinical images, the image size is 65×65; 500 pairs of images are used for the training stage (500 with-signal and 500 without-signal images), and 100 pairs of images are used for the test stage. The simulated lesion is located in the center of the images. The signal orientation range is $[\theta_{\text{min}},\ \theta_{\text{max}}] = [0,\ \pi]$. The signal amplitude range $[a_{\text{min}},\ a_{\text{max}}]$ and the signal scale range $[\sigma_{\text{min}},\ \sigma_{\text{max}}]$ cover the corresponding maximum accessible parameter space according to the background type. For example, $[a_{\text{min}},\ a_{\text{max}}] = [1,\ 255]$ (pixel intensity value), and $[\sigma_{\text{min}},\ \sigma_{\text{max}}] = [1,\ 6]$ with $\sqrt{b} = 2$ for real clinical images.

7.3.2 Numerical instability problem

7.3.2.1 Problem formulation

To verify whether the signal parameters (amplitude, orientation and scale) could be well estimated by the CJO, we can plot the estimated signal parameters vs. true values of signal parameters. One example is shown in Figure 7.3, from which we can see that the relationship between the estimated signal parameters and their true values is nearly linear, and the orientation estimation is less biased than the estimation of the other two parameters.

In this study, we used the AUC and the estimate of its variance [33] as the FOM for the detection performance evaluation. In general, a higher value of AUC means a better detection

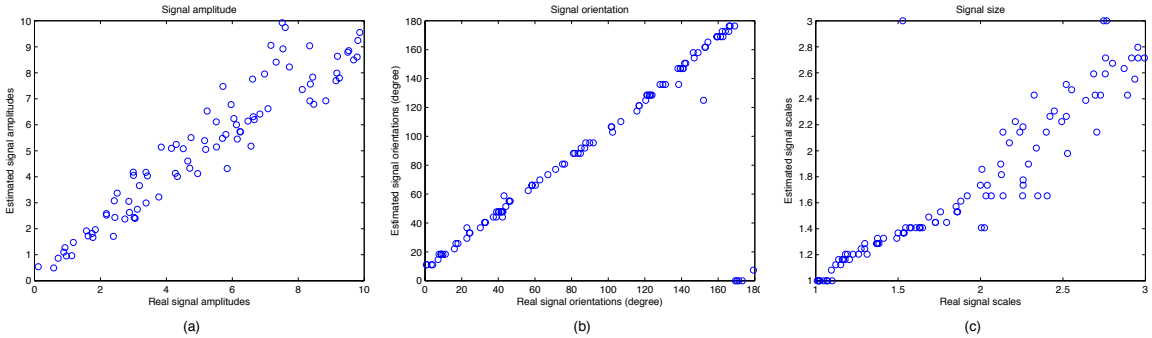


Figure 7.3: (a) Plot of estimated signal amplitudes vs. true values of signal amplitude; (b) Plot of estimated signal orientations vs. true values of signal orientation; and (c) Plot of estimated signal scales vs. true values of signal scale. Here the number of *steerable* channels $K = 3$ and the number of *scale-shiftable* channels $J = 5$.

performance. Moreover, we used the normalized root-mean-square error (NRMSE) [123] to quantify the signal parameter estimation errors to facilitate their comparison under different setups. The NRMSE is the root-mean-square error (RMSE) [124] divided by the range of true values. Lower NRMSE value indicates that the difference between the estimated value and the true value is smaller.

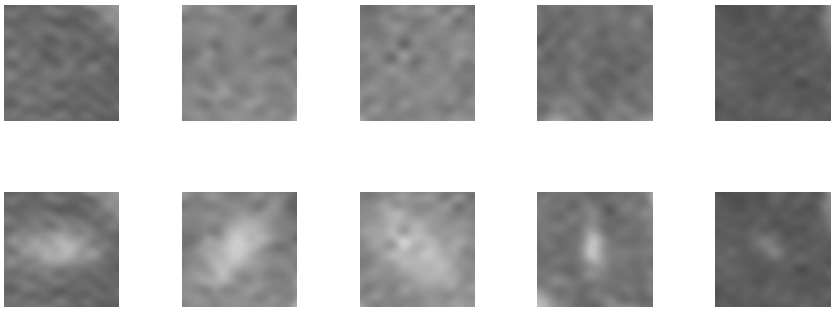


Figure 7.4: First row: examples of real clinical backgrounds which are white matter regions extracted from brain MR axial images of T2 FLAIR sequence. Second row: examples of real clinical backgrounds plus simulated lesions with different signal parameters.

We tested CJO's performances using both the Correlated Gaussian Background (CGB) and the real clinical background. The CGB here was generated by convolving white noise following the distribution $\mathcal{N}$ (0,1) with a 2D Gaussian kernel characterized by $\sigma_{\boldsymbol{b}} = 10$. The real clinical background here was a white matter region, extracted from healthy brain MR axial images of a T2 FLAIR sequence. As said before, the MS is a white matter disease; and in the T2 FLAIR sequence, MS lesions appear as hypersignals (intensity is higher than its surrounding white matter), which

could then be simulated by Eq.(3.3). Examples of white matter regions and those with simulated lesions are shown in Figure 7.4.

Plots of NRMSE and AUC with different combinations of number of *steerable* channels K and number of *scale-shiftable* channels J for the two types of background with the signal shape parameter $\sqrt{b} = 2$ are shown in Figure 7.5 and Figure 7.6.

Viewing the CJO performance results in Figure 7.5 and Figure 7.6, we notice that contrary to expectation, increasing the number of channels yields unstable (even decreasing) performances. We only showed the results for $\sqrt{b} = 2$ here, but this instability problem exists for whatever values of $\sqrt{b}$.

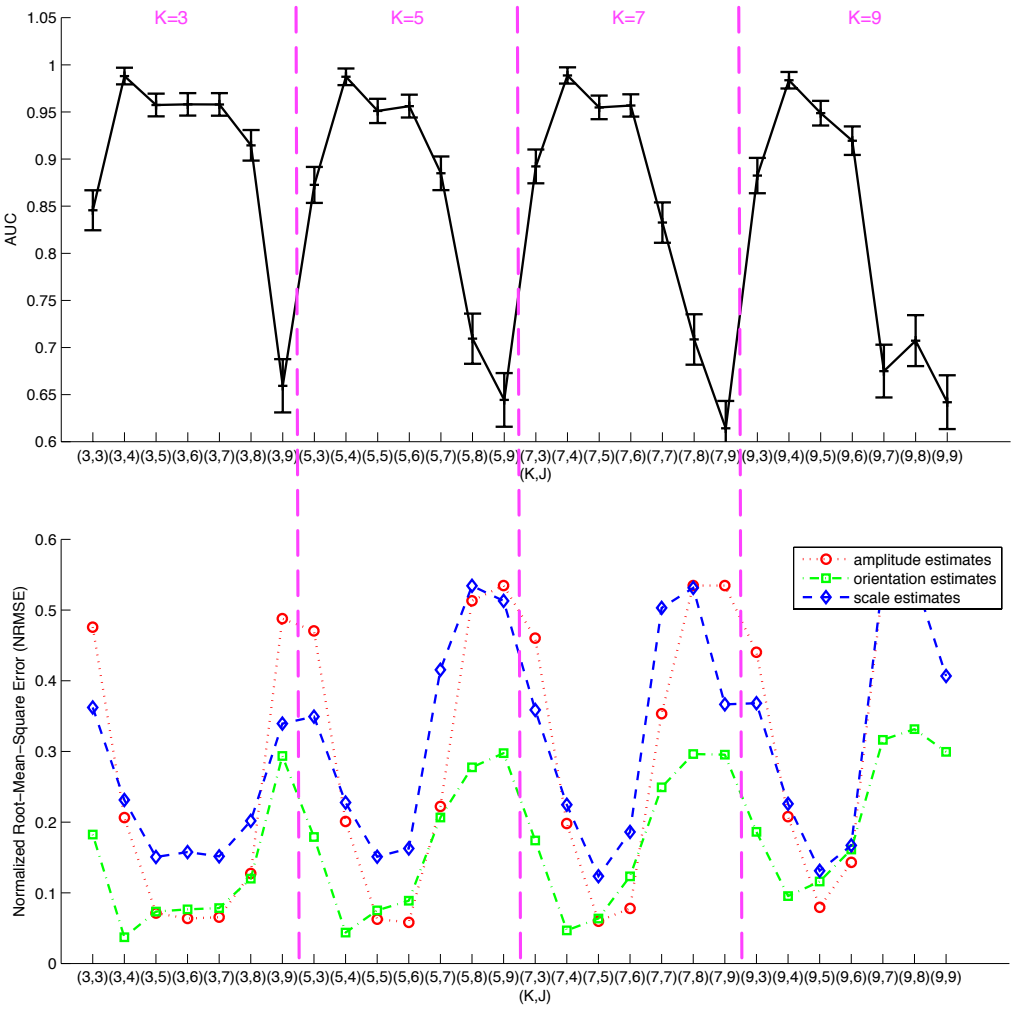


Figure 7.5: Plots of NRMSE and AUC as a function of the number of channels (the number of *steerable* channels is K and the number of *scale-shiftable* channels is J.) for correlated Gaussian background with the signal shape parameter $\sqrt{b} = 2$.

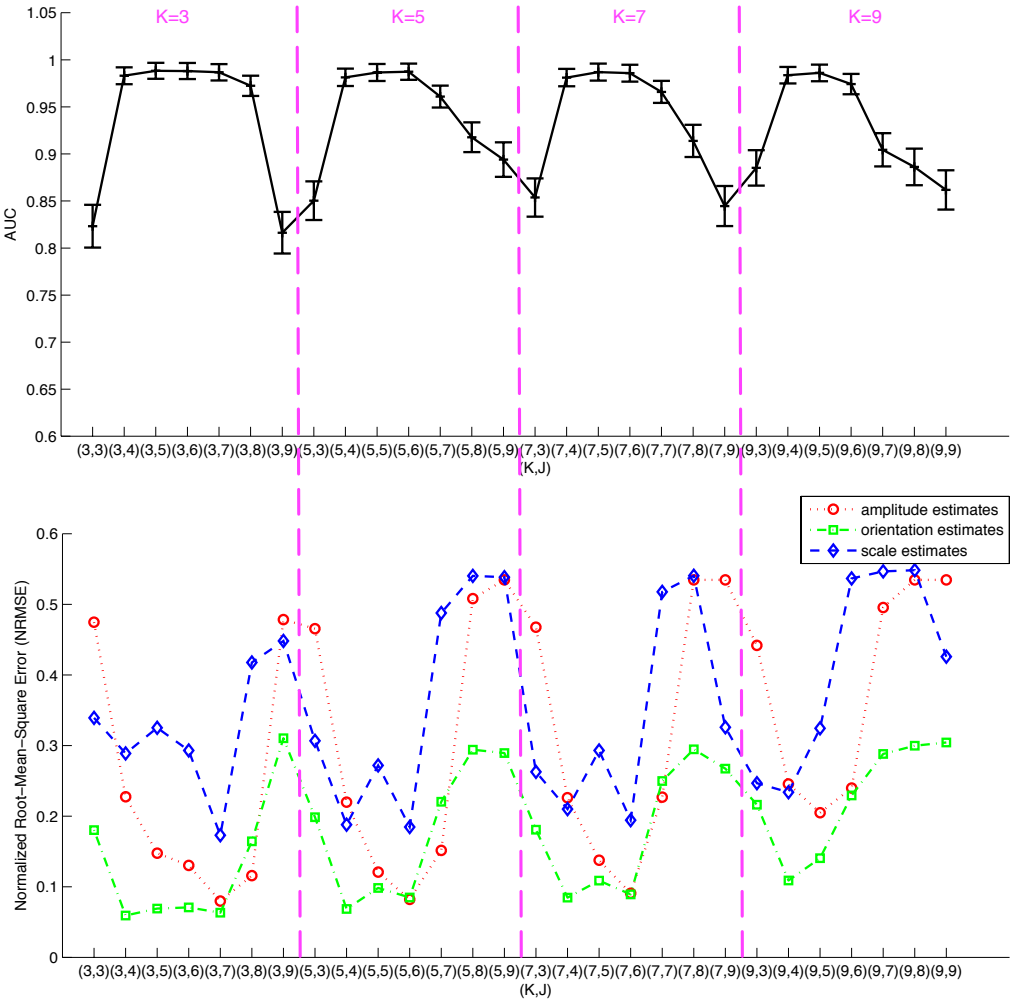


Figure 7.6: Plots of NRMSE and AUC as a function of the number of channels (the number of *steerable* channels is K and the number of *scale-shiftable* channels is J.) for white matter region with the signal shape parameter $\sqrt{b} = 2$.

7.3.2.2 Matrix condition and inversion

In collaboration with a postdoctoral fellow D. GE who also worked on the EQuIMOSe project, after investigating all the aspects in the design and implementation of the CJO, we finally find the cause of this numerical instability by tracing back to the implementation of the classical SKE CHO in the literature. This is because the estimated covariance matrix $\widehat{\Sigma'_b}$ in Eq.(7.31) becomes extremely ill-conditioned when the number of channels is large. The classical SKE CHO (the MATLAB source code of which can be found on the website http:

//www.radiology.arizona.edu/CGRI/IQ/page2/page4/page4.html) also faces this problem regarding the estimated covariance matrix $\widehat{\Sigma'}$ in Eq.(3.22).

This problem has not been revealed before, mainly because the number of channels used in the SKE CHO is normally small. But when we increase the number of channels in the SKE CHO, we can observe that the detection performances of the SKE CHO do not necessarily augment correspondingly for most channel and background sets. The investigation on the numerical instabilities for the SKE CHO will be given in [125] in the near future.

Now let us get back to the CJO. An example of the condition number [1] w.r.t. the number of channels for the correlated Gaussian background (results in the numerical instabilities as shown in Figure 7.5) is given in Figure 7.7. Note that the total number of channels in the CJO is equal to $K \times J$, we observed that:

- the condition numbers grow nearly exponentially by increasing the number of *scale-shiftable* channels J for every number of *steerable* channels K, while the condition numbers grow relatively slowly by increasing the number of *steerable* channels K for every number of *scale-shiftable* channels J;
- when $(K, J) = (3, 9), (5, 8), (5, 9), (7, 8), (7, 9), (9, 7), (9, 8), (9, 9)$, the matrix is ill-conditioned [2] (with high condition numbers $> 10^{11}$).

Note that the condition number evolution w.r.t. the number of channels for white matter region (results in the numerical instabilities as shown in Figure 7.6) yields similar results as commented above even though only correlated Gaussian background results are shown in Figure 7.7.

As a general rule, the inversion of a matrix for the resolution of a linear system $\mathbf{S}\boldsymbol{x} = \boldsymbol{y}$ yields an error of order :

$$|\, \mathbf{S}^{-1}\boldsymbol{y} - \boldsymbol{x}_0 \,| \sim \epsilon \cdot \text{cond}(\mathbf{S}),$$

by noting $\boldsymbol{x}_0$ the exact solution, $\epsilon = 10^{-16}$ the order of float point precision, and $\text{cond}(\mathbf{S})$ the condition number of the matrix $\mathbf{S}$.

1. The condition number is the ratio between the biggest eigenvalue in absolute value and that of the smallest
2. A problem with a high condition number is said to be ill-conditioned

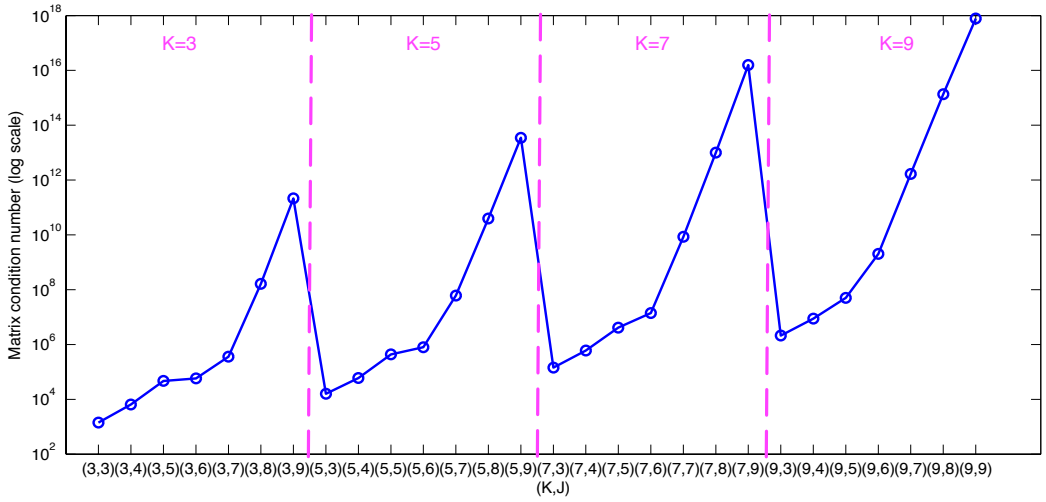


Figure 7.7: Example of the condition number of the matrix $\widehat{\Sigma'_{\boldsymbol{b}}}$ with increasing number of channels (the number of *steerable* channels is K and the number of *scale-shiftable* channels is J) for the correlated Gaussian background. The test is realized with a parameter setup identical to Figure 7.5

7.3.2.3 Tikhonov regularization method

The empirical estimate matrix $\widehat{\Sigma'_{\boldsymbol{b}}}$ is by construction (cf. Eq.(7.31)) non-negative definite and allows the following singular value decomposition (SVD) :

$$\widehat{\Sigma'_{\boldsymbol{b}}} = \mathbf{V}\mathbf{D}\mathbf{V}^{\mathrm{t}}, \tag{7.34}$$

where $\mathbf{V}$ is an orthonormal matrix containing normalized eigenvectors ($\mathbf{V}\mathbf{V}^{\mathrm{t}} = \mathbf{I}$), and $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_p)$ is a diagonal matrix whose main diagonal contains eigenvalues in the descending order $d_1 \geq \ldots \geq d_p > 0$. We recall that the condition number is the ratio d_1/d_p. Its direct inverse writes thus $\widehat{\Sigma'}^{-1} = \mathbf{V}\mathbf{D}^{-1}\mathbf{V}^{\mathrm{t}}$.

