

CHANNELIZED HOTELLING OBSERVER OPTIMIZATION FOR MEDICAL
IMAGE QUALITY ASSESSMENT IN LESION DETECTION TASKS

BY

MIREIA CALVO GONZÁLEZ

Submitted in partial fulfillment of the
requirements for the degree of
Superior Telecommunication Engineering
in the Graduate College of the
Illinois Institute of Technology

Approved Jovan G. Brankov, Ph.D.

Chicago, Illinois
July 2011

ACKNOWLEDGEMENT

I would like to thank Jovan G. Brankov for generously providing the human observer data used in the experiments and for his helpful lessons and comments during the realization of this manuscript.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENT	ii
LIST OF TABLES	iv
LIST OF FIGURES	v
LIST OF SYMBOLS	xi
ABSTRACT	xii
CHAPTER	
1. INTRODUCTION	1
1.1. Image quality assessment	1
1.2. Human vision models	2
1.3. Image quality for lesion detection in medical imaging	3
2. NUMERICAL METHODS FOR LESION DETECTION	5
2.1. Lesion detection tasks	5
2.2. Numerical observers	7
2.3. Channelized Hotelling Observer	9
3. EXPERIMENTS	15
3.1. Myocardial SPECT for perfusion defect detection	15
3.2. Defect detection in different correlated noises	24
4. CONCLUSIONS	36
4.1. General conclusions	36
4.2. Experiment 1	36
4.3. Experiment 2	36
REFERENCES	37

LIST OF TABLES

Table	Page
1. Decision results	5
2. OSEM reconstructed images	
a. Example OSEM reconstructed images with 1 iteration	17
b. Example OSEM reconstructed images with 5 iterations	17

LIST OF FIGURES

Figure	Page
1. Example images in which the MSE is calculated	2
a. Original image	2
b. Distorted image by adding a constant to the original	2
c. Distorted image by adding an alternated constant to the original.	2
2. Contrast sensitivity function3
3. Probability density functions under H_0 and H_1 and the resulting TPF, FNF, FPF and TPF for one value of γ	6
4. Receiver Operating Characteristic	7
5. Band pass filters	10
6. Gabor filters	10
7. Laguerre-Gauss filters	11
8. Difference Of Gaussians filters	11
9. Images with defect present and absent from a myocardial SPECT for perfusion defect detection15
10. Coronal cuts of the heart	16
11. Human observer defect detection performance measured by AUC in relation to $FWHM$18
12. AUC curves of the Human Observer and Numerical Observer when training with iteration 1 and testing with iteration 5 using Band Pass filters and Internal-noise model 620
13. AUC curves of the Human Observer and Numerical Observer when training with iteration 5 and testing with iteration 1 using Band Pass filters and Internal-noise model 620
14. AUC curves of the Human Observer and Numerical Observer when training with iteration 1 and testing with iteration 5 using Gabor filters and Internal-noise model 6	21
15. AUC curves of the Human Observer and Numerical Observer when training with iteration 5 and testing with iteration 1 using Gabor filters and Internal-noise model 6	21

16. AUC curves of the Human Observer and Numerical Observer when training with iteration 1 and testing with iteration 5 using Laguerre-Gauss filter and Internal-noise model 6	22
17. AUC curves of the Human Observer and Numerical Observer when training with iteration 5 and testing with iteration 1 using Laguerre-Gauss filters and Internal-noise model 6	22
18. AUC curves of the Human Observer and Numerical Observer when training with iteration 1 and testing with iteration 5 using Difference Of Gaussians filters and Internal-noise model 6	23
19. AUC curves of the Human Observer and Numerical Observer when training with iteration 5 and testing with iteration 1 using Difference of Gaussians filters and Internal-noise model 6	23
20. Example images with defect present and absent	24
21. Lesion	24
22. Power-law textures for $\beta = 0$ to 3	25
23. Human observers defect detection performance measured by <i>AUC</i> in relation to β	25
24. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Band Pass filters and Internal-noise models 1, 2, 3 and 4	28
25. AUC curves of the Human Observer 1 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Band Pass filters and Internal-noise models 1, 2, 3 and 4.	28
26. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Band Pass filters and Internal-noise models 5, 6, 7 and 8	28
27. AUC curves of the Human Observer 1 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Band Pass filters and Internal-noise models 5, 6, 7 and 8.	28
28. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Band Pass filters and Internal-noise models 9, 10 and 11	28

29. AUC curves of the Human Observer 1 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Band Pass filters and Internal-noise models 9, 10 and 11	28
30. AUC curves of the Human Observer 2 and Numerical Observer when training and testing with images of all textures using Band Pass filters and Internal-noise models 1, 2, 3 and 4	29
31. AUC curves of the Human Observer 2 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Band Pass filters and Internal-noise models 1, 2, 3 and 4.	29
32. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Band Pass filters and Internal-noise models 5, 6, 7 and 8	29
33. AUC curves of the Human Observer 2 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Band Pass filters and Internal-noise models 5, 6, 7 and 8.	29
34. AUC curves of the Human Observer 2 and Numerical Observer when training and testing with images of all textures using Band Pass filters and Internal-noise models 9, 10 and 11	29
35. AUC curves of the Human Observer 2 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Band Pass filters and Internal-noise models 9, 10 and 11	29
36. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Gabor filters and Internal-noise models 1, 2, 3 and 4	30
37. AUC curves of the Human Observer 1 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Gabor filters and Internal-noise models 1, 2, 3 and 4.	30
38. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Gabor filters and Internal-noise models 5, 6, 7 and 8	30
39. AUC curves of the Human Observer 1 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Gabor filters and Internal-noise models 5, 6, 7 and 8.	30

40. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Gabor filters and Internal-noise models 9, 10 and 11	30
41. AUC curves of the Human Observer 1 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Gabor filters and Internal-noise models 9, 10 and 11	30
42. AUC curves of the Human Observer 2 and Numerical Observer when training and testing with images of all textures using Gabor filters and Internal-noise models 1, 2, 3 and 4	31
43. AUC curves of the Human Observer 2 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Gabor filters and Internal-noise models 1, 2, 3 and 4.	31
44. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Gabor filters and Internal-noise models 5, 6, 7 and 8	31
45. AUC curves of the Human Observer 2 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Gabor filters and Internal-noise models 5, 6, 7 and 8.	31
46. AUC curves of the Human Observer 2 and Numerical Observer when training and testing with images of all textures using Gabor filters and Internal-noise models 9, 10 and 11	31
47. AUC curves of the Human Observer 2 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Gabor filters and Internal-noise models 9, 10 and 11	31
48. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Gabor filters and Internal-noise models 1, 2, 3 and 4	32
49. AUC curves of the Human Observer 1 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Laguerre-Gauss filters and Internal-noise models 1, 2, 3 and 4.	32
50. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Laguerre-Gauss filters and Internal-noise models 5, 6, 7 and 8	32

51. AUC curves of the Human Observer 1 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Laguerre-Gauss filters and Internal-noise models 5, 6, 7 and 8 . . .	32
52. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Laguerre-Gauss filters and Internal-noise models 9, 10 and 11	32
53. AUC curves of the Human Observer 1 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Laguerre-Gauss filters and Internal-noise models 9, 10 and 11 . .	32
54. AUC curves of the Human Observer 2 and Numerical Observer when training and testing with images of all textures using Laguerre-Gauss filters and Internal-noise models 1, 2, 3 and 4	33
55. AUC curves of the Human Observer 2 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Laguerre-Gauss filters and Internal-noise models 1, 2, 3 and 4. . .	33
56. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Laguerre-Gauss filters and Internal-noise models 5, 6, 7 and 8	33
57. AUC curves of the Human Observer 2 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Laguerre-Gauss filters and Internal-noise models 5, 6, 7 and 8. . .	33
58. AUC curves of the Human Observer 2 and Numerical Observer when training and testing with images of all textures using Laguerre-Gauss filters and Internal-noise models 9, 10 and 11	33
59. AUC curves of the Human Observer 2 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using Laguerre-Gauss filters and Internal-noise models 9, 10 and 11 . .	33
60. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using Difference of Gaussians filters and Internal-noise models 1, 2, 3 and 4	34
61. AUC curves of the Human Observer 1 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using DOG filters and Internal-noise models 1, 2, 3 and 4.	34

62. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using DOG filters and Internal-noise models 5, 6, 7 and 8	34
63. AUC curves of the Human Observer 1 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using DOG filters and Internal-noise models 5, 6, 7 and 8.	34
64. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using DOG filters and Internal-noise models 9, 10 and 11	34
65. AUC curves of the Human Observer 1 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using DOG filters and Internal-noise models 9, 10 and 11	34
66. AUC curves of the Human Observer 2 and Numerical Observer when training and testing with images of all textures using DOG filters and Internal-noise models 1, 2, 3 and 4	35
67. AUC curves of the Human Observer 2 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using DOG filters and Internal-noise models 1, 2, 3 and 4.	35
68. AUC curves of the Human Observer 1 and Numerical Observer when training and testing with images of all textures using DOG filters and Internal-noise models 5, 6, 7 and 8	35
69. AUC curves of the Human Observer 2 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using DOG filters and Internal-noise models 5, 6, 7 and 8.	35
70. AUC curves of the Human Observer 2 and Numerical Observer when training and testing with images of all textures using DOG filters and Internal-noise models 9, 10 and 11	35
71. AUC curves of the Human Observer 2 and Numerical Observer when training with images where $\beta = 2$ and testing with images of all textures using DOG filters and Internal-noise models 9, 10 and 11	35

LIST OF SYMBOLS

Symbol	Definition
γ	Threshold for the detection decision
γ	Decision- variable noise
ϵ	Internal-noise
β	Power-law exponent

ABSTRACT

Although human observers are the medical imaging main users in lesion detection tasks and, as a consequence, they are the best agents to optimize image quality assessment algorithms with that purpose, their measurements can be in many cases too much costly and time-consuming. A widely used alternative to evaluate imaging devices and algorithms are model observers – that are numerical methods acting as surrogates of the human performance – and, more specifically, it is widely accepted that the Channelized Hotelling Observer correlates well with the human observer when it uses proper channels and introduces an optimal Internal-noise model. For this reason, in this report it is studied how can the Channelized Hotelling Observer estimate the human performance with four different channels and eleven Internal-noise models when it is trained with data from different humans and then tested in other related images the human data is not available so as to ensure that the CHO would be a good human observer surrogate in that situations.

The research is done in two different scenes. The first experiment tries to estimate the human detections in images after a specific iteration of a filter, knowing the human scores for another iteration previously; while in the second experiment the tested data has different but correlated noise than the images used to train the Internal-noise. The results show both experiments are not optimized with the same noise models, but in each one, after using the one that obtains best results, the estimation of the human observer performance is quite accurate. Then, although they need being more studied in different kinds of data and human observers, they obtain good first results to consider their use in image quality assessment for lesion-detection tasks in the future.

CHAPTER 1

INTRODUCTION

1.1. Image quality assessment

In the optimization of any image processing system there is the necessity of evaluating the quality of its final images to ensure they are useful and suitable to work with them. For instance, the goal of image compression is to reduce the amount of data required to store an image while at the same time it is ensured the results are of good quality enough; in image enhancement systems, final images should be of better visual quality than the originals; and taking into account current communication networks, their images are transported by channels that introduce errors, thus they should be evaluated to ensure we can work with the final images they have transported. More specifically, according to the type of data used in this study, in medical imaging we have to ensure that the images representing the human body are useful for clinical purposes like diagnosis or examination of diseases, or for the medical science to, for example, study the anatomy. It can be concluded, to sum up, that this evaluation is related to the understanding of the Human Visual System because the decision that an image has good quality depends on if it is useful for the human, which in most cases will be the final user of the data.

The earliest metrics in image processing to assess this image quality were simple numerical criteria like the Mean Square Error (*MSE*) and the Peak Signal to Noise Ratio (*PSNR*). Being $f(\mathbf{n})$ and $g(\mathbf{n})$ the values of the original and distorted image pixels, $\mathbf{n}=(n_1, n_2)$ the pixels in a Cartesian grid, N the total number of pixel locations in the images, and E the maximum value that a pixel can take (for instance 255 for a 8-bit grayscale), these metrics are defined as

$$MSE [f(\mathbf{n}), g(\mathbf{n})] = \frac{1}{N} \sum_{\mathbf{n}} [f(\mathbf{n}) - g(\mathbf{n})]^2$$

$$PSNR [f(\mathbf{n}), g(\mathbf{n})] = 10 \log_{10} \frac{E^2}{MSE [f(\mathbf{n}), g(\mathbf{n})]}$$

However, since they use simple equations that just compare the difference between the original and final images' values pixel by pixel, they have well-known limitations as they do not represent in most cases the sensitivity of the Human Visual System.

For instance, in *Figure 1* are shown two distorted images from the same reference image (*Figure 1.a*). *Figure 1.b* was created by adding a constant to all its pixels, while *Figure 1.c* added the same constant but randomly alternating the sign of the samples. This is a good example in which, although the Mean Square Error of both images is the same –that is 0.04 –, it is easy to realize the visual quality in the second

case is drastically worse as it is obtained a very noisy image, while the first one just changes the luminance of all its values in the same way.



Figure 1.a. Original image



Figure 1.b. Distorted image by adding a constant to the original ($MSE = 0.04$)



Figure 1.c. Distorted image by adding an alternated constant to the original ($MSE = 0.04$)

After verifying these numerical methods fail in the assessment of image quality, it appears the need of using a model of evaluation that better correlates with the human vision performance.

1.2. Human vision models

These mathematical models define quality metrics estimating the different blocks of processing that occur in the Human Visual System. First, they have a preprocessing stage in which they fix parameters of calibration and registration. They perform a mapping in the images to units of visual frequencies or cycles per degree of

visual angle after receiving input parameters as the viewing distance, screen resolution, fixation depth or eccentricity of the images in the observer's visual field. They also take into account in the first stage the establishing of the point-by-point correspondence between the original and distorted images. In the second block, called frequency analysis, they decompose both images in channels with different spatial frequencies and orientations to mimic the processing occurred in the Human Visual System since neurons in the visual cortex respond selectively to stimuli with particular spatial frequencies and orientations. Another stage represents the contrast sensitivity that can be thought as a band pass filter in each frequency channel. *Figure 2* illustrates an estimation of the Contrast Sensitivity Function (CSF), expressing the spatial frequency in cycles per degree of visual angle. As a consequence, it can be concluded that the visibility of details at a concrete frequency it is related to the viewing distance.

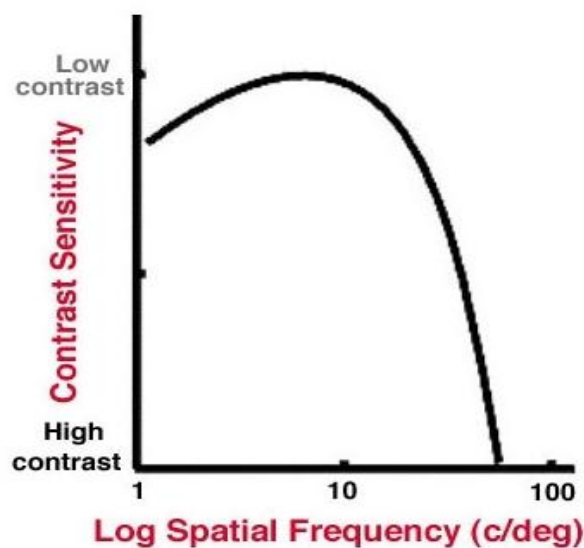


Figure 2. Contrast sensitivity function [2]

Human vision models also take into account the luminance masking that is the perception of lightness or the light adaptation of the human eye. It refers to the fact that the luminance of the original image masks the variations in the resulting image. And finally, they are defined concerning about the contrast masking that is the reduction of visibility of an image component by the appearance of another with similar spatial location and frequency content [1].

1.3. Image quality for lesion detection tasks in medical imaging

In addition, to ensure image quality, that is the goal of the proposed algorithm, it should be used a model that evaluates the quality of the images by the degree they achieve their purpose. More specifically, according to the cases of study in this report, this purpose is to detect lesions in medical images, thus the quality will be judged by the ability of an observer to detect the lesion within the images.

Although human observers are medical imaging main users in lesion detection tasks and, as a consequence, they are the best agents to optimize image quality assessment algorithms with that purpose, their measurements can be in many cases too much costly and time-consuming. That is why numerical observers (also called model observers) acting as surrogates of human observers in the detection of lesions are becoming an important tool in image quality assessment.

