

# Modèle de perception couleur. Application à l'évaluation de qualité d'images

## Color perception model. Application to image quality assessment

**P. Le Callet, D. Barba**

IRCCyN UMR CNRS 6597, École polytechnique de l'université de Nantes, rue Christian Pauc BP 50609, 44306 Nantes Cedex 3

Manuscrit reçu le 15 juin 2004

Résumé et mots clés

La conception de critères objectifs d'évaluation de qualité visuelle des images est d'une grande utilité afin d'optimiser les performances des systèmes de traitement d'images pour le stockage et la transmission de l'information visuelle. Un critère objectif de qualité doit fournir des mesures de qualité en forte correspondance avec le jugement humain. Il est donc nécessaire d'intégrer des modèles des principales propriétés du système visuel humain dans la conception d'un tel critère. Ce papier présente un modèle très élaboré de perception des images couleur par le système visuel humain (SVH). Il prend en compte et modélise de façon cohérente les fonctions clé de la perception : espace de représentation visuelle des couleurs, fonctions de sensibilité aux contrastes des composantes achromatique et chromatiques, décomposition en canaux perceptifs, fonctions de masquage intra et inter composante. Tous ces modèles ont été déduits à partir de résultats de tests psychophysiques et de leur interprétation. L'application du modèle de perception à la conception d'un critère objectif de qualité d'images couleur avec image de référence est ensuite effectuée. Elle se base sur une analyse critique des différentes façons de construire une fonction de cumul des dégradations perçues : cumul inter-composante, cumul inter-canal, cumul spatial. L'ensemble du critère objectif de qualité a été testé sur une base d'images naturelles couleur ayant subies des dégradations de nature et de force différentes. Les performances en terme de corrélation avec le jugement humain moyen est supérieure à 0,94 et 97 % des images sont correctement évaluées au regard de la précision sur les notes subjectives données par l'humain.

Évaluation de qualité, perception couleur humaine, effet de masquage, espace couleur psychovisuel.

Abstract and key words

The design of image visual quality objective criteria is very useful in order to optimize the performances of image processing systems for storage and transmission of visual data. An objective quality criterion should provide quality measures in good correspondence with human judgment. So, it is necessary to introduce the main properties of the human visual system (HSV) in the design of such a criterion.

This paper presents a very elaborated model of color images perception by the HVS. It takes into account and models in a coherent way the key functions of human perception : visual space for color representation, contrast sensitivity functions for achromatic and chromatic components respectively, perceptual channels decomposition, intra and inter pathways masking functions. All these models have been deduced from the interpretation of results of psychophysics experiments. Next, we present an application of the perception model to the design of an objective quality criterion

with full reference for color images. This application is based on a critical analysis of different ways to construct the pooling function of perceived errors including inter-pathway pooling, inter-channel pooling and spatial pooling. The entire objective quality criterion have been tested on a database of natural color images which have been stressed with distortions of different nature and strength. Performances in terms of correlation with mean human judgment is more than 0.94 and 97 % of images are correctly assessed according to the accuracy on the subjective rate given by observers.

Quality assessment, human color perception, masking effect, psychovisual color space.

## 1. Introduction

**S** De nos cinq sens, la vision est celui qui utilise le plus de neurones dans notre cerveau. De longues recherches neurobiologiques ont permis de dégager quelques principes généraux de fonctionnement du système visuel humain (SVH). La connaissance de celui-ci est particulièrement intéressante en traitement d'images dans la mesure où la plupart des applications sont destinées à fournir des images visualisées et exploitées par l'humain. Ainsi, la prise en compte du fonctionnement du SVH doit permettre d'adapter les traitements à l'élément en bout de chaîne. Le SVH apparaît comme un système multicapteurs, intégrant progressivement des informations spatiales, de couleur, de profondeur, de fréquences spatiales et de mouvement de notre environnement (le mouvement n'est pas considéré dans ce papier). La physiologie de l'aire visuelle primaire V1 lui confère un rôle clé de préparation et de codage de tous ces types d'informations provenant d'une scène visuelle sur une même carte spatiale. Cette organisation des informations, résultante de nombreux traitements bas niveaux, constitue une organisation adaptée pour des traitements de plus hauts niveaux tels qu'ils sont réalisés dans les aires visuelles supérieures.

Dans ce papier, nous proposons un modèle de perception des images couleur basé sur une exploitation fine des propriétés du SVH jusqu'à l'aire V1. Des applications de type compression ou tatouage d'images peuvent aisément tirer profit d'un tel modèle. Dans le cas d'une application de compression, le modèle proposé permet, notamment, d'identifier les informations non visibles et donc de s'en affranchir au moment de la compression. Les exploitations d'un tel modèle en tatouage d'images sont multiples, mais on peut citer en particulier la capacité du modèle à fournir des informations propres à régler la force avec laquelle on insère une marque et, de fait, à gérer judicieusement le compromis invisibilité/robustesse. Dans le cadre de ce papier, nous proposons comme illustration d'utiliser ce modèle afin de construire un critère pour évaluer la qualité des images couleur. Un tel critère visant à fournir des notes de qualité en bonne correspondance avec le jugement humain, doit certainement inté-

grer des traitements proches de ceux mis en jeu au sein du SVH, l'exploitation de notre modèle est donc dans ce cas bien justifiée. S'il est courant de trouver des modèles de perception ainsi que des critères de qualité pour les images monochromes dans la littérature, rares sont les propositions pour les images couleur. Ceci peut s'expliquer par le fait que passer d'une composante monochrome à trois composantes pour des images couleur ne revient pas à dupliquer simplement trois traitements. L'originalité de notre modèle de perception tient donc essentiellement dans sa complétude, notamment parce qu'il intègre les problèmes d'interaction en terme de visibilité entre les différentes composantes d'un espace de représentation des couleurs. Au niveau du critère d'évaluation de qualité, nous abordons également le problème non trivial de regroupement des erreurs provenant de composantes couleur différentes afin de construire une note de qualité globale.

Dans la suite de ce papier, nous décrivons d'abord la structure globale du modèle de perception, puis nous revenons précisément sur les fonctions les plus sensibles de celui-ci : espace couleur, fonction de sensibilité aux contrastes (abrégée CSF en anglais pour *Contrast Sensitivity Function*), décomposition en sous-bandes perceptuelles, fonction de masquage intra et inter composante. Enfin, nous présentons l'application d'évaluation de qualité d'image couleur avec les résultats associés.

## 2. Structure globale du modèle de perception

Le modèle que nous proposons s'inspire du VDP (*Visual Difference Predictor*) de Daly [1], avec quelques adaptations d'importance. L'adaptation principale réside dans l'extension du modèle au traitement des images couleur. De manière générale, le modèle permet de projeter les informations dans un espace perceptuel et de les normaliser par rapport à un seuil différentiel de visibilité (défini en fonction des effets de masquage). On retrouve donc toutes les fonctions classiques d'un modèle de

perception (CSF, décomposition en canaux, fonctions de masquage) auxquelles vient s'ajouter en amont une transformation en composantes perceptuelles (les trois composantes de l'espace de représentation des couleurs).

Si on souhaite comparer deux images dans cet espace perceptuel, alors nous devons définir la manière d'opérer. Pour ce faire, nous rajoutons en sortie du modèle de perception un opérateur de différence de manière à fournir des cartes de différences perceptuelles (appelées aussi erreurs perceptuelles). Cette étape apparaît naturellement utile, entre autres, dans le cadre de l'application finale visée : l'évaluation de qualité. Néanmoins, il est tout à fait possible de s'en affranchir afin de ne conserver que les parties du modèle nécessaires à la définition d'une application donnée. La seconde adaptation majeure par rapport au modèle de Daly concerne justement cette étape. Là où Daly s'intéresse à la probabilité de détection des erreurs perceptuelles, nous préférons construire des erreurs perceptuelles normalisées en fonction des seuils différentiels de visibilité. Enfin, d'autres adaptations concernent des choix de paramètres liés aux résultats des expériences de psychophysique menées au laboratoire.

Après une transformation en composantes perceptuelles (composantes de l'espace couleur), chaque composante de l'image originale et de l'image dégradée se voit appliquer respectivement la série de traitements de la figure 1.

**Fonctions de sensibilités aux contrastes 2D** (abrégiées CSF 2D sur la figure 1) : Nous savons que le système visuel est davantage sensible localement aux variations de la luminance par rapport à la luminance moyenne du fond plutôt qu'à la valeur absolue de cette luminance. Le modèle le plus connu pour exprimer cette dépendance est la loi de Weber-Fechner. Cependant, la luminance de fond n'est pas le seul facteur influant sur le seuil différentiel de visibilité d'un signal. D'autres caractéristiques sont essentielles, notamment sa couleur ou encore son contenu en fréquences spatiales. Les CSF sont généralement utilisées pour caractériser ces dépendances. La CSF décrit l'évolution de la sensibilité visuelle, c'est-à-dire l'inverse du contraste d'un signal à son seuil de visibilité, en fonction le plus souvent des fréquences spatiales et de l'orientation du signal à détecter. Les modèles proposés dans la littérature sont élaborés, la plupart du temps, à partir de résultats expérimentaux sur la détection de

signaux sinusoïdaux en utilisant la définition de Michelson pour le contraste. Le fait que les signaux réels soient composés de multiples fréquences implique des variations des seuils de visibilité, ces variations traduisent ce qu'on appelle l'effet de masquage. La définition de base de la CSF est celle pour laquelle le signal à détecter (stimulus) est superposé à un signal constant. Si ce dernier n'est pas constant, la CSF est modifiée par l'effet de masquage, le signal non constant est appelé signal masquant. L'effet de masquage n'est facilement modélisable que si on utilise une décomposition en sous-bandes. Ainsi nous appliquons pour chaque composante une CSF appropriée en entrée du modèle afin de normaliser les signaux par rapport au seuil différentiel de visibilité sur signal constant. La prise en compte des effets de masquage est appliquée dans la suite du modèle en définissant des facteurs d'élévation de seuil de visibilité traduisant la variation de ces seuils. Nous parlons d'élévation de seuil pour respecter le formalisme de Daly. Cela étant, il s'agit d'un abus de langage car dans certaines conditions, on observe plutôt une diminution du seuil différentiel de visibilité (élévation  $< 1$ ). Daly a défini les fonctions d'élévation pour la composante achromatique. Nous proposons des fonctions d'élévation de seuils pour chaque composante couleur de notre espace de représentation perceptuel couleur.