The Tikhonov regularization [126] is particularly adapted to the evaluation of moderate dimension ill-conditioned problems such as Eq.(7.30) (a few channels are necessary to enter the ill-conditioned zone) and provides a direct closed-form solution. Instead of calculating Eq.(7.30) by direct inverse as the solution of a linear problem :

$$\widehat{\Sigma'_{\boldsymbol{b}}} \cdot \boldsymbol{w} = \widehat{\boldsymbol{x}'_0}, \tag{7.35}$$

a bias is introduced deliberately in the solution research by considering the following problem :

$$\min\left\{\frac{1}{2}\|\widehat{\Sigma'_{\boldsymbol{b}}}\cdot\boldsymbol{w}-\widehat{\boldsymbol{x}'_0}\|^2+\frac{1}{2}\eta\|\boldsymbol{w}\|^2\right\}, \tag{7.36}$$

for which the positive coefficient η controls the regularization level of the expected solution.

In the SVD point of view, the regularization method can be applied to replace $\left(\widehat{\Sigma'_{\boldsymbol{b}}}\right)^{-1}$ in Eq.(7.30) with

$$\boldsymbol{w}^{\dagger}=\Sigma^{\dagger}\widehat{\boldsymbol{x}'_0} \tag{7.37}$$

$$\Sigma^{\dagger}=\left(\widehat{\Sigma'}^{\mathrm{t}}\widehat{\Sigma'}+\eta\mathbf{I}\right)^{-1}\widehat{\Sigma'}^{\mathrm{t}}=\mathbf{V}\begin{bmatrix}\frac{d_1}{\eta+d_1^2} & & \\ & \ddots & \\ & & \frac{d_p}{\eta+d_p^2}\end{bmatrix}\mathbf{V}^{\mathrm{t}} \tag{7.38}$$

by choosing $d_1^2\gg\eta\gg d_p^2$ for the ill-conditioned cases. We note that for cases with relatively small condition numbers, the regularization has negligible effects by imposing $d_1^2\gg\eta$: $\Sigma^{\dagger}$ is close to the direct inverse of $\widehat{\Sigma'_{\boldsymbol{b}}}$ and thus the detection performances should not change significantly. On the contrary, for ill-conditioned cases the presence of η avoids the division by zero with the eigenvalues such as d_p.

Although the CJO was used to illustrate the principle, other numerical observers involving an ill-conditioned matrix inversion are equally subject to the phenomenon. The Tikhonov regularization method is then adopted for all our proposed novel numerical observers hereafter.

In the Section 7.3.3, the detection and estimation performances of the CJO using the Tikhonov regularization method, for both correlated Gaussian background and real clinical background, are presented.

7.3.3 Results

For both the NRMSE and the AUC, different signal shape parameters are considered: $\sqrt{b}=1.5,2,3,4$. A bigger value of b means the elliptical signal gets a more elongated shape.

The results with the correlated Gaussian background for $\sqrt{b}=1.5,2,3,4$ are shown in Figure 7.8, Figure 7.9, Figure 7.10 and Figure 7.11, respectively. The results with the real clinical background (white matter regions) for $\sqrt{b}=1.5,2,3,4$ are shown in Figure 7.12, Figure 7.13, Figure 7.14 and Figure 7.15, respectively.

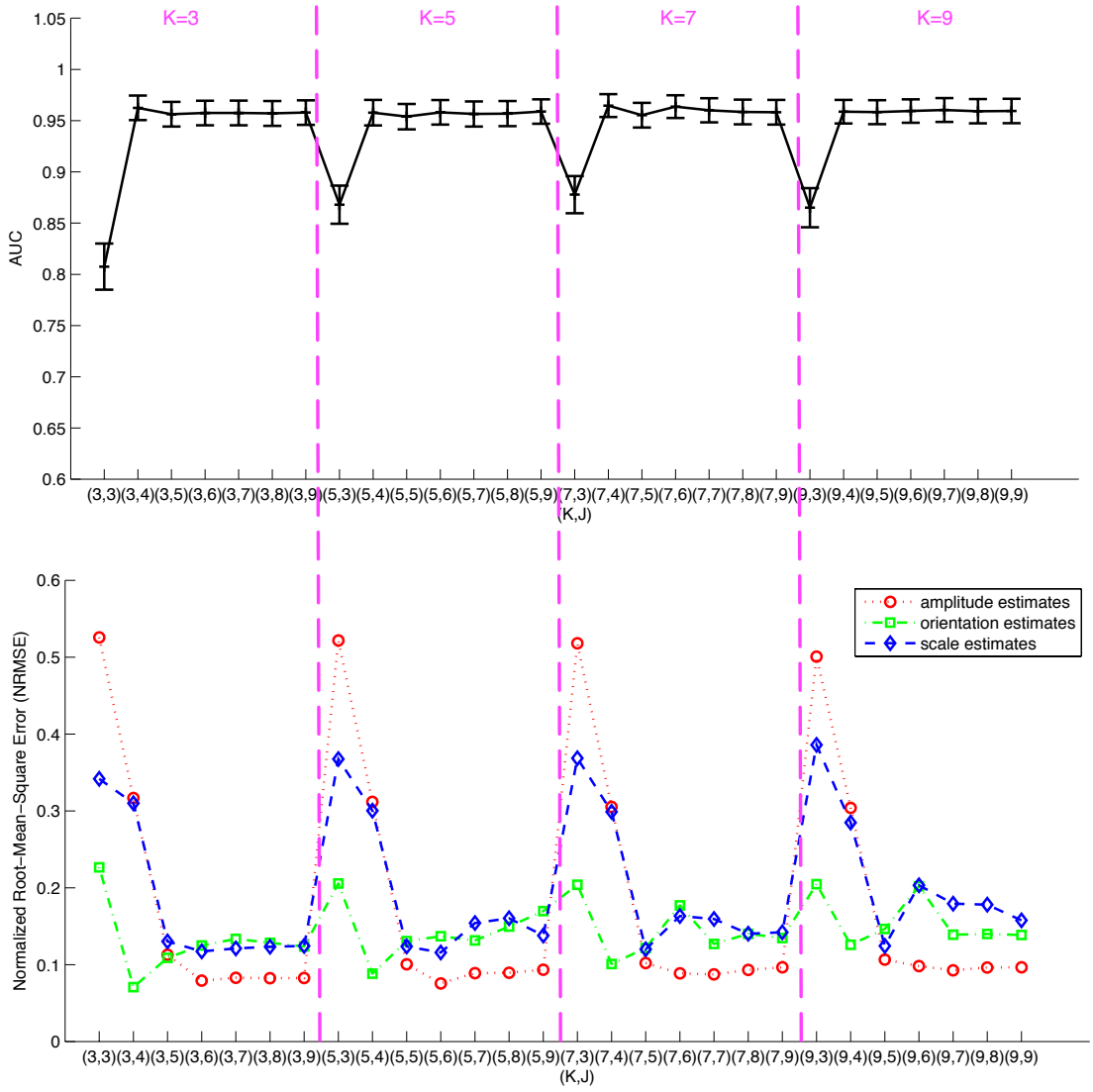


Figure 7.8: Plots of NRMSE and AUC as a function of the number of *steerable* channels K and the number of *scale-shiftable* channels J for correlated Gaussian background when the signal shape parameter $\sqrt{b} = 1.5$, using Tikhonov regularization method.

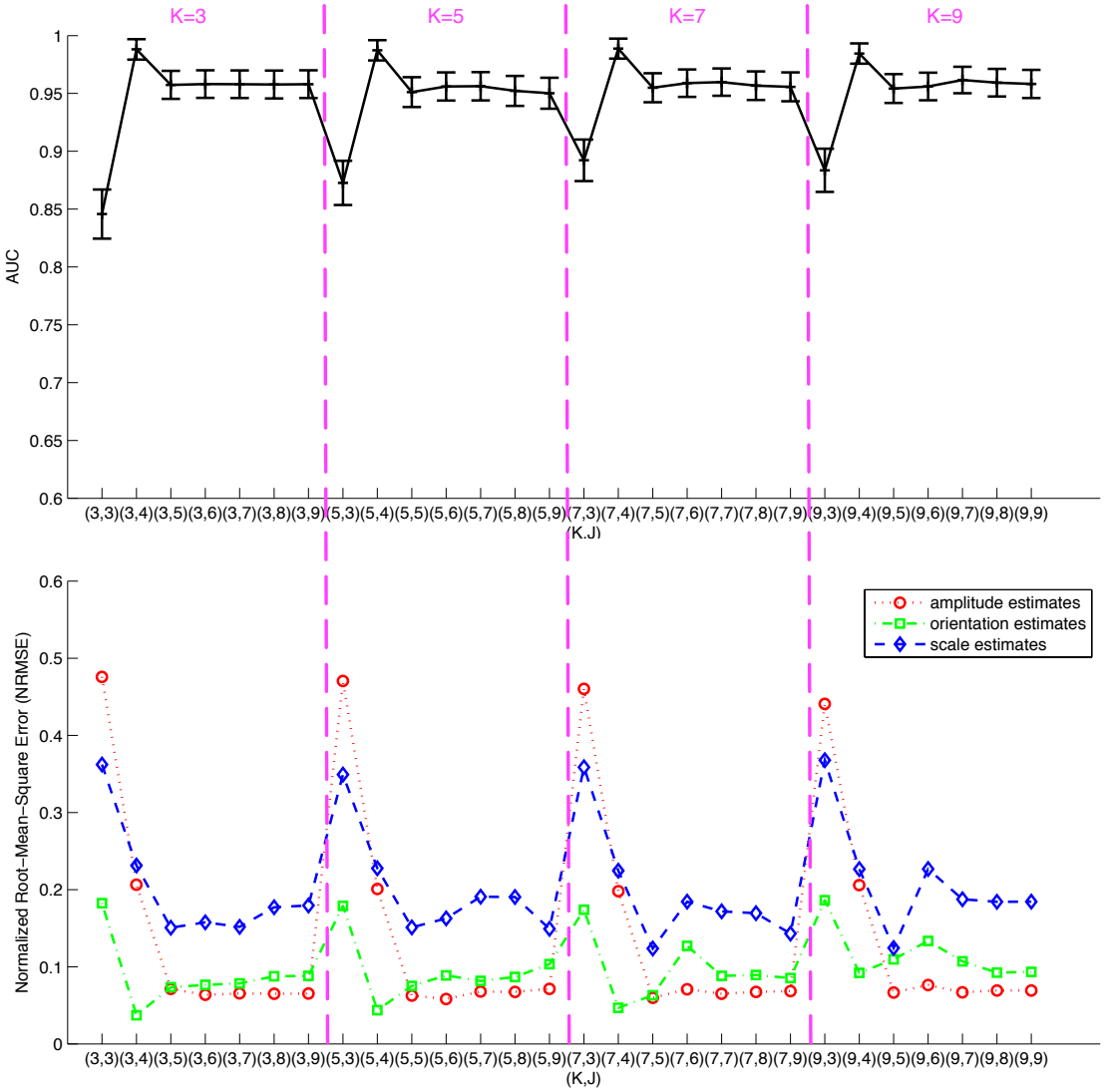


Figure 7.9: Plots of NRMSE and AUC as a function of the number of *steerable* channels K and the number of *scale-shiftable* channels J for correlated Gaussian background when the signal shape parameter $\sqrt{b} = 2$, using Tikhonov regularization method.

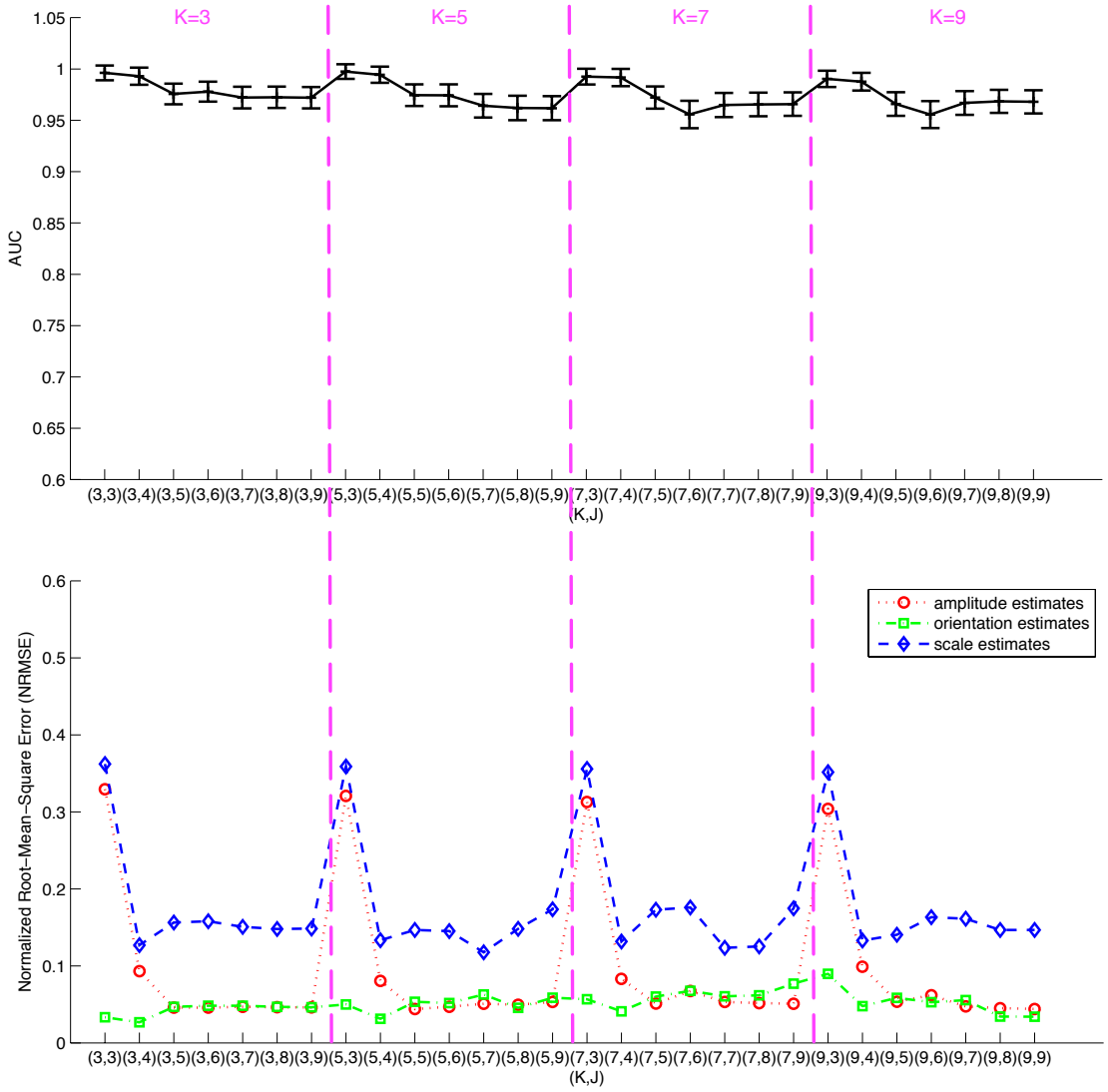


Figure 7.10: Plots of NRMSE and AUC as a function of the number of *steerable* channels K and the number of *scale-shiftable* channels J for correlated Gaussian background when the signal shape parameter $\sqrt{b} = 3$, using Tikhonov regularization method.

