



UCL

Université
catholique
de Louvain

**LABORATOIRE DE TELECOMMUNICATIONS ET
TELEDETECTION**

B-1348 Louvain-La-Neuve

Belgique

LES TRAITEMENTS PERCEPTIFS D'IMAGES NUMERISEES

Serge COMES

Thèse présentée en vue de l'obtention du grade de
Docteur en Sciences Appliquées

Jury composé de Messieurs :

Paul DELOGNE - *Promoteur*
Eric BOURGUIGNAT (CCETT) - *Examineur*
Murat KUNT (EPFL) - *Examineur*
Auguste LALOUX - *Président*
Benoît MACQ - *Examineur*

Juin 1995

Avant-propos

Une thèse de doctorat est souvent considérée comme l'aboutissement d'un travail personnel. Toutefois, le résultat final ne pourrait être atteint sans l'aide et le concours de plusieurs personnes que je tiens, à présent, à remercier :

- Le Professeur Paul Delogne, pour m'avoir accueilli au sein de l'unité ELHY devenue TELE et, surtout, pour le soin et les commentaires judicieux apportés lors de la lecture du premier manuscrit.*
- Le Docteur Benoît Macq, pour l'attention octroyée à tout instant : "C'est bien, Serge, c'est très bien, continue comme ça, tiens bon !", ainsi que pour ses idées, et réflexions (!), toujours originales.*
- Les Professeurs Auguste Laloux et Murat Kunt, et le Docteur Eric Bourguignat, pour avoir accepté de participer au jury de thèse avec enthousiasme.*
- L'équipe TELE dans son ensemble, avec laquelle j'ai partagé 5 ans de recherches et de bonne humeur. Une mention particulière pour les "volontaires" qui se sont prêtés, avec beaucoup de disponibilité, aux tests psychophysiques qui furent parfois longs et ennuyeux : Olivier Bruyn-donckxs, Isabelle Dargent, Christian Kinon, Benoît Maison, Patrick Piscaglia, Benoît Simon, Jean-Philippe Thiran et Xavier Willame.*
- Le feu b207, une expérience souvent imitée mais jamais égale : partagée avec Olivier Poncin et Manu Fontaine.*
- Le Fonds National de la Recherche Scientifique (FNRS), pour m'avoir permis de poursuivre cette thèse durant quatre années dans une totale liberté d'esprit.*
- Mes parents qui m'ont toujours aidé et soutenu.*

Enfin, je tiens à souhaiter beaucoup de succès à tous les chercheurs qui commencent une thèse de doctorat : c'est une belle aventure ...

Serge Comes
15 mai 1995

Table des matières

1	Introduction	1
1.1	Contexte du sujet	1
1.2	Structure de la thèse	5
2	Le système visuel humain	7
2.1	Description physiologique du système visuel humain	9
2.1.1	L'oeil et la rétine	9
2.1.2	Les corps genouilliers latéraux	11
2.1.3	Le cortex visuel	12
2.2	Caractéristiques psychophysiques de la vision	13
2.2.1	Fonctions de sensibilité	14
2.2.2	Phénomènes d'interaction	17
2.3	Interprétation d'une image par le système visuel humain	21
3	Modèle des composantes perceptives	25
3.1	Séparation des canaux	27
3.1.1	Principe d'incertitude	27
3.1.2	Canaux perceptifs	29
3.1.3	Représentation orientée des fonctions de Gabor	32
3.1.4	Application au traitement des images	35
3.2	Modèle perceptif	41
3.2.1	Composante perceptive	41
3.2.2	Masquage	42
3.3	Définition du contraste	43
3.4	Quantification des paramètres de masquage	48
3.4.1	Méthode d'évaluation des paramètres du modèle	48
3.4.2	Sensibilité fréquentielle : paramètre C_{T_0}	51
3.4.3	Pente de la fonction de masquage : paramètre ϵ	57
3.5	Conclusion	61