**Décomposition en canaux perceptuels** : Plusieurs évidences physiologiques révèlent que les cellules du système visuel sont accordées pour la plupart à certains types d'informations visuelles comme la couleur, l'orientation ou la fréquence. Les résultats d'expériences psychophysiques suggèrent également l'existence d'une décomposition spatio-fréquentielle en canaux visuels pour traiter l'information. Ainsi, Sakrison [2] remarque que si l'on présente un stimulus contenant plusieurs composantes fréquentielles, comme par exemple une dent de scie, seule la composante fondamentale fixe le seuil de visibilité. Comme nous l'avons indiqué dans le paragraphe précédent, simuler ce phénomène de canaux visuels ou perceptuels est nécessaire si on souhaite modéliser correctement les effets de masquage. En pratique, nous avons nous-même observé que l'interprétation de résultats de nos expériences de psychophysique était grandement facilitée dès lors que l'on introduisait une décomposition en canaux perceptuels bien adaptée. Cela

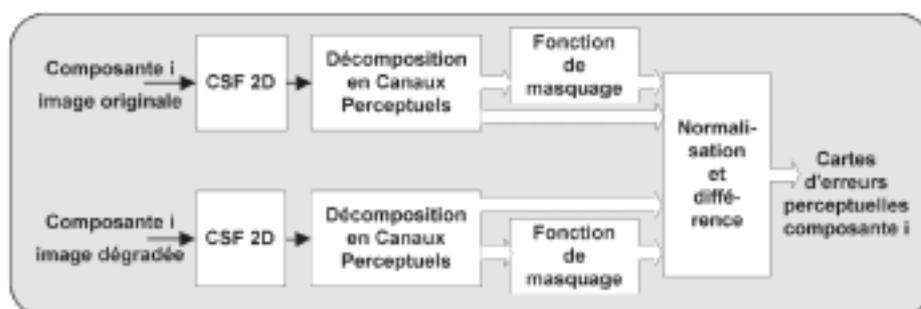


Figure 1. Structure globale du traitement de chaque composante du modèle.

conduit à définir plusieurs filtres qui caractérisent les sous-bandes ou canaux perceptuels. La construction du banc de filtres de la décomposition est sujet à polémique. Bien que certains auteurs [3], [4] réfutent l'hypothèse de canaux séparables dans une représentation polaire, les caractéristiques de la décomposition sont en général décrites en terme de sélectivités radiale et angulaire.

Watson [5] et Daly [6] ont bien détaillé les conditions et propriétés nécessaires des transformations linéaires utilisées pour modéliser la sélectivité spatio-fréquentielle naturelle du système visuel. La plupart des transformations linéaires classiques possèdent certaines de ces propriétés mais rarement toutes. Par exemple, la transformation en ondelettes séparables autorise une sélectivité angulaire selon trois orientations uniquement (0, 90 et 45/135 degrés). La transformée Cortex proposée par Watson [5] est la première transformation qui réunit toutes les conditions. Dans la littérature, on trouve de nombreuses façons de mettre en oeuvre la décomposition en canaux spatio-fréquentiels visant à approcher le comportement du système visuel humain. Cela étant, rares sont les bancs de filtres proposés qui reposent complètement sur des résultats d'expériences de psychophysique et ce, d'autant plus si on s'intéresse aux composantes chromatiques. Daly utilise pour la composante achromatique une décomposition spatio-fréquentielle caractérisée par une sélectivité radiale dyadique et une sélectivité angulaire constante, égale à 30 degrés. Dans notre modèle, nous proposons d'utiliser les décompositions en canaux perceptuels définies à partir de résultats expérimentaux de psychophysique obtenus au laboratoire [7], [8] qui conduisent à 17 canaux pour la composante achromatique et 5 canaux pour chaque composante chromatique. Pour une composante donnée, chaque canal sera indicé par  $(i, j)$  où  $i$  est l'indice de la bande de fréquences radiales et  $j$  l'indice désignant la gamme d'orientations.

**Fonction de masquage:** Nous avons vu que l'effet de masquage traduit la modification de la visibilité d'un signal par la présence d'un autre signal. Cette modification peut aller dans le sens d'une diminution du seuil de visibilité, on parle dans ce cas de facilitation, ou bien dans le sens d'une augmentation du seuil, on parle alors de masquage au sens propre. On comprend donc bien que la prise en compte d'un tel phénomène est un point clé dans bien des applications de traitement d'images. Le principal problème pour modéliser les effets de masquage réside dans la complexité des phénomènes étudiés qui se traduit par une multitude de conditions expérimentales et autant de modèles. Les expériences classiques mesurent la probabilité de détection d'un signal sinusoïdal en présence d'un signal masquant également sinusoïdal. Ces expérimentations portent sur le masquage entre signaux de différentes orientations [9], de différentes fréquences spatiales [10], ou encore entre signaux chromatiques et achromatiques [11], [12], [13]. La plupart de ces expériences conduisent à un modèle du type de celui défini par Legge et Foley [10]. Ce modèle, largement admis, prédit le degré avec lequel un signal cible sinusoïdal est masqué en présence d'un autre signal masquant sinusoïdal. Ce type de modèle a connu

quelques améliorations [14], [9] en introduisant le concept de décomposition spatio-fréquentielle. Il a été ainsi constaté que le masquage dépendait de l'énergie dans un canal et aussi de l'énergie des canaux adjacents en terme d'orientation. Récemment, les modèles de contrôle de gain du contraste ont connu un succès notable car ils prédisent également bien les seuils de visibilité de signaux simples. Initialement développés par Teo et Heeger [15], ces modèles sont sans cesse affinés de manière à expliquer les multiples interactions entre les canaux visuels ou les composantes de l'espace de représentation des couleurs [16], [17]. Dans notre cas, nous explicitons de la manière la plus fine possible les effets de masquage en prenant en compte ces multiples interactions. Le modèle proposé doit s'appuyer sur des expériences qui permettent de bien maîtriser les stimuli proposés aux observateurs. Trop souvent dans la littérature, on trouve des approximations grossières d'auteurs exploitant des résultats d'expériences conduites en utilisant un espace de représentation donné pour produire un modèle dans un autre espace. Afin d'éviter cet écueil, notamment sur tous les aspects de perception couleur, nous avons veillé à mener les expériences de psychophysique nécessaires.

Les étapes décrites précédemment (espace couleur, CSF, décomposition en sous bandes) ne visent qu'à faciliter la modélisation des effets de masquage. Dans un espace de représentation visuelle de l'information à plusieurs canaux et composantes, il est possible de dissocier les différentes origines des effets de masquage :

- le masquage intra canal explicite les relations entre des signaux traités par le même canal ;
- le masquage inter canal caractérise les influences entre des signaux traités par des canaux différents mais sur la même composante ;
- le masquage inter composante traduit les interactions entre des signaux portés par des composantes différentes.

Pour le masquage intra composante (impliquant tant l'effet de masquage intra canal qu'inter canal), nous devons définir une fonction par canal et par composante liée uniquement au contenu de cette composante. On calcule ainsi l'élévation du seuil de visibilité  $T_{i,j,c}^{intra}(m,n)$  en chaque site  $(m,n)$  de chaque canal  $(i, j)$  de la composante  $c$ . Pour le masquage inter composante (non visible sur la figure 1), nous avons également besoin de définir les fonctions permettant de prendre en compte les interactions entre les composantes. Ainsi, on définit l'élévation du seuil de visibilité  $T_{i',j',c' \rightarrow i,j,c}^{inter}(m,n)$  en chaque site  $(m,n)$ , liée à l'action du canal  $(i', j')$  de la composante  $c'$  sur le canal  $(i, j)$  de la composante  $c$ .

Enfin, il faut combiner toutes les fonctions d'élévation du seuil de visibilité de manière à obtenir la variation totale du seuil de visibilité  $T_{i,j,c}(m,n)$  en chaque site  $(m,n)$  de chaque canal  $(i, j)$  de chaque composante  $c$ . Pour cela nous utilisons un modèle multiplicatif. En effet nous avons vu précédemment qu'une telle approche était cohérente avec les résultats expérimentaux. La grandeur  $T_{i,j,c}(m,n)$  décrit donc l'élévation relative des seuils de visibilité due au contenu de l'image par rapport à la situation d'un fond uniforme. On a alors :

$$T_{i,j,c}(m,n) = T_{i,j,c}^{intra}(m,n) \cdot \prod_{i'} \prod_{j'} \prod_{c'} T_{i',j',c' \rightarrow i,j,c}^{inter}(m,n). \quad (1)$$

Les définitions des fonctions  $T_{i,j,c}^{intra}(m,n)$  et  $T_{i',j',c' \rightarrow i,j,c}^{inter}(m,n)$  sont décrites dans ce papier.

**Normalisation et différence :** l'étape de normalisation consiste à diviser les sorties des filtres par  $T_{i,j,c}$  en chaque site  $(m,n)$  dans la mesure où les sorties sont déjà normalisées par le contraste seuil sur fond uniforme grâce à l'application de la CSF 2D. Dans son modèle, Daly propose de diviser la sortie des filtres par la plus petite des deux valeurs de  $T_{i,j,c}(m,n)$  calculées respectivement sur l'image originale et sur l'image dégradée. Nous pensons qu'il est préférable de normaliser les sorties des filtres de chaque image par leur propre élévation relative de seuil  $T_{i,j,c}(m,n)$ . Cela respecte les conditions de perception pour chacune d'elles. Au final, on calcule la différence entre les sorties normalisées de l'image originale et l'image dégradée, ce qui nous fournit 27 cartes d'erreurs perceptuelles normalisées (17 pour la composante achromatique et cinq pour chacune des deux composantes chromatiques).

### 3. Choix d'un espace de représentation des couleurs

Dans la mesure où nous souhaitons obtenir un modèle relativement proche du comportement du SVH, il est nécessaire d'adopter un espace de représentation des couleurs le plus réaliste possible. Pour cela, nous nous sommes appuyés sur des observations neurobiologiques [18] ainsi que sur des résultats d'expériences de psychophysique.