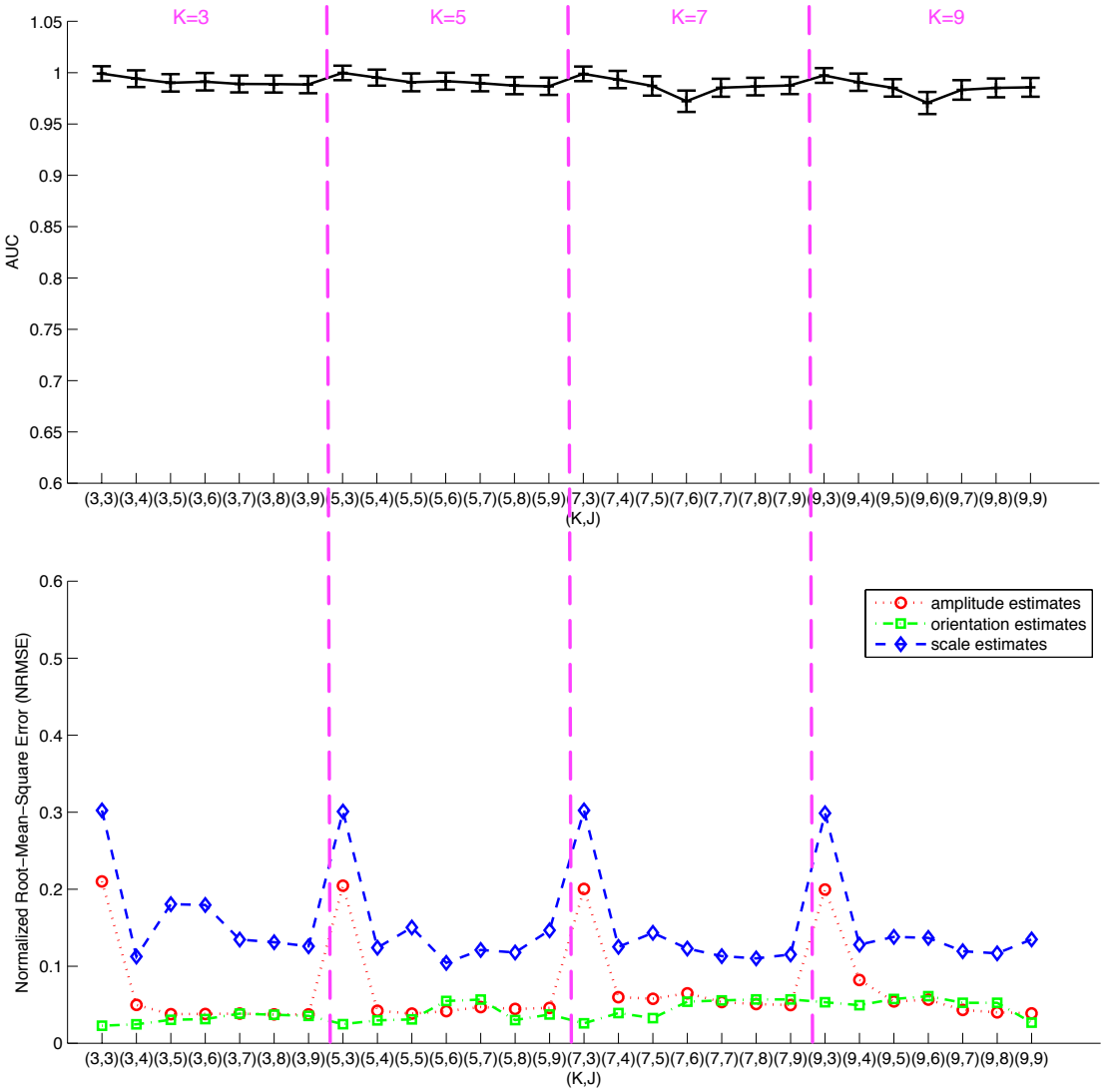


Figure 7.11: Plots of NRMSE and AUC as a function of the number of *steerable* channels K and the number of *scale-shiftable* channels J for correlated Gaussian background when the signal shape parameter $\sqrt{b} = 4$, using Tikhonov regularization method.

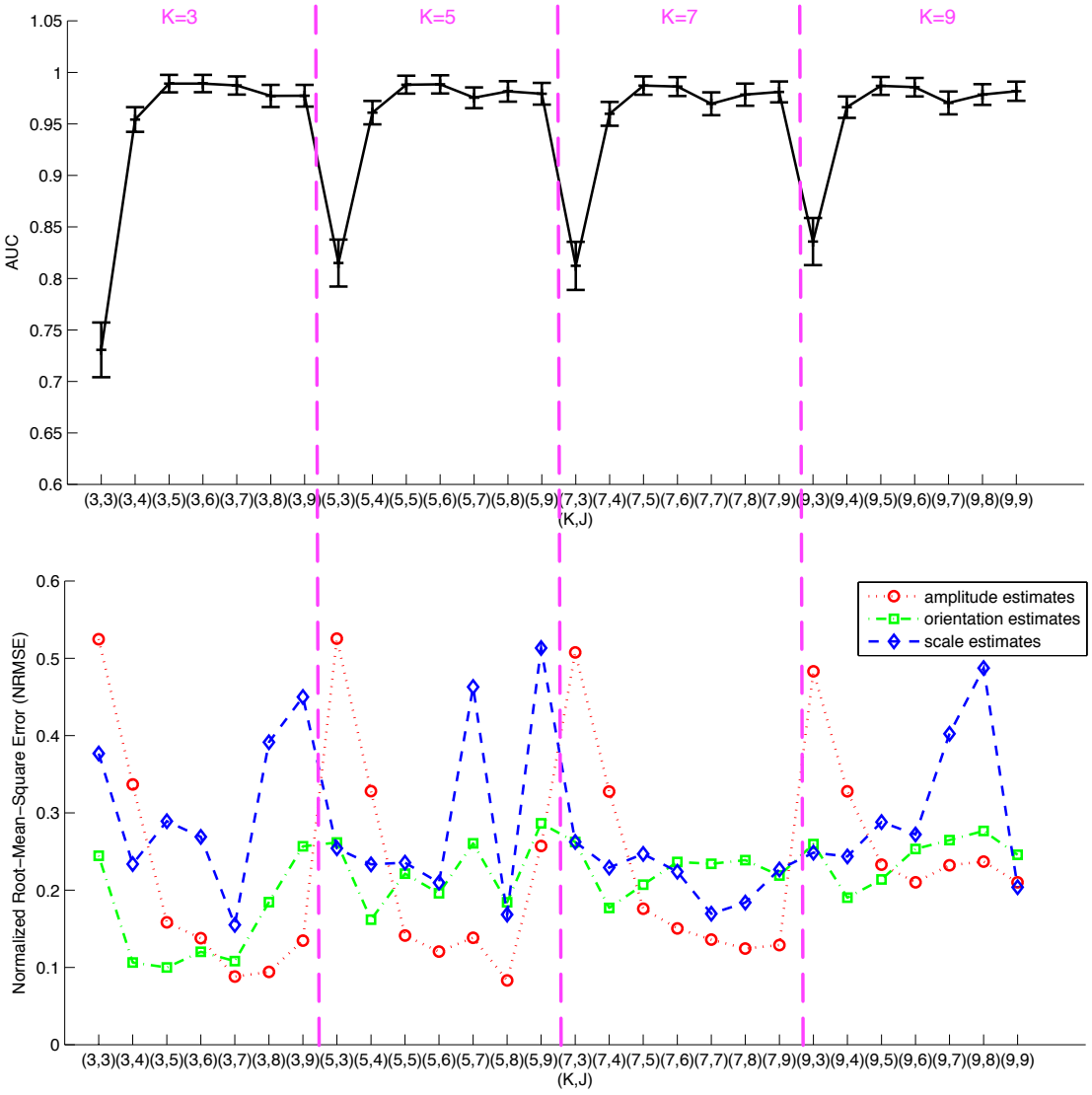


Figure 7.12: Plots of NRMSE and AUC as a function of the number of *steerable* channels K and the number of *scale-shiftable* channels J for real clinical background when the signal shape parameter $\sqrt{b} = 1.5$, using Tikhonov regularization method.

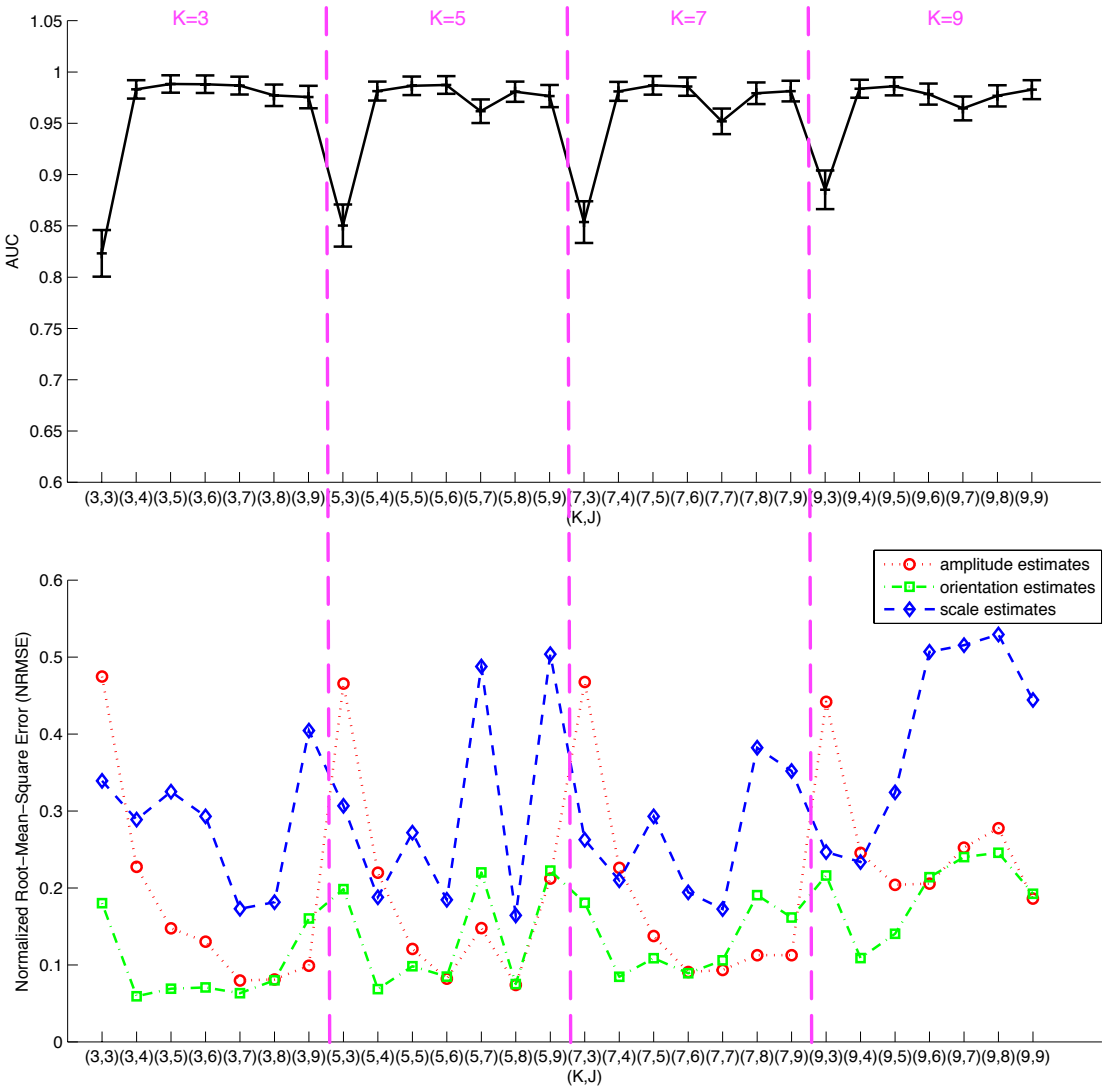


Figure 7.13: Plots of NRMSE and AUC as a function of the number of *steerable* channels K and the number of *scale-shiftable* channels J for real clinical background when the signal shape parameter $\sqrt{b} = 2$, using Tikhonov regularization method.

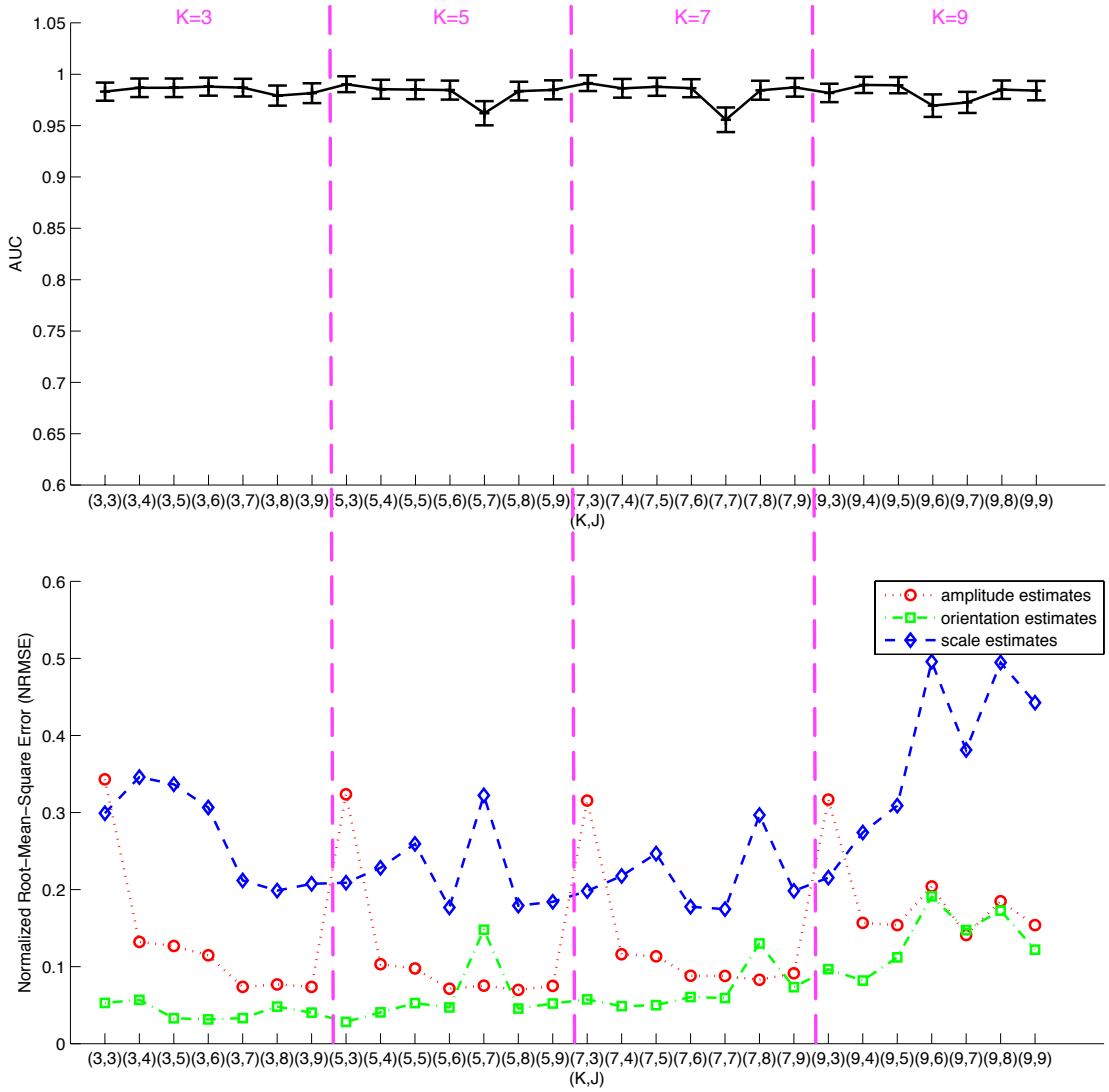


Figure 7.14: Plots of NRMSE and AUC as a function of the number of *steerable* channels K and the number of *scale-shiftable* channels J for real clinical background when the signal shape parameter $\sqrt{b} = 3$, using Tikhonov regularization method.

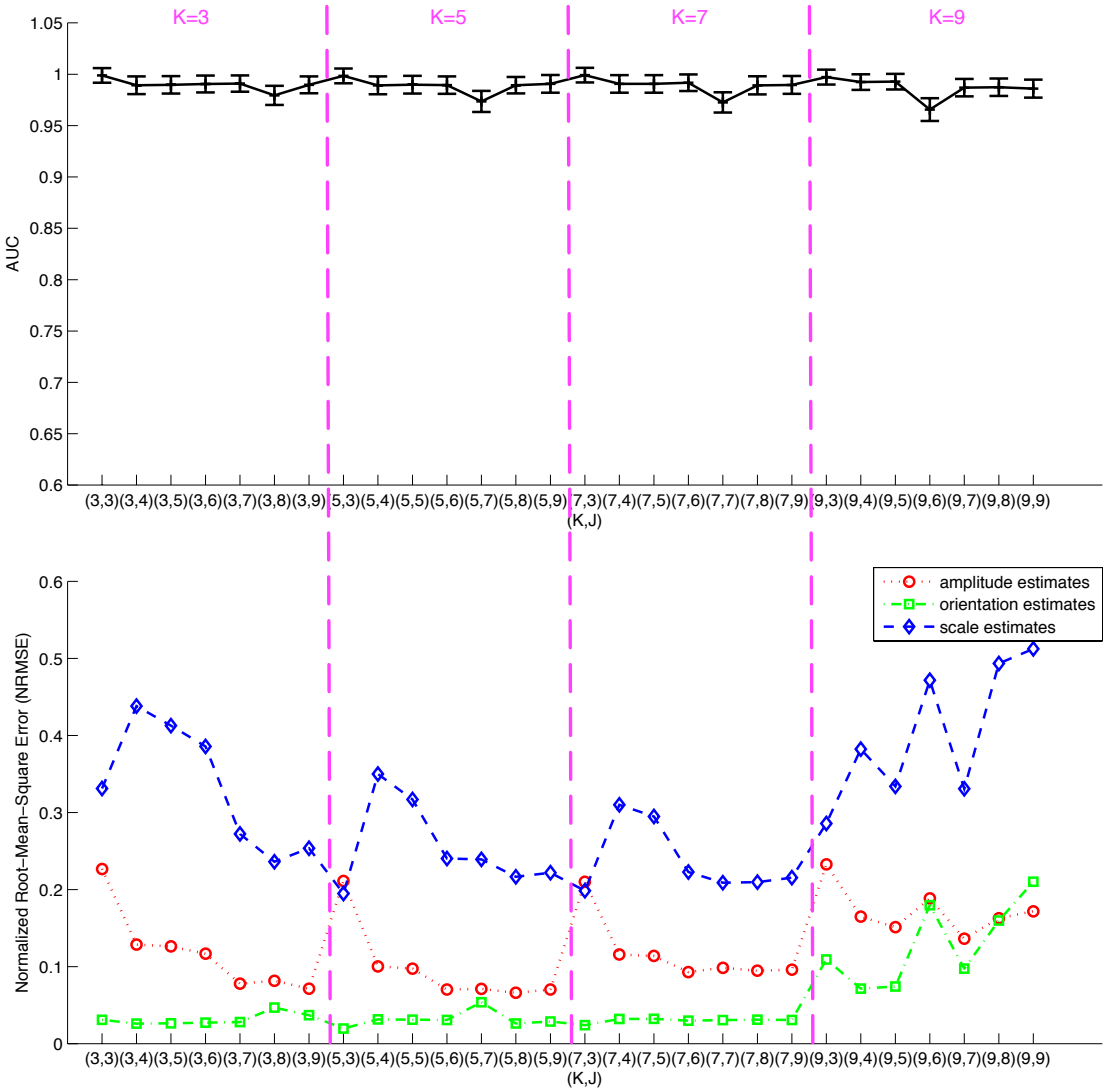


Figure 7.15: Plots of NRMSE and AUC as a function of the number of *steerable* channels K and the number of *scale-shiftable* channels J for real clinical background when the signal shape parameter $\sqrt{b} = 4$, using Tikhonov regularization method.