4 Critère objectif de qualité visuelle	63
4.1 Evaluation de la qualité d'une image	65
4.1.1 Méthodes subjectives	65
4.1.2 Mesures objectives de la qualité visuelle	66
4.2 Critères de qualité perceptifs	67
4.2.1 Variantes pondératives du MSE	68
4.2.2 Mesures d'activité spatiale	69
4.2.3 Décomposition en multi-canaux	70
4.2.4 Approches sémantiques	71
4.2.5 Discussion	72
4.3 Critère de qualité basé sur les composantes perceptives	73
4.3.1 Schéma global	73
4.3.2 Décomposition perceptive	74
4.3.3 Fonction de masquage	76
4.3.4 Définition du MPSNR	78
4.4 Optimisation du critère	79
4.4.1 Tests psychophysiques	80
4.4.2 Identification du critère	86
4.5 Application à la comparaison de la qualité d'images bruitées	90
4.6 Conclusion	93
5 Codage perceptif	95
5.1 Compression d'images	97
5.1.1 Codage de source	98
5.1.2 Mesures des performances	100
5.2 Codage décorrélatif	102
5.2.1 Codage prédictif	102
5.2.2 Transformations orthogonales	103
5.2.3 Transformations sous-bandes	104
5.2.4 Caractérisation du bruit de codage	106
5.3 Transformation perceptive	111
5.3.1 Extraction de caractéristiques physiques	111
5.3.2 Décompositions en canaux perceptifs	112
5.4 Quantification perceptive	116
5.4.1 Pondération fréquentielle	116
5.4.2 Masquage	117
5.4.3 Quantification vectorielle perceptive	117
5.5 Conclusions	118

6 Transformation sous-bande non-séparable	119
6.1 Position du problème	121
6.2 Sous-échantillonnage bidimensionnel	122
6.2.1 Passe-bande admissible	122
6.2.2 Application à la décomposition hexagonale	124
6.3 Représentation polyphase	128
6.4 Optimisation d'un banc de filtres sous contraintes	129
6.4.1 Critère de reconstruction parfaite	131
6.4.2 Conditions de symétrie	135
6.4.3 Optimisation des coefficients des filtres	137
6.5 Synthèse directe du banc de filtres	137
6.5.1 QMF 4 à 4	140
6.5.2 QMF 4 à 4 orthogonal	145
6.5.3 QMF 2 à 2	147
6.5.4 Aucune condition de symétrie	150
6.5.5 Comparaison des résultats	151
6.5.6 Transformation orientée	155
6.5.7 Conclusions	157
6.6 Synthèse du banc de filtres par la transformation de McClellan	158
6.6.1 Position du problème	158
6.6.2 Transformation de McClellan	159
6.6.3 Conditions de reconstruction parfaite	161
6.6.4 Exemples de transformations perceptives	165
6.6.5 Conclusions	169
7 Quantification perceptive	171
7.1 Quantification de coefficients sous-bandes	173
7.1.1 Définitions	173
7.1.2 Optimisation de la quantification	174
7.1.3 Application d'un modèle perceptif	175
7.2 Pondération fréquentielle	176
7.3 Quantification masquée	183
7.3.1 Auto-masquage	183
7.3.2 Quantification vectorielle masquée	187
7.4 Conclusions	189
8 Post-traitement perceptif	191
8.1 Restauration d'images	193
8.2 Description du post-traitement perceptif	194
8.2.1 Décomposition perceptive et reconstitution	194

8.2.2	Fonction de masquage	195
8.3	Résultats expérimentaux	197
8.4	Conclusions	199
9	Conclusions générales	201
	ANNEXES	205
A	Echantillonnage 2-D	207
A.1	Définitions	207
A.2	Sous-échantillonnage d'un signal sur un sous-réseau	209
B	Transformée en Z à 2-D	211
C	Transformation paraunitaire hexagonale	213
C.1	QMF 4 à 4	213
C.1.1	Conditions de reconstruction parfaite	214
C.1.2	Matrices R_i et $D_i(z_1, z_2)$	214
C.2	QMF 4 à 4 orthogonal	216
C.2.1	Conditions de reconstruction parfaite	216
C.2.2	Matrices R_i et $D_i(z_1, z_2)$	216
C.3	QMF 2 à 2	217
C.3.1	Conditions de reconstruction parfaite	217
C.3.2	Matrices R_i et $D_i(z_1, z_2)$	218
C.4	Aucune condition de symétrie	220
D	Réalisation de la représentation polyphase	221
D.1	Banc de filtres QMF 4 à 4 de type pair-pair	221
D.1.1	Etage d'analyse	221
D.1.2	Etage de synthèse	223
D.1.3	Reconstruction parfaite	226
D.2	Structure QMF 4 à 4 orthogonale	228
E	Schémas des transformations perceptives	231
E.1	Synthèse directe du banc de filtres	231
E.2	Synthèse du banc de filtres par la transformation de McClellan	233
F	Conversions d'unités de la fréquence spatiale	237
	BIBLIOGRAPHIE	241

