

# Pertinence des modèles du SVH pour la tâche de détection de lésions SEP sur des IRMs

Lu ZHANG<sup>1</sup>, Christine CAVARO-MÉNARD<sup>1</sup>, Patrick LE CALLET<sup>2</sup>

<sup>1</sup>LISA, EA 4094, University of Angers  
62 avenue Notre Dame du Lac, 49000 Angers, France

<sup>2</sup>IRCCyN, UMR-6597, University of Nantes  
Rue Christian Pauc, BP50609, 44306 Nantes cedex 3, France

lu.ge@etud.univ-angers.fr, christine.menard@univ-angers.fr,  
patrick.lecallet@univ-nantes.fr

**Résumé** – Il est largement admis que la qualité des images médicales doit être évaluée dans le contexte d’une tâche diagnostique spécifique. Dans ce papier, différents modèles numériques du SVH sont comparés pour une utilisation optimale de ces modèles dans l’évaluation objective de la qualité d’images médicales.

**Abstract** – It is widely accepted that the quality of medical images should be evaluated in the context of a specific diagnostic task. In this paper, different HVS models are compared for the optimal use of these models in the objective medical image quality assessment.

## 1 Introduction

L’évaluation de la qualité des images médicales est indispensable pour la comparaison, voire l’optimisation des systèmes d’imagerie, tels que les systèmes d’acquisition, les systèmes de visualisation et les algorithmes de traitement d’images. Par exemple, dans le cadre de la télémédecine, une compression irréversible est essentielle pour optimiser le temps de transmission des données image. Mais quel algorithme et quel taux de compression choisir pour assurer la qualité diagnostique [6] ?

L’analyse de la fidélité diagnostique fait appel à des études nécessitant l’évaluation par des radiologues de la lisibilité diagnostique sur un ensemble d’images. Dans la pratique, ces études sont longues et difficiles à organiser. En revanche, les approches objectives sont plus rapides et reproductives. Elles se répartissent en trois familles: les mesures basées sur le signal (tel que le PSNR<sup>1</sup>) ; les modèles d’observateurs [1] ; et les modèles du Système Visuel Humain (SVH) [3]. Les travaux présentés dans ce papier se limitent aux modèles du SVH.

La fonction du SVH se divise en deux étapes : la sensation et la perception. La sensation est la fonction d’une chaîne d’événements biochimiques et neurologiques commençant par le heurt du stimulus (lumière) sur les photorécepteurs de l’œil. La perception est le processus qui se reflète dans une déclaration représentant la connaissance ou la compréhension de la réalité provoquée par l’information observée.

Notons que les modèles du SVH existants concernent seulement la modélisation de la sensation ; le niveau de la perception reste difficile à modéliser. De plus, tous les modèles ont

été établis à partir de résultats d’expériences menées par des observateurs naïfs (sans connaissance médicale). D’où notre question : Les modèles du SVH, sont-ils adaptés au domaine médical où les utilisateurs finaux sont des radiologues experts en interprétation de ces images ?

Dans la littérature, peu de travaux ont été menés pour répondre à cette question. Bien que plusieurs modèles HVS aient été comparés pour des images naturelles, leur comparaison dans le contexte d’images médicales n’est pas un exercice simple, étant donné les spécificités et l’enjeu clinique des images médicales. Cette comparaison nécessite le développement d’une étude spécifique. Dans ce papier, nous étudions la pertinence des modèles du SVH pour l’évaluation objective de la qualité des images médicales, en comparant leurs performances avec celles des radiologues dans le cadre de la tâche de détection, au niveau de la sensation et de la perception.

## 2 Modèles du SVH ciblés

Quatre des modèles du SVH les plus connus ont été sélectionnés pour cette étude : le modèle VDP<sup>2</sup> [3] ; le modèle HDR-VDP<sup>3</sup> [5] ; le modèle VCGC<sup>4</sup> [8] ; et le modèle SSO<sup>5</sup> [7].