7.3.4 Discussion

By comparing Figure 7.9 to Figure 7.5, and Figure 7.13 to Figure 7.6, we can see that after using Tikhonov regularization method, the CJO's performances are non-degrading (monotonous) while increasing the number of channels. Thus this result confirms that Tikhonov regularization method can resolve the ill-conditioned matrix inversion problem, the main cause of the numerical instability. It is important to achieve stable performance results with respect to the number of channels for a given dataset such that the choice of optimal parameters should not be influenced by numerical instability. In that sense, the Tikhonov regularization method achieves that goal by obtaining a stable asymptotic performance that characterizes the upper limit of linear observers.

Throughout all the results, we observe that:

- Contrary to the number of *scale-shiftable* channels J, the number of *steerable* channels K does not influence a lot the detection and estimation performance. The detection performance tends to be stable when J reaches 4 or 5. Since fewer number of channels requires fewer calculation complexities, the optimal channel numbers could be $(K, J) = (3, 4)$ or $(3, 5)$ here for this given dataset.
- In general, more important estimation errors (bigger NRMSE values) lead to lower values of AUC, viz. worse detection task performance (e.g., $(K, J) = (3, 3), (5, 3), (7, 3), (9, 3)$).
- Among different parameters, the CJO estimates better the signal orientation than the signal amplitude and scale. And low orientation estimation error helps to gain a really good detection task performance ($AUC > 0.95$), even though amplitude and scale estimation errors remain higher (cf. Figure 7.10, Figure 7.11, Figure 7.14 and Figure 7.15). Ultimately the orientation estimation seems to have the greatest effect on the detection performance among the three unknown signal parameters.
- For both types of background, the CJO's estimation and detection performances are better for more elongated signal (bigger value of b).
- The CJO performs better on the correlated Gaussian background than on the real clinical background. This is understandable considering the background model used in the CJO is indeed the correlated Gaussian background and the complexity of the problem in real clinical situations.

7.4 Conclusion

In this chapter, we proposed a novel SKS MO - Channelized Joint detection and estimation Observer (CJO), that treats signal with variable amplitude, orientation and scale. Compared to the

few existing SKS MO, the CJO extends the range of variable signal parameters. The preliminary performance results show that the detection task performances of the CJO are desirable both for the correlated Gaussian background and the white matter background of MR images of T2 FLAIR sequence.

In the meantime, we presented in this section the idea of using Tikhonov regularization method to address the ill-conditioned matrix inversion problem, the main cause of the numerical instability. This can be generalized to all numerical observers requiring the inverse of an ill-conditioned matrix.

8

PCJO for detection-localization task on single-slice

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 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 [120]. 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 [122]. Thus, it is impossible to design channels that are shiftable in position and in scale at the same time. Moreover, localization task of multiple signals on an image is more challenging .

However, in order to represent more realistically the routine diagnostic process, the localization task is highly important, especially if the signal locations have significantly impact on the subsequent therapeutic interventions (e.g. biopsy or surgery).

To solve this problem, we propose a novel Perceptually relevant CJO (PCJO) by taking advantage of the state-of-the-art in HVS modeling and our CJO. The only known signal parameter for the PCJO is the signal shape. The PCJO also involves the perception stage to some extent.

To validate the task performance of the PCJO compared to that of human observers, we also designed a free-response subjective experiment in this chapter.

This work has been published in the journal paper [127].

8.1 Structure diagram of the PCJO

The structure diagram of the PCJO is shown in Figure 8.1. The PCJO consists of two parts:

1. Candidate selection (outlined by a red dashed line in Figure 8.1) 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);
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);

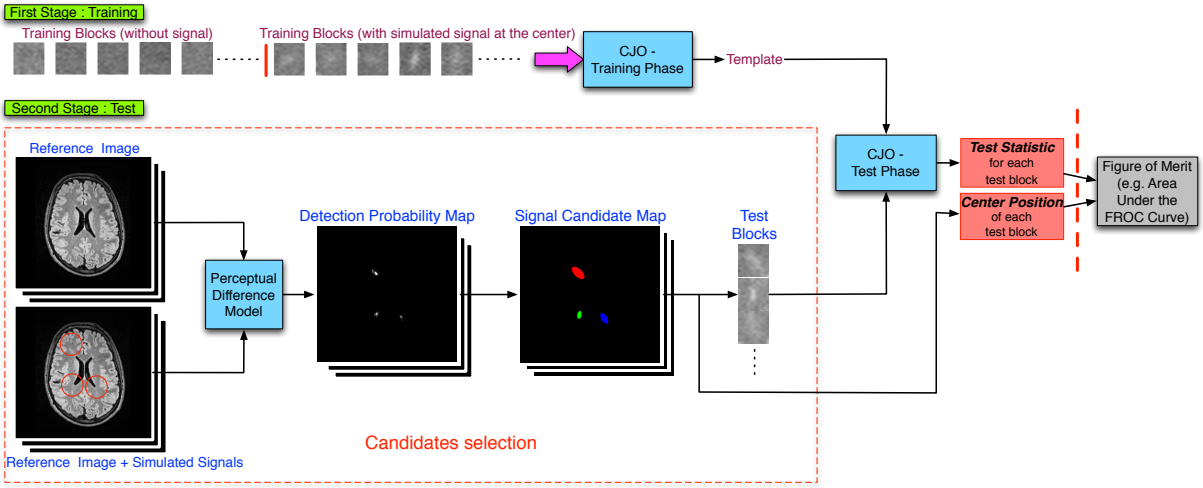


Figure 8.1: The structure diagram of the proposed Perceptually relevant Channelized Joint Observer (PCJO).

– 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.

8.1.1 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 [115] showed that the Visible Differences Predictor (VDP) [79] 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 8.3). 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 4.1. 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.(3.3)) as *Distorted image*, as Jackson did in [76]. Thus, the VDP actually predicts the probability that the signal is visible in radiologists' perceptual domain. Remind the different steps in VDP models: 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.

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:

$$b_{ij} = \begin{cases} 1 & \text{if } p_{ij} \geq T_p \\ 0 & \text{else} \end{cases} \tag{8.1}$$

   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. Eliminating certain objects by thresholding the number of pixels for each labeled object, such as $\pi\sqrt{b}\sigma^2_{\text{min}}$.

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 [1] $\sqrt{M} \times \sqrt{M}$ centered around each object center is extracted from the distorted image. These test blocks are the inputs for the test stage of the CJO, and the center positions are saved as the localization response of the PCJO.

1. This size has to be adapted to the lesion size of the studied pathology.

8.1.2 Application of the CJO on candidates

Our last study [115] also showed that there is still a gap between the VDP's performance and radiologists' *perception* performance. One important reason is 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.

The outputs of the PCJO are the test statistics and the center positions of the test blocks. Then an FROC or AFROC FOM could be used to characterize the PCJO detection-localization task performance.

8.2 Comparisons of PCJO with other numerical observers

Table 8.1 gives a comparison of our PCJO with several existing numerical observers. We can see that the PCJO is the only one which tackles both the detection and localization task problems among the five listed numerical observers. The PCJO needs least signal information and integrates most HVS features. In addition, it allows multiple signals on the image.

8.3 Human observer study for performance evaluation

8.3.1 Studied pathology

To study the PCJO task performance, we selected MS as the studied pathology. As said in Section 1.1.2, the main reason is that for this particular pathology, orientation, size, position and number of lesions are all considered as diagnostic criteria [17], 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 [18]. 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.

Table 8.1: Comparison of our PCJO with existing model observers.

	Task	Signal information	Allowed # signals per image	Included HVS features
CHO [52]	Detection	Full knowledge about all the signal parameters	one	Frequency and orientation selectivity of HVS
Eckstein's model [57, 58, 59]	Detection	Signal size and shape are unknown, while other signal parameters are known exactly	one	Frequency and orientation selectivity of HVS
Goossens's SKS CHO [67]	Detection	Signal orientation is unknown, while other signal parameters are known exactly	one	Frequency and orientation selectivity of HVS
Park's model [15]	Localization	Signal position is unknown, while other signal parameters are known exactly	one	Frequency and orientation selectivity of HVS
PCJO	Detection & localization	Signal amplitude, orientation, scale, position are unknown, while signal shape is known exactly	multiple	Many HVS characteristics: nonlinear response of photoreceptors to luminance; variations in visual sensitivity as a function of spatial frequency; spatial frequency & orientation selectivity; visual masking properties; visual psychometric function.

8.3.2 Experimental images

From a retrospective database of the 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 after the study, 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 per pixel. 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 [128]. We used Eq.(3.3) to model the MS lesions on this sequence, since our precedent study [113] 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.

8.3.3 Human observer study

We conducted a free-response subjective experiment [129] in order to assess human performances.

8.3.3.1 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.

8.3.3.2 Experimental protocol

According to [130], 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

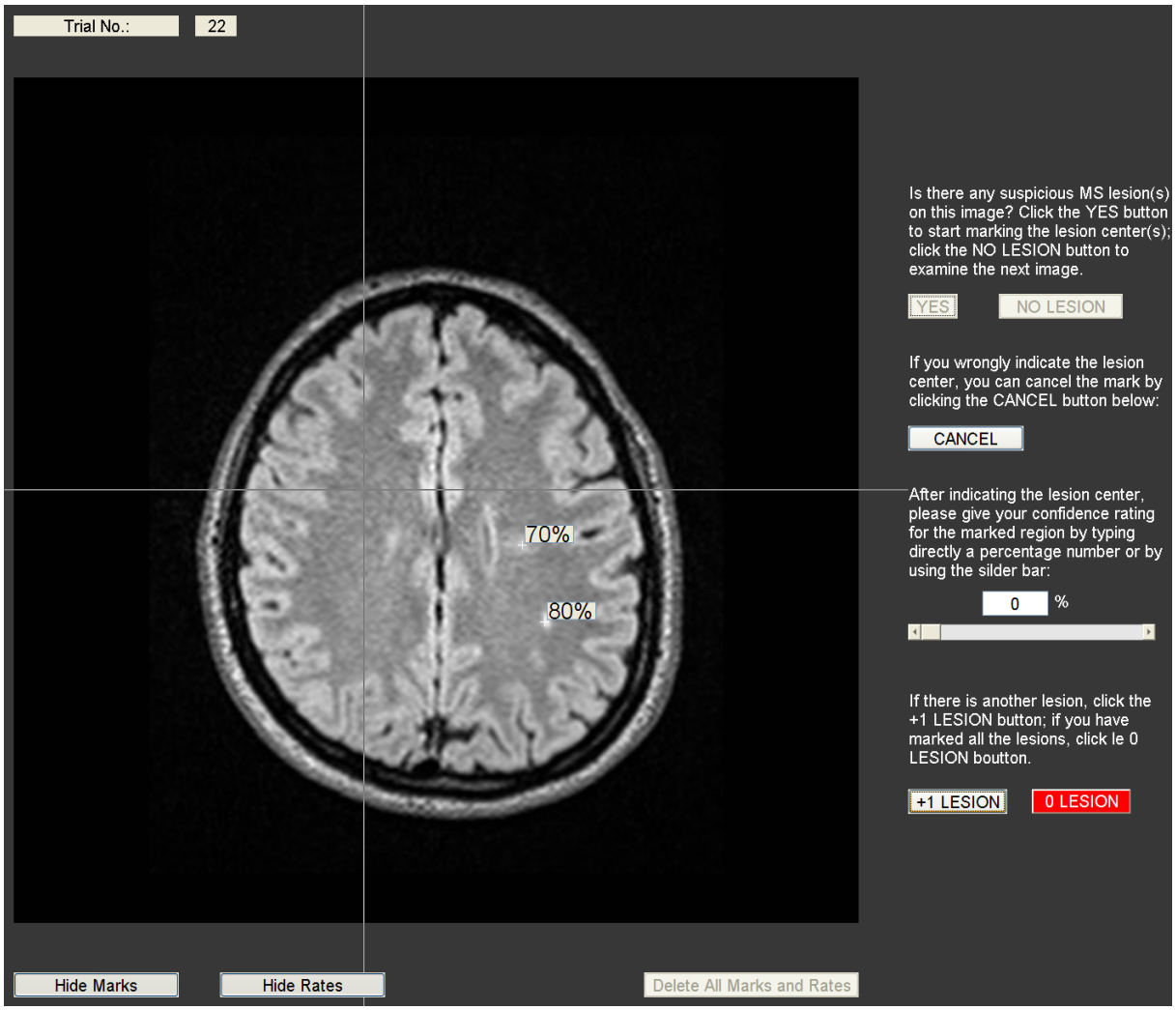


Figure 8.2: GUI of the free-response subjective experiment for the performance evaluation of the PCJO.

(same as the radiology reading environment at the Hospital of 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 (grey level=55, which had been adjusted by the same neuroradiologist who did not participate in the experiment) 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). They

were even not told the maximum possible number of lesions on each image (only knew the number of lesions ≥ 0). The experiment was conducted in a controlled environment, thus the readers could not modify the images.

8.3.3.3 Experiment's graphical user interface (GUI)

For this experiment, we developed a special GUI (shown in Figure 8.2).

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 lesion 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 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.

8.3.4 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 images with lesions, read by radiologists in our free-response subjective experiment (cf. Section 8.3.2), fed the VDP as distorted images.

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.(8.1) was set to 0.9.

The training blocks for the CJO were constructed by extracting white matter regions from the stacks 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 our experiment, 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 $[\sigma_{\text{min}},\ \sigma_{\text{max}}] = [1,\ 6]$, and the lesion shape was fixed ($\sqrt{b} = 2$ in this paper). The possible signal amplitude was $[a_{\text{min}},\ a_{\text{max}}] = [1,\ 255]$ (pixel intensity value). The signal orientation range was $[\theta_{\text{min}},\ \theta_{\text{max}}] = [0,\ \pi]$.

8.3.5 Performance evaluation method

The JAFROC1 [44] (cf. Section 2.3.3) 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.

8.4 Results and discussion

8.4.1 PCJO performance w.r.t. radiologists' performances

The JAFROC1 FOMs of the six radiologists are shown in Table 8.2. The DBM-MRMC significance testing is conducted to examine JAFROC1 FOM differences between each pair of radiologists. The significance level is set to $p < 0.05$. The p-values are shown in Table 8.3. Table 8.2 and Table 8.3 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 8.4.3 for more discussions).

Table 8.2: 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

Table 8.3: The p-values from the DBM-MRMC significance testing to examine JAFROC1 FOM differences between each pair of the radiologists. The p-values smaller than 0.05 are highlighted in red. Radiologists 3-6 are denoted by Rad 3, Rad 4, Rad 5 and Rad 6 in this table.

p-values	expert 1	expert 2	Rad3	Rad 4	Rad 5	Rad 6
expert 1						
expert 2	0.1240					
Rad 3	<0.0001	<0.0001				
Rad 4	<0.0001	<0.0001	0.2176			
Rad 5	<0.0001	<0.0001	0.2176	0.2115		
Rad 6	<0.0001	<0.0001	0.9835	0.2254	0.9704	

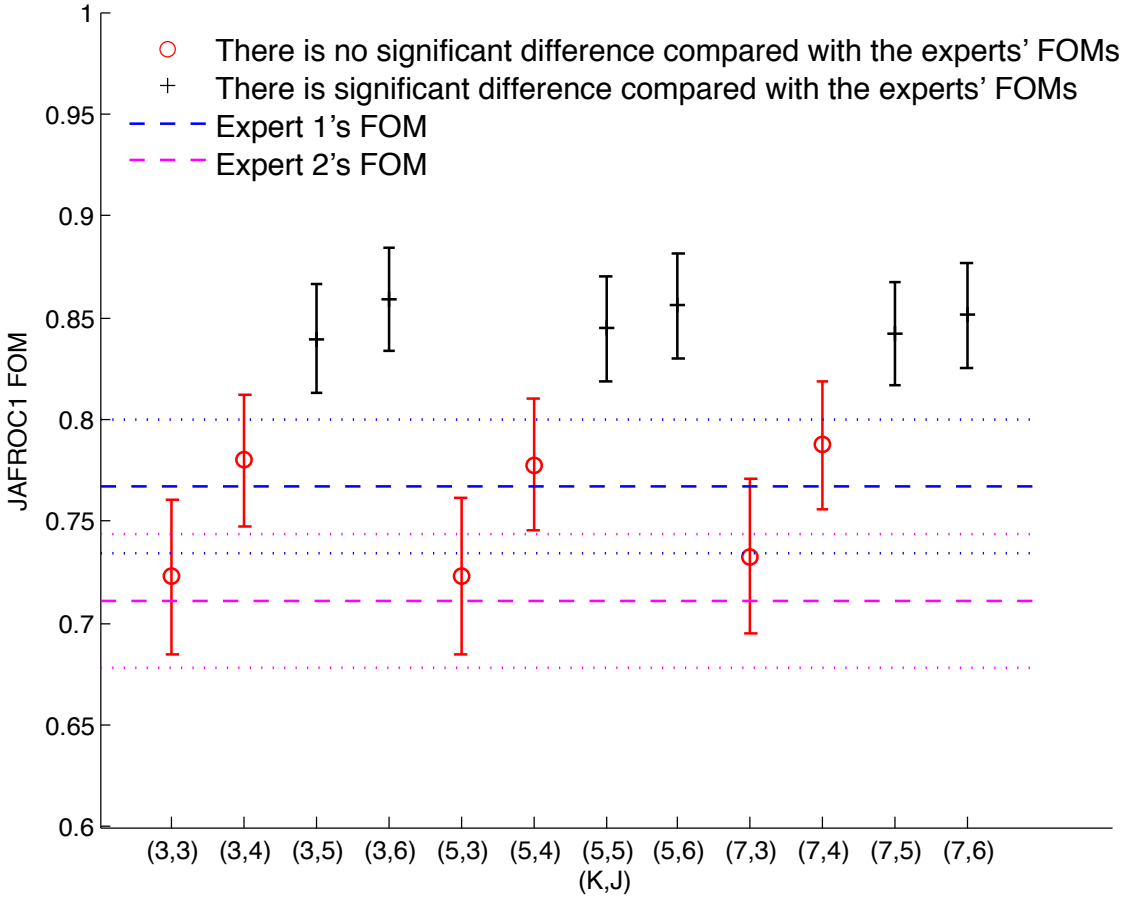


Figure 8.3: JAFROC1 FOMs of the PCJO for different combinations of K and J. The p-values for examining JAFROC1 FOM differences between the PCJO and the experts are also calculated, and those greater than 0.05 are highlighted in red, which means that the corresponding JAFROC1 FOMs of the PCJO are not significantly different from those of the experts.