Liste des abbréviations

AMSD	(Adaptive Morphological Subband Decomposition) Décomposition sous-bande morphologique adaptative
AFB	(Asymmetrical Filter Banks) Transformation sous-bande basée sur un banc de filtres asymétriques
CCIR	Comité Consultatif International pour les Radiocommunications
DCT	(Discrete Cosine Transform) Transformation en cosinus discrète
DFT	(Discrete Fourier Transform) Transformation de Fourier discrète
DPCM	(Difference Pulse Code Modulation) Modulation différentielle d'impulsions codées
DVS	(Digital Video System) Système de visualisation d'images numériques en temps réel
FIR	(Finite Impulse Response) Réponse impulsionnelle finie
FFT	(Fast Fourier Transform) Transformation de Fourier rapide
HVS	(Human Visual System) Système visuel humain
JPEG	(Joint Picture Expert Group) Standard de compression d'images fixes
LBG	(Linde Buzo Gray) Algorithme de quantification vectorielle développé par Linde, Buzo et Gray
LGN	(Lateral geniculate Nucleus) Corps genouilliers latéraux
LOT	(Lapped Orthogonal Transform) Transformation orthogonale à recouvrement
KLT	(Karhunen-Loeve Transform) Transformation de Karhunen-Loeve
LSB	(Least Significant Bit) Bit le moins significatif
MLP	(Multi Layer Perceptron) Perceptron multi-couches
MOS	(Mean Opinion Score) Score reflétant l'opinion moyenne pour des tests subjectifs
MPEG	(Moving Picture Expert Group) Standard de codage d'images mobiles
MPSNR	(Masked Peak Signal to Noise Ratio) Rapport signal à bruit masqué de crête
MSE	(Mean Square Error) Erreur quadratique moyenne
PEST	(Parametric Estimation by Sequential Testing) Méthode d'estimation de paramètres par des tests séquentiels
PSNR	(Peak Signal to Noise Ratio) Rapport signal à bruit de crête
QMF	(Quadrature Mirror Filter) Filtres miroir en quadrature
RBC	(Recursive Block Coding) Codage récursif en bloc
RMSE	(Root Mean Square Error) Racine carrée de l'erreur quadratique moyenne
SNR	(Signal to Noise Ratio) Rapport signal à bruit

Notations

$f(.)$	Fonction scalaire dans le domaine spatial
$F(.)$	Fonction scalaire dans le domaine de Fourier
$\mathbf{v}$	Vecteur
$\mathbf{M}$	Matrice
Γ	Réseau d'échantillonnage
Cte	Constante
$\det[\mathbf{M}]$	Déterminant d'une matrice $\mathbf{M}$

Unités

b/p	Bit/pixel
c/d	Cycle par degré
c/H	Cycle par hauteur d'image
dB	Décibel
Hz	Hertz

Chapitre 1

Introduction

1.1 Contexte du sujet

La visualisation des images a souvent profité des limitations de notre système visuel humain. Le 28 décembre 1895, Louis Lumière organise la première démonstration publique de son "cinématographe" qui projette 24 images par seconde sans que l'oeil ne se rende compte de l'effet cadencé grâce aux propriétés de persistance rétinienne. Cinquante ans plus tard, la télévision nous a permis de recevoir des images de l'autre bout du monde en direct grâce, entre autres, à la limitation de la résolution spatiale du système visuel humain : l'image est balayée séquentiellement suivant plus de 500 lignes horizontales. L'apparition des images numériques, ces dernières années, a ouvert de nouvelles perspectives : d'une part, la qualité de la transmission est mieux contrôlable et, d'autre part, la représentation des images sous forme d'une matrice de points autorise des traitements plus diversifiés, en particulier ceux concernant l'intégration d'effets perceptifs.

L'objectif de la thèse est d'étudier l'application de propriétés perceptives au traitement d'images numériques. Ce projet est ambitieux et motivant car il confronte deux mondes fort éloignés : l'étude du système visuel humain (HVS) qui se rattache à la neurophysiologie et à la psychophysique et le traitement d'images qui fait partie du domaine de l'informatique et des télécommunications.