Les modèles étudiés incluent des fonctions (relativement élaborées) modélisant des caractéristiques du SVH. Les principales fonctions sont présentées dans le TAB. 1. Deux images

<sup>2</sup>Visible Difference Predictor

<sup>3</sup>High Dynamic Range Visible Difference Predictor

<sup>4</sup>Visual Contrast Gain Control model

<sup>5</sup>Spatial Standard Observer

<sup>1</sup>Rapport signal/bruit de crête

TAB. 1: Comparaison des quatre modèles du SVH.

Fonctions	VDP	HDR-VDP	VCGC	SSO
Nonlinéarité de l'amplitude	unités arbitraires	unités JND	—	—
Fonction de Sensibilité au Contraste (FSC)	1 filtre	plusieurs filtres	1 filtre	1 filtre
Décomposition	filtres cortex	filtres cortex	filtres Gabor	—
Cumul des canaux	cumul des probabilités	cumul des probabilités	cumul Minkowski	—
Cumul Spatial	cumul des probabilités	cumul des probabilités	cumul Minkowski	cumul Minkowski
Fonction masquage	intra-canaux	intra-canaux	inter-canaux	—

(l'image de référence et l'image de distorsion) sont analysées par ces fonctions qui fournissent une carte des différences justes perceptibles (JND<sup>6</sup>) entre ces deux images. Ces différences visuelles sont synthétisées pour construire une note globale.

Ce type de modèles est efficace pour évaluer la visibilité d'une distorsion proche du seuil de visibilité, ce qui est le cas sur de nombreuses pathologies, comme la Sclérose En Plaques (SEP) où les contrastes des lésions sont proches du seuil de visibilité.

Dans le cadre de l'évaluation de la qualité d'images pour la tâche de détection de lésions, les entrées de ces modèles sont en général une image d'un volontaire sain et cette même image avec des lésions simulées [4]. Cette conception s'appuie sur le raisonnement des radiologues qui ont souvent une image saine en tête lors de l'interprétation de l'examen. La note du modèle du SVH quantifie alors la probabilité de visibilité (détection) de la lésion, probabilité que nous avons comparée à celle des radiologues dans l'étude présentée ci-après.

### 3 Protocole d'évaluation

#### 3.1 Modalité et Pathologie

Pour pouvoir évaluer la performance de la tâche de détection, l'anomalie de la pathologie à étudier ne doit pas être trop visible. Nous avons choisi la Sclérose en Plaque (SEP) dont les lésions sont subtiles et difficiles à détecter.

Le diagnostic de la SEP est largement basé sur l'analyse d'IRM (imagerie par résonance magnétique). Dans certaines circonstances un diagnostic de la SEP peut être confirmé avec un seul examen d'IRM. Nous avons donc choisi l'IRM comme modalité d'imagerie à étudier.

Notons que la séquence la plus sensible pour la détection des lésions de SEP est la séquence T2 FLAIR<sup>7</sup> où la lésion de SEP apparaît en hyper-signal<sup>8</sup>.

#### 3.2 Images

Afin d'étudier l'influence de l'information anatomique sur la détection d'une lésion, nous avons utilisé deux types de fonds:

- une coupe entière : IRM cérébrale d'une séquence T2 FLAIR d'un volontaire sain (cf. image à gauche dans FIG. 1) ;
- une zone homogène : zone de matière blanche extraite de la coupe entière pour conserver les mêmes caractéristiques (cf. image à droite dans FIG. 1).

Les lésions de SEP simulées sont modélisées par des fonctions gaussiennes elliptiques bidimensionnelles. Notons que la simulation des lésions nous permet de maîtriser parfaitement les caractéristiques des lésions. La forme, la taille, l'orientation et l'emplacement des lésions ont été vérifiés par un spécialiste en neuroradiologie. En utilisant 14 niveaux de contraste et 6 emplacements, nous avons généré 84 images au total.