So, to optimize the correlation between human and numerical observers in lesion detection tasks in medical images, this manuscript revises some lesion detection theory in which the numerical observers defined after are based and it goes into the Channelized Hotelling Observer by defining its components and studying two particular cases in which this criterion is judged by comparing the results to different human data.

CHAPTER 2

NUMERICAL METHODS FOR LESION DETECTION

2.1. Lesion detection tasks

In a lesion detection task, as in any binary decision, the observer provides a score to classify the object into one of two classes of hypotheses: H_0 and H_1 ; representing lesion absent or present, respectively. D_0 and D_1 refer to the decisions made by the observer of being in each hypothesis.

As the observations are made in data obtained through some imaging system and contaminated by noise, they will not always be correct and four different situations, represented in *Table 1*, can exist in each experiment [3][13]. The first possible outcome is the True Positive (TP), also called Hit, which refers to the decision of lesion present when it is. The False Positive (FP), or False Alarm, makes the same decision but in an image it has not lesion. In the False Negative (FN), or Miss, the observer decides the hypothesis of lesion absent when it should have detected it and, finally, in the True Negative (TN) there is not lesion and the algorithm decides H_0 .

	Hypothesis	Decision	
True Positive (TP)	H_1	D_1	Hit
False Positive (FP)	H_0	D_1	False alarm
False Negative (FN)	H_1	D_0	Miss
True Negative (TN)	H_0	D_0	-

Table 1. Decision results

The performance of an observer is fully specified by the probability of two of the scenarios above. Hence, to see how well a detector works these parameters should be defined likewise in a statistical perspective. Therefore, being $p(\mathbf{x}|H_0)$ the probability of lesion absent and $p(\mathbf{x}|H_1)$ the probability of lesion present, the True Positive Fraction refers to the probability of the observer of making hits, the False Positive Fraction the probability of false alarms, the False Negative Fraction the probability of misses and the True Negative Fraction the probability of making a decision of no lesion when the lesion is absent.

The values of these fractions depend on a decision threshold or criterion γ as the observer decides D_0 when the test statistic is lower than the threshold and decides lesion present when it is higher. *Figure 3* represents these concepts.

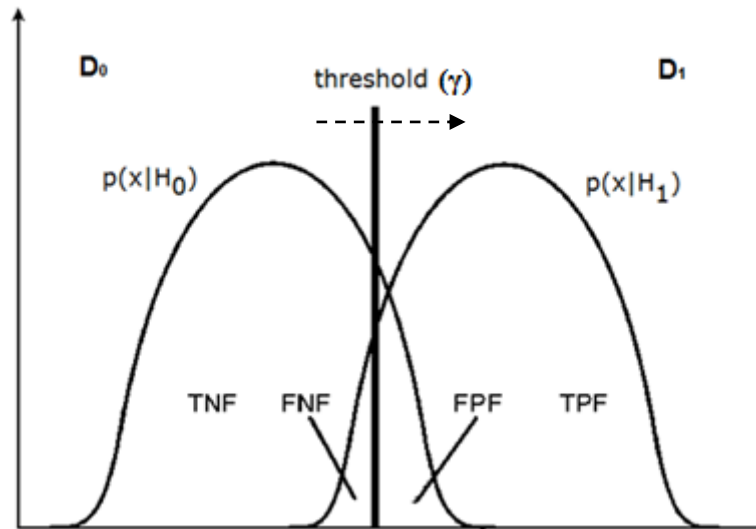


Figure 3. Probability density functions under H_0 and H_1 and the resulting TPF, FNF, FPF and TPF for one value of γ [15].

A widely used performance measure defined using the True Positive Fraction and the False Positive Fraction is the Receiver Operating Characteristic or ROC curve. It is computed changing the threshold γ from $-\infty$ to $+\infty$ and plotting the TPF over the FPF for all the values. In *Figure 3* and *Figure 4* the discontinuous arrow represents the direction in which the threshold is moved and the ROC curve plotted.

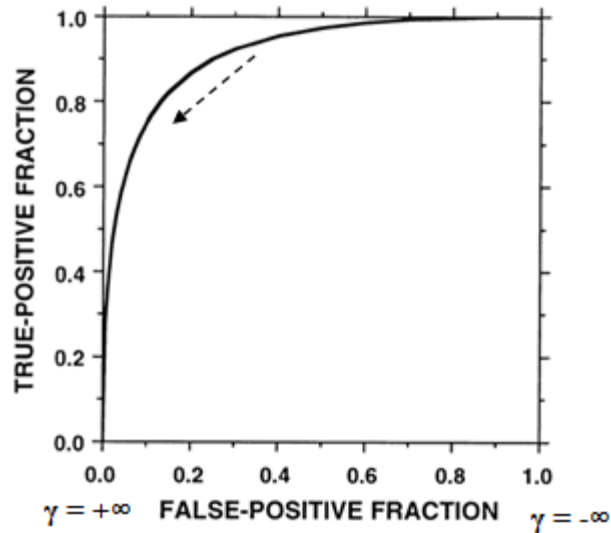


Figure 4. Receiver Operating Characteristic [16].

It is quite simple to see that, related to the overlap of the probability density functions, a bad detector would be one having a ROC curve close to the 45° line, while the best possible would be formed by the left and top borders of the square because that would mean that for all the threshold values it performs 100 % correct detections.

Another well-known performance measure is the Area Under the ROC curve (*AUC*). It ranges from 0 to 1 and it is defined as

$$AUC = \int_0^1 dFPF \text{ TPF}(FPF)$$

Both figures of merit represent a measure of the overlap of the probability density functions, and so the separability of the hypothesis. However, the *AUC* computes the average TPF over all FPF giving a more general idea of the probability of correct detections of the algorithm. Accordingly, preferable detectors would be those with a higher *AUC*.

However, in this report the goal is to fit the *AUC* of our system with the Area Under the Curve of the human observer it is being estimated.

2.2.Numerical observers

In the images used in this report, the quality is defined depending on the degree a human observer can detect a lesion correctly on them. Consequently, it would make sense assessing this quality by computing the Area Under the ROC Curve of some human observer, but this is, in most cases, costly and time consuming and that is why model observers have gained a lot of importance lately. They are numerical methods

that can perform lesion detection tasks acting as surrogates of human observers, trying to mimic the human observer performance, if their parameters are properly adjusted.

All model observers perform a scalar test statistic $t = T(\mathbf{g})$ where \mathbf{g} is the raw image and $T(\cdot)$ the discriminant function. When t is higher than some threshold, they decide the image belongs to H_1 , while when t is lower, H_0 is the hypothesis selected. Then, by classifying the number of images correctly detected for the different values of the threshold it can compute the ROC curve and the *AUC*.

As there are many numerical observers, they can be classified whether their discriminant function is optimal or suboptimal [3]. We refer to optimal observers when they are statistically efficient because they perform the best possible detection between all possible observers, even better than the human observer; while suboptimal model observers adjust their parameters to estimate the human performance.

Early efforts concentrated on the first group and more specifically on the Ideal or Bayesian Observer [4]. It is the best possible because it obtains the highest *AUC* between all observers (meaning it computes maximum number of True Positives for any False Positives) by perfectly performing the likelihood ratio test as it has all the information needed. The likelihood ratio is defined as the probability of having lesion over the probability of not having it:

$$t(\mathbf{g}) = \frac{p(\mathbf{g}|H_1)}{p(\mathbf{g}|H_0)}$$

However, the ideal observer may not be possible to calculate because if the noise is not Gaussian or the lesion or background on which it is superimposed are random, we may not have information enough about the data statistics to calculate the likelihood ratio or the detection performance associated with it.

Thus, linear observers -by definition, observers that use linear discriminant functions- should be considered to facilitate the performance of the test statistic. In this case, they make calculations by functions with the following form, being \mathbf{w} an $M \times 1$ template:

$$t = \mathbf{w}^T \mathbf{g}$$

When the ensemble mean and covariance of \mathbf{g} are known, the best possible linear observer that maximizes the Signal to Noise Ratio (*SNR*) between all possible linear observers is the Hotelling observer. It uses the following template, being \mathbf{K}_g the ensemble data covariance and $\Delta \bar{\mathbf{g}}$ the difference in the mean data vector under the two hypotheses:

$$\mathbf{w}_{\text{Hot}} = \mathbf{K}_g^{-1} \Delta \bar{\mathbf{g}}$$

Moreover, when the lesion and the background are nonrandom and the noise distribution is Gaussian this calculation becomes even simpler and, given that the linear observer's test statistic is a linear weighted sum of many random variables, the

Gaussian assumption is, because of the central-limit theorem, usually a valid assumption.

Under these assumptions, the two hypotheses can be redefined as Gaussian distributed signals where \mathbf{n} represents the Gaussian noise, \mathbf{b} the known background and \mathbf{d} the defect or lesion:

$$H_0: \quad \mathbf{g} = \mathbf{b} + \mathbf{n}$$

$$H_1: \quad \mathbf{g} = \mathbf{b} + \mathbf{d} + \mathbf{n}$$

The Signal to Noise Ratio and the Area Under the ROC curve can be then calculated and related according to:

$$SNR = \frac{\langle \mathbf{g} \rangle_1 - \langle \mathbf{g} \rangle_0}{\sqrt{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_0^2}}$$

$$AUC = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{SNR}{2}\right)$$

where $\langle \mathbf{x} \rangle_i$ and σ_i are the mean value and the variance of each hypothesis i .

Because of the large dimensionality of current images it may be necessary to use channels to preserve the information in the data while enabling the observer performance. Then, a Channelized Hotelling Observer (CHO) can be designed to perform an optimal linear observer in spite of the reduced dimensionality. Any linear channel model can be represented by an $M \times P$ matrix \mathbf{U} whose columns are the channel profiles \mathbf{u}_p , transforming the raw data \mathbf{g} to the output vector, or feature vector \mathbf{x} :

$$\mathbf{x} = \mathbf{U}^T \mathbf{g}$$

In addition, these channels can be chosen to suit different purposes: optimal or suboptimal. They could be designed to be efficient, giving minimal loss of detectability, or to estimate the human observer performance by imitating the Human Visual System processing.

The goal of this study is to estimate the human performance because it will be the final user of the images to detect lesions. Hence, the next chapter goes into the suboptimal use of the Channelized Hotelling Observer to assess image quality.

2.3. Channelized Hotelling Observer

The Channelized Hotelling Observer is a numerical observer that acts as a surrogate of human observers in the assessment of image quality in lesion detection tasks and it is the cascade of two linear operators that can be combined into one [5][6].

The first one is the channeling operator (\mathbf{U}). It measures the feature vectors of the image by applying filters that estimate the Human Visual System intending to reflect the response of neurons in the primary visual cortex. Being \mathbf{f} the image, the feature vector is defined as

$$\mathbf{x} = \mathbf{U}\mathbf{f}$$

In this paper the experiments are tested with the four most common channeling operators:

- **CHO-BP:** Rotationally symmetric, square profile, non-overlapping, band pass filters with cutoff frequencies of $\frac{1}{32}$, $\frac{1}{16}$, $\frac{1}{8}$, $\frac{1}{4}$ and $\frac{1}{2}$ cycles per pixel, resulting in four channels.

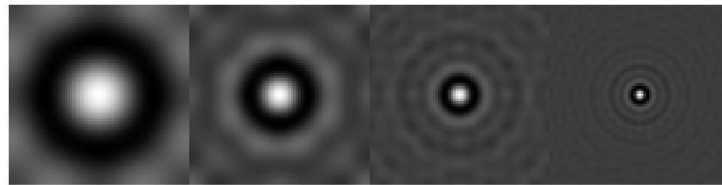


Figure 5. Band pass filters

- **CHO-GB:** Gabor filters with three spatial frequency bands of $\frac{1}{8}$, $\frac{1}{4}$ and $\frac{1}{2}$ cycles per pixel and eight orientations from 0 to π , resulting in 24 channels.

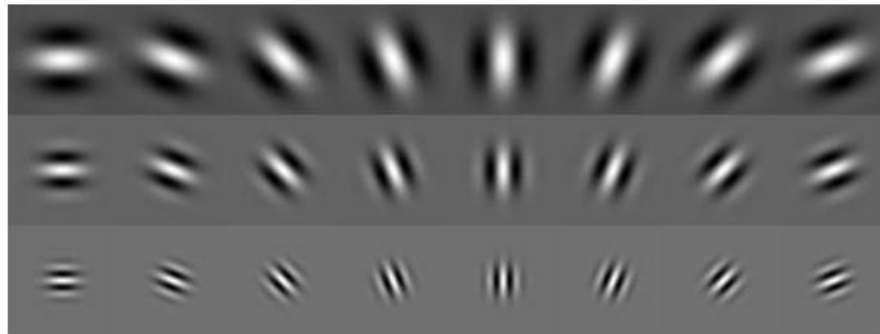


Figure 6. Gabor filters

- **CHO-LG:** Laguerre-Gauss polynomials of order 0 through 6 with vertical, horizontal and rotationally invariant orientations, resulting in 18 channels.

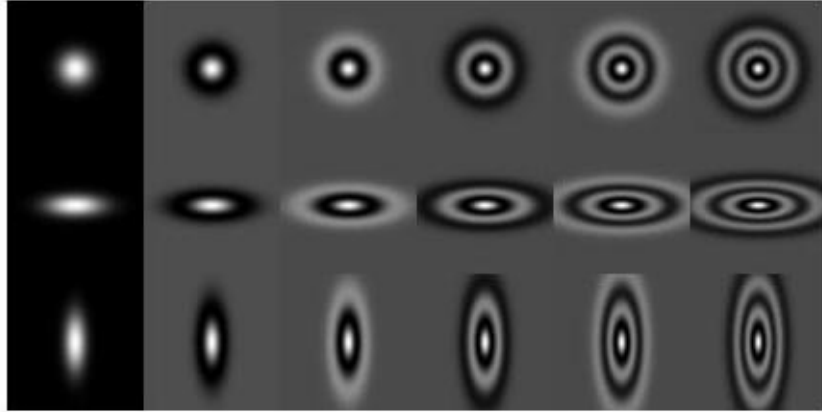


Figure 7. Laguerre-Gauss filters

- **CHO-DOG:** Rotationally symmetric, overlapping, difference-of-Gaussians (DOG) profiles, resulting in 10 channels.

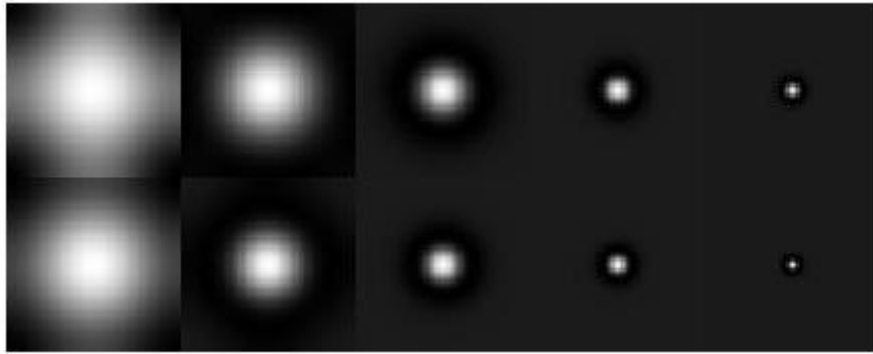


Figure 8. Difference Of Gaussians filters

All channels are designed to have non-zero values on a 71×71 pixel window centered at the lesion location and they are normalized, being M the number of channels, as follows:

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M]^T$$

$$\mathbf{u}_i^T \mathbf{u}_i = 1, \quad i = 1, 2, \dots, M$$

As defined later, the Channelized Hotelling Observer introduces an Internal-noise model to better estimate human performance inaccuracies and avoid outperforming the human observer. It is a noise vector $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{K}_{\text{int}})$ injected into all channel outputs, becoming:

$$x_i = \mathbf{u}_i^T \mathbf{f} + \varepsilon_i, \quad i = 1, 2, \dots, M$$

The second operator is the Hotelling Observer, which computes a test statistic for choosing between the two hypotheses (lesion absent H_0 or present H_1) based on the observed feature vector \mathbf{x} computing the likelihood ratio test and comparing the result λ to a threshold γ and then choosing H_1 when λ is greater than γ or H_0 if it is lower than γ .