Les traitements ou transformations d'images sont effectués en vue de la visualisation, du stockage ou de la transmission de ces images et ils se

basent souvent sur des *modèles statistiques*.

- Dans le domaine du *codage de source*, l'objectif est de compresser efficacement le signal en vue de pouvoir le stocker ou le transmettre sur un canal de transmission. Sur base de la théorie de l'information de Shannon [131], l'image subit dans un premier temps une transformation décorrélatrice basée, par exemple, sur la DCT [2] ou la décomposition en sous-bandes [26]. Les coefficients transformés sont ensuite quantifiés de manière à minimiser une fonction de distorsion, la variance de l'erreur de quantification dans le cas des quantificateurs de Lloyd-Max [54]. Enfin, un codage entropique élimine la redondance résiduelle par un codage efficace des échantillons quantifiés.
- Dans le domaine de la *restauration d'images*, ces dernières sont classiquement modélisées par un modèle auto-régressif markovien et les dégradations sont supposées causées par un effet de filtrage et un bruit additif [3, 129].
- Enfin, la *qualité des images* est habituellement évaluée objectivement à l'aide du calcul du rapport signal à bruit (SNR) qui est défini comme le rapport entre la variance du signal et la variance du bruit.

Les modèles statistiques sont populaires parce qu'ils permettent une analyse théorique des problèmes et qu'ils conduisent à des algorithmes de traitement d'images efficaces et réalisables dans un grand nombre de cas.

Quelles que soient les méthodes de traitements ou les applications visées, *les images seront toujours observées et jugées par le récepteur final: l'être humain*. Cette constatation peut paraître évidente, mais elle prend tout son sens lorsqu'on remarque que les aspects perceptifs sont peu pris en considération dans le domaine du traitement d'images. Plusieurs raisons peuvent expliquer cette aberration. Le fonctionnement du système visuel humain (HVS) est décrit par des modèles neurophysiologiques au niveau cellulaire ou microscopique et par des modèles psychophysiques pour les aspects macroscopiques en définissant des notions de sensibilité. Ces modèles ne s'adaptent pas toujours aux réalités du traitement d'images. Le système visuel humain est complexe et encore mal connu. Seul le fonctionnement des premiers niveaux du traitement de l'information visuelle est actuellement modélisé et bien maîtrisé. La validation des modèles est réalisée par des tests psychophysiques ou des expériences neurophysiologiques. Ceux-ci sont appliqués à des phénomènes perceptifs locaux sous des conditions

expérimentales bien définies. L'adaptation ou l'extrapolation des résultats aux applications de traitement d'images n'est pas triviale.

Le **choix du modèle perceptif** qui sera intégré au traitement d'images est primordial : il doit tout à la fois représenter fidèlement le fonctionnement du système visuel humain et favoriser le traitement efficace des images en termes de réalisation, d'interprétation et de qualité des résultats. Il est impossible de développer un modèle qui s'adapte à toutes les situations. Ainsi nous limiterons sa portée aux *applications suivantes*:

- Nous nous intéressons en particulier au *codage perceptif* qui doit atteindre deux objectifs orthogonaux : la réduction de la redondance ou de la corrélation pour une compression efficace et l'augmentation de la pertinence pour une qualité visuelle optimale. Les techniques de codage actuelles allient ces deux aspects avec plus ou moins de bonheur. Par exemple, la combinaison d'une transformation sous-bande ou DCT avec une pondération fréquentielle décorrèle le signal tout en préservant les signaux les plus sensibles visuellement [74, 154]. Par contre, la structure séparable de ces transformations produit du bruit non masqué qui peut être gênant. De même, le codage par objets [106, 127] qui tente de modéliser l'interprétation des images par le système visuel humain n'est pas toujours très efficace aux faibles facteurs de compression.
- La compression n'est pas le seul objectif du codage d'images. La visualisation des images est tout aussi cruciale et elle met en oeuvre des traitements basés sur des *méthodes d'analyse perceptive*. D'une part, il est indispensable d'évaluer la qualité des images codées à l'aide d'un critère perceptif de qualité plus efficace que la mesure peu représentative du rapport signal-à-bruit (SNR). D'autre part, les normes de codage d'images comme JPEG ou MPEG qui seront utilisées dans un proche avenir intègrent peu d'éléments perceptifs. Un post-traitement perceptif des images décodées au récepteur est utile, voire indispensable, pour des facteurs de compression élevés. Nous souhaiterions utiliser le même modèle perceptif pour les méthodes de codage et d'analyse, bien que ces dernières soient moins restrictives étant donné que les contraintes de compression n'entrent pas en compte.
- Le choix du modèle perceptif dépend du niveau de qualité désiré. Si la qualité des images est élevée, les défauts sont imperceptibles ou peu gênants. Dans ce cas, les modèles de détection sont tout indiqués car