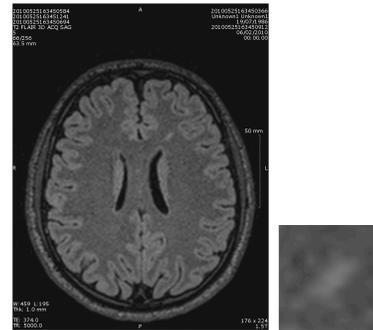


FIG. 1: Exemples d'images test : l'image à gauche est une coupe entière avec une lésion simulée sur le coin supérieur droit de la coupe; l'image à droite est une zone homogène contenant une lésion simulée au centre de la zone.

#### 3.3 Expériences subjectives

Nous supposons ici que la détection d'un hyper-signal est un processus de sensation (même un observateur naïf peut le faire après un entraînement). Mais la détection d'un signe cliniquement significatif nécessite des connaissances médicales et implique le processus de reconnaissance et d'interprétation, qui est l'étape de perception dans une certaine mesure. Dans la suite, la sensation désigne la détection d'un hyper-signal simple, et la perception désigne la détection d'un signe cliniquement significatif.

Afin d'étudier les tâches de sensation et de perception, avec

<sup>6</sup>Just Noticeable Difference

<sup>7</sup>Fluid Attenuated Inversion Recovery

<sup>8</sup>Ayant une intensité plus haute que le tissu entourant

ou sans information anatomique, trois radiologues ont participé aux quatre expériences décrites ci-dessous :

1. Expérience 1 (tâche de sensation sur une coupe entière): Dans chaque essai, la coupe entière avec une lésion simulée superposée était affichée. Nous avons demandé aux radiologues s'ils détectaient un hyper-signal sur la coupe.
2. Expérience 2 (tâche de sensation sur une zone homogène): Dans chaque essai, une zone homogène, avec une lésion simulée superposée était affichée. Nous avons demandé aux radiologues s'ils détectaient un hyper-signal sur la zone.
3. Expérience 3 (tâche de perception sur une coupe entière): Dans chaque essai, la coupe utilisée dans l'Expérience 1 avec une lésion simulée superposée était affichée. Nous avons demandé aux radiologues s'ils détectaient un signe cliniquement significatif sur la coupe.
4. Expérience 4 (tâche de perception sur une zone homogène): Dans chaque essai, la zone utilisée dans l'Expérience 2 avec une lésion simulée superposée était affichée. Nous avons demandé aux radiologues s'ils détectaient un signe cliniquement significatif sur la zone.

## 4 Méthode d'évaluation

Pour évaluer la performance de détection des modèles du SVH, nous avons utilisé l'Aire Sous la Courbe ROC<sup>9</sup> (AUC) comme figure de mérite.

Dans le cadre de l'analyse ROC, la note globale sortie d'un modèle du SVH, représentant une probabilité de détection (désignée par  $\lambda$ ), est comparée à un seuil  $\lambda_c$ . La décision du modèle est considérée comme "positive" (lésion-présente) si  $\lambda > \lambda_c$ , sinon comme "négative" (lésion-absente). Le décision peut donc être Vrai Positif (TP), Vrai Négatif (TN), Faux Positif (FP) ou Faux Négatif (FN) selon un *gold standard*<sup>10</sup>, comme indiqué dans le TAB. 2. En variant le seuil  $\lambda_c$ , nous obtenons alors une courbe ROC qui donne le taux de TPs en fonction du taux de FPs, comme illustré dans la FIG. 2.

TAB. 2: Vrai Positif (TP), Vrai Négatif (TN), Faux Positif (FP) ou Faux Négatif (FN).