More specifically, if it is a signal-known-exactly (SKE) with Gaussian noise, the test statistic can be expressed

$$t(\mathbf{x}) = \mathbf{w}^T(\mathbf{x} + \boldsymbol{\varepsilon})$$

where

$$\mathbf{w} = \mathbf{K}^{-1}\mathbf{U}\Delta\bar{\mathbf{f}}$$

being $\mathbf{K} = \mathbf{K}_{\text{ext}} + \mathbf{K}_{\text{int}}$, where \mathbf{K}_{ext} expresses the noise covariance originating in the data and not in the visual system and it is defined as:

$$\mathbf{K}_{\text{ext}} = \frac{1}{2} \sum_{j=0}^1 \mathbf{U} \left[\langle [\mathbf{f} - \langle \mathbf{f} \rangle_j] [\mathbf{f} - \langle \mathbf{f} \rangle_j]^T \rangle_{H_j} \right] \mathbf{U}^T$$

In addition, some models show that human observers inconsistencies can be described by a Decision-variable noise model, which is a noise $\gamma \sim N(\mathbf{0}, \sigma_\gamma^2)$ injected into the test statistic:

$$t(\mathbf{x}) = \mathbf{w}^T(\mathbf{x} + \boldsymbol{\varepsilon}) + \gamma$$

Going into to the Internal-noise models (and the Decision-variable noise), they represent the inconsistencies of the human eye as well as the variance in detections caused by factors like tiredness and mood of the human observer, or weather inside and outside the room, among others.

Some of them have some parameters to adjust empirically by computing the minimum error iteratively between the human observer and the numerical observer results. The first model does not use any Internal or Decision-variable noise, models from 2 to 7 have Internal-noise but no Decision-variable noise, models 8 to 10 have Decision-variable noise but no Internal-noise and model 11 incorporates both noises. These are the models presented in this paper:

- **Model 1:** No Internal or Decision-variable Noises

$$\mathbf{K}_{\text{int}} = \mathbf{0}$$

$$\sigma_\gamma^2 = 0$$

Number of parameters to adjust: 0

- **Model 2:** Quantization Noise

$$[\mathbf{K}_{\text{int}}]_{i,j} = \frac{Q^2}{12}$$

being

$$Q = \frac{\max\langle f \rangle_{H_0 \cup H_1} - \min\langle f \rangle_{H_0 \cup H_1}}{256}$$

Number of parameters to adjust: 0

- **Model 3:** Uniform-variance Internal Noise

$$[\mathbf{K}_{\text{int}}]_{i,j} = \sigma^2$$

Number of parameters to adjust: 1

- **Model 4:** Non-uniform Internal Noise variance, proportional to External Noise variance

$$[\mathbf{K}_{\text{int}}]_{i,j} = p[\mathbf{K}_{\text{ext}}]_{i,j}$$

Number of parameters to adjust: 1

- **Model 5:** Uniform Internal Noise variance, proportional to the maximum External Noise variance

$$[\mathbf{K}_{\text{int}}]_{i,j} = p \max_j [\mathbf{K}_{\text{ext}}]_{j,j}$$

Number of parameters to adjust: 1

- **Model 6:** Non-uniform Internal Noise variance, proportional to External Noise standard deviation

$$[\mathbf{K}_{\text{int}}]_{i,j} = p \sqrt{[\mathbf{K}_{\text{ext}}]_{i,j}}$$

Number of parameters to adjust: 1

- **Model 7:** Non-uniform compound noise

$$[\mathbf{K}_{\text{int}}]_{i,j} = p_1 [\mathbf{K}_{\text{ext}}]_{i,j} + p_2 \max_j [\mathbf{K}_{\text{ext}}]_{j,j}$$

Number of parameters to adjust: 2

- **Model 8:** Constant variance Decision-variable Noise

$$\sigma_y^2 = \sigma^2$$

Number of parameters to adjust: 1

- **Model 9:** Decision-variable variance proportional to the External-Noise standard deviation

$$\sigma_y^2 = p \sigma_{\text{ext}}$$

Number of parameters to adjust: 1

- **Model 10:** Decision-variable variance proportional to the External Noise variance

$$\sigma_y^2 = p\sigma_{ext}^2$$

Number of parameters to adjust: 1

- **Model 11:** Combination

$$[\mathbf{K}_{int}]_{i,j} = p_1[\mathbf{K}_{ext}]_{i,j}$$

$$\sigma_y^2 = p_2\sigma_{ext}^2$$

Number of parameters to adjust: 2

To adjust the Internal Noise, as it is said above, the training was based on minimizing the Mean Square Error. First, by searching the minimum error on a coarse grid ranged over 14 orders of magnitude from 10^{-7} to 10^{10} to ensure the best fit between the human observer and the numerical observer *AUCs*. Then, the search was done in a finer grid, by focusing to a range that spanned four orders of magnitude where the minimum was located. Finally, at each consecutive iteration, the search range was reduced by an order of magnitude for a total of ten iterations.

CHAPTER 3

EXPERIMENTS

3.1. Myocardial SPECT for perfusion defect detection

In the first experiment, the images are from the myocardium, a muscle of the heart in charge of pumping the blood through the circulatory system, which, consequently, when it has a defect, it does not give enough blood to the heart. The data that better evaluate the presence of that defect are the images taken when the patient is practicing some exercise since under the realization of effort the heart needs more activity and it is easier to see if some coronary arteries do not work properly. This defect is also called ischemia. To proceed with the evaluation, the patient is given an intravenous radioactive tracer (Thallium-201, Isonitriles or Tetrofosmin marked with Technetium-99m) to study the status of myocardial perfusion at rest, during physical exertion or after drug administration to simulate the effort [7]. In *Figure 9* noise free images with the defect present and absent are represented and in the first image the defect is located.

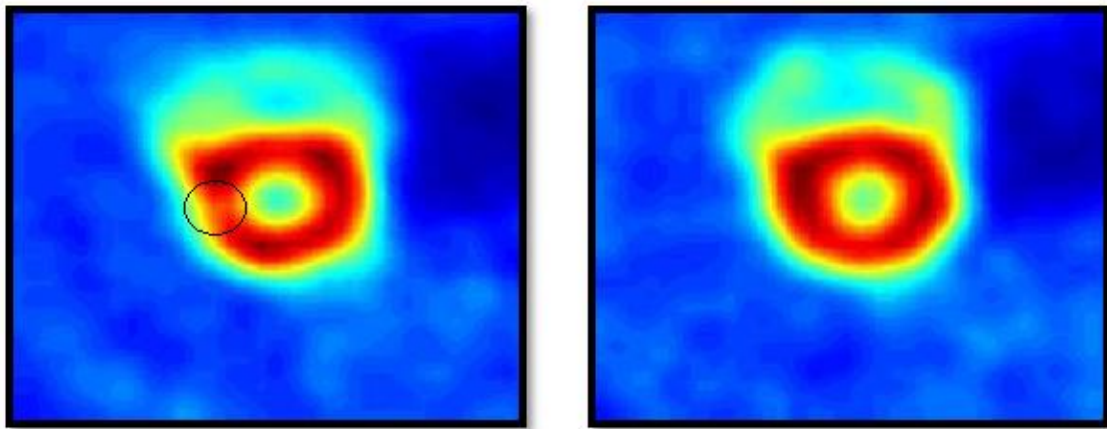


Figure 9. Images with defect present (left) and absent (right) from a myocardial SPECT for perfusion defect detection.

In addition, they are cross-sectional images and specifically, coronal cuts (see *Figure 10* [8]) in which the colors going from red to blue refer to the intensity of blood flowing from highest to lowest. Then, the lesion, as it can be seen in *Figure 8* is the lack of blood flux in a certain location of the myocardium.

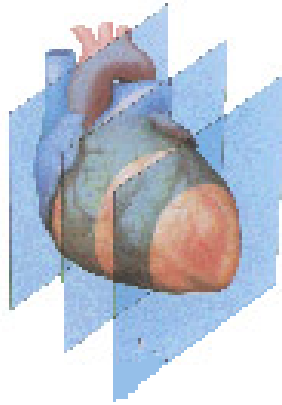


Figure 10. Coronal cuts of the heart [9].

The images had been extracted by Single-Photon Emission Computed Tomography (SPECT) that is an Emission Tomography technique. This means it uses radioactive materials to image properties of the body's physiology and, as any ET technique, it is based on the Tracer Principle, which claims radioactive and nonradioactive materials participate in an organism in the same way and as the first ones can be detected by their emission of gamma rays, we can track the flow and distribution of some substances. Specifically, the SPECT uses radiopharmaceuticals labeled with a single-photon emitter (like the Tetrofosmin marked with Technetium-99m named before) that means they emit a gamma ray in each radioactive decay event [9].

More specifically, the images used in this report were reconstructed using the ordered-subsets expectation-maximization (OSEM) algorithm with one and five effective iterations [10]. The images were also low-pass filtered with three-dimensional Gaussian filters with different full-widths at half-maximum (*FWHM*): 0, 1, 2, 3, 4 or 5 pixels. The combination of all the parameters yields in 12 different strategies of which we had 100 noisy image realizations: 50 with defect present and 50 with defect absent. In *Table 2* a realization of some combinations is represented.

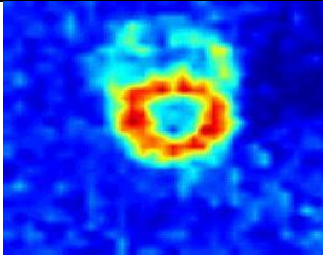
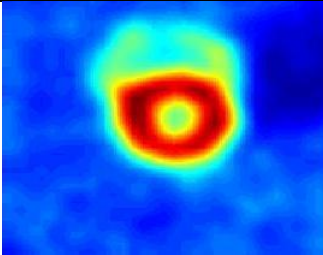
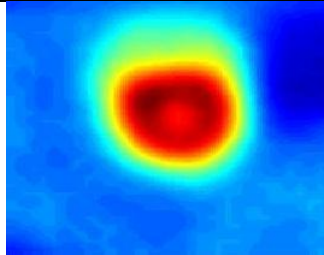
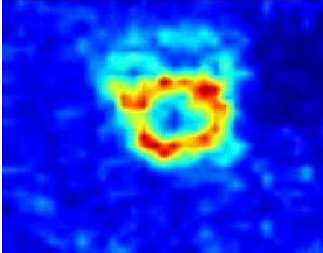
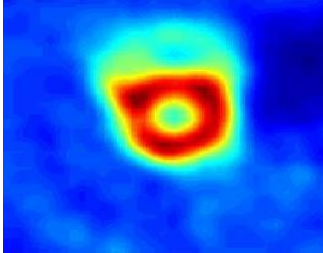
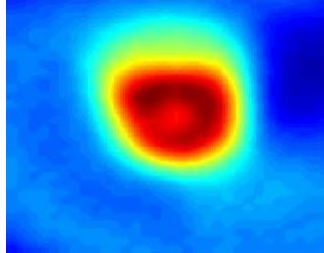
<i>FWHM</i> (pixels)	0	2	4
Lesion absent			
Lesion present			

Table 2a. Example of OSEM reconstructed images with 1 effective iteration

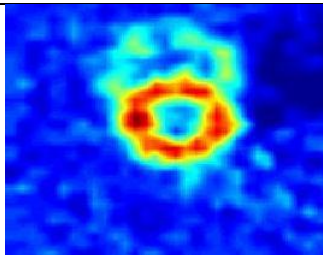
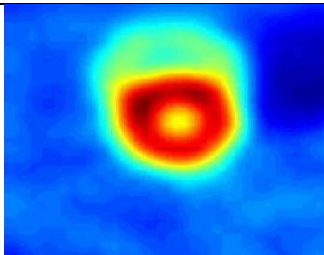
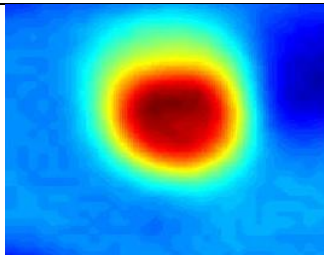
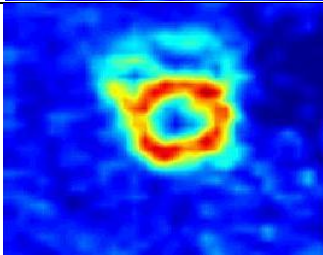
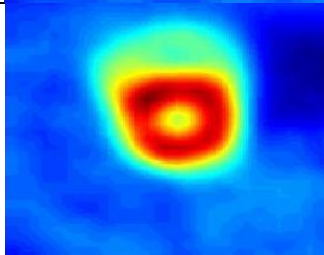
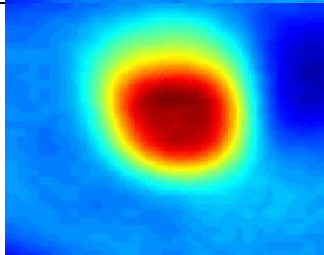
<i>FWHM</i> (pixels)	1	3	5
Lesion absent			
Lesion present			

Table 2b. Example of OSEM reconstructed images with 5 effective iterations

a) *Human Observer Data*

The first step in the experiment was obtaining the human observer scores. Two medical physicists evaluated the defect visibility for all the images, one by one, at every combination of the number of iterations and *FWHM* of the filter on a six-point scale. Then, it was computed the mean of the AUC curves between the observers for both iterations by calculating the Area Under the Curve of each value of *FWHM* comparing the human scores with the ground truth with ROCKIT [11]. In *Figure 11* the AUC curves for the Human Observer data in the different iterations (iteration 1 in blue and iteration 5 in green) are represented by relating the *AUC* with each *FWHM* value.

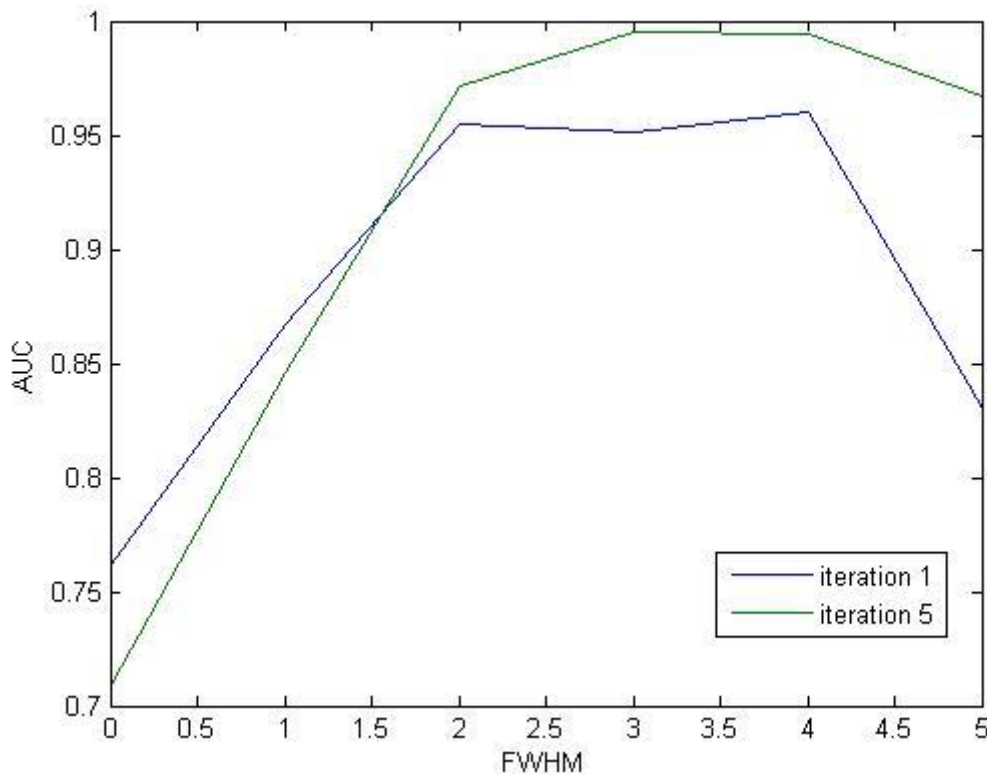


Figure 11. Human observer defect detection performance measured by *AUC* in relation to *FWHM*

b) *Numerical Observer Data*

For each noise model and channeling operator two different evaluations were done. However, both evaluations were based on dividing the study in two parts, in the use of two sets of data: trained and tested images. The first set is used to optimize the Internal-noise parameters with the human data available to ensure that the best match with them is found; while the tested images are used to evaluate the performance of the Numerical Observer by using the optimized Internal-noise obtained in the training session.

In the first evaluation, also called comparison 1, the images used in the training and testing sets are of the same kind but different noise realizations. Although the ability of the Numerical Observer to fit the Human Observer data is a necessary, it is not a sufficient property to be useful. Actually, there is no need to apply a Numerical Observer on images of the same kind as those belonging to the training set since the Human Observer performance for that reconstruction method is already available. So, by testing with images that were not present in the training phase but they are produced by the same reconstruction algorithm the Numerical Observer is satisfying a general train-testing paradigm but it is not demonstrating usefulness for its applications.