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 8.3. 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.

8.4.2 PCJO performance w.r.t. 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. Chapter 7) 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 8.1: the CHO [52] and Goossens's SKS CHO [67]. 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 8.1.1 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 8.4 and Table 8.5, 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 8.4: 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	9	0.8586	0.0265
11	0	2	21	0.8484	0.0273
11	$\pi/4$	6	3	0.8676	0.0242
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
11	$\pi/2$	10	21	0.8434	0.0262
51	0	6	3	0.8675	0.0242
51	0	6	9	0.8644	0.0248
51	0	6	21	0.8621	0.0251
51	$\pi/4$	2	3	0.8643	0.0253
51	$\pi/4$	2	9	0.8564	0.0269
51	$\pi/4$	2	21	0.8475	0.0277
51	$\pi/2$	10	3	0.8656	0.0244
51	$\pi/2$	10	9	0.8536	0.0260
51	$\pi/2$	10	21	0.8485	0.0264
131	0	10	3	0.8656	0.0244
131	0	10	9	0.8536	0.0260
131	0	10	21	0.8488	0.0264
131	$\pi/4$	2	3	0.8639	0.0254
131	$\pi/4$	2	9	0.8559	0.0269
131	$\pi/4$	2	21	0.8478	0.0277
131	$\pi/2$	10	3	0.8656	0.0244
131	$\pi/2$	10	9	0.8636	0.0260
131	$\pi/2$	10	21	0.8488	0.0264

Table 8.5: 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	2	21	0.8222	0.0287
11	6	3	0.8641	0.0248
11	6	9	0.8613	0.0251
11	6	21	0.8576	0.0249
11	10	3	0.8626	0.0249
11	10	9	0.8606	0.0251
11	10	21	0.8587	0.0248
51	2	3	0.8523	0.0277
51	2	9	0.8585	0.0254
51	2	21	0.8400	0.0261
51	6	3	0.8661	0.0248
51	6	9	0.8657	0.0246
51	6	21	0.8567	0.0251
51	10	3	0.8651	0.0247
51	10	9	0.8643	0.0248
51	10	21	0.8515	0.0256
131	2	3	0.8562	0.0272
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
131	6	21	0.8644	0.0244
131	10	3	0.8662	0.0245
131	10	9	0.8625	0.0252
131	10	21	0.8604	0.0250

8.4.3 Discussion

It will be interesting to further investigate the radiologists' results (cf. Table 8.2). 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 8.6, 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 8.6, 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 [113]. 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-signal, 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 reading 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 8.6: 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.

8.5 Conclusion

In this chapter, we have proposed a novel numerical observer called the Perceptually relevant Channelized Joint Observer (PCJO). To our best knowledge, 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 might present on one image, and the number of lesions is unknown.

Compared to other numerical observers in the literature,

- the PCJO greatly extends the range of variable signal parameters by extending the SKS MOs;
- the PCJO considers the perception stage of the HVS using a training process;
- the PCJO addresses the difficulty of multiple-signal localization 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 performances of radiologists. We then used the JAFROC1 FOM to evaluate the PCJO performance. We found that for the studied images (axial MRI slices of the T2 FLAIR sequence), there is not a significant difference between the PCJO performance and human performance. We also compared the PCJO with other mathematical observers using the 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.

We conclude that 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.

9

msPCJO for detection-localization task on multi-slice

Three-dimensional (3D) volumetric visualization of MRI and CT data of the brain, the internal organs and the spine has become the standard for routine patient diagnostic care [131]. Compared to single-slice image, volumetric (multi-slice) image provides more anatomical information, which allows for a better distinction between true lesions and noise or background structure.

Thus radiologists would very probably change their behavior and perform differently when they could make their diagnoses on multiple slices (viewing a stack in its entirety by scrolling back and forth). During our subjective experiment for the validation of the proposed numerical observer PCJO (cf. Section 8.3), we noticed that without adjacent slices, radiologists less experienced in MS image reading often mistook cerebral cortex for lesion. That might explain why Radiologist 3-6 made a lot of FP marks (cf. Table 8.6). Two concrete examples are shown in Figure 9.1, where radiologists marked the hyper-signal in the red circle as a lesion candidate, whereas they considered the hyper-signal as the cerebral cortex in the presence of the adjacent slice.

This discovery inspires us to further investigate the multi-slice diagnoses and explore multi-slice (or 3D) numerical observers which have a potential to be more clinically relevant.

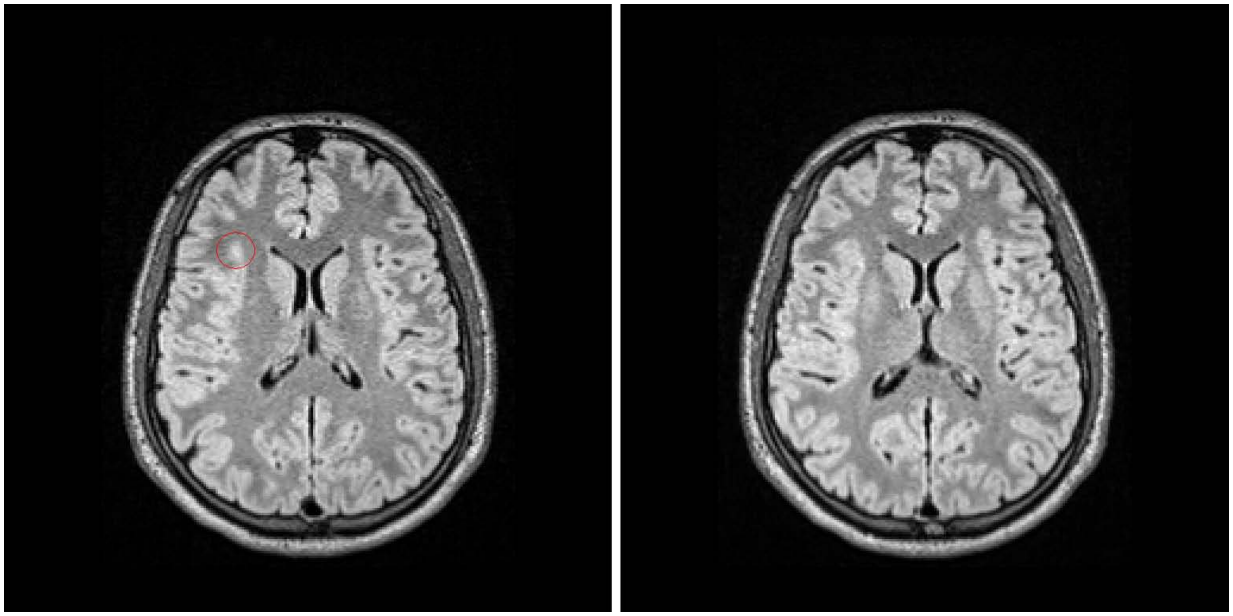
In the literature, multi-slice numerical observers have been proposed in the framework of MOs, but not yet in the frame of HVS models. According to the overview on the existing multi-slice MOs in Section 3.4, we remark that:

1. existing multi-slice MOs perform only SKE detection task;
2. existing multi-slice MOs detect only symmetrical signal;
3. the signal is always located at the centre of the volumetric (or 3D) image.

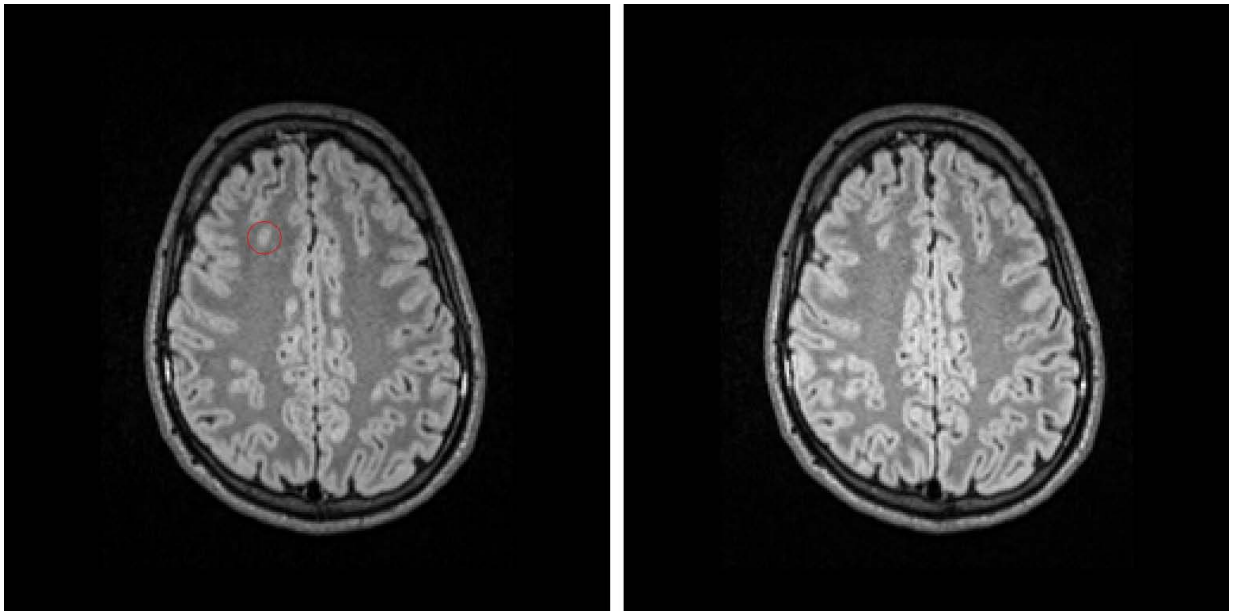
In this chapter, we will extend the PCJO to a multi-slice numerical observer (Section 9.1). We introduce the asymmetrical signal and the SKS detection task into the multi-slice numerical observer, capable of estimating signal parameters (location, amplitude, orientation and size). In addition, we will conduct a multi-slice subjective experiment in which the adjacent slices will be shown to the radiologists for the performance comparison (Section 9.2).

Figure 9.1: Two examples of two adjacent cerebral slices: the red circles indicate where the cerebral cortex may be confused with a lesion if only the left slice is provided (single-slice image).

(a)



(b)



9.1 Multi-slice PCJO (msPCJO)

9.1.1 3D signal model

To add simulated MS lesion on volumetric MR images, firstly we construct a 3D signal model based on the 2D Gaussian signal model in Eq.(3.3):

$$[\boldsymbol{x}]_{\boldsymbol{p}} = a \, \exp\left(-\frac{1}{2} \left(\boldsymbol{p}-\boldsymbol{q}\right)^{\mathrm{t}} \mathbf{B}^{\mathrm{t}} \mathbf{D}^{-1} \mathbf{B} \left(\boldsymbol{p}-\boldsymbol{q}\right)\right) \tag{9.1}$$

where $[\boldsymbol{x}]_{\boldsymbol{p}}$ denotes the intensity value of the added signal at the 3-D coordinate $\boldsymbol{p}$. The signal intensity attenuation is hereby modeled by a Gaussian function of peak amplitude a, centered at $\boldsymbol{q}$. Without loss of generality, the diagonal matrix $\mathbf{D}$ in Eq.(9.2) specifies three parameters, respectively the equatorial radii σ_1 and σ_2 (along the x and y axes) and the polar radius σ_3 (along the z-axis), in order to determine the shape of the ellipsoid, whereas the rotation matrix $\mathbf{B}$ given in Eq.(9.3) specifies the spatial orientation transformation of the current ellipsoid w.r.t. the image coordinate system [132]. We used here the x-y-z convention with independant Euler angles ϕ, θ, and ψ. [1]

$$\mathbf{D} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix} \tag{9.2}$$

$$\mathbf{B} = \begin{bmatrix} \cos\theta\cos\psi & -\cos\phi\sin\psi + \sin\phi\sin\theta\cos\psi & \sin\phi\sin\psi + \cos\phi\sin\theta\cos\psi \\ \cos\theta\sin\psi & \cos\phi\cos\psi + \sin\phi\sin\theta\sin\psi & -\sin\phi\cos\psi + \cos\phi\sin\theta\sin\psi \\ -\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta \end{bmatrix} \tag{9.3}$$

The visualization of a 3D signal example for which $\phi = \theta = 0, \psi = \pi/4$ is presented as volumetric slices in Figure 9.2.

1. The rotations are applied firstly around the x-axis by ϕ, secondly around the y-axis by θ and thirdly around the z-axis by ψ. Each basic rotation appears counter-clockwise and the axis about which it occurs points toward the observer.

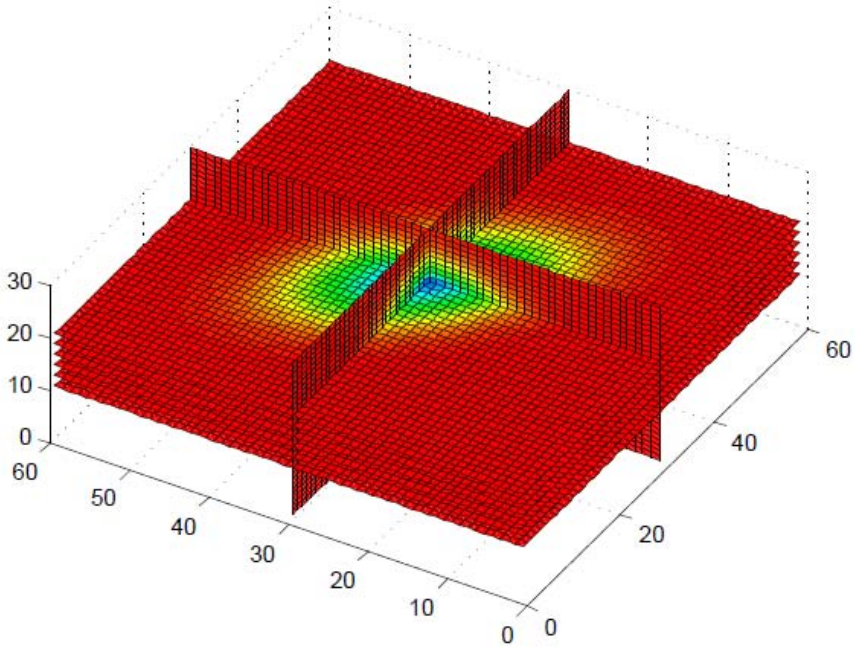


Figure 9.2: The visualization of a 3D signal example as volumetric slices.

9.1.2 Extension of the PCJO to the msPCJO

The promising task performances of the PCJO on 2D images (cf. Section 8.4) encourage us to extend it to the multi-slice case, as a more realistic modeling of the diagnostic process in clinical routine.

The structure diagram of the multi-slice PCJO (msPCJO), shown in Figure 9.3, is similar to that of the PCJO. The msPCJO also consists of two steps:

1. Candidate selection (outlined by red dashed lines in Figure 9.3) to generate several 2D test blocks for each volumetric (3D) image (each volumetric image corresponds to one patient), and then to output the 3D center position of each test block on the image;
2. Application of the 2D CJO on candidates to calculate a test statistic for each test block.

The difference between the msPCJO and PCJO lies in the first step - candidate selection, which will be explained in Section 9.1.2.1.