		Gold Standard	
		lésion-présent	lésion-absent
Décision du modèle	positive	TP	FP
	négative	FN	TN

Notons qu'il est tacitement admis qu'un *gold standard* reste complexe à définir, surtout lorsque la modalité étudiée constitue elle-même le *gold standard*, ce qui est le cas en partie pour la SEP. Dans ce cas, une alternative consiste à prendre comme *gold standard* le consensus des experts médicaux. De

<sup>9</sup>en anglais, Receiver Operating Characteristic

<sup>10</sup>l'état exact de l'image (positive ou négative)

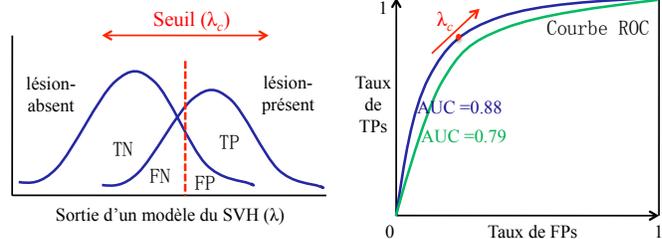


FIG. 2: Courbe ROC résulte du seuil variant  $\lambda_c$ .

ce fait, nous considérons la décision majoritaire des 3 radiologues comme *gold standard* pour chaque expérience. L'AUC mesure alors la distance entre la performance des modèles du SVH et celle des radiologues. Ainsi, plus la valeur de l'AUC sera proche de 1 et plus la performance du modèle sera proche de celle des radiologues.

## 5 Résultats

Les courbes ROC des modèles sont montrées dans la FIG. 3. Pour chaque modèle et chaque expérience, la moyenne et l'écart-type de l'AUC sont calculés avec la méthode proposée dans [2], et présentés FIG. 4. Le test *t* a été utilisé pour identifier si les différences significatives entre les modèles et les expériences (TAB. 3 et TAB. 4).

Les résultats montrent que la performance du VDP est significativement meilleure que celle des autres modèles pour la sensation (FIG. 4a et TAB. 3) ; la performance du VDP est également meilleure pour la perception (FIG. 4b et TAB. 3). De plus, FIG. 4b et TAB. 4 montrent que la performance des modèles du SVH est plus proche de celle des radiologues lorsque l'information anatomique est accessible.

Dans tous les cas, le modèle SSO est toujours le moins performant. Cela indique que la modélisation de la sélectivité de fréquence spatiale et d'orientation (une caractéristique importante du SVH), non prise en compte par le SSO, est essentielle pour l'inspection des images médicales.

TAB. 3: Comparaison des AUCs des modèles du SVH, les quatre symboles de chaque cellule désignent les quatre expériences. "o" correspond à p-value < 0.05 (différence significative) et "-" correspond à p-value  $\geq$  0.05 (pas de différence significative). Par exemple, dans la cellule VDP vs. VCGC, "ooo-" signifie qu'il y a une différence significative entre les modèles VDP et VCGC dans les Expérience 1-3, mais qu'il n'y a pas de différence significative entre les deux modèles dans l'Expérience 4.

	VDP	HDR-VDP	VCGC	SSO
VDP				
HDR-VDP	oooo			
VCGC	ooo-	oooo		
SSO	oooo	oooo	oooo	

FIG. 3: Courbes ROC des modèles du SVH pour les quatre expériences.