Then, it is needed a Numerical Observer trained not only to predict Human Observer performance on images reconstructed by one algorithm, but also to estimate lesion detection performance as a measure of image quality for reconstruction methods not yet evaluated by a human study. Therefore, to be useful, the CHO must accurately predict human performance over a wide range of image reconstruction parameter settings for which human data are not available. Thus, the NO must exhibit good generalization properties. In the comparison 2, the study was based on the train-testing generalization most representative of the practical use of the Numerical Observer. In this experiment, after the adjustment of the Internal-Noise models on a broad set of images, the CHO was then tested on a different, but equally broad, set of images. Specifically, for the realization of this report, the CHO was trained using 50 images (half with lesion present and half with lesion absent) for every value of *FWHM* and one iteration of OSEM, and then tested using the rest of the images for every value of *FWHM* with five iterations of OSEM. Then the process was repeated with the roles of one and five iterations reversed. By this procedure it can be seen how, by knowing the Human Observer data of an specific effective iteration, other iterations can be estimated.

c) *Results*

Best results were obtained using the Internal-noise model 6 that estimates the internal noise proportional to the standard deviation of the external noise. In *Figures 12-19* the resulting AUC curves are represented, being the red one the Human Observer performance and the blue curve the Numerical Observer estimation.

Figures 12-13 have been obtained using Band Pass filters as channeling operator, *Figures 14-15* using Gabor filters, *Figures 16-17* with Laguerre-Gauss filters and *Figures 18-19* for Difference of Gaussians filters. For each channeling operator, first figures are the result of training with iteration 1 and testing with iteration 5, while second figures have been obtained after training with iteration 5 and then testing in iteration 1.

- *CHO-BP*

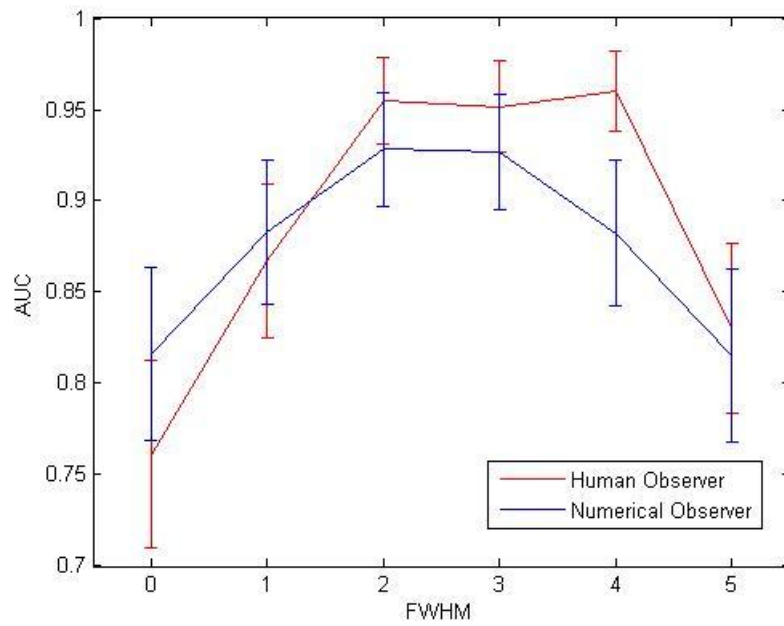


Figure 12. AUC curves of the Human Observer (red) and Numerical Observer (blue) when training with iteration 1 and testing with iteration 5 using Band Pass filters and Internal-noise model 6.

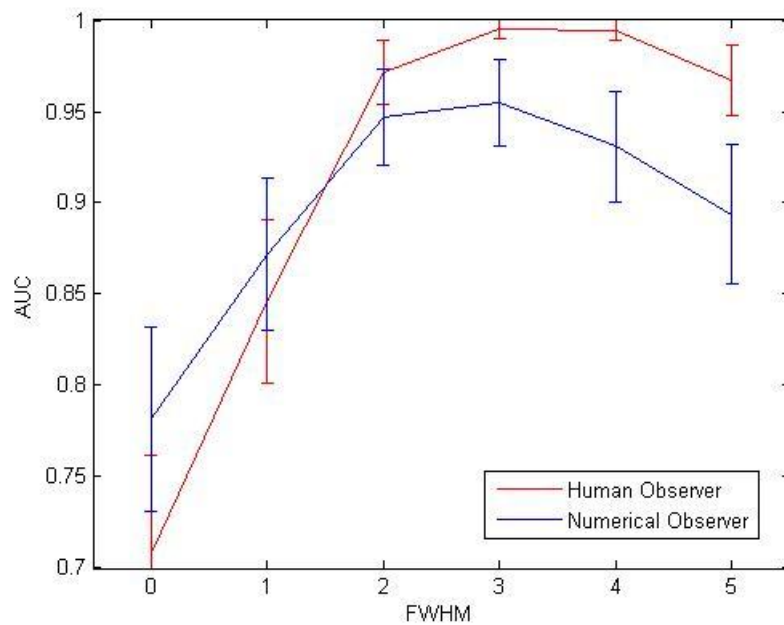


Figure 13. AUC curves of the Human Observer (red) and Numerical Observer (blue) when training with iteration 5 and testing with iteration 1 using Band Pass filters and Internal-noise model 6.

- *CHO-GB*

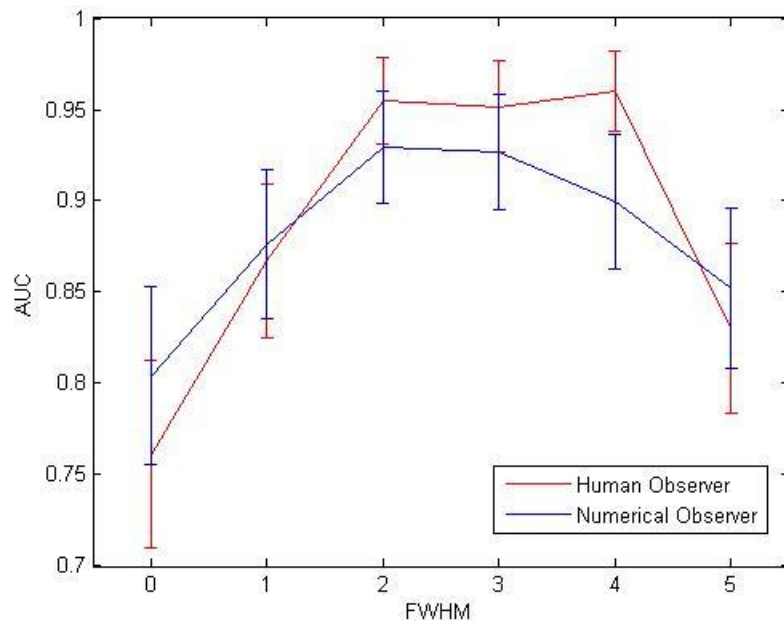


Figure 14. AUC curves of the Human Observer (red) and Numerical Observer (blue) when training with iteration 1 and testing with iteration 5 using Gabor filters and Internal-noise model 6.

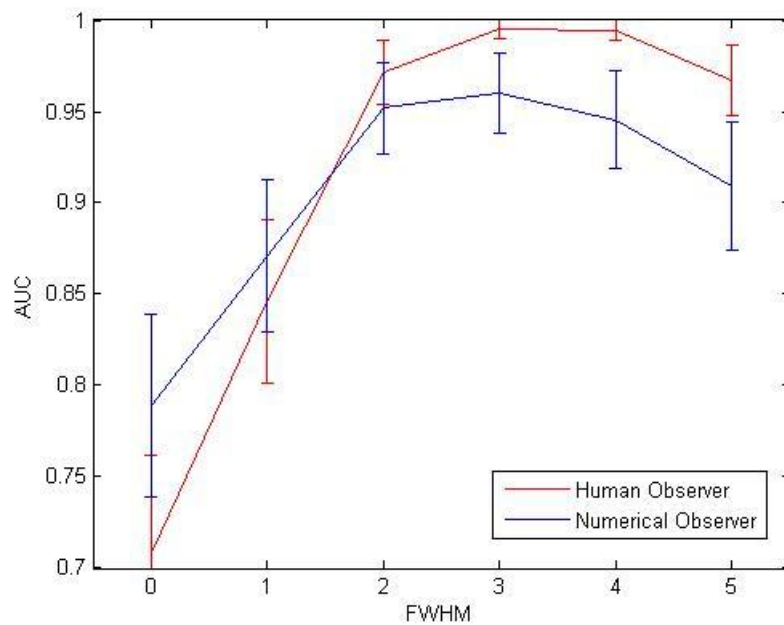


Figure 15. AUC curves of the Human Observer (red) and Numerical Observer (blue) when training with iteration 5 and testing with iteration 1 using Gabor filters and Internal-noise model 6.

- *CHO-LG*

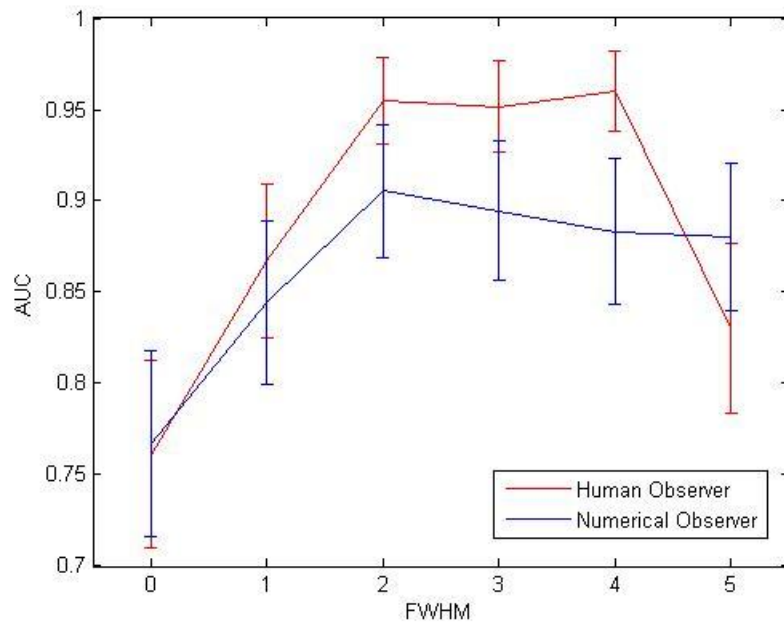


Figure 16. AUC curves of the Human Observer (red) and Numerical Observer (blue) when training with iteration 1 and testing with iteration 5 using Laguerre Gauss filters and Internal-noise model 6.

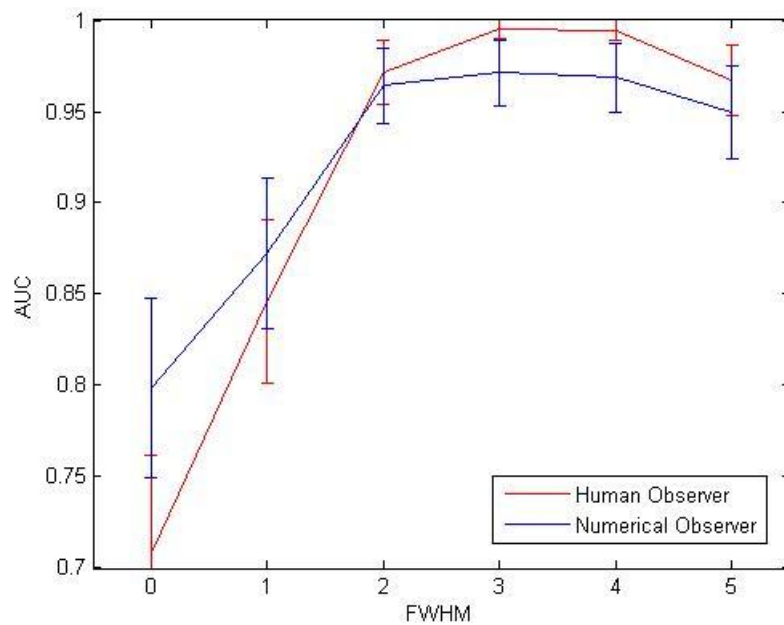


Figure 17. AUC curves of the Human Observer (red) and Numerical Observer (blue) when training with iteration 5 and testing with iteration 1 using Laguerre Gauss filters and Internal-noise model 6.

- *CHO-DOG*

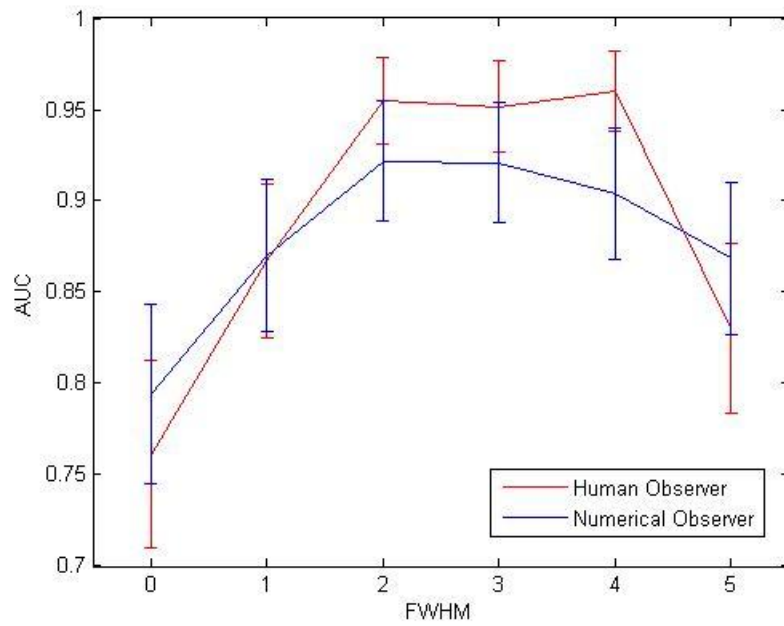


Figure 18. AUC curves of the Human Observer (red) and Numerical Observer (blue) when training with iteration 1 and testing with iteration 5 using Difference of Gaussians filters and Internal-noise model 6.

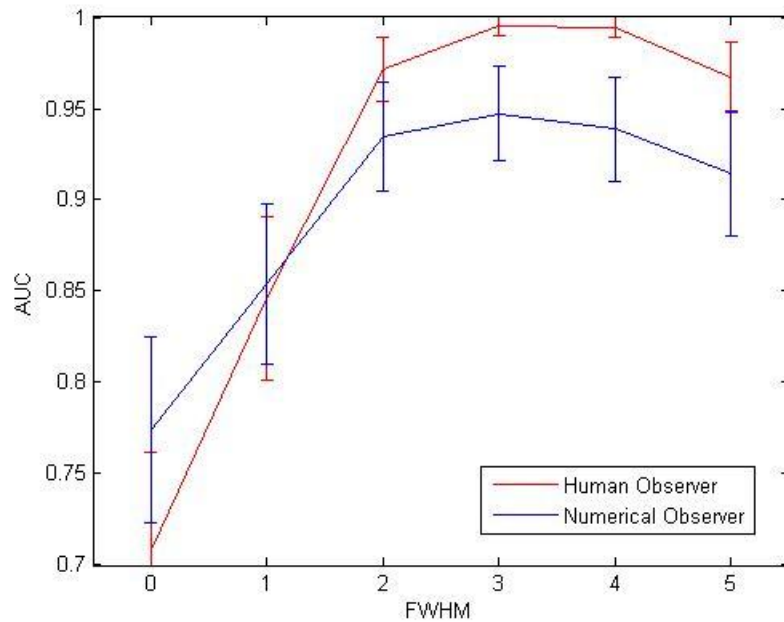


Figure 19. AUC curves of the Human Observer (red) and Numerical Observer (blue) when training with iteration 5 and testing with iteration 1 using Difference of Gaussians filters and Internal-noise model 6.

3.2. Defect detection in different correlated noises

In the second experiment, it is investigated the signal-known-exactly (SKE) detection in correlated noise having a power-law power spectrum where the considered defect is an additive focal Gaussian bump signal in a known location of the image. In *Figure 20* two example images (first one with lesion and the second without) are represented. In the first one the defect is marked and, then, in *Figure 21* this defect is displayed with noise reduced.