Lorsqu'un flux lumineux rencontre la rétine, il est transformé en plusieurs impulsions électriques. La rétine est tapissée d'une multitude de capteurs réagissant à différentes longueurs d'ondes et à différentes intensités lumineuses. On distingue deux familles de cellules photoréceptrices, les bâtonnets et les cônes. Les bâtonnets sont utiles exclusivement pour la perception des scènes de faibles luminances (conditions scotopiques). Les études sur les cônes ont permis de dégager trois grandes populations qui se distinguent par leur sensibilité. On observe tout d'abord les cônes dits S (pour *small*) dont la sensibilité maximale est située autour de 420 nm (plutôt bleu). Les deux autres familles de cônes, notées M et L (pour *medium* et *large*), ont leur maximum de sensibilité autour de 531 et 558 nm respectivement. Les cônes sont dédiés à la perception photopique, c'est-à-dire pour des conditions de luminance moyenne ou forte. La répartition des photorécepteurs sur la rétine n'est pas uniforme. La fovéa, qui correspond à une zone d'environ un degré de champ visuel, possède une densité de cônes importante alors

qu'elle est dépourvue de bâtonnets. Cette densité chute fortement en zone périphérique alors que la densité des bâtonnets y est importante. Les signaux issus de plusieurs bâtonnets convergent vers une seule cellule, ce qui augmente la sensibilité mais réduit la résolution. Au contraire, plusieurs cellules encodent l'information de chaque cône, ce qui confère à la fovéa une bien meilleure capacité en résolution spatiale que la zone périphérique. Cet aspect illustre le fait que la taille et la répartition des photorécepteurs influent sur la résolution spatiale maximale du système visuel. Dans la rétine, les cônes M et L sont beaucoup plus nombreux que les cônes S ; ce qui a des conséquences sur la résolution spatiale des signaux chromatiques par rapport aux signaux achromatiques.

Les cellules photoréceptrices transmettent leur signal à un ensemble de couches de cellules, véritables neurones rétiniens, qui assurent un premier traitement de l'information. On trouve quatre types de cellules : les cellules horizontales, les cellules bipolaires, les cellules amacines et les cellules ganglionnaires. Les interconnexions entre ces cellules sont à l'origine de la notion de champ récepteur, très importante pour la compréhension des mécanismes de perception du système visuel. Le champ récepteur d'un neurone est défini comme étant la zone pour laquelle la lumière influence la réponse du neurone. Ce concept n'est pas limité aux cellules de la rétine mais s'étend à tous les neurones des organes suivants du système visuel. Il est admis que l'information perçue par la rétine au niveau des capteurs L, M et S est ainsi combinée pour former trois composantes distinctes. La définition de ces composantes perceptuelles détermine un espace perceptuel de représentation couleur. Plusieurs espaces sont décrits dans la littérature présentant quelques différences sur la manière de combiner les réponses des capteurs L, M et S. L'analyse des phénomènes de perception des couleurs a inspiré la définition de nombreux modèles. Ceux qui ont trouvé le plus de justifications tant au niveau physiologique qu'au niveau psychophysique sont les modèles basés sur la théorie des signaux antagonistes. Selon cette théorie, les signaux issus des trois types de cônes L, M, S sont combinés de manière à définir trois nouvelles composantes, une composante achromatique et deux composantes purement chromatiques. La plupart du temps, cela revient à définir une transformation linéaire :

$$\begin{pmatrix} A \\ Cr1 \\ Cr2 \end{pmatrix} = [T] \cdot \begin{pmatrix} L \\ M \\ S \end{pmatrix}, \quad (2)$$

où A, Cr1 et Cr2 représentent les trois composantes perceptuelles et L, M et S les signaux issus des trois types de cônes de la rétine. Pour obtenir l'espace LMS, il faut utiliser une transformation des signaux lumineux issus des trois types de phosphore du moniteur (rouge, vert et bleu), en passant par l'espace CIE  $x'y'$  qui prend en compte les modifications de Judd [19]. Il est important de noter que, partant d'une image numérique RGB, la transformation ainsi proposée est fortement dépendante du matériel utilisé.

Des expériences physiologiques ont montré l'existence de cellules sensibles à des signaux antagonistes. Certains champs récepteurs des cellules rétiniennes ou du corps géniculé latéral (CGL) sont sensibles à des couples de signaux, l'un excitateur l'autre inhibiteur. Les neurones excités par les cônes « rouges » L sont en général inhibés par les cônes « verts » M, et les neurones excités par les cônes « bleus » S sont souvent inhibés par une combinaison des cônes L et M, soit un mélange « jaune ». De Valois et De Valois [20] ont proposé un modèle physiologique de construction des trois composantes en définissant les signaux arrivant sur les zones excitatrice et inhibitrice des champs récepteurs. Pour cela, ils se basent sur les proportions des différents types de cônes qui leur permettent de définir une transformation linéaire entre les signaux L, M et S, et les trois composantes de l'espace antagoniste. Faugeras *et al.* [21] s'appuient davantage sur les courbes d'absorption des différents types de cônes pour définir leur transformation. Enfin, nous trouvons des modèles dont les axes sont entièrement déduits d'expériences psychophysiques. Ces expériences consistent à mesurer la modification de la perception d'un signal chromatique porté par un axe a priori, en fonction de la direction dans le plan chromatique d'un signal dit de masquage. Il apparaît que le seuil de visibilité du stimulus chromatique augmente lorsque le stimulus et le signal masquant sont sur le même axe de l'espace chromatique. Par contre, lorsque les directions sont différentes, le seuil de visibilité du stimulus n'est presque pas modifié par rapport à une configuration sans signal masquant. Parmi les modèles déterminés à partir de ces expériences, nous pouvons citer le modèle de Krauskopf *et al.* [22], celui de Flanagan *et al.* [23] et celui de Webster *et al.* [24]. Ces modèles étant très proches, nous ne donnons que la matrice de transformation proposée par Krauskopf.

$$[T]_{Krauskopf} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ -0.5 & -0.5 & 1 \end{pmatrix}. \quad (3)$$

Dans ses travaux de thèse [25] menés dans notre laboratoire, L. Bédât a validé l'espace de représentation défini par Krauskopf *et al.* [22] pour des images visualisées dans les conditions recommandées par l'ITU [26]. Afin de valider cet espace, L. Bédât a mené une série d'expériences en utilisant l'effet d'adaptation du système visuel. Le principe de ces expériences repose sur le fait qu'une exposition à un signal porté par un axe du plan chromatique augmente le seuil de visibilité d'un stimulus coloré. L'augmentation du seuil est maximale dans la direction de désensibilisation, elle est minimale sur l'axe psychovisuel orthogonal. Pour la désensibilisation, deux procédures distinctes ont été utilisées : l'une temporelle, l'autre spatiale. Lors de ces expériences, Krauskopf a uniquement utilisé une procédure de désensibilisation temporelle. L. Bédât observe que les résultats fournis par les deux procédures confirment la pertinence de l'espace défini par Krauskopf dans les conditions d'observation standard. Ce constat remarquable a notam-

ment permis de préférer l'approche de Krauskopf à d'autres approches, comme celles de Webster *et al.* [24] ou Cavanagh *et al.* [27] validées par une seule procédure expérimentale. Nous avons donc retenu l'espace de Krauskopf, les trois composantes A, Cr1 et Cr2 définissent ainsi notre espace de représentation des couleurs.

Dans cet espace, la composante A représente la composante achromatique. La composante Cr1 prend ses valeurs sur un axe rouge-vert, la composante Cr2 prend ses valeurs sur un axe jaune-violet. La comparaison de ces composantes avec les composantes  $L^*$ ,  $a^*$  et  $b^*$  définies par le Comité international de l'électricité (CIE) révèle que A et Cr1 sont respectivement proches des composantes  $L^*$  et  $a^*$ , en revanche on note une forte divergence de la composante Cr2 avec la composante  $b^*$ .

## 4. Fonctions de sensibilité aux contrastes

### 4.1. Fonction de sensibilité aux contrastes de la composante achromatique

Pour la composante achromatique, on dispose de plusieurs fonctions de sensibilité dans la littérature. Nous avons choisi de reprendre celle définie par Daly [1]. Celle-ci donne la sensibilité  $S_A$  en fonction de la fréquence radiale  $\omega$  en cpd (cycle par degré), l'orientation  $\theta$  en degrés, le niveau  $l$  d'adaptation en luminance en  $cd/m^2$ , la surface  $s$  de l'image en degrés<sup>2</sup>, la distance  $d$  d'observation en mètres et l'excentricité  $e$  en degrés. On a ainsi l'expression suivante :

$$S_A(\omega, \theta, l, s, d, e) = P. \min[S(\frac{\omega}{bw_a \cdot bw_e \cdot bw_\theta}, l, s), S(\omega, l, s)], \quad (4)$$

où le paramètre  $P$  désigne la sensibilité maximale. Il varie d'un observateur à un autre mais nous avons conservé la valeur 250 utilisée par Daly. Les paramètres  $bw_a$ ,  $bw_e$ ,  $bw_\theta$  assurent la prise en compte des changements de la largeur de bande en fonction de la distance, de l'excentricité, de l'orientation respectivement. Leurs expressions sont données par :

$$bw_a = 0.856 \cdot d^{0.14} \quad (5)$$

$$bw_e = \frac{1}{1 + 0.24 \cdot e} \quad (6)$$

$$bw_\theta = 0.15 \cdot \cos(4 \cdot \theta) + 0.85 \quad (7)$$

et  $S(\omega, l, s)$  est définie par :

$$S(\omega, l, s) = ((3.23.(\omega^2.s)^{-0.3})^5 + 1)^{-1/5} . A_l . 0.9. \omega \cdot e^{-(B_l.0.9.\omega)} \cdot \sqrt{1 + 0.06.e^{B_l.0.9.\omega}} \quad (8)$$

$$A_l = 0.801.(1 + 0.7/l)^{-0.2} \quad (9)$$

$$B_l = 0.3.(1 + 100/l)^{0.15} \quad (10)$$

Avec ces valeurs de paramètres, on obtient la fonction de sensibilité de la composante achromatique représentée sur la figure 2.

#### 4.2. Fonctions de sensibilité des mécanismes de perception des composantes Cr1 et Cr2

À partir des données expérimentales collectées au laboratoire, nous avons pu définir une CSF pour chaque composante chromatique. Les résultats expérimentaux suggèrent un comportement de type filtrage passe-bas, qui diffère de celui observé pour la composante achromatique, ce qui implique une autre formulation analytique. Concernant l'évolution de la sensibilité en fonction de l'orientation, les résultats expérimentaux indiquent une sensibilité plus faible pour les directions obliques comme on l'observe pour la composante achromatique. De même que pour la composante achromatique, la valeur moyenne de la luminance de fond  $L_0$  influe sur la sensibilité maximale et la largeur de bande de la fonction de sensibilité. Les résultats expérimentaux montrent que lorsque  $L_0$  augmente la sensibilité maximale diminue alors que la largeur de bande augmente. Toutefois les variations observées pour différentes valeurs de  $L_0$  restent faibles : de l'ordre de 5 % sur la sensibilité maximale et de 10 % sur la largeur de bande. Nous négligeons donc l'influence de ce paramètre en utilisant les données obtenues pour une luminance de fond  $L_0$  de 22 cd/m<sup>2</sup>.