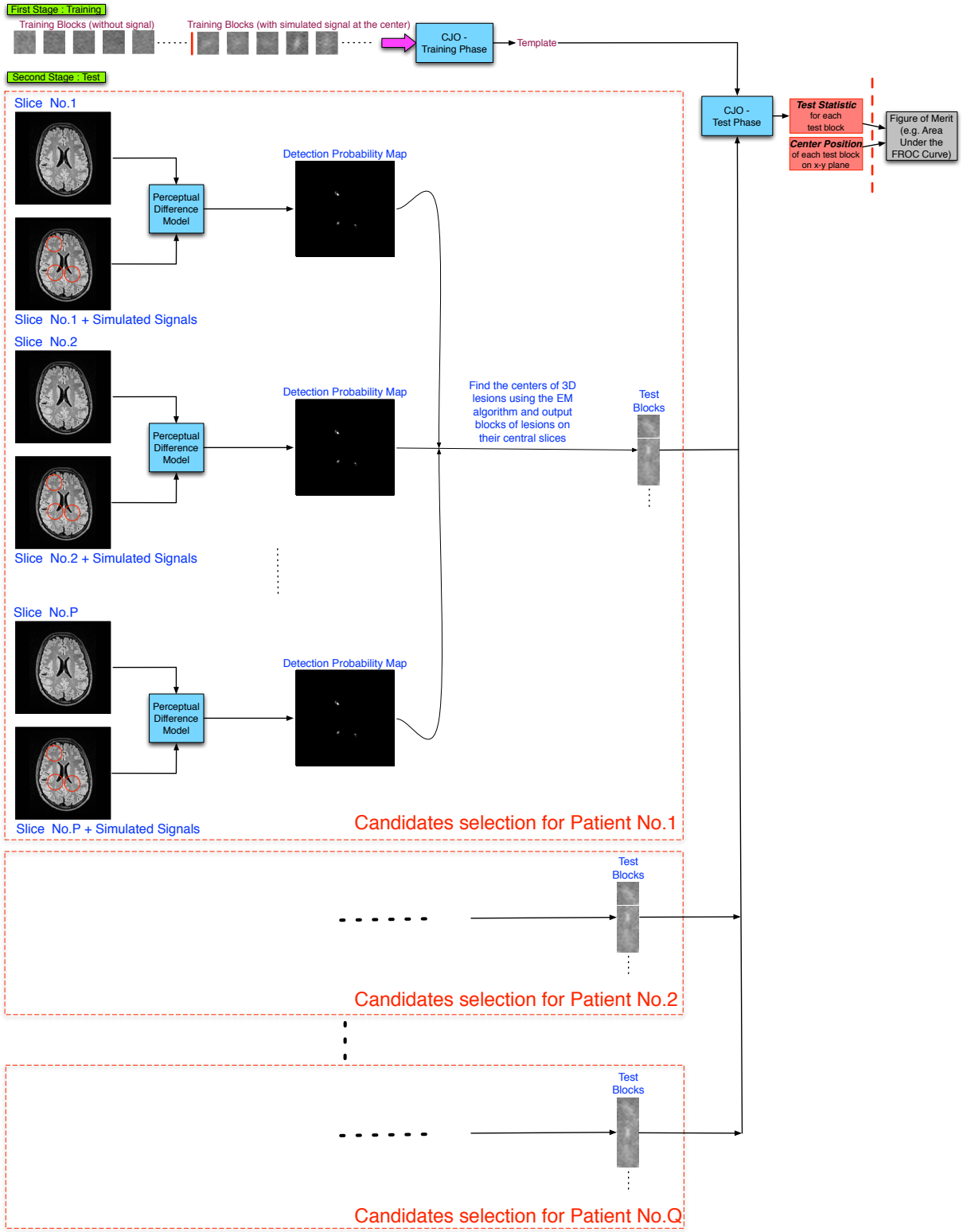


Figure 9.3: Structure diagram of the multi-slice PCJO (msPCJO).

9.1.2.1 Candidate selection in the msPCJO

Many radiologists view the 3D volume first then look at slices, others look at slices directly, depending on the modality and the pathology of interest. In both way, they have to process the volumetric image slice by slice. Thus we assume that radiologists interpret multi-slice images by pre-processing the volumetric image slice-after-slice (in x-y plane) to locate the lesion candidates on each slice and then integrate information across the slices to make the final decisions.

Under this assumption, the VDP (HVS model) is applied and outputs a detection probability map for each slice. 3-D coordinates with detection probability superior to the given threshold T_p ($T_p = 0.9$ in this study) are preselected. These selected coordinates $\{\boldsymbol{x}_n\}_{1,\ldots,N}$ are then readily to be classified based on a Gaussian mixture model (GMM) since the signal intensity is itself a gaussian function (cf. Eq.(9.1)) and the problem here is to inverse the mixture (de-mix) of a linear superposition of unknown numbers of Gaussian distributions. We tested here one of the well-known methods of classification based on the GMM: the expectation-maximization (EM) [133] algorithm.

For a given class number I, the EM algorithm draws confidence ellipsoids (centers and covariances). The Bayesian Information Criterion (BIC) is then computed to assess the number of classes. Indeed, the minimization of the BIC solves the over-fitting problem [2] by using a penalization term on the number of free parameters k (directly related to the number of classes) :

$$\mathrm{BIC} = -2 \sum_{n=1}^{N} \log \left\{ \sum_{i=1}^{I} \pi_i \mathcal{N}(\boldsymbol{x}_n \,|\, \mu_i, \Sigma_i) \right\} + k \log(N) \tag{9.4}$$

9.1.2.2 Application of the CJO in the msPCJO

For each 3D lesion center, a 2D block (of fixed dimension $\sqrt{M} \times \sqrt{M} = 65 \times 65$ in this study) on the corresponding x-y plane is extracted, upon which is applied the 2D CJO algorithm as detailed in the Chapter 7. The above-described procedure needs to be applied on the multi-slice images of several patients to allow for the calculation of the figure of merit.

2. Likelihood tends to infinity if one class is created for each data $\boldsymbol{x}_n$

9.2 Multi-slice subjective experiment for performance evaluation

9.2.1 Subjective experiment

To evaluate the msPCJO's detection-localization task performance, we designed a multi-slice free-response subjective experiment. The studied pathology was MS and the studied modality was MRI, the same as the performance evaluation study for the PCJO (cf. Section 8.3.1).

We used the same 10 healthy subjects' brain MR axial images of the T2 FLAIR sequence as in Section 8.3.2, but we chose 50 contiguous slices among the 3D stack of each subject. The slice thickness is 1mm and the voxel size in the 3D image is 1mm $\times$1mm $\times$1mm. Note that the 90 images (single slices) used in the performance evaluation study for the PCJO (cf. Section 8.3.1) are part of the 500 slices (10 subjects $\times$ 50 contiguous slices/subject). The purpose is to compare observers' task performances in both cases of multi-slice and single-slice.

3D synthetic MS lesions were added on these 3D volumetric images. Note that for the present 3D signal model from Eq.(9.1), a different elliptical hyper signal (as an intersection of the ellipsoid volume and the slice plane) is added upon each 2D slice. To shorten the experiment time, the maximum number of lesions is 15 per patient.

The participants, the experimental protocol and the setup of the VDP and the CJO are also the same as in the performance evaluation study for the PCJO (cf. Section 8.3.3).

The process of the experiment is as follows:

- Before the experiment, radiologists were told that the number of lesions in the stack of each subject is smaller than 15; and there is no lesion on the first and the last slices.
- For each patient, radiologists could always scroll up and down the slices via the up and down arrow keys on the keyboard. The slice number is shown on the right side of the GUI.
- When they detected a first suspicious lesion, they went to the central slice of this lesion and pressed the spacebar on the keyboard. This action made appeared a horizontal line and a vertical line on the image whose point of intersection followed the cursor to indicate the positions on the image, as shown in Figure 9.4.
- Then a dialog box popped up to ask radiologists the confidence rate (number contained between 0 and 100), as shown in Figure 9.5.
- Radiologists repeated the above procedure to mark all the lesions in the 3D stack, and click the "Finish" button to finish the experiment for one patient.
- Radiologists repeated the above procedure for the 10 patients.

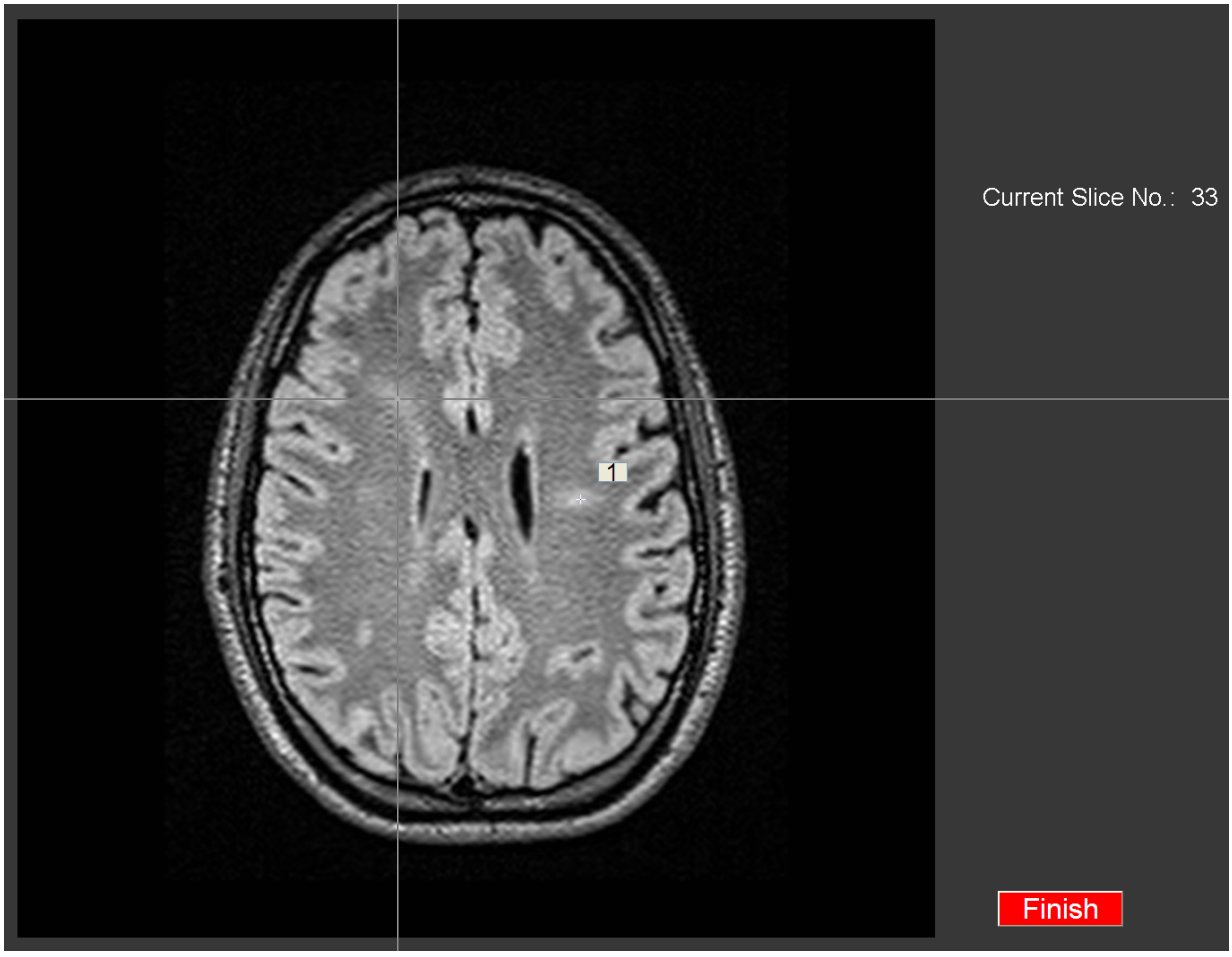


Figure 9.4: GUI of the multi-slice free-response subjective experiment for the performance evaluation of the msPCJO, where a horizontal line and a vertical line appear on the image and their point of intersection followed the cursor to indicate the lesion positions on the image.

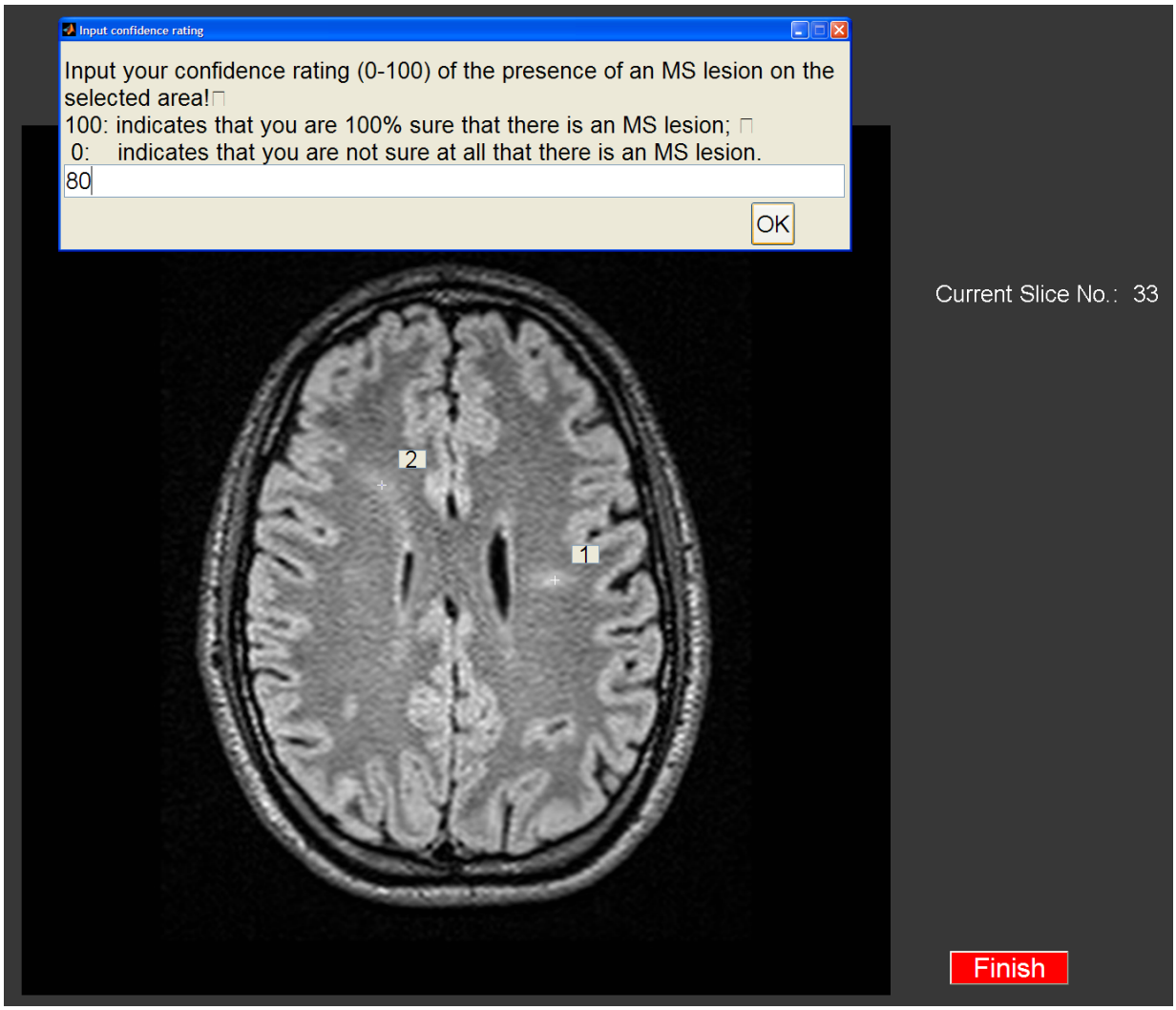


Figure 9.5: GUI of the multi-slice free-response subjective experiment for the performance evaluation of the msPCJO, where a dialog box popped up to ask radiologists the confidence rate (number contained between 0 and 100).

9.2.2 Participants

Except Radiologist 2 (namely expert 2), all other five radiologists having participated in the experiment for the performance evaluation of the PCJO have also participated in this multi-slice experiment. Thus we keep the same identification number for the radiologists as in Section 8.3.3.1. Radiologist 1 is an MS expert, with 21 years' experience; Radiologists 3-6 are not MS experts, with respectively 6, 3, 8 and 5 years' experience.

9.2.3 Performance evaluation method

We still use the software "JAFROC 4.1" to evaluate the msPCJO's performance w.r.t. that of radiologists, and the DBM-MRMC significance testing integrated in the software was conducted to examine JAFROC1 FOM differences between observers.

9.3 Results and discussion

9.3.1 JAFROC1 FOMs and p-values

The JAFROC1 FOMs of the five radiologists are shown in Table 9.1. The DBM-MRMC significance testing is conducted to examine JAFROC1 FOM differences between each pair of radiologists. The significance level is set to $p < 0.05$. The p-values are shown in Table 9.2.

Table 9.1: The JAFROC1 FOMs of the five radiologists for the multi-slice experiment. The JAFROC1 FOMs of the expert is highlighted in red.

	JAFROC1 FOM	Standard Error
Radiologist 1 (expert 1)	0.9267	0.0161
Radiologist 3	0.7837	0.0636
Radiologist 4	0.9192	0.0420
Radiologist 5	0.6547	0.0852
Radiologist 6	0.6320	0.0995

We then calculate the JAFROC1 FOMs for the msPCJO 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 radiologists are also calculated. Part of the results are illustrated in Table 9.3 and Table 9.4, noting that the results tend to be stable when $J > 4$ and K barely affects the JAFROC1 FOMs.

Table 9.2: The p-values from the DBM-MRMC significance testing to examine JAFROC1 FOM differences between each pair of the radiologists. The p-values smaller than 0.05 are highlighted in red. Radiologists 3-6 are denoted by Rad 3, Rad 4, Rad 5 and Rad 6 in this table.

	Expert 1	Rad3	Rad 4	Rad 5	Rad 6
Expert 1					
Rad 3	0.1489				
Rad 4	0.9499	0.2637			
Rad 5	0.0278	0.2866	0.0322		
Rad 6	0.0177	0.2114	0.0206	0.8506	

Table 9.3: The JAFROC1 FOMs of the msPCJO for the multi-slice experiment, w.r.t. different number of *steerable* channels K and number of *scale-shiftable* channels J.