ils sont bien connus et facilement validables par des tests de détection. Par contre, si la qualité des images est médiocre, ces modèles ne sont plus applicables car la non-linéarité du système visuel humain empêche l'extrapolation des résultats à des défauts bien visibles. Les modèles d'interprétation du signal visuel seraient plus adéquats, mais ils ne sont pas encore assez développés pour être appliqués au traitement d'images. Le choix du modèle se limitera donc aux applications à *qualité visuelle élevée*. A partir de ce choix, une hypothèse peut être formulée pour caractériser une image codée à qualité élevée. Celle-ci sera modélisée comme le résultat de *la superposition d'un bruit additif sur l'image originale*. Cette hypothèse est un élément important de la justification du modèle perceptif décrit au chapitre 3, en particulier concernant le modèle de masquage.

- Nous limiterons les investigations au traitement des *images monochromes et fixes*. Les effets perceptifs intégrant des signaux de couleur sont beaucoup plus complexes, en particulier les effets d'interférence ou de masquage entre couleurs. D'autre part, les modèles perceptifs temporels [163] sont peu adaptés au codage des images mobiles basé sur la compensation de mouvement. De plus, des interactions spatio-temporelles sont à prendre en considération.

En conséquence de ces limitations, l'étude présentée est spécialement dédiée aux images de qualité élevée, monochromes et fixes. Toute extrapolation aux images fortement détériorées et/ou colorées et/ou mobiles est à considérer avec beaucoup de prudence, étant donné que les hypothèses formulées ne seront plus nécessairement valables.

Sur base de ces considérations, nous avons développé le **modèle des composantes perceptives**. Le principe est de décomposer le signal suivant des canaux perceptifs accordés selon la position dans l'image, la fréquence radiale et l'orientation. Les composantes perceptives résultantes possèdent une sensibilité influencée par les effets de masquage. Perceptivement, ce modèle est adéquat car il décrit le traitement subi par les signaux visuels dans les premiers niveaux du système visuel humain. Les phénomènes de masquage sont largement étudiés dans la littérature avec des résultats concordants. De plus, ce modèle s'applique facilement au traitement d'images car les composantes perceptives peuvent être obtenues par une opération de filtrage, suivie d'un traitement adéquat de ces composantes pour tenir compte de leur sensibilité spécifique.

1.2 Structure de la thèse

La thèse est divisée en trois parties : la modélisation, l'analyse et le codage.

La première partie concerne la **définition du modèle perceptif** qui sera appliqué à l'analyse et au codage d'images monochromes et fixes de qualité élevée.

Le chapitre 2 est une *description succincte du système visuel humain* selon une approche "traitement des signaux visuels". Le chemin des données suivi par le signal visuel et les transformations qu'il subit font référence aux connaissances neurophysiologiques. D'autre part, la sensibilité aux différentes caractéristiques visuelles est étudiée à partir des résultats de tests psychophysiques.

Le chapitre 3 décrit le *modèle des composantes perceptives*. On s'attache à justifier le choix de la décomposition perceptive et du modèle de masquage, en s'appuyant sur le principe d'incertitude et les propriétés neurophysiologiques et psychophysiques de la vision. Une attention particulière est consacrée à l'étude de l'orientation. Afin d'appliquer notre modèle de masquage au traitement d'images, la notion de contraste a été étendue aux signaux aléatoires quelconques, tandis que les paramètres du modèle résultent de tests psychophysiques appliqués à des stimuli de bruit gaussien coloré par un canal perceptif, qui sont plus cohérents avec le bruit de quantification que les gratings habituellement utilisés dans ce type d'expériences.

La deuxième partie concerne l'**analyse perceptive** qui a pour objectif de valider le modèle des composantes perceptives de manière globale avant de l'appliquer au codage perceptif, en développant un critère objectif de qualité et un schéma de post-traitement pour lesquels les conditions d'application sont moins contraignantes.