Nous supposons que les évolutions de la sensibilité en fonction de l'orientation et de la fréquence spatiale sont deux phénomènes séparables. Ainsi, la fonction de sensibilité 2D sur la composante Cr1 est donnée par :

$$S_{Cr1}(\omega, \theta) = s_{Cr1}(\omega) \cdot H_{Cr1}(\theta), \quad (11)$$

avec :

$$s_{Cr1}(\omega) = \frac{G}{1 + (\frac{\omega}{\omega_c})^\gamma}. \quad (12)$$

On retrouve bien un comportement passe-bas,  $G$  représentant le gain statique,  $\omega_c$  en cy/deg, et  $\gamma$  permettant de définir la bande passante et la pente, et

$$H_{Cr1}(\theta) = (1 - k.\sin(2.\theta)) \quad (13)$$

pour prendre en compte l'évolution de la sensibilité en fonction de l'orientation.

Finalement, à partir des résultats des expériences de psychophysique, nous obtenons l'expression analytique suivante pour la fonction de sensibilité sur la composante Cr1 (voir figure 3 pour une représentation graphique) :

$$S_{Cr1}(\omega, \theta) = \frac{33}{1 + (\frac{\omega}{5.52})^{1.72}} \cdot (1 - 0.27.\sin(2.\theta)). \quad (14)$$

Pour la composante Cr2, nous retenons le modèle suivant :

$$S_{Cr2}(\omega, \theta) = \frac{5}{1 + (\frac{\omega}{4.12})^{1.64}} \cdot (1 - 0.24.\sin(2.\theta)). \quad (15)$$

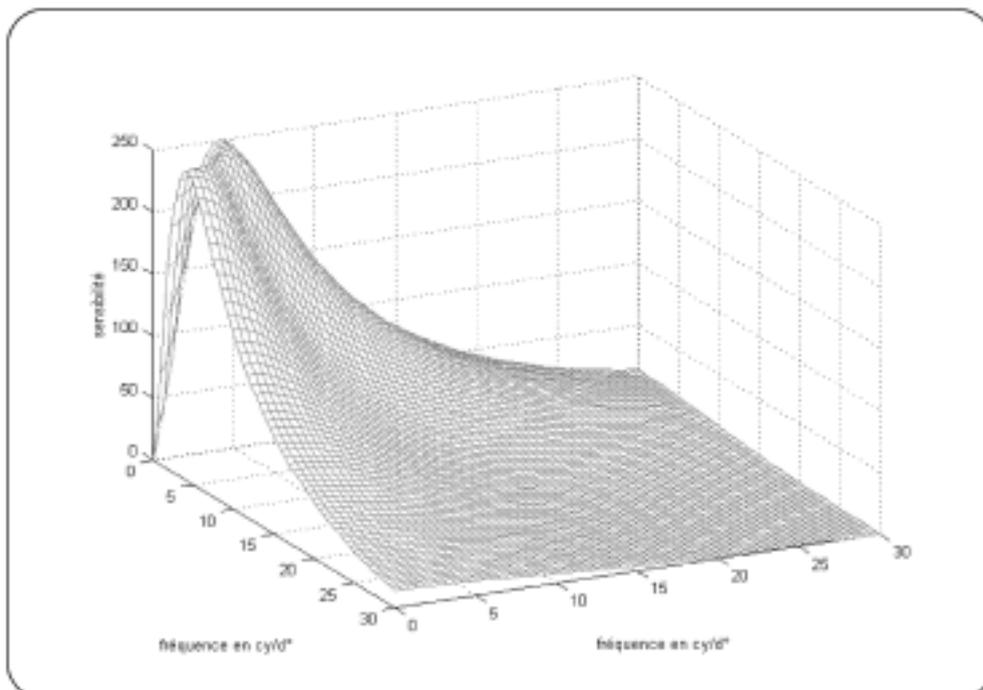


Figure 2. Fonction de sensibilité  $S_A$  aux contrastes pour la composante achromatique.

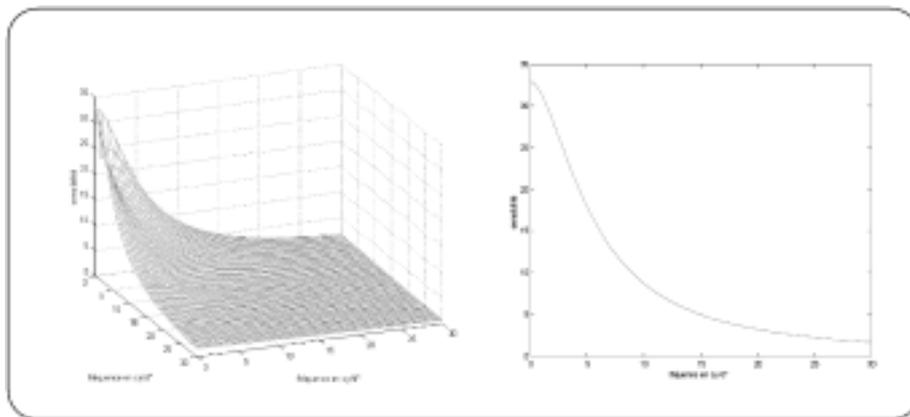


Figure 3. Fonction de sensibilité  $S_{Cr1}$  pour la composante  $Cr1$  : à gauche CSF 2D, à droite CSF dans la direction horizontale.

## 5. Décomposition en canaux perceptuels



### 5.1. Décomposition en canaux perceptuels de la composante A

Le pavage proposé (figure 4) pour la composante A est dans son principe analogue à celui de Watson, mais présente des différences notables qui trouvent leur justification dans les résultats des expériences psychophysiques menées au laboratoire. En effet, contrairement à la plupart des décompositions proposées dans la littérature, celle que nous proposons s'appuie sur un ensemble de mesures expérimentales. Le principe utilisé pour mesurer les largeurs de bandes des canaux repose sur la méthode de masquage proposée par Daugman [3]. Cette méthode consiste à présenter simultanément un signal masqué et un signal masquant, ce dernier étant fixé à un contraste nettement supérieur au seuil de visibilité. On fait varier un des paramètres du signal masqué (l'orientation, la fréquence spatiale) autour du paramètre correspondant au signal masquant, et on mesure alors le seuil de visibilité du signal masqué. Si le signal masquant et le signal masqué excitent le même canal perceptuel, alors la présence du signal masquant augmentera de manière significative le seuil de visibilité du signal masqué. En revanche, si les deux signaux excitent des canaux différents, la présence du signal masquant aura une influence plus réduite sur le seuil de visibilité du signal masqué. En jouant sur la valeur du paramètre modifié du signal masquant, on obtient une série de mesures du seuil de visibilité permettant de définir une fonction de masquage. La bande passante est alors fixée à mi-hauteur du maximum de la fonction de masquage. La répétition de ce principe pour plusieurs fréquences et orientations du signal masquant, a permis de préciser les largeurs de bande de tous les canaux de la représentation.

Dans les conditions normalisées de visualisation à six fois la hauteur de l'écran, on distingue quatre domaines, de fréquences radiales indicés de I à IV. Le domaine I correspond aux fréquences spatiales comprises entre 0 et 1.5 cpd (cycles par degré), le domaine II aux fréquences comprises entre 1.5 et 5.7 cpd, le domaine III aux fréquences comprises entre 5.7 et 14.2 cpd et le domaine IV aux fréquences comprises entre 14.2 et 28.2 cpd. La sélectivité angulaire dépend du domaine de fréquences considéré. Il n'y a pas eu de sélectivité mise en évidence pour les basses fréquences (domaine I). Pour le domaine II, il a été mesuré une sélectivité angulaire de  $45^\circ$ ; on définit donc

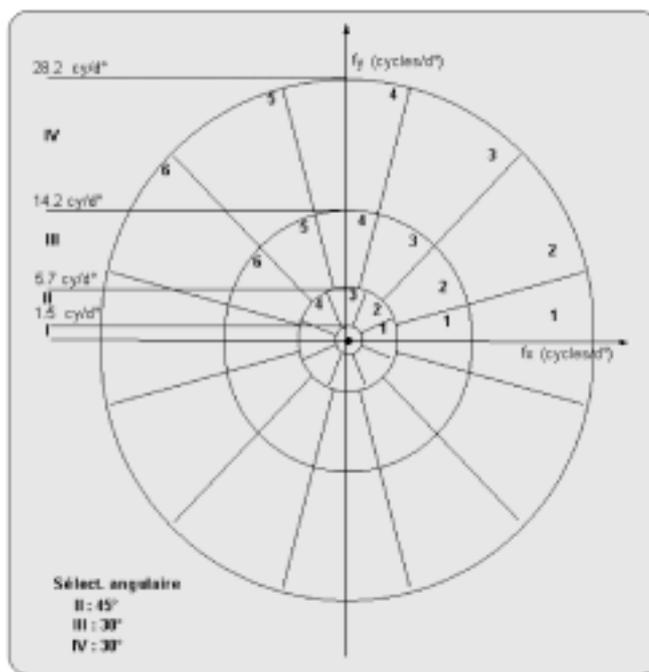


Figure 4. Pavage fréquentiel de la décomposition en canaux perceptuels de la composante A.

quatre canaux orientés pour ce domaine, indicés de 1 à 4. Pour les domaines III et IV, la sélectivité angulaire mesurée est de 30°; six canaux orientés, indicés de 1 à 6, sont donc définis. Dans la suite de ce papier, on nommera chaque canal en faisant référence aux indices du domaine, ou bande radiale, et à l'orientation. Ainsi, le canal correspondant au domaine de fréquences radiales III et d'orientation 2 sera appelé canal ou sous-bande III,2.

Les décompositions proposées dans la littérature ont été synthétisées, la plupart du temps, dans un objectif de codage d'images. Ceci implique des contraintes de sous échantillonnage critique visant à limiter la redondance des informations à transmettre. Dans un cadre plus général, ces contraintes n'ont pas lieu d'être, d'autant plus qu'elles ne permettent pas de respecter fidèlement le comportement du système visuel humain. Le recouvrement entre les filtres illustre en partie la forte redondance de la représentation des signaux par notre système visuel. Dès lors que l'on connaît les caractéristiques du recouvrement, il est possible de mettre en œuvre celui-ci en jouant sur les raideurs des filtres dans les zones de transition.