(K,J)	JAFROC1 FOM	Standard Error
(3,3)	0.7076	0.1320
(3,4)	0.7645	0.1057
(3,5)	0.7820	0.0991
(5,3)	0.7029	0.1336
(5,4)	0.7622	0.1094
(5,5)	0.7785	0.1033
(7,3)	0.7099	0.1324
(7,4)	0.7622	0.1102
(7,5)	0.7797	0.1019

Table 9.4: The p-values from the DBM-MRMC significance testing to examine JAFROC1 FOM differences between the msPCJO with different (K, J) settings and radiologists. Radiologists 3-6 are denoted by Rad 3, Rad 4, Rad 5 and Rad 6 in this table. The p-values smaller than 0.05 are highlighted in red, which means that the corresponding JAFROC1 FOMs of the msPCJO are significantly different from those of the radiologists.

(K,J)	Expert 1	Rad3	Rad 4	Rad 5	Rad 6
(3,3)	0.0157	0.3990	0.0196	0.5578	0.4026
(3,4)	0.0731	0.8316	0.0874	0.2240	0.1426
(3,5)	0.1095	0.9846	0.1292	0.1590	0.0973
(5,3)	0.0136	0.3708	0.0171	0.5929	0.4321
(5,4)	0.0691	0.8116	0.0828	0.2339	0.1497
(5,5)	0.1012	0.9538	0.1198	0.1707	0.1053
(7,3)	0.0168	0.4135	0.0210	0.5406	0.3883
(7,4)	0.0691	0.8116	0.0828	0.2339	0.1497
(7,5)	0.1039	0.9640	0.1228	0.1667	0.1025

9.3.2 Detection rate, # TPs and # FPs

In Table 9.5, we present the number of TP marks, the number of FP marks for the laxest criterion (confidence rating > 0) on all the slices of the 10 patients for both the radiologists and the msPCJO, as well as the detection rate (number of TP marks divided by the total number of simulated lesions, that is 86). Note that the signal locations are calculated in the "candidate selection" step in the msPCJO schema, thus the number of *steerable* channels K and the number of *scale-shiftable* channels J (in the CJO part) do not influence the number of TP marks, the number of FP marks and the detection rate.

Table 9.5: The detection rate, number of TP marks, number of FP marks for the laxest criterion (confidence rating > 0) on all the slices of 10 patients. The total number of simulated lesions is 86. The results of the msPCJO is highlighted in red.

	Detection Rate	# TP marks	# FP marks
Radiologist 1 (expert 1)	0.8953	77	2
Radiologist 3	0.8372	72	3
Radiologist 4	0.9186	79	2
Radiologist 5	0.7209	62	5
Radiologist 6	0.6163	53	5
msPCJO	0.8256	71	4

9.3.3 Discussion

9.3.3.1 On radiologists' performances

It is interesting to inquire into the changes in radiologists' performances from single-slice experiment to multi-slice experiment.

Comparing Table 9.1 to Table 8.2, we find that both experts and radiologists improved their detection-localization performances with the help of volumetric information (adjacent slices): Expert 1's JAFROC1 FOM is increased from 0.7672 to 0.9267; Radiologists 3-6's JAFROC1 FOMs are increased from around 0.47 to above 0.63. The biggest surprise is that Radiologist 4 attained similar performance to that of Expert 1. The p-value results in Table 9.2 confirm that there is no significant difference between Radiologist 3-4 and Expert 1, while Expert 1 and Radiologist 4 have a significantly higher performance than Radiologist 5-6.

Comparison between Table 9.5 and Table 8.6 also reveals that when volumetric images were provided, all the five radiologists had various degree of improvement on the detection rate and

reduction on the number of FP marks for the laxest criterion. Especially Radiologist 4 detected even slightly more lesions than Expert 1.

Besides, in the course of the subjective experiments, we observed that all the radiologists gave higher confidence notes in the multi-slice experiment than in the single-slice experiment. They explained to us that with the access to adjacent slices, they can be more confident about the presence of MS lesions.

In general terms, the results show the importance of the volumetric information for the radiologists to detect and localize abnormalities. Radiologists less experienced in the diagnosis of MS lesions on MR images may also reach a good performance as an expert when volumetric information is available.

9.3.3.2 On the msPCJO's performances

Similar to the results in Section 8.4.1, Table 9.3 indicates that the number of *steerable* channels K hardly influences the msPCJO FOM, while the number of *scale-shiftable* channels J does.

Table 9.3 and Table 9.4 reveals that:

1. There is no significant difference between the msPCJO FOMs and those of the five radiologists ($p > 0.05$), except when $J = 3$ the msPCJO FOMs (around 0.7) are significantly lower than those of Expert 1 and Radiologist 4 (around 0.9).
2. With $(K, J) = (3, 5)$ (not much calculation burden), the msPCJO attains its maximum detection-localization task performance (around 0.78), which is nearest to that of Radiologist 3 (medium performance level among five radiologists), but still lower than those of Expert 1 and Radiologist 4 (top performance level among five radiologists).

From the detection rate point of view (cf. Table 9.5), the msPCJO almost detects as many lesions as Radiologist 3 does, but fewer than Expert 1 and Radiologist 4 do. This indicates that the "candidate selection" step (which controls the detection rate of the msPCJO) needs to be improved if we want to approach the top performance level of radiologists. Towards that goal, it would be essential to investigate the mechanisms that radiologists employ to combine the volumetric information in the tasks of detection and localization of lesions on a multi-slice image.

9.4 Conclusion

In this chapter, we have proposed a novel numerical observer called the multi-slice PCJO (msPCJO) in order to predict human performance in a more clinical relevant situation.

Compared to other multi-slice numerical observers in the literature, the msPCJO greatly extends the range of variable signal parameters and diagnostic tasks by including an SKS MO (CJO) and an HVS model (VDP). To our best knowledge, we are the first ones to propose such a numerical observer that can perform the joint detection-localization of multiple 3D signals with random amplitude, orientation, size, location and number of lesions.

We also conducted a multi-slice free-response experiment close to the clinical paradigm to get the detection-localization task performance of radiologists when volumetric information is available. We found that for our studied modality and pathology, the msPCJO can predict human performance in some degree: close to the medium performance level of radiologists, but still some distance to the top performance level of radiologists. This implies that a deeper investigation on the radiologists' interpretation process of multi-slice images is necessary to achieve a better performance (in terms of approximation to radiologists' possible top performance level).

In this preliminary study, we had only 10 patients in the multi-slice free-response experiment. This leads to lower statistical power when statistical analysis (such as the JAFROC1 and the DBM-MRMC) was performed. Therefore, more patients are absolutely needed for further validation of the msPCJO.

Conclusion of Part III

This part of thesis was dedicated to our proposed novel numerical observers.

The CJO extends the range of variable signal parameters of the existing SKS MO based on the JDE theory. It is expected to be more clinically relevant in case of one-signal detection on single-slice images.

Regarding the PCJO, in the "candidate selection" stage, the integration of an HVS model helps to treat the entire single-slice image with anatomical structures and to overcome many problems associated with the computational feasibility for the localization task; in the "application of the CJO on candidates" stage, the inherent training mechanism in the CJO contributes to develop the expertise of the PCJO by accumulating experience when reading more medical images.

The msPCJO is an extension of the PCJO on multi-slice images. It has the same advantages as the PCJO, thus the msPCJO can treat more complex task than the existing numerical observers while extending the task scope to joint detection-localization and the unknown signal parameters scope to amplitude, orientation, size, location and the number of signals in the volumetric image.

Two subjective experiments were conducted for validating the PCJO and the msPCJO. The JAFROC1 FOM was used for comparing their detection-localization task performance to that of radiologists. The results show that the two numerical observers predict well the human performance. Thus our models are promising for the evaluation of medical imaging systems from end-user's point of view.

10

Conclusion

10.1 Brief summary

With the growth of medical imaging technologies, the question of their assessment on the impact and benefit on patient care is rising. Development and design of those medical imaging technologies should take into account the concept of image quality assessment as it affects the diagnostic ability of radiologists. While the subjective quality image assessment is costly and time-consuming, there is a great need to evaluate the image quality objectively, but from the radiologists' perspective. Recently emerged new trend is the numerical observer in the frame of *task-based approach*.

This thesis contributed various new aspects and methods to this topic. After a careful review of the existing numerical observers, we proposed several novel numerical observers while considering several human factors involved in image analysis and interpretation, e.g. sensation and perception issues, expertise influence and diagnostic analysis pipeline (detection, localization, characterization), etc.. Our novel numerical observers greatly extended the range of variable signal parameters and the scope of diagnostic tasks. They reached the radiologists' task performance in more and more realistic clinical conditions.

Note that our objective to develop predictive numerical observers of human task performance is not to substitute radiologists in the daily diagnosis. Assuming that we have two medical imaging techniques or systems, the objective of the predictive numerical observers is to come up with, without using radiologists, the better technique or system for which radiologists can have a better diagnostic performance.

10.2 Contributions and list of publications

The main contributions and novelties of this research that were discussed in detail in this thesis, are:

1. A specialized study of the influence of radiologists' expertise and anatomical information on the perception of MR images (detailed in Chapter 5). The exploration of these aspects provided further insight into human visual system processing for the diagnostic task. This study provided the inspiration and argument for the numerical observer modeling for MRI image quality assessment by addressing the perspective of radiologist diagnostic performance. Published in SPIE MI conference [113].
2. A first comparison between the diagnostic task performance of Human Visual System (HVS) models (one family of numerical observers, cf. Chapter 4) and that of radiologists in the

context of MR images (detailed in Chapter 6). The characteristics of medical images require different comparison methods, compared to natural images. Our comparison method can be then applied on images of other medical imaging modalities. It's important to know the existing models' performance in order to use them correctly and improve them according to their application. Published in SPIE EI conference [114] and received the Student Scholar Award in MIPS XIV Conference [115].

3. A novel Signal Known Statistically (SKS) Model Observer (MO) - Channelized Joint detection and estimation Observer (CJO) (detailed in Chapter 7). The CJO is an extension of the SKS MO proposed by Goossens in [120], which can detect a signal with variable parameters (amplitude, orientation and size) on a single-slice image. The CJO widens the range of variable signal parameters of the existing SKS MOs for the detection task on single-slice images, and involves the estimation task (part of the characterization task).
4. A novel numerical observer - Perceptually relevant Channelized Joint Observer (PCJO) for detecting and localizing multiple signals with variable parameters (amplitude, orientation, size, location and the number of signals on an image) on a single-slice image (detailed in Chapter 8). The existing numerical observers for the localization task only consider the image with at most one signal [15], so far none considers the image with more than one signal. The PCJO greatly expands the limited task scope of existing numerical observers while integrating part of the human visual processing aspects. It is much more clinically relevant and its detection-localization task performance on single-slice images is close to that of radiologists. Published in IEEE TRANS. MEDICAL IMAGING [115].
5. A novel framework and implementation of the multi-slice PCJO (msPCJO) for the detection and localization of multiple signals with variable parameters (amplitude, orientation, size, location and the number of signals in an image) on a multi-slice/3D image (detailed in Chapter 9). Existing multi-slice/3D MOs narrow the studied task down to the Signal Known Exactly (SKE) detection task of one symmetrical signal located at the centre of the 3D image. In our framework, the application of the PCJO to the multi-slice/3D case leads to numerical observers that are more clinically realistic.
6. The design and implementation of two validation studies (free-response subjective experiments) for the PCJO and the msPCJO by comparing the numerical observers' detection-localization task performance to that of radiologists. Note that although the Free-response ROC (FROC) / Alternative Free-response ROC (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 on images that contain multiple signals).

So far, our research work resulted in 6 published papers (1 journal publication and 5 conference publications), as bellow:

[Zhang et al., 2012] L. Zhang, C. Cavaro-Ménard, and P. Le Callet (2012). A Perceptually relevant Channelized Joint Observer (PCJO) for the detection-localization of parametric signals. *IEEE Transactions on Medical Imaging*, June 2012.

[Zhang et al., 2012a] L. Zhang, C. Cavaro-Ménard, and P. Le Callet (2012). Evaluation of HVS models in the application of medical image quality assessment. In *IS&T/SPIE Electronic Imaging*, Burlingame, California, USA, 22 - 26 January 2012

[Zhang et al., 2012b] L. Zhang, C. Cavaro-Ménard, and P. Le Callet (2012). Key issues and specificities for the objective medical image quality assessment. In *VPQM*, Scottsdale, Arizona, USA, 19 - 20 January 2012

[Zhang et al., 2011a] L. Zhang, C. Cavaro-Ménard, and P. Le Callet (2011). The effects of anatomical information and observer expertise on abnormality detection task. In *SPIE Medical Imaging*, Orlando, Florida, USA, 12-17 February 2011

[Zhang et al., 2011b] L. Zhang, C. Cavaro-Ménard, and P. Le Callet (2011). Using AUC to study perceptual difference model suitability for the detection task on MR image. In *MIPS XIV Conference*, Dublin, Ireland, 9-12 August 2011 (Student Scholar Award)

[Cavaro-Ménard et al., 2010] C. Cavaro-Ménard, L. Zhang, and P. Le Callet (2010). Diagnostic quality assessment of medical images : Challenges and trends. In *2nd European Workshop on Visual Information Processing (EUVIP)*, Paris, France, 4-6 July 2010 (invited paper)

10.3 Directions for future work

Several directions are open for future researches:

1. The preliminary validation results of our novel numerical observers demonstrate their great potential. Thus we are considering their application on the quantification of medical imaging system performance and the comparison of different image processing algorithms or different medical displays, as shown in Figure 10.1. When the display system is modeled by the display model in the VDP part, with regard to different image processing algorithms, the PCJO or the msPCJO will give different outcomes, based on which the different performances of these algorithms can be evaluated. Similarly, when the image processing

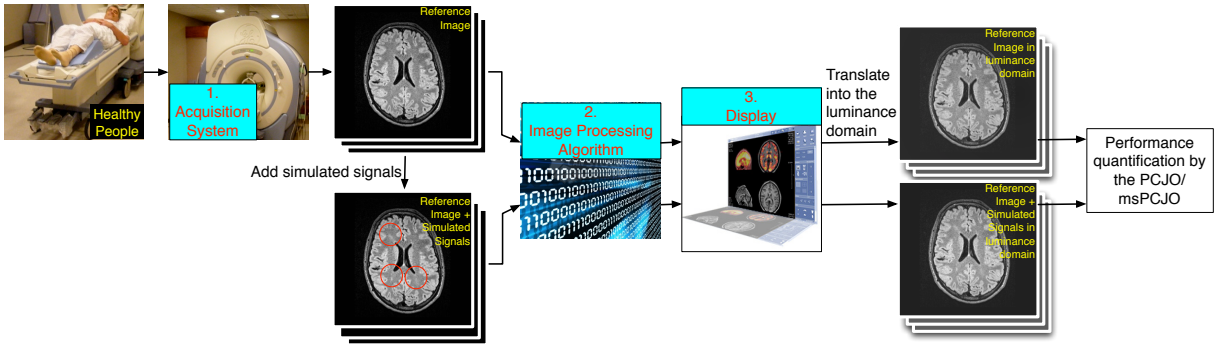


Figure 10.1: Using the PCJO to quantify the medical imaging system performance: the reference images for the VDP and the training images for the CJO can be obtained from healthy subjects, then the distorted images for the VDP can easily be produced using a signal model.

algorithm is given, we can compare the performances of different displays by building their corresponding display models.

2. There is a lot of room for improvement of the CJO/PCJO: improving the steerable/shiftable channels to incorporating the decomposition mechanisms of the HVS; ameliorating the candidate select algorithm based on the VDP output *detection probability map*; including the variability of the signal shape; considering the characterization of signals, etc..

   There is also a lot of room for improvement of the msPCJO: extending the CJO to 3D CJO; improving the candidate select algorithm based on the VDP output *detection probability map* on each slice; or incorporating factors in human temporal perception. The objective is to integrate better the information in the z direction or the spatial correlation of the background and signal in the adjacent slices for the multi-slice perception.

   These future extensions could make numerical observers more anthropomorphic and more clinically relevant, to better predict the human performance.

3. Although in this thesis, we restricted the scope to the MRI modality and the Multiple Sclerosis (MS) pathology, the framework of the proposed numerical observers can be applied on all digital imaging modalities. It will be interesting to apply the proposed numerical observer frameworks on other imaging modalities and/or pathologies. Each medical imaging modality has its own specific features corresponding to the studied physical and physiological phenomena. It may be needed to change the background model, the signal model, and the derivations of signal parameter estimation and the test statistic calculation according to the studied modality and pathology.

Bibliography

[1] B. Erickson, "Irreversible compression of medical images," *Journal of Digital Imaging*, vol. 15, no. 1, pp. 5–14, 2002.

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] L. Dickinson, H. Ahmed, C. Allen, J. Barentsz, B. Carey, J. Futterer, S. Heijmink, P. Hoskin, A. Kirkham, A. Padhani *et al.*, "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.

[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: addi-

tional 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*, pp. 3601–3617, 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] L. Zhang, C. Cavaro-Ménard, and P. L. Callet, "Key issues and specificities for the objective medical image quality assessment," in *Sixth International Workshop on Video Processing and Quality Metrics (VPQM)*, vol. 7966, January 2012.

[17] 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.

[18] 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.

[19] 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.

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

[21] 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.

[22] 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.

[23] 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.

[24] C. Metz, "Roc analysis in medical imaging: a tutorial review of the literature," *Radiological physics and technology*, vol. 1, no. 1, pp. 2–12, 2008.

[25] B. Everitt and A. Skrondal, *The Cambridge dictionary of statistics*. Cambridge University Press Cambridge, 2006, vol. 4.

[26] J. Claassen, "The gold standard: not a golden standard," *BMJ*, vol. 330, no. 7500, p. 1121, 2005.