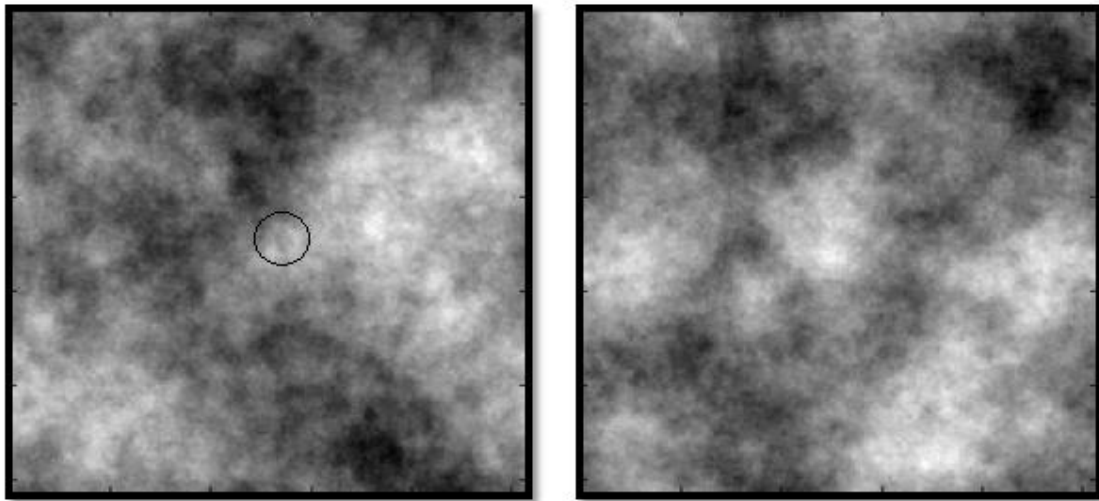


Figure 20 . Example images with defect present (left) and absent (right).

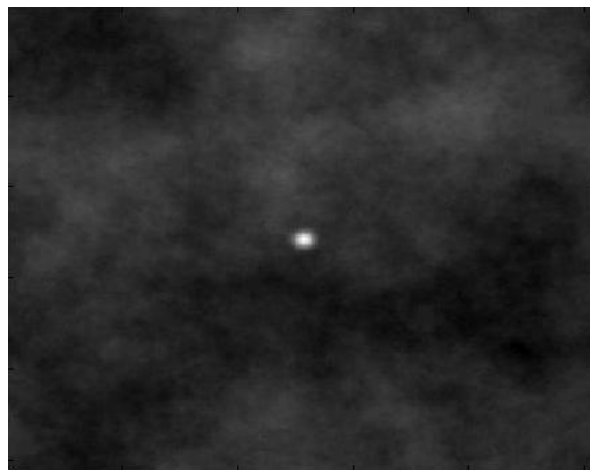


Figure 21. Defect or lesion

There are four possible background textures and they are characterized by a power-law function $S(u,v)$:

$$S(u,v) = C_{\beta}(u^2 + v^2)^{-\beta/2}$$

where β is the power-law exponent, and C_β is a normalizing constant set to yield a pixel standard deviation that is 20% of the mean background. These different textures are represented in *Figure 22* depending on the parameter β going from 0 to 3 [12].

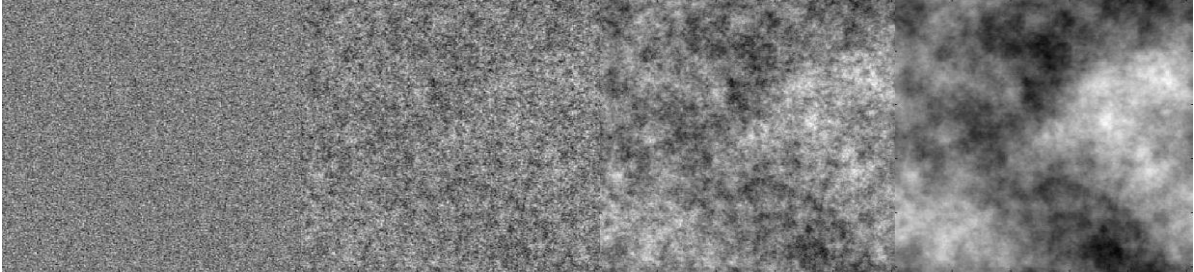


Figure 22. Power-law textures for $\beta = 0$ to 3 from left to right

Of each kind of background there were 2000 images: 1000 with the defect present and 1000 with it absent.

a) Human Observer Data

In this case, 2 different human observer scores were obtained to compute the experiment. They evaluated the defect visibility for all the images, one by one, with the order of the backgrounds and the images in each background randomized on a 0-1 scale. Then, it was computed the AUC curve of each observer by calculating the Area Under the Curve of each value of β comparing the human scores with the ground truth with ROCKIT [11]. In *Figure 23* the AUC curves for the different human observers are represented by relating the AUC with each β value.

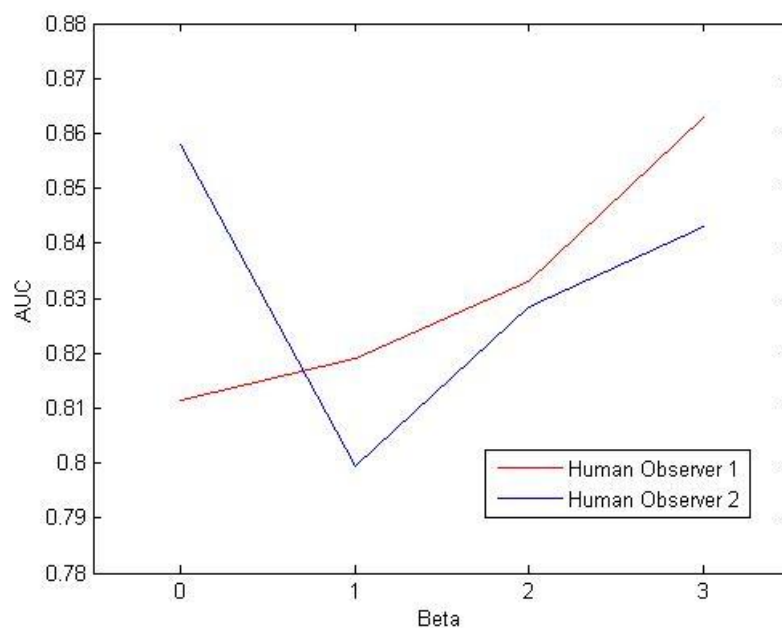


Figure 23. Human observers defect detection performance measured by AUC in relation to β

b) *Numeric Observer Data*

As it has done in the first experiment, two evaluations have been computed for each noise model and channeling operator combination depending on the images belonging to the training and testing sets.

In this case, the first comparison, which was previously defined as the one that trains and tests with the same kind of images but different noise realizations, was performed by training with the same number of images of each background in the training and testing sets. More specifically, it used 1000 images of each texture (half with lesion and half without) to train and then test the algorithm.

However, the goal of the study was to prove that by training with one specific kind of noise it is possible to estimate the Human Observer performance in front of other correlated noises. Therefore, in the second comparison it was trained with all the images of one texture, specifically with $\beta = 2$, and then tested in all available images in order to see how the Human Observer performance can be estimated by knowing it in a correlated noise.

c) *Mean and variance of the Area Under the Curve: Jackknife method*

It must be also pointed out how the mean and variance values for the *AUC* were estimated in this experiment: the Jackknife method [14]. It is based on the fact that the mean of a group of estimations is a good approximation of the total estimation. In this case it was applied by calculating 5 Area Under the Curves in groups of “leave-400 images-out” and then computing the mean of the 5 results. In other words, 5 different *AUCs* using 1600 images were calculated and the final *AUC* was extracted as the mean of all the groups. The *AUC* of each one is defined as:

$$\widehat{AUC}_{(i)} = \widehat{AUC}(\dots, x_{i-1}, x_{i+1}, \dots)$$

Then, the Jackknife estimation of the mean for the *AUC* was calculated by applying

$$\widehat{AUC}_{(.)} = \frac{1}{n} \sum_{i=1}^n \widehat{AUC}_{(i)} = \frac{1}{5} \sum_{i=1}^5 \widehat{AUC}_{(i)}$$

Likewise, the Jackknife estimation of the variance of the estimated *AUC* is

$$\text{Var}[\widehat{AUC}] = \frac{n-1}{n} \sum_{i=1}^n (\widehat{AUC}_{(i)} - \widehat{AUC}_{(.)})^2 = \frac{4}{5} \sum_{i=1}^5 (\widehat{AUC}_{(i)} - \widehat{AUC}_{(.)})^2$$

As we make 5 groups of *AUCs*, there can be 120 different possible combinations of choosing them. Accordingly, the final mean and variance will be the mean of the results of all the combinations.

d) Results

In *Figures 24-35* are represented the AUC curves of the Human Observer (red) and the Numerical Observer using different Internal-Noise models and Band Pass filters as channeling operator, in *Figures 36-47* with Gabor filters, *Figures 48-59* for Laguerre-Gauss filters and in *Figures 59-71* using Difference of Gaussians filters.

For each channeling operator, there are represented 12 graphs, 6 for each Human Observer. Then, for instance, *Figures 24-29* are for the first Human Observer, while *Figures 30-35* are for the estimation of the second HO when Band Pass filters are used.

For each combination of channeling operator and human observer the graphs in the left are those obtained when training and testing with images of all textures (first comparison), while the graphs in the right have been obtained by training with images where $\beta = 2$ and testing with images of all different textures (second comparison). Thus, using Band Pass filters in Human Observer 1, *Figures 24, 26* and *28* are the results of the first comparison, while *Figures 25, 27* and *29* are the resulting curves of the second comparison.

Finally, in each graph the Human Observer AUC curve is represented in red and in other colors the Numerical Observer estimations using Internal-noise models 1,2,3 and 4 in the first graphs (like in *Figure 24* and *25*), models 5, 6, 7 and 8 in the second ones (like *Figure 26* and *27*) and models 9, 10 and 11 in the last graphs (like in *Figures 28* and *29*).

- *CHO-BP*

Human Observer 1 : comparison 1 (left) and comparison 2 (right)

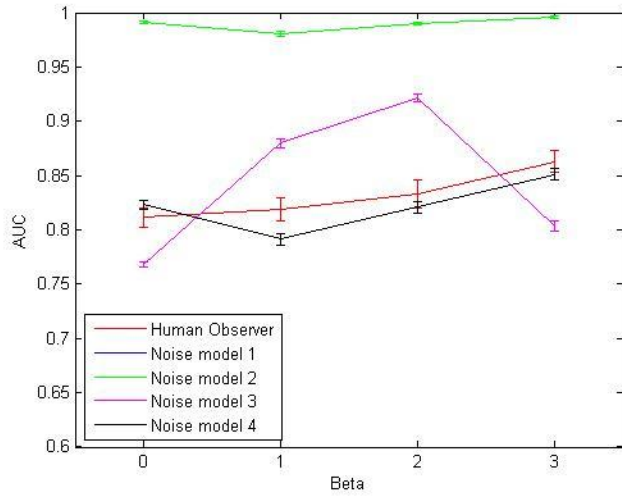


Figure 24

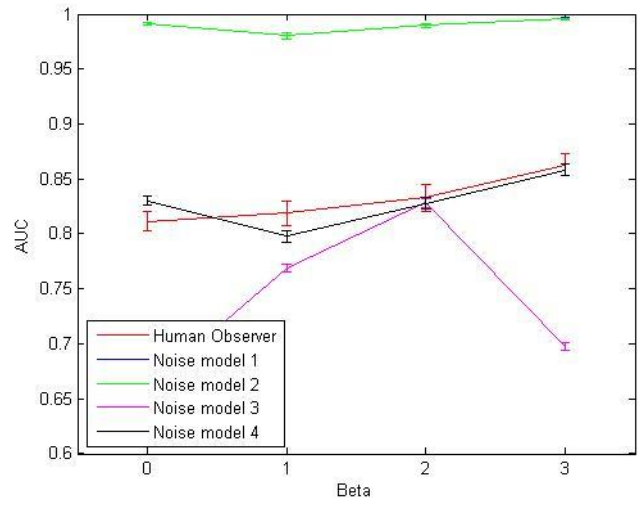


Figure 25

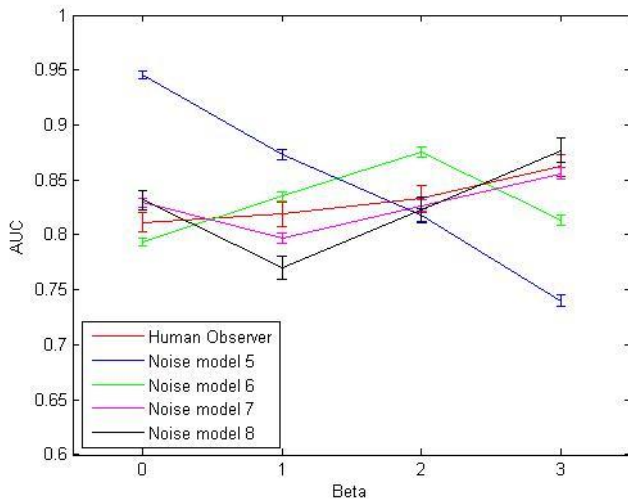


Figure 26

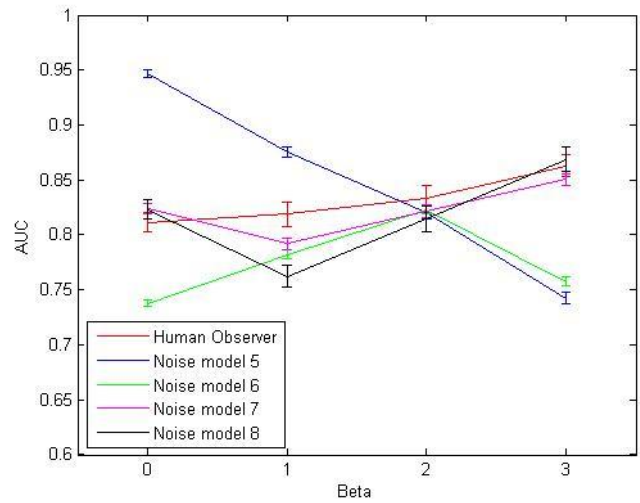


Figure 27

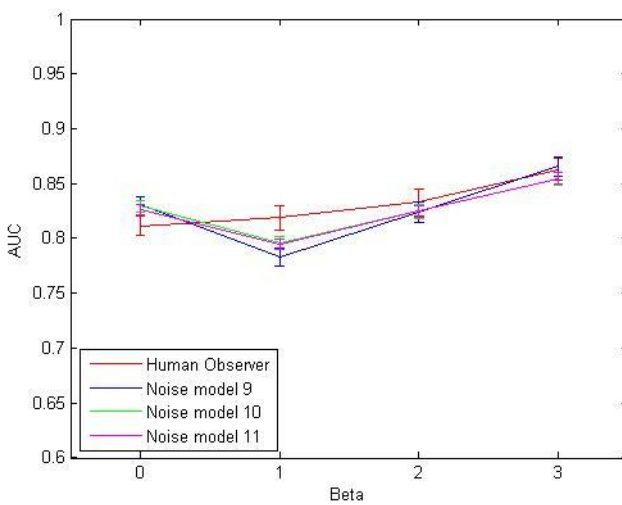


Figure 28

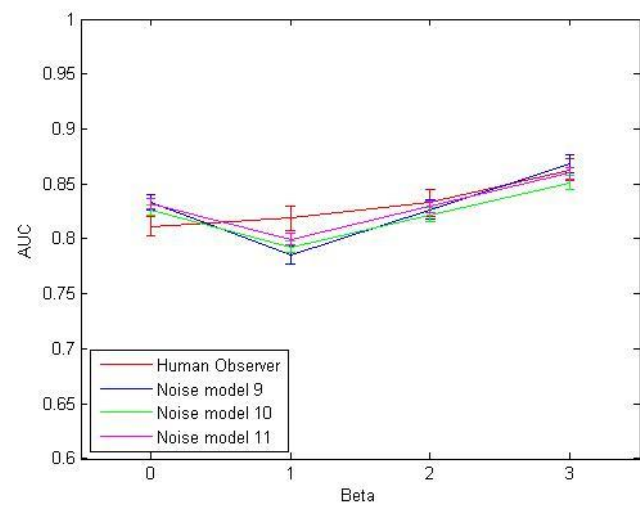


Figure 29

Human Observer 2: comparison 1 (left) and comparison 2 (right)