Afin de définir analytiquement les filtres de notre décomposition, appelés filtres Cortex, permettant un réglage correct dans les bandes de transition, nous avons défini un nouveau modèle s'inspirant de la transformée Cortex de Daly [1]. Un filtre Cortex est obtenu en multipliant un filtre « Fan » par un filtre « Dom ». Les filtres « Dom » sont obtenus par soustraction de filtres « Mesa ». Les filtres « Mesa » proposés sont modélisés dans leur zone de transition par une fenêtre de Hanning, ainsi le filtre est complètement décrit par la fréquence à mi-hauteur  $\omega_c$  et la largeur de la zone de transition  $t_w$  de la manière suivante :

$$mesa(\omega) = \begin{cases} 1 & \text{pour } \omega < \omega_c - \frac{t_w}{2} \\ \frac{1}{2} \cdot (1 + \cos(\frac{\pi \cdot (\omega - \omega_c + \frac{t_w}{2})}{\frac{t_w}{2}})) & \text{pour } \omega_c - \frac{t_w}{2} \leq \omega \leq \omega_c + \frac{t_w}{2} \\ 0 & \text{pour } \omega > \omega_c + \frac{t_w}{2} \end{cases} \quad (16)$$

Les filtres « Fan » sont également construits à l'aide d'une fenêtre de Hanning et sont complètement décrits par l'orientation à mi hauteur  $\theta_c$  et la largeur de la bande de transition  $\theta_w$ . Ainsi, le filtre « Fan » pour l'orientation  $\theta_i$  est donné par :

$$fan_i(\theta) = \begin{cases} 1 & \text{pour } |\theta - \theta_i| < \theta_c - \frac{\theta_w}{2} \\ \frac{1}{2} \cdot (1 + \cos(\frac{\pi \cdot (\theta - \theta_c + \frac{\theta_w}{2})}{\frac{\theta_w}{2}})) & \text{pour } \theta_c - \frac{\theta_w}{2} \leq |\theta - \theta_i| \leq \theta_c + \frac{\theta_w}{2} \\ 0 & \text{pour } |\theta - \theta_i| > \theta_c + \frac{\theta_w}{2} \end{cases} \quad (17)$$

Les paramètres des nouveaux filtres sont déterminés à partir des résultats expérimentaux. Contrairement à ce que l'on peut trou-

ver dans la littérature, la décomposition proposée est ainsi entièrement modélisée à partir de résultats expérimentaux. Il en résulte plusieurs propriétés remarquables :

- le découpage radial est non dyadique,
- la sélectivité angulaire varie avec la bande de fréquences radiales,
- le rapport entre largeur de transition et fréquence à mi-hauteur varie d'une couronne à une autre (ce rapport est constant chez Daly),
- la largeur de la bande de transition des filtres « Fan » ne couvre pas toute la largeur de bande et varie d'un domaine à un autre.

## 5.2. Décomposition en canaux perceptuels des composantes chromatiques

Pour déterminer la décomposition en canaux perceptuels de manière expérimentale, nous avons repris le principe proposé par Daugman énoncé précédemment pour la composante achromatique en supposant que ce principe est également applicable aux composantes Cr1 et Cr2. Pour réaliser des séries d'expériences psychophysiques, nous nous sommes basés sur le protocole de tests à choix forcé, considéré comme le protocole conduisant aux mesures les plus précises et les plus stables. Chaque test consiste à présenter à un observateur simultanément trois images sur un écran (voir figure 5), dans des conditions normalisées d'observation. L'image du centre, position r, correspond toujours au signal masquant seul, l'une des images latérales (position a ou b) est une recopie de l'image du centre. L'autre image latérale (b ou a) correspond au signal masquant auquel on a superposé le stimulus à détecter. L'observateur doit obligatoirement désigner une position (a ou b) de l'image incluant le stimulus. Selon la validité de la réponse de l'observateur, on modifie l'amplitude du stimulus de façon à atteindre

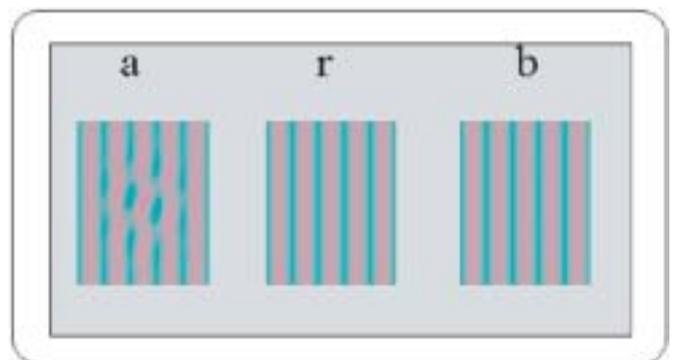


Figure 5. Exemple d'affichage de test à choix forcé de visibilité d'un stimulus porté par Cr1 (amplitude volontairement exagérée ici) superposé à un signal masquant porté par Cr1 à variation horizontale. Le stimulus n'est présent ici qu'en position a et est apodisé spatialement de manière à contrôler son extension spatiale (limitation à la zone fovéale).

rapidement le seuil différentiel de visibilité de celui-ci dans les conditions testées. Dans l'exemple proposé sur la figure 5, le signal masquant et le stimulus ont des orientations différentes. Ainsi en jouant sur l'orientation du stimulus, on peut déterminer la sélectivité angulaire d'un canal en traçant la variation du seuil de visibilité en fonction de l'orientation. Avec le même principe de test mais en faisant varier les fréquences radiales des stimuli, on détermine la sélectivité radiale des canaux.

Nous avons ainsi mis en évidence des phénomènes sélectifs en orientation et en fréquence radiale sur les composantes Cr1 et Cr2. Concernant Cr1, ce constat rejoint les résultats expérimentaux de plusieurs auteurs [24], [28], qui ont remarqué des similarités de traitement entre cette composante chromatique et la composante achromatique concernant la sélectivité angulaire. Nos résultats indiquent que dans les mêmes gammes de fréquences spatiales, les sélectivités angulaires seraient identiques sur les composantes chromatiques et la composante achromatique. Il n'existe pas à notre connaissance dans la littérature, d'élément de comparaison dans l'espace que nous utilisons. Concernant la sélectivité angulaire, on peut tout de même citer les travaux de Bradley, Switkes et De Valois [29], dans lequel les mesures sont effectuées en utilisant une procédure de désensibilisation temporelle. Les auteurs annoncent une sélectivité angulaire de  $58^\circ$  à 2 cpd sur un axe chromatique rouge-vert. Cette valeur est supérieure à celle que nous avons mesurée mais la différence peut s'expliquer par le choix de l'espace couleur. En effet, les auteurs ont synthétisé leurs signaux chromatiques en modulant avec la même amplitude les tensions des canons rouge et vert : l'axe ainsi obtenu diffère significativement de l'axe Cr1.

La décomposition en canaux perceptuels est donc très proche de celle définie pour la composante achromatique. En ce qui concerne le pavage du plan fréquentiel, il est même identique mais limité aux deux premières couronnes (voir figure 6).

Pour la synthèse des filtres, nous reprenons les modèles analytiques des filtres « Mesa » et « Dom » définis respectivement par les relations 16 et 17.

Nous avons ainsi déterminé les valeurs des paramètres des filtres « Mesa » à partir des résultats expérimentaux, les largeurs

de transition obtenues sont plus importantes que sur la composante achromatique.

Pour les paramètres des filtres « Fan », le modèle analytique proposé pour les filtres ne nous permet pas d'obtenir ici une mise en correspondance parfaite avec les données expérimentales. Cela étant, fixer la largeur de la bande de transition dans la couronne II des deux composantes chromatique à la valeur de la largeur de bande totale ( $\theta_w = 45^\circ$ ) reste un bon compromis.

## 6. Fonction de masquage

### 6.1. Modélisation du masquage pour la composante achromatique

Le modèle de masquage intra composante de la composante achromatique défini par Daly n'est pas le plus complet qu'on puisse trouver dans la littérature, notamment parce qu'il ne prend en compte ni les effets de masquage inter canal ni les effets de facilitation. Cependant, les paramètres ont été optimisés à partir d'un nombre important de données expérimentales [6], or nous ne disposons pas de résultats expérimentaux équivalents au laboratoire. Nous conservons donc ce modèle pour la composante achromatique.

L'élévation du seuil de visibilité en chaque site  $(m, n)$  de chaque canal  $(i, j)$  est définie par la relation :

$$T_{i,j}(m,n) = (1 + (k_1 \cdot (k_2 \cdot |f_{i,j}(m,n)|)^s)^b)^{1/b}, \quad (18)$$

où  $f_{i,j}(m,n)$  désigne la sortie de la décomposition en canaux perceptuels. Les courbes expérimentales donnant le seuil de visibilité d'un signal à détecter en fonction de l'amplitude d'un signal masquant peuvent être caractérisées par deux comportements asymptotiques lorsqu'on les représente sur des échelles logarithmiques. Sur l'un deux, correspondant à des contrastes de signal masquant faibles, la pente est nulle. Sur l'autre (contraste de signal masquant fort), on observe une pente positive indiquant une élévation du seuil de visibilité. Cette pente, donnée par le paramètre  $s$ , dépend de la nature du signal masquant. Si on normalise le seuil de visibilité du signal masqué et du signal masquant par rapport au seuil de visibilité sur fond uniforme – ce qui est le cas ici lorsqu'on applique la CSF – la principale hypothèse du modèle proposé est qu'alors une seule fonction peut décrire l'effet de masquage pour toutes les fréquences lorsqu'on considère le masquage intra canal uniquement.

### 6.2. Modélisation du masquage pour les composantes chromatiques

Pour définir le modèle, nous disposons de données expérimentales collectées par L. Bédard. Celui-ci a mesuré le seuil de détec-

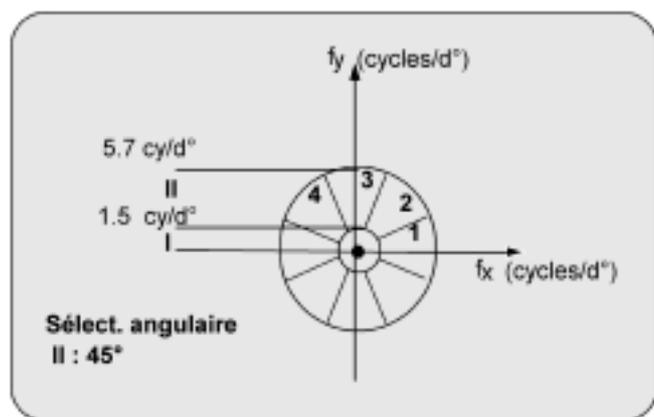


Figure 6. Pavage fréquentiel de la décomposition en canaux perceptuels des composantes chromatiques.

tion d'un signal sinusoïdal sur une composante chromatique en présence d'un autre signal sinusoïdal sur la même composante en faisant varier: la fréquence du signal à détecter, celle du signal, l'amplitude du signal masquant. Les expériences psychophysiques menées sont comparables à celles décrites à la section 5.2 avec le même protocole de test à choix forcé. Des exemples de résultats d'élévation du seuil de visibilité pour la composante Cr1 sont représentés sur la figure 7. La forme générale des courbes obtenues est systématiquement la même, délimitant deux zones: une zone de facilitation pour les amplitudes faibles du signal masquant et une zone de masquage pour les amplitudes plus importantes du signal masquant.