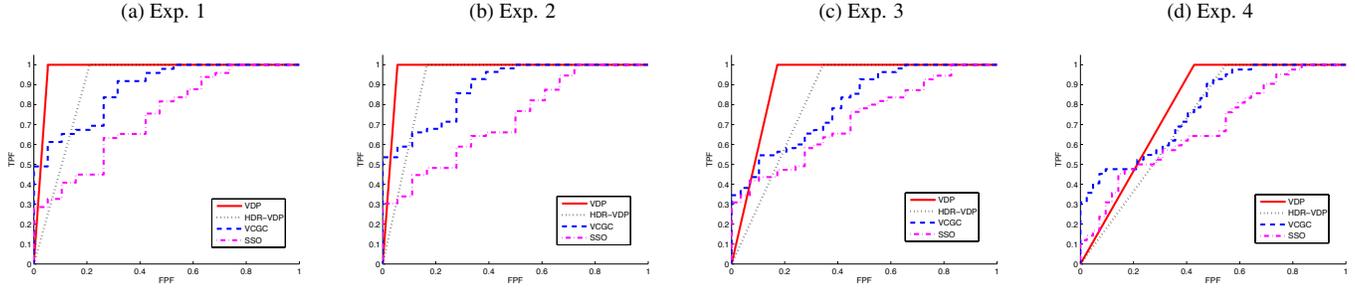
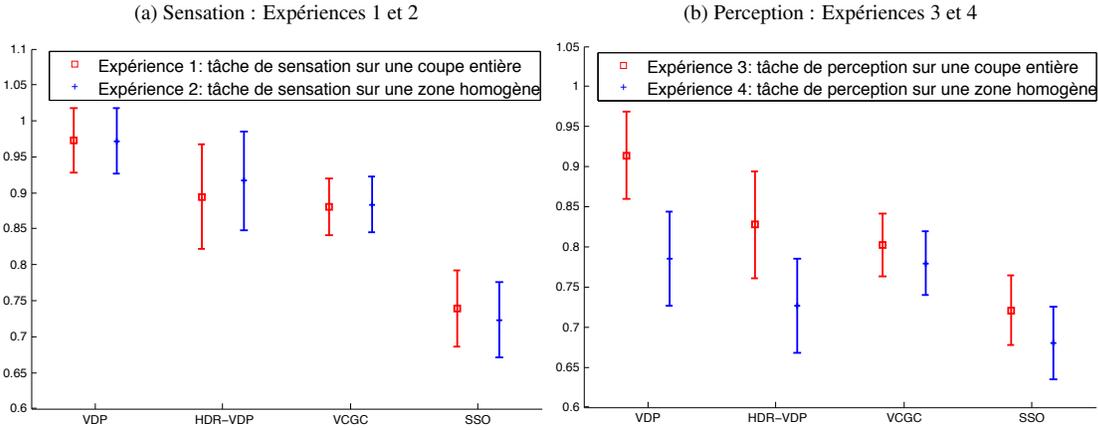


FIG. 4: Comparaison des AUCs.



TAB. 4: Comparaison des AUCs des quatre expériences, les quatre symboles dans chaque cellule désignent les modèles VDP, HDR-VDP, VCGC et SSO dans cet ordre.

	Exp. 1	Exp. 2	Exp. 3	Exp. 4
Exp. 1	□	□	□	□
Exp. 2	+	+	+	+
Exp. 3	○	○	○	○
Exp. 4	○	○	○	○

## 6 Conclusion

Cette étude montre que la performance du modèle VDP est la plus proche de celle des radiologues dans la tâche de détection, au niveau de la sensation et de la perception. Le VDP modélise correctement la sensation des radiologues mais n'apparaît pas assez satisfaisant pour la perception. D'autres travaux de recherche sont nécessaires pour mieux modéliser la perception.

## References

[1] H.H. Barrett and K.J. Myer. *Foundations of Image Science*. John Wiley and Sons, Inc., Hoboken, New Jersey, USA, 2004.

[2] Gallas BD. One-shot estimate of MRMC variance: AUC. *Academic Radiology*, 13:353–362, 2006.

[3] S. Daly. *The visible differences predictor: an algorithm for the assessment of image fidelity*, volume 13, pages 179–206. MA: MIT Press, 1993.

[4] 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*, volume 3036, page 64ñ73, 1997.

[5] R. Mantiuk, S. Daly, K. Myszkowski, and H.P. 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*, pages 204–214, 2005.

[6] Amine Nai-Ali and Christine Cavaro-Menard. *Compression of Biomedical Images and Signals*. John Wiley & Sons, Inc., 2007.

[7] A.B. Watson. The spatial standard observer: A new tool for display metrology. In *Information Display*, volume 23, pages 12–15, 2007.

[8] A.B. Watson and J.A. Solomon. Model of visual contrast gain control and pattern masking. In *Journal of the Optical Society of America A*, volume 14, pages 2379–2391, 1997.