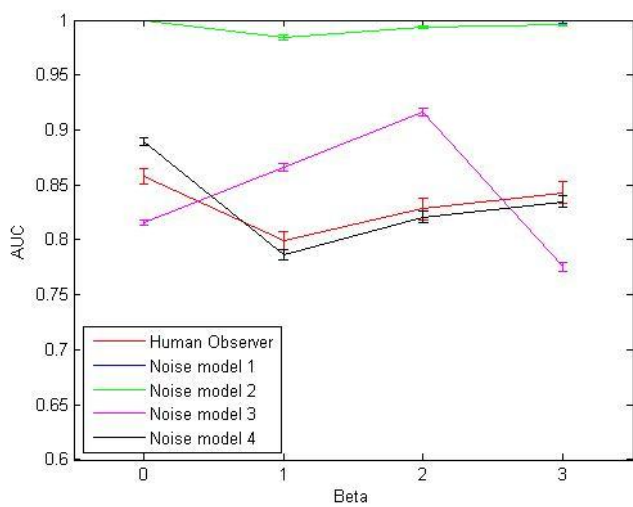


Figure 30

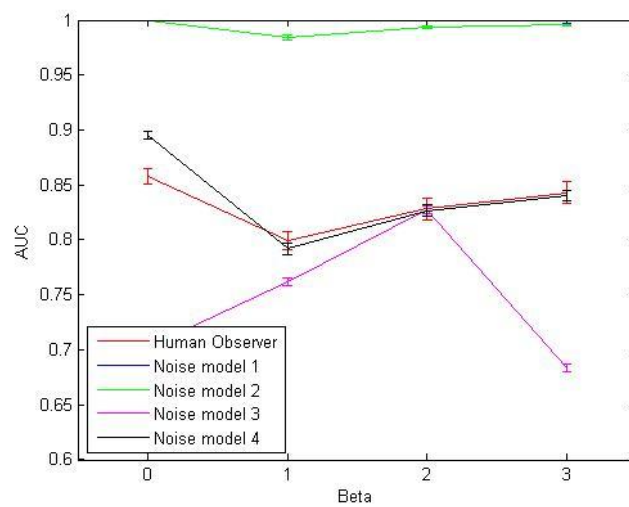


Figure 31

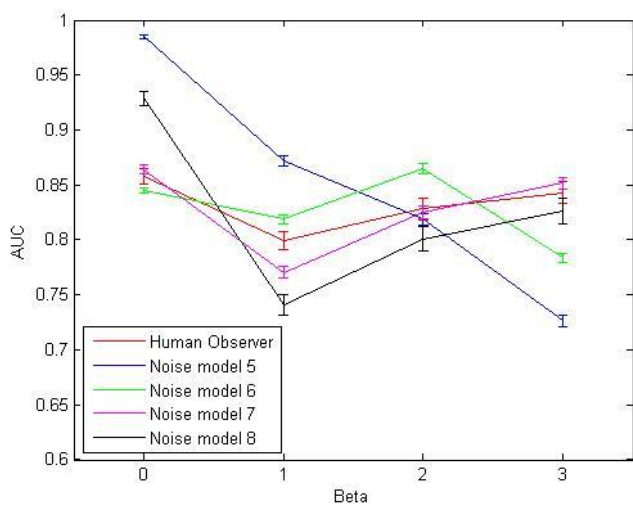


Figure 32

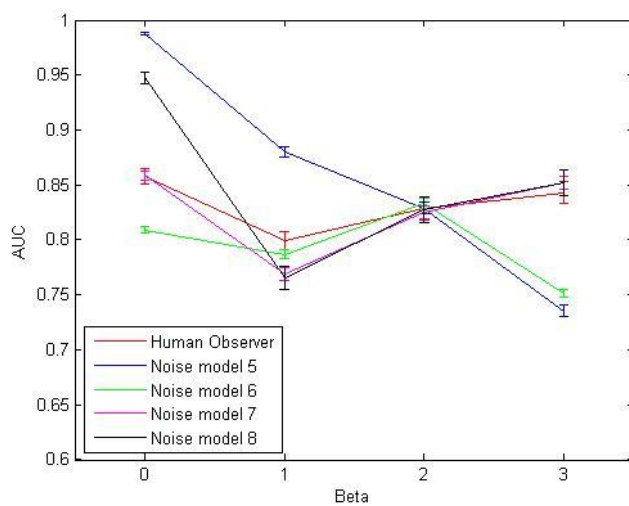


Figure 33

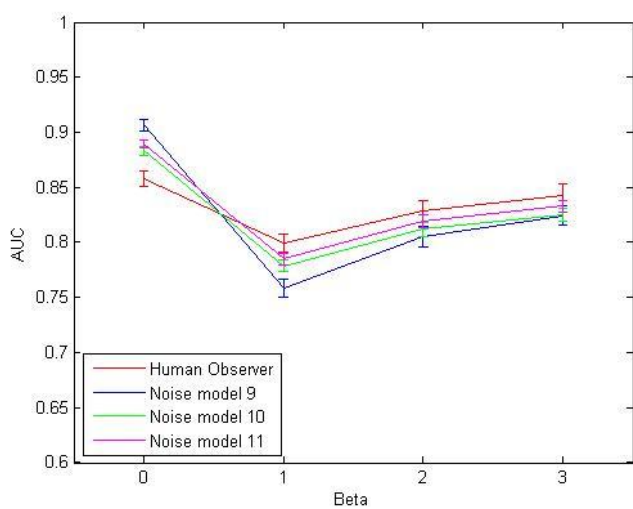


Figure 34

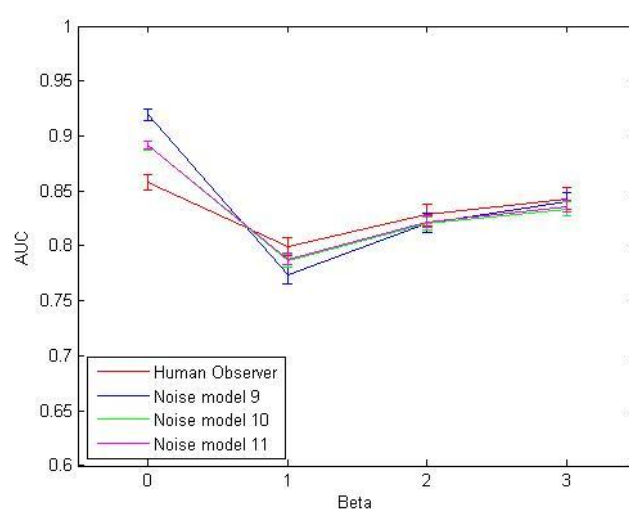


Figure 35

- *CHO-GB*

Human Observer 1: comparison 1 (left) and comparison 2 (right)

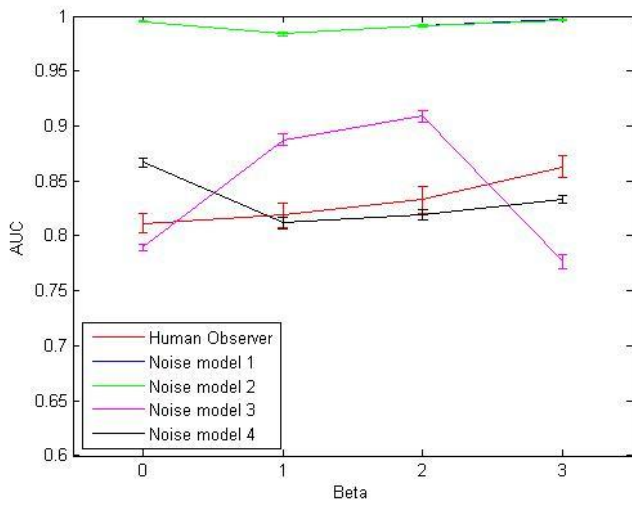


Figure 36

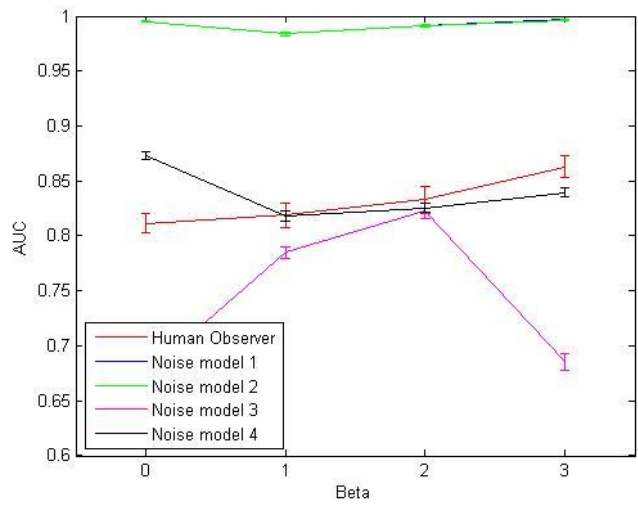


Figure 37

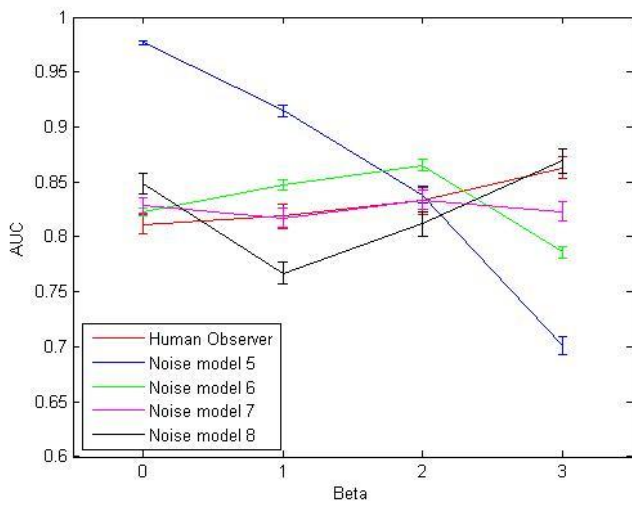


Figure 38

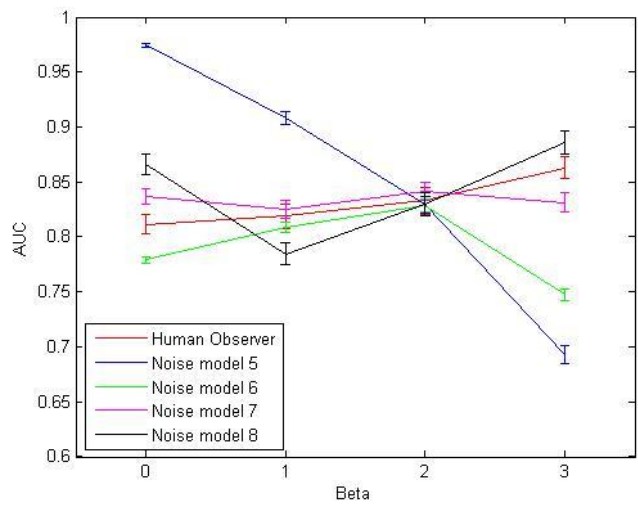


Figure 39

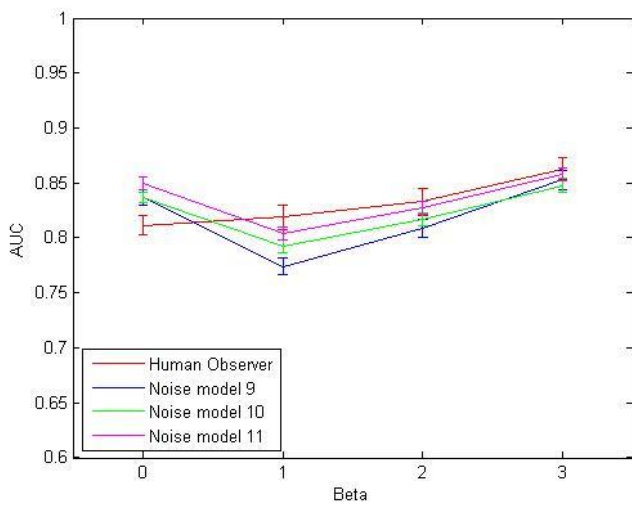


Figure 40

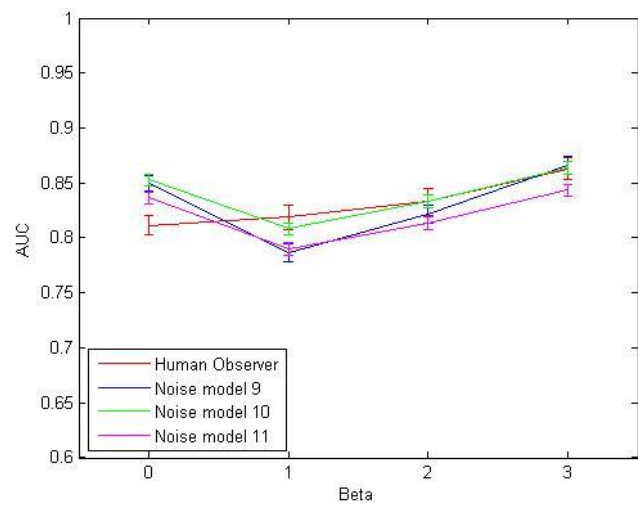


Figure 41

Human Observer 2: comparison 1 (left) and comparison 2 (right)