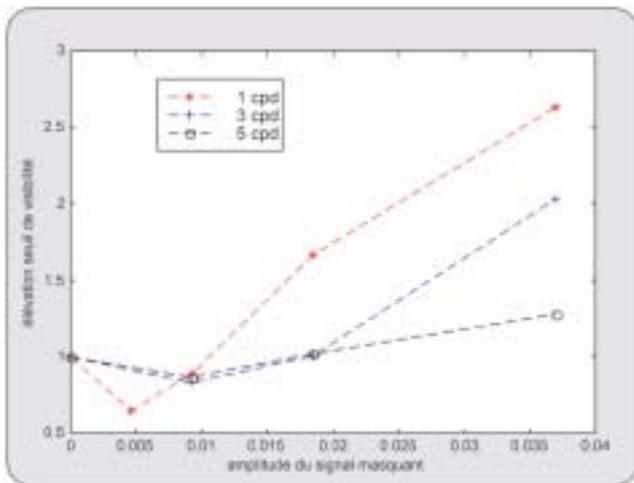


Figure 7. Élévation du seuil de visibilité d'un signal sinusoïdal porté par la composante Cr1 en fonction de l'amplitude d'un signal masquant sinusoïdal porté par la composante Cr1.

Concernant le masquage inter canal, les données expérimentales ne mettent pas en évidence d'influences significatives entre les canaux adjacents radialement. En revanche, aucune expérience n'a été menée en faisant varier l'orientation des signaux, ce qui ne nous permet pas de définir un modèle prenant en compte une interaction entre canaux adjacents angulairement. Nous proposons donc une fonction d'élévation du seuil de visibilité modélisant uniquement les effets intra canaux mais intégrant les effets de facilitation. La forme analytique de cette fonction en chaque site  $(m,n)$  de chaque canal  $(i,j)$  est définie par la relation :

$$T_{i,j,c}^{intra} = \frac{1 + a \cdot |f_{i,j,c}(m,n)| + b \cdot |f_{i,j,c}(m,n)|^2}{1 + c \cdot |f_{i,j,c}(m,n)|}, \quad (19)$$

où  $f_{i,j,c}(m,n)$  désigne la sortie de la décomposition en canaux perceptuels dans la composante chromatique considérée. Les paramètres  $a$ ,  $b$  et  $c$  dépendent de la composante et du canal. À partir des résultats expérimentaux, on peut définir une fonction générique qui prédit l'élévation du seuil de visibilité, caractérisée par trois paramètres: sa pente  $p$  en l'infini, son minimum

$m$  et l'amplitude  $x_0$  du signal masquant pour laquelle on obtient le minimum d'élévation. Nous pouvons chercher les relations qui lient les paramètres  $a$ ,  $b$  et  $c$  du modèle aux paramètres  $p$ ,  $m$  et  $x_0$  de la courbe expérimentale. On obtient ainsi, que la valeur du paramètre  $c$  est la racine positive de l'équation du second degré suivante :

$$p \cdot x_0^3 \cdot c^2 + (m \cdot x_0 + p \cdot x_0^2 - x_0) \cdot c + m - 1 = 0, \quad (20)$$

les deux autres paramètres sont alors donnés par les relations suivantes :

$$b = p \cdot c \quad (21)$$

$$a = p \cdot [1 - (x_0 \cdot c + 1)^2] + c \quad (22)$$

La figure 8 représente l'allure de l'élévation relative du seuil de visibilité en fonction de l'amplitude du signal masquant.

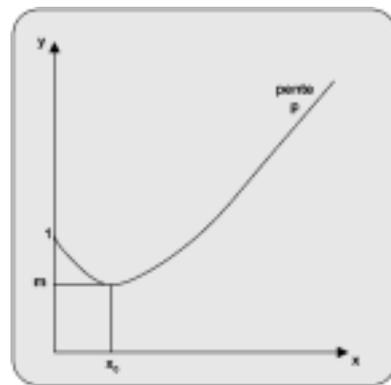


Figure 8. Courbe générique décrivant les effets de masquage intra canal pour les composantes chromatiques.

### 6.3. Modélisation du masquage inter composante

#### 6.3.1. Détermination des interactions significatives

Comme nous l'avons déjà précisé, les composantes A, Cr1 et Cr2 ne sont pas indépendantes. En effet, nous constatons des évolutions du seuil de visibilité de signaux portés par une composante en fonction de l'amplitude de signaux portés par une autre composante. Dans le cas de signaux sinusoïdaux, les évolutions dépendent également de la fréquence du signal à détecter et de celle du signal masquant, de sorte qu'il est préférable de caractériser les interactions entre les canaux perceptuels de deux composantes. Nous disposons de 17 canaux pour la composante achromatique et de 5 canaux pour chaque composante chromatique, il nous faut donc caractériser  $(17 * 5 + 17 * 5 + 5 * 5) * 2$  soit 390 fonctions d'interaction. Fort heureusement, les résultats expérimentaux de L. Bédard suggèrent qu'il est possible de réduire le nombre de fonctions d'interaction, car certaines actions d'un canal d'une composante sur un autre canal d'une

autre composante ne sont guère significatives. Ainsi, il est possible d'écarter les interactions suivantes :

- les actions entre les composantes A et Cr2,
- les actions entre la bande radiale IV de la composante A et les composantes Cr,
- les actions de la bande radiale II de chaque composante Cr sur les autres bandes radiales de l'autre composante Cr,
- les actions de la bande radiale III de la composante A sur la composante Cr1,
- les actions entre deux canaux d'orientations différentes dans une même bande radiale.

De plus, en considérant que les interactions entre deux canaux de même orientation sont identiques quelle que soit l'orientation, il ne reste finalement qu'à définir les 14 fonctions d'interactions indiquées dans le Tableau 1.

Tableau 1. Fonctions de masquage prises en compte dans le modèle.

	composante « masquée »		
	A	Cr1	Cr2
A	intra	I sur I et II,n II,n sur I et II,n	négligée
Cr1	I sur I, II,n et III,n II,n sur I, II,n II sur III,n	intra	I sur I et II,n
Cr2	négligée	I sur I et II,n	intra

### 6.3.2. Modèles

Les allures des courbes expérimentales obtenues par L. Bédet donnant l'évolution du seuil de visibilité d'un signal sinusoïdal en fonction de l'amplitude d'un signal masquant sont de deux natures. La première allure rappelle la forme des courbes obtenues pour le masquage intra composante avec une zone de facilitation et une zone de masquage. Nous utilisons donc le même modèle analytique, noté modèle **A**, donnant la fonction d'élévation du seuil du canal  $(i, j)$  de la composante  $c$  au site  $(m, n)$  en fonction de la sortie de la décomposition en canaux perceptuels au site  $(m, n)$  du canal  $(i', j')$  de la composante  $c'$  :

$$T_{i',j',c' \rightarrow i,j,c}^{inter}(m,n) = \frac{1 + a \cdot |f_{i',j',c'}(m,n)| + b \cdot |f_{i',j',c'}(m,n)|^2}{1 + c \cdot |f_{i',j',c'}(m,n)|} \quad (23)$$

La seconde allure de courbes relevées ne traduit qu'un effet de masquage avec une pente nulle à l'infini. Il faut donc définir un nouveau modèle, noté modèle **B**. La fonction d'élévation du seuil que nous proposons pour ce modèle peut être approximée par la forme analytique suivante :

$$T_{i',j',c' \rightarrow i,j,c}^{inter}(m,n) = a - b \cdot \exp(-c \cdot |f_{i',j',c'}(m,n)|) \quad (24)$$

Cette fonction est caractérisée par trois paramètres : sa limite  $l$  en l'infini, la dérivée  $d$  à l'origine et sa valeur à l'origine, ici constante et égale à 1. Les relations qui permettent alors d'obtenir les paramètres du modèle  $a, b$  et  $c$  sont les suivantes :

$$a = l \quad (25)$$

$$b = a - 1 \quad (26)$$

$$c = \frac{d}{b} \quad (27)$$

À titre indicatif, le modèle retenu et les valeurs de paramètres de celui-ci pour chaque interaction significative (cf. Tableau 1) apparaissent dans le Tableau 2 pour les interactions entre composantes chromatiques.

Tableau 2. Valeurs des paramètres du modèle de masquage entre composantes Cr.

	Cr1 sur Cr2		Cr2 sur Cr1	
	I sur I	I sur II	I sur I	I sur II
modèle	A	B	A	B
a	0.136	1.8	0.09	1.16
b	0.004	0.8	0.1	0.16
c	0.196	0.02	0.82	0.048

## 7. Application à l'évaluation de qualité d'images couleur

Le modèle de perception proposé permet d'obtenir des cartes d'erreurs perceptuelles entre deux images couleur. À partir de ces cartes, on peut combiner toutes les erreurs perceptuelles afin d'obtenir une appréciation globale de qualité d'une image donnée par rapport à sa référence (voir figure 9). Cette étape, appelée

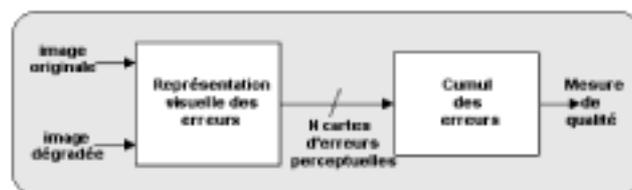


Figure 9. Structure globale d'un critère objectif de qualité d'image avec référence complète.

cumul d'erreurs, est délicate car elle met en jeu des traitements encore mal identifiés au niveau du système visuel. L'ensemble formé du modèle de perception et du cumul d'erreurs constitue un critère d'évaluation de qualité d'images avec référence complète.