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

[28] S. Mason and N. Graham, "Areas beneath the relative operating characteristics (roc) and relative operating levels (rol) curves: Statistical significance and interpretation," *Quarterly Journal of the Royal Meteorological Society*, vol. 128, no. 584, pp. 2145–2166, 2002.

[29] 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

[30] D. D. Dorfman, K. S. Berbaum, and R. V. Lenth, "Multireader, multicase receiver operating characteristic methodology: a bootstrap analysis," *Academic Radiology*, vol. 2, no. 7, pp. 626–633, 1995.

[31] D. D. Dorfman, K. S. Berbaum, R. V. Lenth, Y. F. Chen, and B. A. Donaghy, "Monte carlo validation of a multireader method for receiver operating characteristic discrete rating data: factorial experimental design," *Academic radiology*, vol. 5, no. 9, pp. 591–602, 1998.

[32] C. A. Roe and C. E. Metz, "Dorfman-berbaum-metz method for statistical analysis of multireader, multimodality receiver operating characteristic data: validation with computer simulation," *Academic radiology*, vol. 4, no. 4, pp. 298–303, 1997.

[33] B. D. Gallas, "One-shot estimate of MRMC variance: AUC," *Academic Radiology*, vol. 13, pp. 353–362, 2006.

[34] B. D. Gallas, G. A. Pennello, and K. J. Myers, "Multireader multicase variance analysis for binary data," *J. Opt. Soc. Am. A*, vol. 24, pp. B70–B80, 2007.

[35] B. D. Gallas, A. Bandos, F. W. Samuelson, and R. F. Wagner, "A framework for random-effects roc analysis: Biases with the bootstrap and other variance estimators," *Communications in Statistics - Theory and Methods*, vol. 38, pp. 2586–2603, 2009.

[36] N. A. Obuchowski, "New methodological tools for multiple-reader roc studies1," *Radiology*, vol. 243, no. 1, pp. 10–12, 2007.

[37] S. L. Hillis, "A comparison of denominator degrees of freedom methods for multiple observer roc analysis," *Statistics in medicine*, vol. 26, no. 3, pp. 596–619, 2007.

[38] H. H. Barrett, M. A. Kupinski, and E. Clarkson, "Probabilistic foundations of the mrmc method," in *Proc. SPIE*, vol. 5749, 2005, pp. 21–31.

[39] M. A. Kupinski, E. Clarkson, and H. H. Barrett, "A probabilistic model for the mrmc method, part 2: Validation and applications," *Academic radiology*, vol. 13, no. 11, pp. 1422–1430, 2006.

[40] L. Popescu, "Nonparametric roc and lroc analysis," *Medical physics*, vol. 34, p. 1556, 2007.

[41] P. Khurd and G. Gindi, "Rapid computation of lroc figures of merit using numerical observers (for spect/pet reconstruction)," *Nuclear Science, IEEE Transactions on*, vol. 52, no. 3, pp. 618–626, 2005.

[42] 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.

[43] ——, "Estimation ROC curves and their corresponding ideal observers," in *Proc. SPIE*, vol. 6515, 2007, pp. 651 504–1–651 504–7.

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

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

[46] ——, *The Handbook of Medical Image Perception and Techniques*. Cambridge University Press, 2010, ch. Recent developments in free-response methodology, pp. 216–239.

[47] J. P. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," *J. Opt. Soc. Am. A*, vol. 9, no. 5, pp. 649–658, May 1992. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-9-5-649

[48] F. O. Bochud, C. K. Abbey, and M. P. Eckstein, "Statistical texture synthesis of mammographic images with clustered lumpy backgrounds," *Opt. Express*, vol. 4, no. 11, pp. 33–43, 1999.

[49] 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.

[50] 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.

[51] E. C. M. A. Kupinski, J. W. Hoppin and H. H. Barrett, "Ideal-observer computation in medical imaging with use of Markov-chain Monte Carlo techniques," *Journal of the Optical Society of America A*, vol. 20, pp. 430–438, 2003.

[52] 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.

[53] K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," *Journal of the Optical Society of America A*, vol. 4, p. 2447–2457, 1987.

[54] 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.

[55] 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.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

[60] Y. Zhang, B. T. Pham, M. Eckstein, and S. Barbara, "Evaluation of JPEG 2000 encoder options: human and model observer detection of variable signals in x-ray coronary angiograms," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 613–632, 2004.

[61] C. Lartizien, P. Kinahan, and C. Comtat, "Volumetric model and human observer comparisons of tumor detection for whole-body positron emission tomography< sup> 1</sup>," *Academic radiology*, vol. 11, no. 6, pp. 637–648, 2004.

[62] H. C. Gifford and M. A. King, "A comparison of human and model observers in multislice LROC studies," *IEEE Transactions on Medical Imaging*, vol. 24, no. 2, pp. 160–169, 2005.

[63] H. Gifford and M. King, "Implementing visual search in human-model observers for emission tomography," in *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*. IEEE, 2009, pp. 2482–2485.

[64] ——, "Em clustering for holistic search in human-model observers," in *Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE*. IEEE, 2010, pp. 3584–3587.

[65] 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.

[66] 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.

[67] B. Goossens, L. Platiša, E. Vansteenkiste, and W. Philips, "The use of steerable channels for detecting asymmetrical signals with random orientations," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2010, pp. 76 270S–76 270S.

[68] ——, "Design of Model Observers for joint Detection and Estimation of Random Parametric Signals in Images," 2012, in preparation.

[69] L. Platiša, B. Goossens, E. Vansteenkiste, S. Park, B. D. Gallas, A. Badano, and W. Philips, "Channelized hotelling observers for the assessment of volumetric imaging data sets," *Journal of the Optical Society of America A*, vol. 28, no. 6, pp. 1145–1163, 2011.

[70] H. Liang, S. Park, B. D. Gallas, K. J. Myers, and A. Badano, "Image browsing in slow medical liquid crystal displays," *Academic Radiology*, vol. 15, pp. 370–382, 2008.

[71] J. S. Kim, P. Kinahan, C. Lartizien, C. Comtat, and T. Lewellen, "A comparison of planar versus volumetric numerical observers for detection task performance in whole-body PET imaging," *Nuclear Science, IEEE Transactions on*, vol. 51, pp. 34–40, 2004.

[72] P. E. K. C. Lartizien and C. Comtat, "Volumetric model and human observer comparisons of tumor detection for whole-body positron emission tomography," *Academic Radiology*, vol. 11, pp. 637–648, 2004.

[73] M. Chen, J. Bowsher, A. Baydush, K. Gilland, D. DeLong, and R. Jaszczak, "Using the hotelling observer on multislice and multi- view simulated SPECT myocardial images,," *Nuclear Science, IEEE Transactions on*, vol. 49, pp. 661–667, 2002.

[74] S. Young, S. Park, S. K. Anderson, A. Badano, K. J. Myers, and P. Bakic, "Estimating breast tomosynthesis performance in detection tasks with variable-background phantoms," in *Proceedings of SPIE*, vol. 7258, 2009, p. 72580O.

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

[76] 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. 64ñ73.

[77] 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

[78] D. Huo, D. Xu, Z. 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

[79] 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.

[80] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital images and human vision*. MIT Press, 1993, pp. 163–178.

[81] 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.

[82] 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, 1997, pp. 2379–2391.

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

[84] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. LaMantia, J. O. McNamara, and L. E. White, *Neuroscience, 4th ed.* Sinauer Associates, 2008.

[85] A. Ninassi, "De la perception locale des distorsions de codage a l'appreciation globale de la qualite visuelle des images et videos. apport de l'attention visuelle dans le jugement de qualite," Ph.D. dissertation, Université de Nantes, 2009.

[86] H. R. R. A. Normann, B. S. Baxter and P. J. Anderton, "Photoreceptor Contributions to Contrast Sensitivity: Applications in Radiological Diagnosis," *IEEE transactions on systems, man, and cybernetics*, vol. 13, pp. 944–953, 1983.

[87] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science, 4th ed.* McGraw-Hill, New York, 2000.

[88] E. Peli, "Contrast in Complex Images," *Journal of the Optical Society of America*, vol. 7, pp. 2032–2040, 1990.

[89] J. Capó-Aponte, L. Temme, H. Task, A. Pinkus, M. Kalich, A. Pantle, and C. Rash, "Visual perception and cognitive performance," *Helmet-Mounted Displays: Sensation, Perception and Cognitive Issues*, pp. 335–390, 2009.

[90] F. W. Campbell and J. G. Robson, "Application of fourier analysis to the visibility of gratings," *Journal of Physiology*, vol. 197, pp. 551–566, 1968.

[91] T. Mayhew and D. Astle, "Photoreceptor number and outer segment disk membrane surface area in the retina of the rat: stereological data for whole organ and average photoreceptor cell," *Journal of neurocytology*, vol. 26, no. 1, pp. 53–61, 1997.

[92] D. Sagi and S. Hochstein, "Lateral inhibition between spatially adjacent spatial-frequency channels?" *Attention, Perception, & Psychophysics*, vol. 37, no. 4, pp. 315–322, 1985.

[93] P. G. J. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*. SPIE Press, 1999.

[94] G. E. Legge and J. M. Foley, "Contrast Masking in Human Vision," *Journal of the Optical Society of America*, vol. 70, pp. 1458–1471, 1980.

[95] E. Simoncelli and D. Heeger, "A model of neuronal responses in visual area mt," *Vision research*, vol. 38, no. 5, pp. 743–761, 1998.

[96] F. T. Qiu and R. von der Heydt, "Figure and Ground in the Visual Cortex: V2 Combines Stereoscopic Cues with Gestalt Rules," *Neuron*, vol. 47, p. 155ñ166, 2005.

[97] A. B. Watson, "Separate Visual Pathways for Perception and Action," in *Computer Vision, Graphics and Image Processing*, vol. 39, 1987, pp. 311–327.

[98] R. Eriksson, B. Andren, and K. Brunnstroem, "Modelling the perception of digital images : A performance study," in *IS&T/SPIE Conference on Human Vision and Electronic*, vol. 3299, 1998, pp. 88–97.

[99] J. M. Foley, "Human luminance pattern mechanisms: masking experiments require a new model," *Journal of the Optical Society of America*, vol. 11, pp. 1710–1719, 1994.

[100] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proceedings of SPIE*, vol. 2179, 1994, p. 127–139.

[101] S. Klein, "Measuring, estimating, and understanding the psychometric function: A commentary," *Attention, Perception, & Psychophysics*, vol. 63, no. 8, pp. 1421–1455, 2001.

[102] A. Logvinenko, Y. Tyurin, and M. Sawey, "A test for psychometric function shift," *Behavior research methods*, pp. 1–13, 2012.

[103] K. Zchaluk and D. Foster, "Model-free estimation of the psychometric function," *Attention, Perception, & Psychophysics*, vol. 71, no. 6, pp. 1414–1425, 2009.

[104] M. A. Miranda and D. B. Henson, "Perimetric sensitivity and response variability in glaucoma with single-stimulus automated perimetry and multiple-stimulus perimetry with verbal feedback," *Acta Ophthalmologica*, vol. 86, pp. 202–206, 2008.

[105] A. Watson and D. Pelli, "Quest: A bayesian adaptive psychometric method," *Attention, Perception, & Psychophysics*, vol. 33, no. 2, pp. 113–120, 1983.

[106] F. Wichmann and N. Hill, "The psychometric function: I. fitting, sampling, and goodness of fit," *Attention, Perception, & Psychophysics*, vol. 63, no. 8, pp. 1293–1313, 2001.

[107] H. Strasburger, "Converting between measures of slope of the psychometric function," *Attention, Perception, & Psychophysics*, vol. 63, no. 8, pp. 1348–1355, 2001.

[108] C. Kaernbach, "Adaptive threshold estimation with unforced-choice tasks," *Attention, Perception, & Psychophysics*, vol. 63, no. 8, pp. 1377–1388, 2001.

[109] S. Park, A. Badano, B. Gallas, and K. Myers, "A contrast-sensitive channelized-hotelling observer to predict human performance in a detection task using lumpy backgrounds and gaussian signals," in *Proceedings of SPIE*, vol. 6515, 2007, p. 65150V.

[110] C. Abbey and M. Eckstein, "High human-observer efficiency for forced-localization tasks in correlated noise," in *Proceedings of SPIE*, vol. 7627, 2010, p. 76270R.

[111] C. Nodine, H. Kundel, S. Lauver, and L. Toto, "Nature of expertise in searching mammograms for breast masses," *Academic radiology*, vol. 3, no. 12, pp. 1000–1006, 1996.

[112] E. Krupinski and R. Weinstein, "Changes in visual search patterns of pathology residents as they gain experience," in *Proceedings of SPIE*, vol. 7966, 2011, p. 79660P.

[113] 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, pp. 79 661G–79 661G–8.

[114] L. Zhang, C. Cavaro-Menard, and P. Le Callet, "Evaluation of hvs models in the application of medical image quality assessment," in *Proceedings of SPIE*, vol. 8291, 2012, p. 82910T.

[115] 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.

[116] J. B. Jared, W. B. Jackson, P. Beebee, D. A. Jared, D. K. Biegelsen, J. O. Larimer, J. Lubin, and J. L. Gille, "X-ray image system design using a human visual model," in *Physics of Medical Imaging SPIE*, 1993, pp. 9758–9765.

[117] R. Dosselmann and X. Yang, "A comprehensive assessment of the structural similarity index," *Signal, Image and Video Processing*, vol. 5, no. 1, pp. 81–91, 2011.

[118] T. Luo and X. Mou, "Divisive normalization in channelized hotelling observer," in *Proceedings of SPIE/IS&T Electronic Imaging, Human Vision and Electronic Imaging XV*, 2010, pp. 76 270E–76 270E–8.

[119] A. Watson and J. Ahumada, "A Standard Model for Foveal Detection of Spatial Contrast," *Journal of Vision*, vol. 5, pp. 717–740, 2005.

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

[121] E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, "Shiftable multiscale transforms," *Information Theory, IEEE Transactions on*, vol. 38, no. 2, pp. 587–607, 1992.

[122] E. Simoncelli and W. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Image Processing, 1995. Proceedings., International Conference on*, vol. 3. IEEE, 1995, pp. 444–447.

[123] M. Anderson and W. Woessner, *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*, 2nd ed. Academic Press, 1992.

[124] J. Armstrong and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting*, vol. 8, no. 1, pp. 69–80, 1992.

[125] D. Ge, L. Zhang, C. Cavaro-Menard, and P. Le Callet, "Numerical stability issues on the channelized hotelling observers under different background assumptions," unpublished.

[126] A. Tihonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Math*, vol. 4, pp. 1035–1038, 1963.

[127] L. Zhang, C. Cavaro-Ménard, P. Le Callet, and J. TANGUY, "A perceptually relevant channelized joint observer (pcjo) for the detection-localization of parametric signals," *IEEE Transactions on Medical Imaging*, vol. pp, p. 1, 2012.

[128] 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.

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

[130] 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.

[131] I. Bankman, *Handbook of medical image processing and analysis*. Academic Press, 2008.

[132] T. Herter and K. Lott, "Algorithms for decomposing 3-d orthogonal matrices into primitive rotations," *Computers & graphics*, vol. 17, no. 5, pp. 517–527, 1993.

[133] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *JSTOR*, vol. 39, no. 1, pp. 1–38, 1977.

l'université nantes angers le mans
PÔLE DE RECHERCHE ET D'ENSEIGNEMENT SUPÉRIEUR

université angers

Thèse de Doctorat

Lu Zhang-Ge

Modèles Numériques pour l'Évaluation Objective de la Qualité d'Images Médicales

Numerical Observers for the Objective Quality Assessment of Medical Images

Résumé

L'évaluation de la qualité des images médicales est essentielle pour optimiser un système d'imagerie. Dans le cadre d'approches basées sur la tâche, les modèles numériques proposés ont des limites: pour la tâche de localisation, la plupart des modèles ont besoin de la connaissance *a priori* des paramètres du signal; les modèles dédiés à la détection de signaux dans une image 3D se limitent aux signaux symétriques.
Dans cette thèse, nous proposons de nouveaux modèles numériques: le CJO pour la détection sur une coupe d'un signal dont l'amplitude, l'orientation et la taille ne sont pas connus *a priori*; le PCJO et le msPCJO pour la détection-localisation de plusieurs signaux paramétriques d'amplitude, d'orientation, de taille et d'emplacement variables sur une coupe (PCJO) ou en 3D (msPCJO). Deux expériences ont été conçues et mises en œuvre pour la validation de nos modèles sur des images IRM présentant des lésions de SEP. Les résultats indiquent que nos modèles sont performants et prometteurs pour l'évaluation de systèmes d'imagerie.

Mots clés

Évaluation de la qualité d'images médicales, Modèle Numérique, SVH, VDP, JAFROC, IRM.

Abstract

Medical image quality assessment is critical for comparing and optimizing medical imaging system. In the frame of *task-based approach*, numerical observers proposed for the objective medical image quality assessment have several limitations: e.g. for the multi-signal localization, most need *a priori* knowledge of signal parameters; multi-slice MOs narrow the task down to the detection of one symmetrical signal at the image centre.
In this thesis, we propose novel numerical observers: the CJO for the detection of one parametric signal with random amplitude, orientation and size on single-slice; the PCJO and the msPCJO for the detection-localization of multiple parametric signals with random amplitude, orientation, size and location on single-slice and multi-slice, respectively. Two observer studies were designed and implemented for their validation on MR images with MS lesions, and the JAFROC FOM results indicate that they are promising for predicting radiologists' task performance, ultimately for evaluating medical imaging systems.

Key Words

Medical image quality assessment, Model Observer, HVS, VDP, JAFROC, MRI.