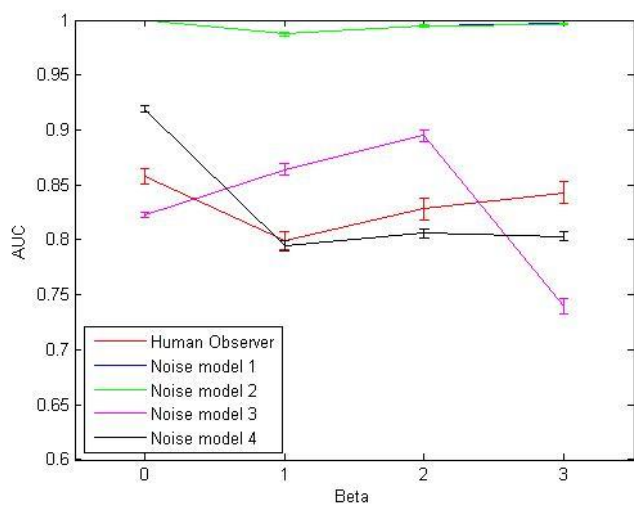


Figure 42

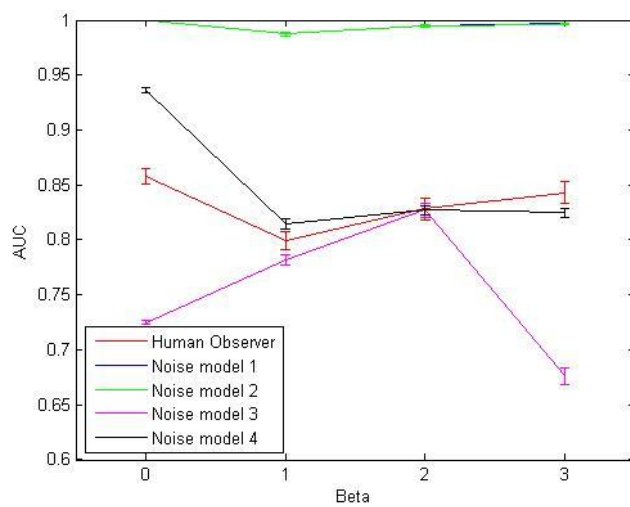


Figure 43

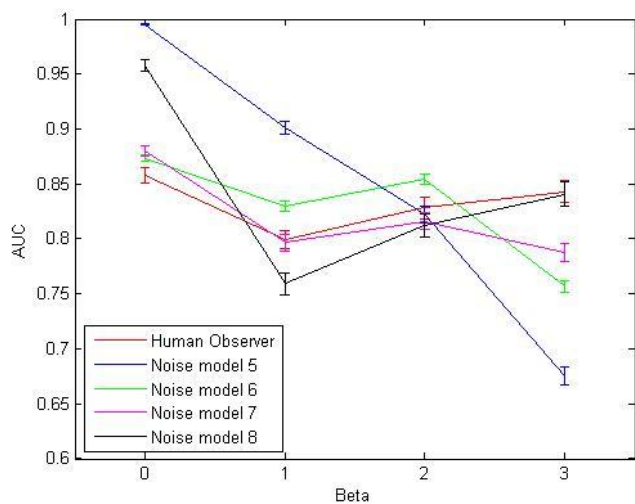


Figure 44

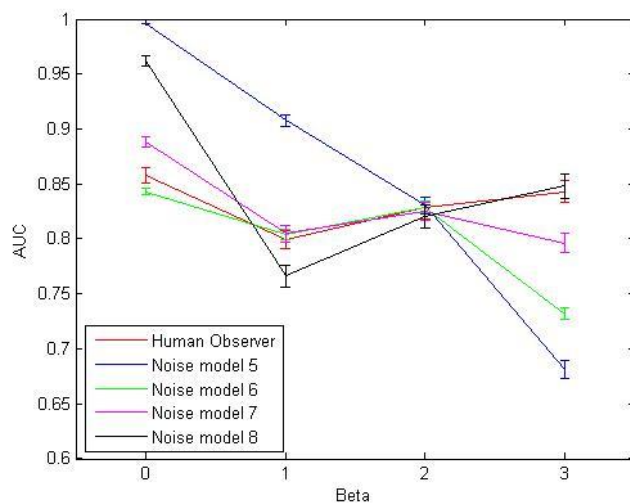


Figure 45

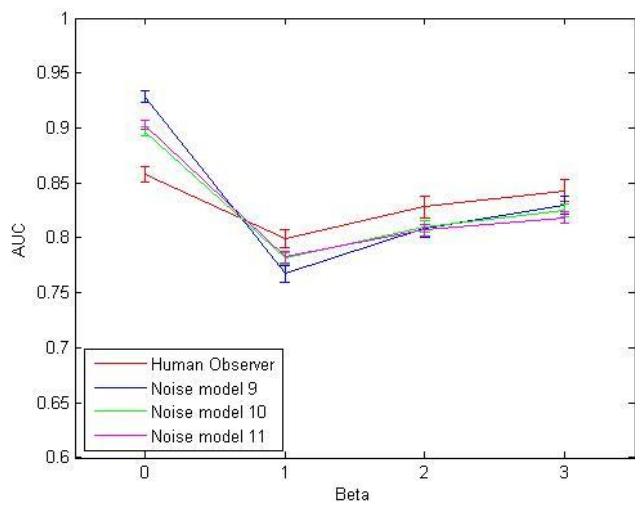


Figure 46

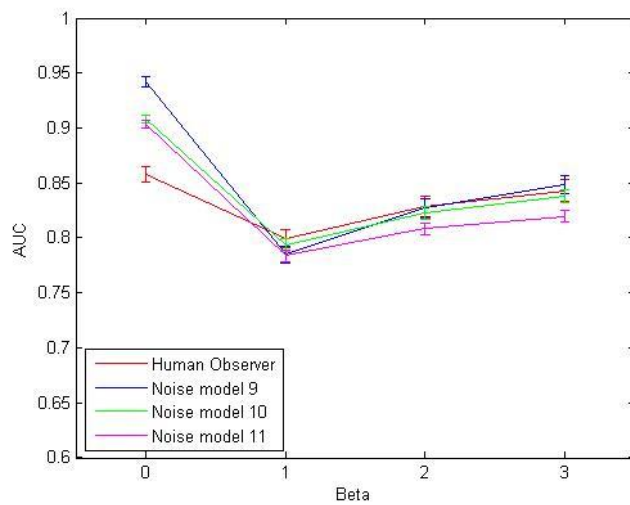


Figure 47

- CHO-LG

Human Observer 1: comparison 1 (left) and comparison 2 (right)

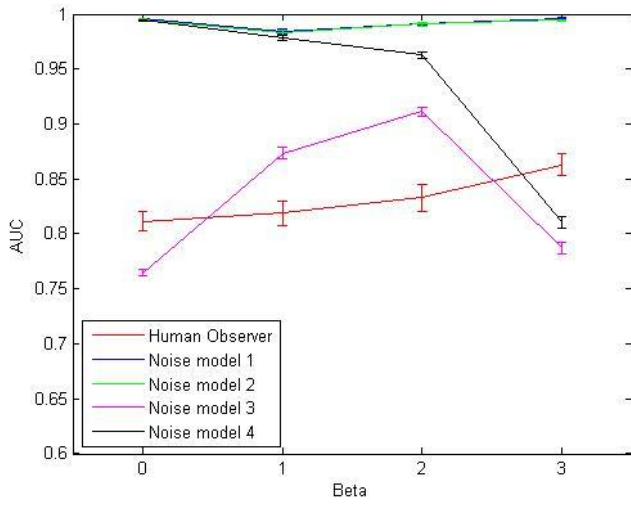


Figure 48

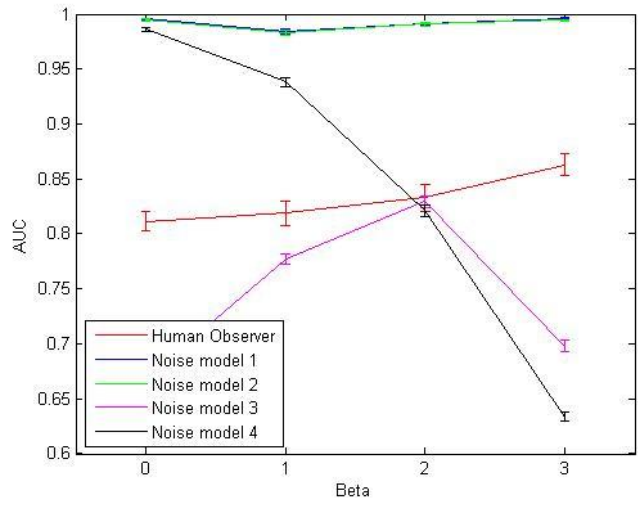


Figure 49

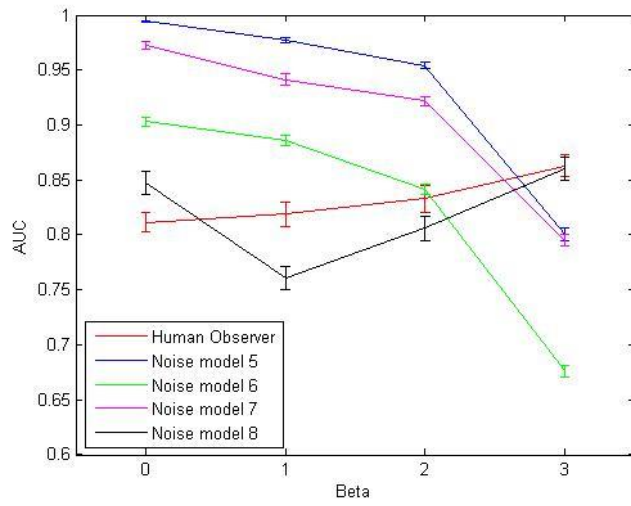


Figure 50

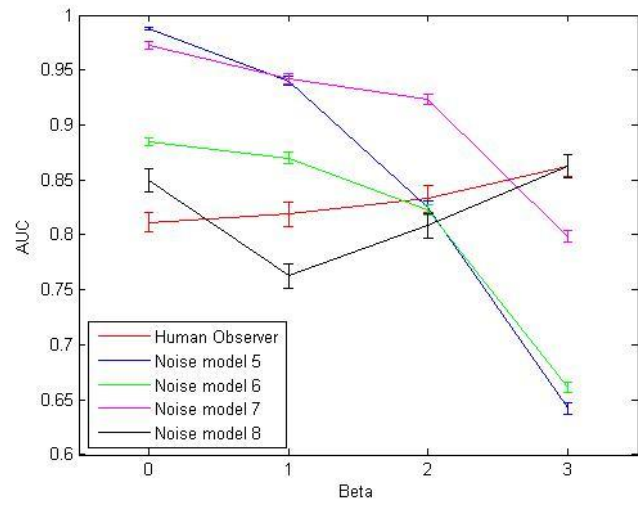


Figure 51

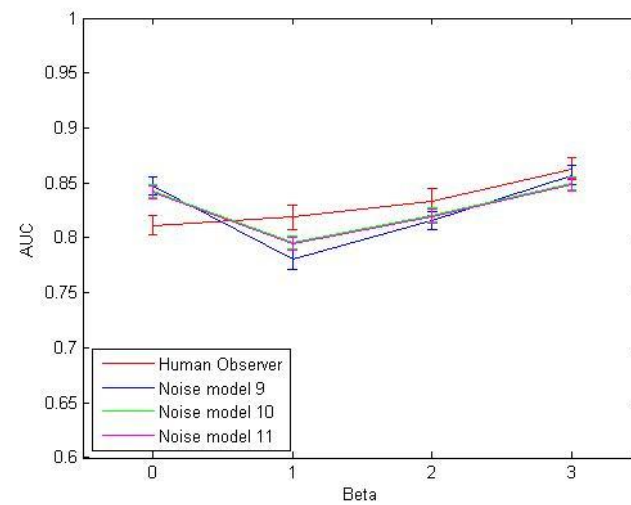


Figure 52

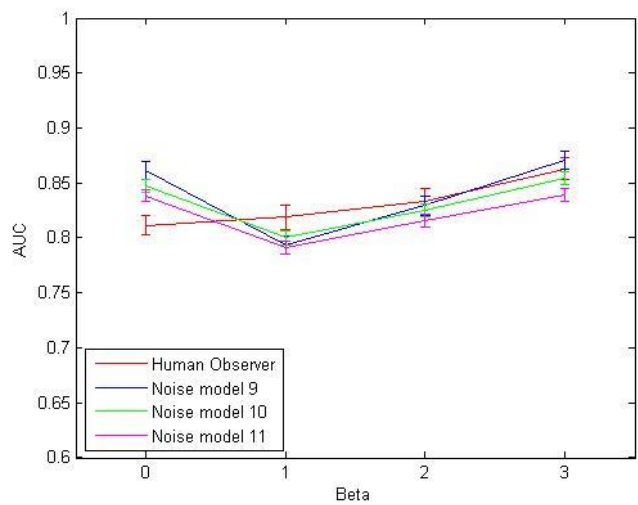


Figure 53

Human Observer 2: comparison 1 (left) and comparison 2 (right)

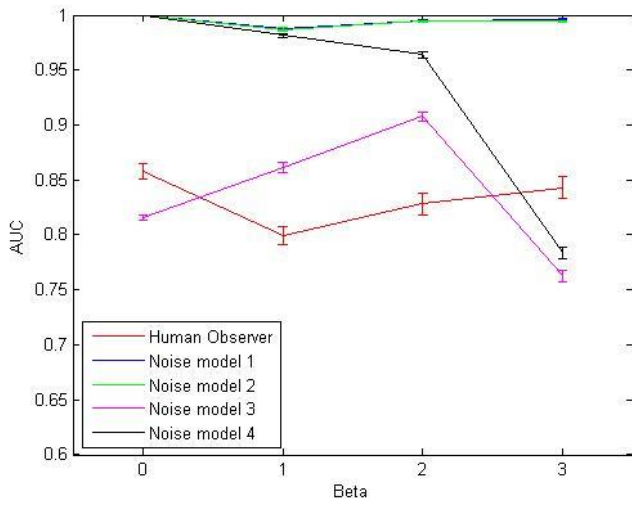


Figure 54

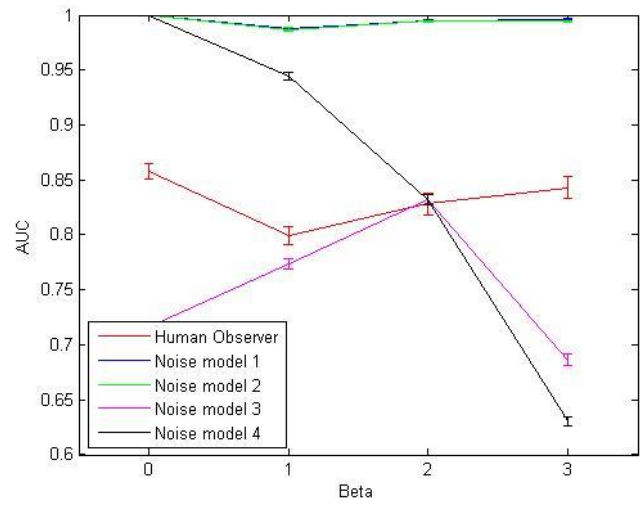


Figure 55

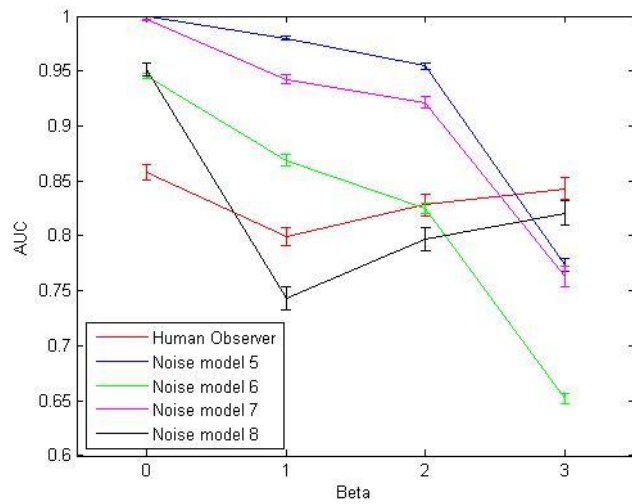


Figure 56

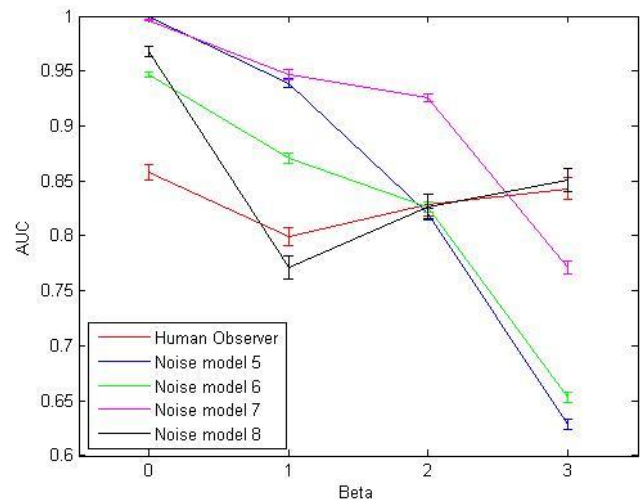


Figure 57

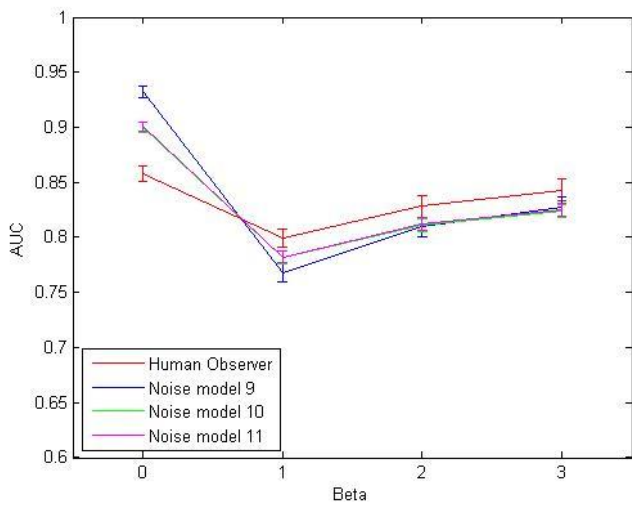


Figure 58

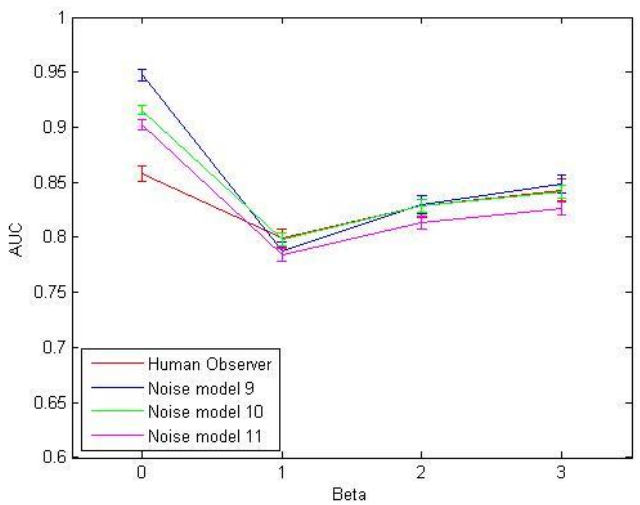


Figure 59

- *CHO-DOG*

Human Observer 1: comparison 1 (left) and comparison 2 (right)