## 7.1. Cumul d'erreurs et validation du modèle de masquage

### 7.1.1. Cumul d'erreurs pour images monochromes

Dans le cas d'un modèle de perception comme le nôtre, les erreurs sont non seulement réparties spatialement mais également selon les canaux et les composantes. Le cumul implique donc une fusion à différents niveaux (spatial, fréquentiel, composante). Il peut être décomposé en une succession de traitements de nature différente. Nous répertorions ainsi trois types de cumul :

- le cumul fréquentiel qui recouvre le regroupement des informations issues des différents domaines fréquentiels,
- le cumul spatial qui recouvre le regroupement des informations issues des différents sites dans le plan image,
- le cumul inter composante qui recouvre le regroupement des informations issues des trois composantes de l'espace de représentation des couleurs.

Nous avons étudié précédemment la manière d'opérer au mieux le cumul d'erreurs dans le cas d'images monochromes [30], [31].

Nous avons étudié trois aspects :

- l'adéquation entre complexité du modèle de perception et complexité de la fonction de cumul,
- l'ordre des différents types de cumul
- la manière d'opérer selon le type de cumul.

Sur le premier aspect, nous avons montré [31] que pour évaluer tous les types de dégradations et donc développer un critère d'évaluation de qualité robuste, il est nécessaire de disposer d'un critère complet (bon modèle de perception des erreurs et bon modèle de cumul) pour évaluer la qualité. À partir d'une base de données d'images notées par des observateurs en utilisant un jugement catégoriel à 5 catégories, on peut par exemple évaluer les performances d'un critère de qualité en calculant le coefficient de corrélation linéaire, noté  $cc$ , entre les notes fournies par le critère et celles fournies par les humains. Un critère évolué au niveau du calcul des erreurs perceptuelles mais avec une technique de cumul simple (de type sommation de Minkowski de l'ensemble des erreurs perceptuelles) conduisait dans le meilleur des cas ici à un  $cc$  de 0.8754 alors qu'avec notre technique de cumul la plus évoluée, on atteignait un  $cc$  de 0.9485. De même, un modèle simpliste pour le calcul des erreurs perceptuelles (application d'une CSF sans modèle de masquage) associé à une bonne méthode de cumul conduit à un  $cc$  de 0.7832 seulement. En ce qui concerne le second aspect, l'ordre des opérations de cumul, nous avons montré [31], qu'il fallait effectuer systématiquement le cumul fréquentiel avant le cumul spatial. En effet, cet ordre traduit bien la stratégie du système visuel humain qui consiste à percevoir les dégradations d'abord localement en prenant en compte toutes les composantes fréquentielles (cumul

fréquentiel), puis à regrouper l'ensemble des dégradations perçues sur tous les sites de l'image (cumul spatial) afin de déterminer une impression globale. Le cumul fréquentiel a lui-même été décomposé de manière à distinguer les cumuls fréquentiels angulaire et radial. La figure 10 présente la structure optimisée des différents cumuls dans le cas d'images monochromes.



Figure 10. Structure globale de la fonction de cumul d'erreurs dans le cas d'images monochromes.

Pour le dernier aspect, nous avons défini une méthode *ad hoc* pour chacun des cumuls. Pour le cumul fréquentiel angulaire, une pondération selon les amplitudes est suffisante. Cela est réalisé sous la forme d'une sommation de Minkowski, caractérisée par son exposant. Pour le cumul fréquentiel radial, nous avons observé que la gêne associée aux dégradations était fortement dépendante de chaque bande radiale combien même les grandeurs perceptuelles étaient normalisées. Une pondération linéaire à la sortie du cumul angulaire entre les bandes radiales est bien adaptée. Enfin, le cumul spatial, le plus délicat, a fait l'objet d'une étude à part entière approfondie [30]. Nous avons proposé une approche originale pondérant les dégradations selon deux critères (un basé sur la densité et l'autre sur la structure).

### 7.1.2. Cumuls d'erreurs pour les images couleur et validation du modèle de masquage

Dans le cas des images couleur, nous devons définir une méthode et une position pour réaliser le cumul inter composante. Une étude [32] nous a permis d'avoir une idée de la participation à la gêne de chacune des composantes. Nous proposons un modèle de type combinaison linéaire pour effectuer le cumul inter composantes afin de rendre compte des différences d'importance entre les composantes. Nous avons en particulier remarqué que seules les distorsions sur la composante achromatique permettaient d'obtenir des corrélations correctes dans un contexte de dégradations multiples (engendrées par plusieurs type de codeurs). Ainsi, le cumul inter composantes est très sensible aux dégradations utilisées. Concernant la position de cette fonction de cumul par rapport aux autres, trois configurations ont été envisagées :

- La première approche (ordre 1) consiste à positionner le cumul inter composantes en amont de tous les autres cumuls. Cela revient à considérer que les erreurs réparties sur les trois composantes sont fusionnées localement dans chaque sous bande.

- La seconde approche (ordre 2) insère le cumul inter composantes entre le cumul fréquentiel et le cumul spatial. Dans ce cas les erreurs sont d'abord cumulées au niveau fréquentiel sur chaque composante, puis le cumul inter composantes est réalisé en chaque site avant d'opérer le cumul spatial.
- La troisième approche (ordre 3) place le cumul inter composantes en dernière position. Les cumuls fréquentiel et spatial sont d'abord réalisés. Le cumul inter composantes revient alors à pondérer les gênes globales perçues sur chacune des composantes. Nous avons retenu l'ordre 1 : cumul inter composantes, cumul fréquentiel puis cumul spatial (figure 11). Toutefois, l'analyse des résultats fournis par les possibilités testées montrent que nous aurions pu également choisir l'ordre 2 ; le choix a été guidé par la réduction du nombre de paramètres du modèle final.

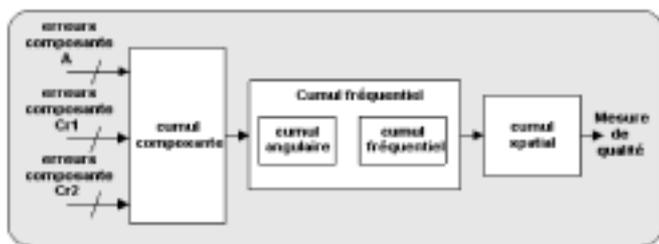


Figure 11. Structure globale de la fonction de cumul d'erreurs dans le cas d'images couleur.

Afin d'apprécier l'apport de notre modèle de masquage intra composante chromatique et inter composante, nous avons procédé à une étude [33] visant à comparer le cc obtenu avec deux critères, l'un n'intégrant pas le modèle de masquage complet contrairement au second, toutes choses restant identiques par ailleurs. Nous avons observé une valeur de 0.93 pour le cc avec le modèle le plus complet alors que si on n'utilise pas de modèle de masquage, le cc tombe à 0.88.

## 7.2. Performances

Une fois notre critère complètement défini, nous devons qualifier ses performances. Nous avons donc constitué une base de données d'évaluations subjectives de dix images de scènes naturelles. Ces images, de taille 512 par 512 pixels, ont été sélectionnées parmi celles communément utilisées par les concepteurs de méthodes de compression d'images. Pour chacune d'elles des versions plus ou moins dégradées ont été générées en utilisant trois techniques de codage différentes :

- le codage JPEG, codage basé bloc utilisant la transformée en cosinus discrète ;
- le codage JPEG2000, codage utilisant la transformée en ondelettes ;
- le codage dénommé LAR [34], codage basé sur une segmentation en régions d'intérêts réaliste par rapport à la structure de l'image à coder.

Le choix de ces méthodes est motivé par deux raisons. La première est que ces techniques sont assez représentatives des méthodes de codage de la littérature : les deux premières techniques sont normalisées et la troisième est basée sur une approche par segmentation en régions. La seconde raison de ce choix, est que les distorsions engendrées par ces méthodes sont de nature et d'aspect visuel différents. En utilisant plusieurs taux de compression, nous aboutissons au total à 140 images à évaluer. Les évaluations subjectives ont été menées dans les conditions normalisées définies par l'ITU : distance de visualisation égale à six fois la hauteur d'écran, utilisation d'une procédure DSIS (*Double Stimulus Impairment Scale*) et une échelle de notation à cinq catégories. Sur cette échelle, la note 5 correspond à la meilleure qualité (pas de dégradations visibles) et la note 1 à la qualité la plus mauvaise (dégradations visibles et très gênantes). Chaque image a été notée deux fois par 14 observateurs âgés de 20 à 45 ans ayant une vision normale (sans ou avec correction optique) y compris pour la perception des couleurs. Un critère de qualité efficace doit satisfaire aux conditions suivantes :

- précision des prédictions,
- monotonie des prédictions,
- cohérence des prédictions.

Afin d'évaluer ces trois aspects, le Video Quality Expert Group (VQEG) a défini cinq mesures que nous allons décrire. Deux mesures rendent compte de la précision de prédiction d'un critère. La première (mesure 1) est l'erreur (entre les notes données par les observateurs, notées MOS et les notes données par le critère notées  $MOS_p$ ) quadratique moyenne pondérée par l'intervalle de confiance à 95 %. La seconde mesure (mesure 2) est définie par l'erreur quadratique moyenne. La mesure de la monotonie (mesure 3) consiste à calculer le coefficient de corrélation linéaire entre les MOS et les  $MOS_p$ . Pour évaluer la cohérence d'une métrique de qualité visuelle, deux mesures sont introduites. L'une (mesure 4) détermine un rapport entre le nombre de configurations mal évaluées par le critère et le nombre total de mesures. Une configuration  $i$  est considérée comme mal évaluée si le critère suivant est satisfait :

$$|MOS(i) - MOS_p(i)| > 2 \cdot \text{Erreur\_standard\_MOS}(i). \quad (28)$$

L'autre mesure introduite est le coefficient de Kappa (mesure 5). Les valeurs de cette mesure sont au plus égales à 1. En général, on considère que le critère d'évaluation de qualité visuelle est satisfaisant à partir de la valeur 0.4.

Nous évaluons les performances de chaque critère à l'aide des 5 mesures définies par le groupe VQEG. Cette évaluation est conduite sur une base constituée des 40 autres images de la base I et des 25 images de la base II. Le Tableau 3 récapitule les résultats obtenus.

Nous pouvons conclure que notre critère est précis relativement à la précision des notes subjectives. L'analyse de la mesure 3 permet d'apprécier la monotonie. Nous obtenons une valeur de coefficient de corrélation supérieur à 0.94, ce qui est remar-

Tableau 3. Performances du critère de qualité.

mesure 1	mesure 2	mesure 3	mesure 4	mesure 5
0.723	0.454	0.941	0.03	0.53

quable compte tenu de la variété des dégradations utilisées. La mesure 5 est nettement supérieure à 0.4. La mesure 4 est particulièrement significative puisqu'elle indique que près de 97 % des images sont correctement évaluées par le critère au regard de la précision sur les notes subjectives. Figure 12, nous présentons deux images dégradées et leur référence respective, pour chaque image nous précisons la note de qualité donnée par les observateurs (MOS), le PSNR et la note fournie par notre critère (MOSp). La combinaison choisie d'images souligne les défauts du PSNR puisque ce critère indique que l'image dégradée *avion* serait de bien meilleure qualité que l'image dégradée *mandrill* alors que le jugement des observateurs est clairement l'opposé (*mandrill* est très proche de la note 5 alors qu'*avion* est proche de la note 1).

## 8. Conclusion

Nous avons présenté dans ce papier un modèle de perception d'images couleur. La définition d'un espace perceptuel a permis d'explicitier finement les différents types d'effets de masquage à partir de résultats d'expériences de psychophysique. Notamment, nous avons pu détailler les interactions qui subsistent entre les différentes composantes d'un espace de représentation couleur. Un tel modèle peut servir dans bien des applications de traitement d'image où l'humain est en fin de chaîne. Nous avons proposé ici une application d'évaluation de qualité d'images. Dans ce cadre, nous nous sommes intéressé à la participation des composantes chromatiques dans l'élaboration d'un jugement subjectif. Nous avons également évalué l'apport de notre modèle de masquage dans ce type d'application. Enfin, les performances du critère d'évaluation de qualité sont remarquables au regard de la variété et de la gamme des distorsions utilisées. Bien sûr, on peut imaginer des métriques plus simples pour évaluer la qualité dans des contextes plus limités (compression JPEG uniquement par exemple) mais la démarche que



Figure 12. Exemples de couple d'images référence, dégradée. En haut image avion MOS = 1.38, PSNR = 26.14 dB, MOSp = 1.17.  
En bas image mandrill MOS = 4.51, PSNR = 24.09 dB, MOSp = 4.16.

nous avons proposé permet d'obtenir une métrique générique et robuste ne nécessitant aucun *a priori* sur le système dégradant. Bien entendu, les perspectives d'application du modèle sont nombreuses. Outre les exploitations en compression, un schéma de tatouage d'images couleur peut bénéficier du modèle, notamment parce que les interactions entre composantes peuvent modifier la visibilité d'une marque. Au niveau du modèle lui-même, nous poursuivons actuellement son extension de manière à définir les régions d'intérêt perceptuelles dans des images.

## Références

- [1] S. DALY, "The visible differences predictor : An algorithm for the assessment of image fidelity", in *SPIE Vol. 1666*, pp. 2-15, 1992.
- [2] D. J. SAKRISON, "On the role of the observer and a distortion measure in image transmission", *IEEE Transactions on Communications COM-25*(11), pp. 1251-1267, 1977.
- [3] J. G. DAUGMAN, "Spatial visual channels in the fourier plane", *Vision research* **24**(9), pp. 891-910, 1984.
- [4] G. C. PHILIPS and H. R. WILSON, "Orientation bandwidths of spatial mechanisms measured by masking", *Journal of the Optical Society of America* **1**(2), pp. 226-232, 1984.
- [5] A. B. WATSON, "The cortex transform : Rapid computation of simulated neural images", *computer vision, graphics and image processing* **39**, pp. 311-327, 1987.
- [6] S. DALY, "A visual model for optimizing the design of image processing algorithms", in *IEEE Int. Conf. on Image Processing*, pp. 16-20, 1994.
- [7] A. SÉNANE, A. SAADANE, and D. BARBA, "The computation of visual components in a subband decomposition of images", in *International Conference on Signal Processing Applications and Technology*, (Santa-clara), october 1993.
- [8] P. LE CALLET, A. SAADANE, and D. BARBA, "Orientation selectivity of achromatic and opponent-colour channels", in *ECVP European Conference on Visual Perception, journal Perception Vol. 28*, (Trieste), august 1999.
- [9] J. M. FOLEY, "Human luminance pattern-vision mechanisms : Masking experiments require a new model", *J. Opt. Soc. Am. A* **11**(6), pp. 1710-1719, 1994.
- [10] G. E. LEGGE and J. M. FOLEY, "Contrast masking in human vision", *J. Opt. Soc. Am.* **70**(12), pp. 1458-1471, 1980.
- [11] A. B. Eugene SWITKES and K. K. D. VALOIS, "Contrast dependence and mechanisms of masking interactions among chromatic and luminance gratings", *Journal of the Optical Society of America* **5**(7), pp. 1149-1159, 1988.
- [12] G. R. COLE, C. F. S. III, and R. E. KRONAUER, "Visual interactions with luminance and chromatic stimuli", *J. Opt. Soc. Am. A* **7**(1), pp. 128-140, 1990.
- [13] M. A. LOSADA and K. T. MULLEN, "The spatial tuning of chromatic mechanisms identified by simultaneous masking", *Vision Res.* **34**(3), pp. 331-341, 1994.
- [14] J. M. FOLEY and G. M. BOYNTON, "A new model of human luminance pattern vision mechanisms : Analysis of the effects of pattern orientation, spatial phase and temporal frequency", in *SPIE Vol. 2054*, pp. 32-42, 1994.
- [15] P. C. TEO and D. J. HEEGER, "Perceptual image distortion", in *ICIP*, pp. 982-986, 1994.
- [16] A. B. WATSON and J. A. SOLOMON, "Model of visual contrast gain control and pattern masking", *Journal of the Optical Society of America* **14**(9), pp. 2379-2391, 1997.
- [17] N. GRAHAM and A. SUTTER, "Normalization : Contrast-gain control in simple (fourier) and complex (non-fourier) pathways of pattern vision", *Vision Research* **40**(20), pp. 2737-2761, 2000.
- [18] B. A. WANDELL, *Foundations of Vision*, Sinauer Associates, 1995.
- [19] D. B. JUDD, "Report of u.s. secretariat committee on colorimetry and artificial daylight", Tech. Rep. 7, bureau central de la CIE, 1951.
- [20] R. L. DE VALOIS and K. K. DE VALOIS, "A multi-stage color model", *Vision Research* **33**(8), pp. 1035-1065, 1992.
- [21] O. D. FAUGERAS, *Digital Color Image Processing and Psychophysics Within the Framework of Human Visual Model*. PhD thesis, University of Utah, 1976.
- [22] J. KRAUSKOPF, D. R. WILLIAMS, and D. W. HEELEY, "Cardinal directions of color space", *Vision Res.* **22**, pp. 1123-1131, 1982.
- [23] P. CAVANAGH, P. FLANAGAN, and O. E. FAVREAU, "Independent orientation-selective mechanisms for cardinal directions of colour space", *Vision Research* **30**(5), pp. 769-778, 1990.
- [24] M. A. WEBSTER, K. K. de VALOIS, and E. SWITKES, "Orientation and spatial-frequency discrimination for luminance and chromatic gratings", *J. Opt. Soc. Am. A* **7**(6), pp. 1034-1049, 1990.
- [25] L. BÉDAT, *Aspects Psychovisuels de la Perception Des Couleurs. Application Au Codage D'images Couleur Fixes Avec Compression de L'information*. PhD thesis, Université de Nantes, IRESTE, 1998.
- [26] ITU, "Méthodologie d'évaluation subjective de la qualité des images de télévision", tech. rep., ITU-R Recommendation BT. 500-10, Geneva, 2000.
- [27] P. CAVANAGH, P. FLANAGAN, and O. E. FAVREAU, "Independent orientation-selective mechanisms for cardinal directions of colour space", *Vision Research* **30**(5), pp. 769-778, 1990.
- [28] A. LÜSCHOW and H. C. NOTHDURFT, "Pop-out of orientation but no pop-out of motion at isoluminance", *Vision Res.* **33**(1), pp. 91-104, 1993.
- [29] A. BRADLEY, E. SWITKES, and K. DE VALOIS, "Orientation and spatial frequency selectivity of adaptation to color and luminance gratings", *Vision Res.* **28**(7), pp. 841-856, 1988.
- [30] P. Le CALLET, A. SAADANE, and D. BARBA, "Frequency and spatial pooling of visual differences for image quality assessment of still images", in *SPIE Human vision and Electronic Imaging, Vol. 3959*, (San Jose), january 2000.
- [31] P. LE CALLET and D. BARBA, "Image quality assessment : from site errors to a global appreciation of quality", in *Picture Coding Symposium*, (Seoul), april 2001.
- [32] P. LE CALLET and D. BARBA, "Perceptual color image quality metric using adequate error pooling for coding scheme evaluation", in *SPIE Human vision and Electronic Imaging*, (San Jose), january 2002.
- [33] P. LE CALLET, *Critères objectifs avec référence de qualité visuelle des images couleur*. PhD thesis, Université de Nantes, IRESTE, 2001.
- [34] O. DÉFORGES and J. RONSIN, "Locally adaptative method for progressive still image coding", *IEEE International Symposium on Signal Processing and its Applications*, 1999.



Patrick **Le Callet**

Docteur en traitement d'images, ingénieur en électronique et informatique, ancien étudiant de l'École Normale Supérieure de Cachan, Patrick Le Callet est également agrégé de génie électrique. Maître de conférence à l'université de Nantes, il effectue des travaux de recherche liés à l'exploitation de la modélisation du système visuel humain dans des applications de traitement d'images. Ses centres d'intérêts actuels sont l'évaluation de qualité des images et de la vidéo (avec référence, référence réduite et sans référence), le tatouage d'image et la construction et l'exploitation de zones d'intérêt perceptuel pour les techniques de compression d'images.



Dominique **Barba**

A obtenu un doctorat de 3<sup>ème</sup> cycle en Télécommunications en 1972 à l'université de Rennes 1 et un Doctorat es Sciences Mathématiques en Informatique à l'université de PARIS VI en 1981. Professeur à l'IRESTE/École polytechnique de l'université de Nantes, ses activités de recherche concernent pour l'essentiel le traitement numérique des images et de la vidéo dans les domaines de l'analyse (segmentation, reconnaissance des formes), de la synthèse ou à celui du codage des images et de la vidéo avec compression d'information. Dans ce dernier domaine, il s'est toujours appliqué à développer des modèles de perception humaine et à les exploiter dans les traitements pour en optimiser visuellement les performances. Il est l'auteur ou co-auteur de plus de 300 papiers de recherche et brevets et est membre de nombreuses sociétés savantes (IEEE, EURASIP, SPIE, SEE, etc.) et relecteur dans les principales revues et colloques internationaux et nationaux de traitement d'images et de la vidéo.