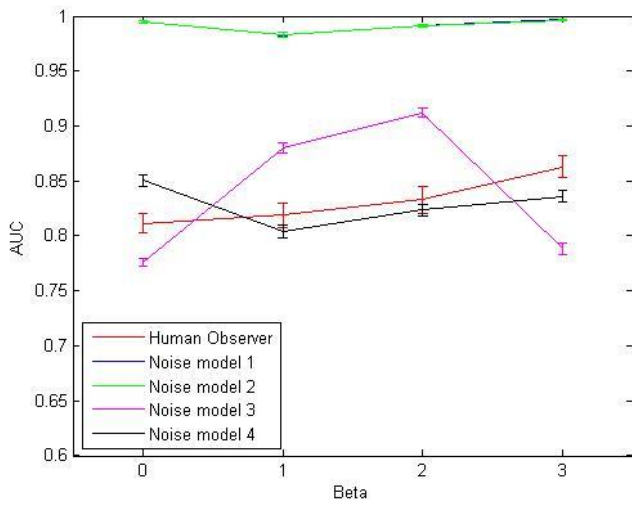


Figure 60

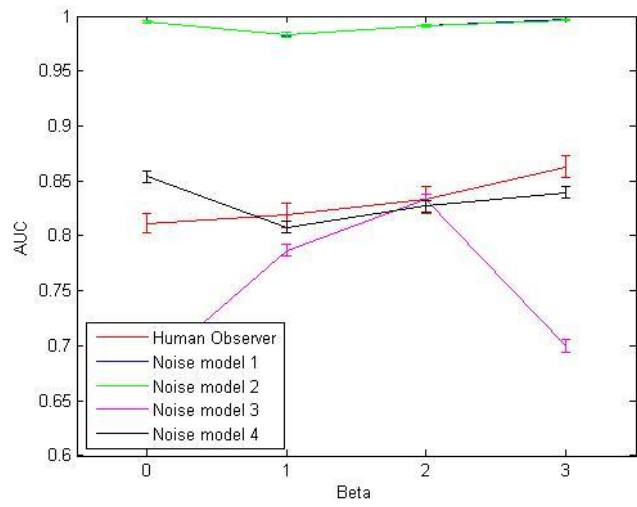


Figure 61

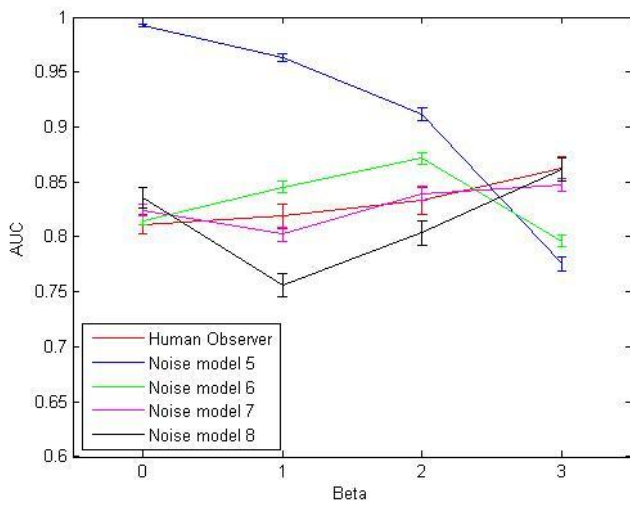


Figure 62

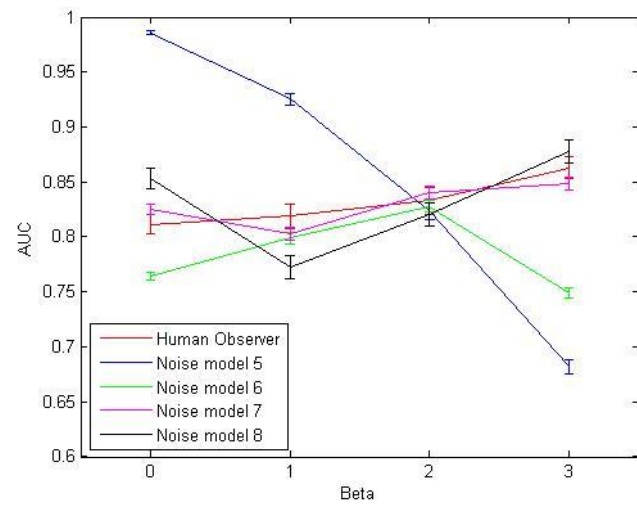


Figure 63

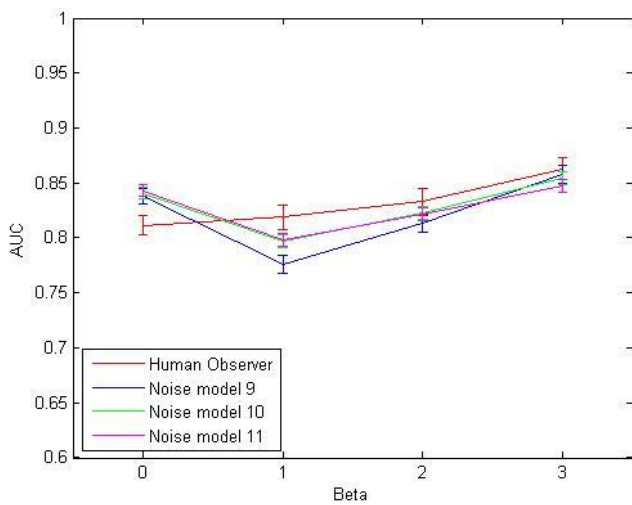


Figure 64

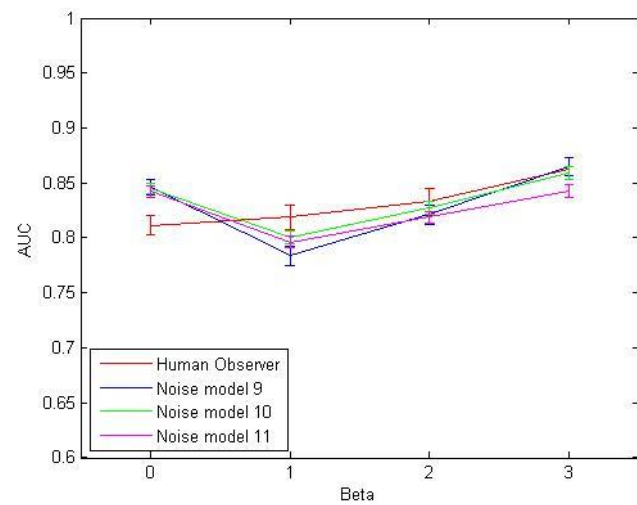


Figure 65

Human Observer 2: comparison 1 (left) and comparison 2 (right)

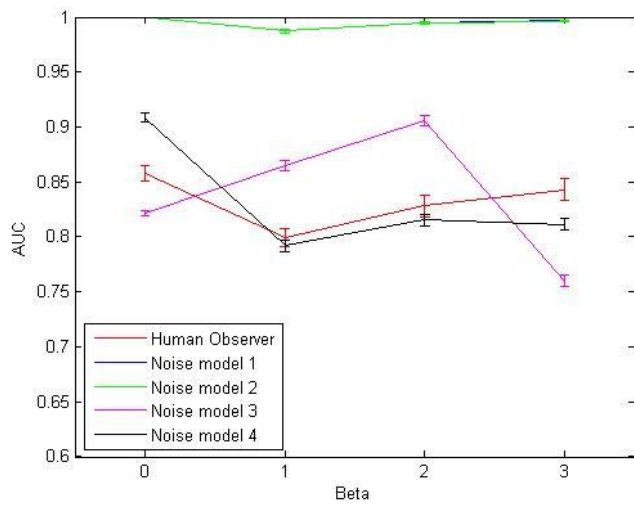


Figure 66

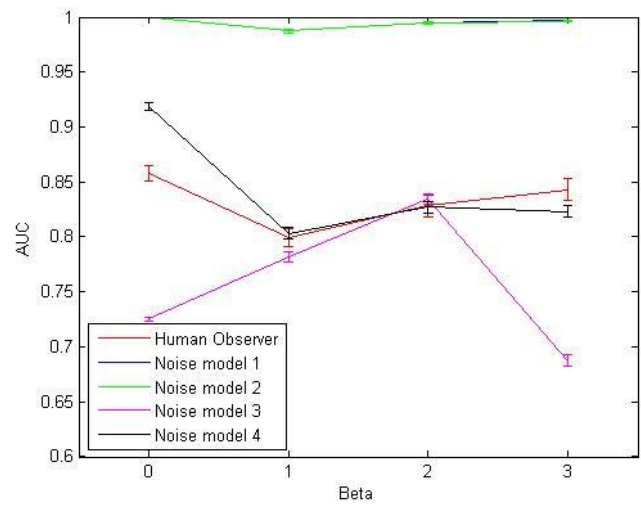


Figure 67

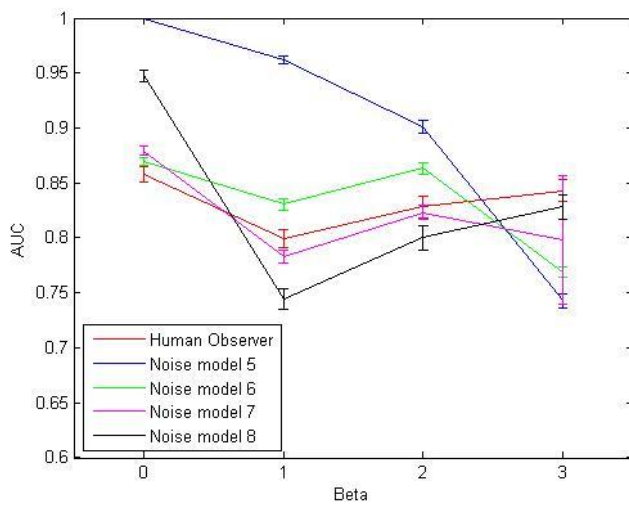


Figure 68

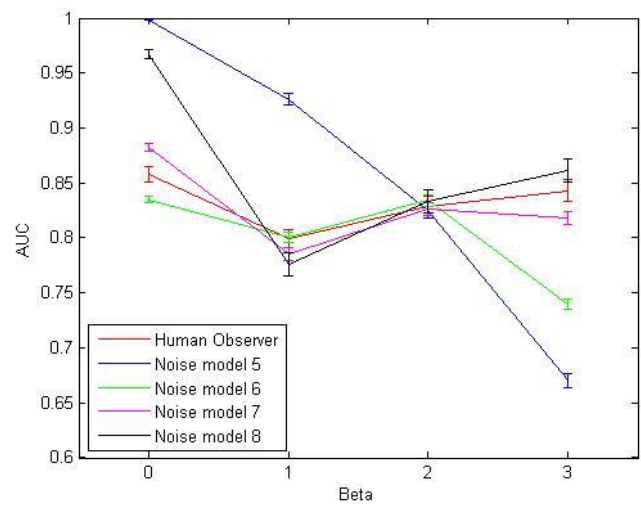


Figure 69

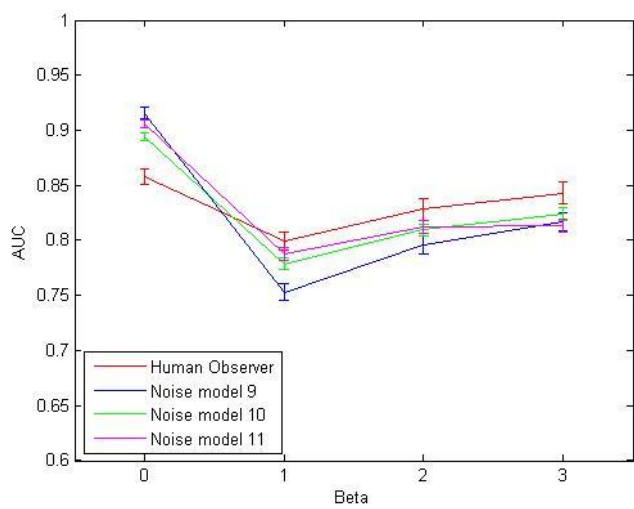


Figure 70

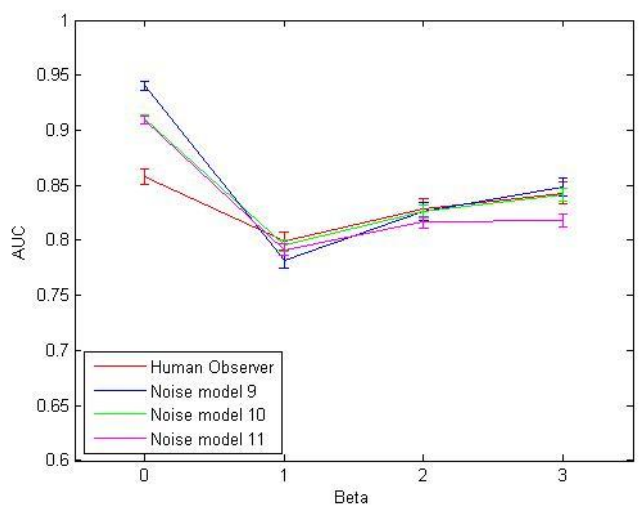


Figure 71

CHAPTER 4

CONCLUSIONS

4.1.General conclusions

In this report, it was studied the optimization of the Channelized Hotelling Observer in predicting Human Observer performance for four different channeling operators and eleven Internal-noise models in two different lesion detection tasks.

In summary, some proposed models can produce accurate predictions of the human performance. However, although most of them perform well to unseen images of the same kind, it is a necessary that they also do when we test them in different reconstructed images in an equally broad range to be able to do a suitable estimation when the human data to assess image quality is not available.

In the case of those that have demonstrated good results in both comparisons, further studies will be needed to fully validate them. However, this study suggests that the CHO with, for instance, the sixth and tenth noise models are promising methods for the calculation of Human Observer performance prediction in lesion detection tasks; while models 1, 2 or 5 do not obtain accurate estimations so they are prone to be discarded in future studies.

4.2.Experiment 1

The best results for all the channeling operators were obtained with the Model 6 that proposes an internal noise proportional to the standard deviation of the external noise. Although all of them have similar results, we can conclude the performance of the Laguerre-Gauss filters is not as good as the others and the Difference of Gaussians seems to be the one obtaining best results.

4.3.Experiment 2

In the case of trying to estimate the human observer performance for images with different correlated noises than the kind of noise used in the training set, the best results for all the channeling operators were obtained for the Model 10 that proposes the decision variable noise variance is proportional to the external noise variance. All channels have similar results but in some cases Band Pass and Difference of Gaussians filters obtain better correlation with the HO than the other two filters evaluated.

Other models that obtain good results are models 9 or 11 and worst results are obtained for models 1, 2 and 5.

REFERENCES

- [1] A. Bovik, *The Essential Guide to Image Processing*, Academic Press, 2009. Chapter 21: Image Quality Assessment.
- [2] Visual acuity.
<http://webvision.med.utah.edu/book/part-viii-gabac-receptors/visual-acuity/>
- [3] H. H. Barrett and K. J. Myers, *Foundations of Image Science*, NY: Wiley, 2003. Chapter 13: Statistical decision theory and Chapter 14: Image Quality.
- [4] H. H. Barrett, J. Yao, J.P. Rolland, K. J. Myers, "Model observers for assessment of image quality", *Proc. Natl. Acad. Sci. USA*, Vol. 90, pp. 9758-9765, November 1993.
- [5] J. G. Brankov, "Optimization of the Channelized Hotelling Observer Internal-Noise Model in a Train-Testing Paradigm".
- [6] J. G. Brankov, Y. Yang, L. Wei, I. El-Naqa and M. N. Wernick, "Learning a Channelized Observer for Image Quality Assessment", *IEEE Trans. Med. Imaging*, Vol. 28, No.7, July 2009.
- [7] PubMed.
<http://www.ncbi.nlm.nih.gov/pubmed/>
- [8] Ischemic Heart Disease
<http://www.fac.org.ar/ccvc/llave/c134/mut.php>
- [9] M. N. Wernick and J. N. Aarsvold, *Emission Tomography*, 2004. Chapter 2: Introduction to Emission Tomography.
- [10] H.M. Hudson and R.S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data", *IEEE Trans. Med. Imaging*, Vol. 4, pp. 601-609, 1994.
- [11] ROCKIT.
<http://metz-roc.uchicago.edu/MetzROC>
- [12] C. K. Abbey and M. P. Eckstein, "High human-observer efficiency for forced-localization tasks in correlated noise".
- [13] S. M. Kay, *Fundamentals of Statistical Signal Processing*, Vol 2, 1998. Chapter 4: Deterministic Signals.
- [14] R. O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, Second Edition, 2001. Chapter 9: Algorithm-independent machine learning.
- [15] TPF,FPF,TNF and FNF graph.
<http://edoc.hu-berlin.de/dissertationen/zaspel-uta-2003-09-26/HTML/chapter2.html>
- [16] ROC curve graph.
<http://radiology.rsna.org/content/212/3/817.full>