Le chapitre 4 est consacré au *critère objectif de qualité* où nous proposons de remplacer l'approche basée sur une mesure du SNR par un critère basé sur notre modèle des composantes perceptives, en mettant l'accent sur la non-linéarité du système visuel humain et la localisation de la vision. Les paramètres du critère sont identifiés grâce aux résultats de tests psychophysiques comparant des images codées au seuil de visibilité.

Le chapitre 8 propose une méthode de *post-traitement perceptif* qui s'applique aux images décodées au récepteur. Il s'agit d'un résumé des travaux entrepris avec des mémorants et en collaboration avec l'Ecole Polytechnique

Fédérale de Lausanne (EPFL). Ce chapitre est placé en fin de texte car la combinaison d'un codage décorrélatif suivi d'un post-traitement perceptif est une solution alternative au codage perceptif intégré.

La troisième partie concerne le **codage perceptif** d'images. L'idée est d'intégrer notre modèle perceptif tant à la transformation décorrélatrice qu'à la quantification. Par rapport aux applications d'analyse d'images, cette intégration doit se réaliser en concordance avec un objectif de compression du signal.

Le chapitre 5 survole les *techniques de codage actuelles* en mettant l'accent sur les liens avec notre modèle perceptif. En particulier, le bruit de quantification produit par les transformations séparables DCT et sous-bande est analysé et commenté. Une comparaison des différentes propositions intégrant des aspects perceptifs permet de tracer les voies qui seront poursuivies dans les deux chapitres suivants consacrés à la transformation et à la quantification perceptives.

Le chapitre 6 étudie de nouvelles *transformations du signal* qui sont à la fois *décorrélatives et perceptives*. La transformation sous-bande non-séparable séparant les fréquences et les orientations est un bon candidat pour décomposer le signal suivant les canaux perceptifs du modèle des composantes perceptives. Toutefois, certaines difficultés apparaissent par rapport aux transformations sous-bandes séparables. L'échantillonnage multidimensionnel, la forme géométrique des sous-bandes et la reconstruction parfaite imposent des contraintes qui seront étudiées dans le cadre de deux techniques de synthèse de bancs de filtres. La première méthode synthétise directement le banc de filtres à partir des conditions bidimensionnelles et elle est appliquée à la transformation hexagonale. La seconde se base sur la transformation de McClellan pour synthétiser le banc de filtres à partir d'un prototype unidimensionnel et elle est appliquée aux transformations orientées obtenues par une cascade de bancs de filtres à deux canaux.

Le chapitre 7 s'attache à la *quantification perceptive* des coefficients transformés. L'objectif est double : d'une part, proposer une méthode de quantification basée sur les propriétés de masquage et, d'autre part, comparer les différentes propositions de transformations perceptives sous l'effet de la quantification.

Chapitre 2

Le système visuel humain

Le système visuel humain est considéré comme notre organe sensitif le plus important et le plus évolué. Sa faculté à nous faire évoluer dans un environnement tridimensionnel nous procure une sensation de liberté qu'aucun autre organe ne pourrait offrir. Pour arriver à un tel niveau, notre système visuel humain traite une quantité d'information gigantesque (il peut identifier jusqu'à 100.000 objets différents en 100 ms!), grâce à la chaîne de transmission constituée d'un récepteur (l'oeil et la rétine), d'un canal de transmission via le nerf optique et d'un outil de traitement, cortex visuel. Les tentatives de modélisation du système visuel humain (HVS) sont, à l'heure actuelle, bien insatisfaisantes, essentiellement à cause de nos connaissances encore très réduites du fonctionnement du cerveau, c'est-à-dire la manière dont un stimulus est géré à travers son réseau constitué de 10^{11} neurones.

L'objectif de ce chapitre est de résumer le fonctionnement du système visuel humain en mettant en évidence les caractéristiques qui serviront de base à un modèle adapté au traitement d'images. Deux aspects seront abordés : l'étude neurophysiologique et l'étude psychophysique de la vision.

La neurophysiologie appliquée à la vision s'intéresse au fonctionnement du système visuel humain à l'échelle microscopique. Elle analyse comment chaque cellule parcourue par l'information visuelle traite cette dernière ou modifie ses caractéristiques. C'est l'objet du paragraphe 2.1. Comme complément indispensable, la psychophysique de la vision s'intéresse à l'aspect "macroscopique" du système visuel humain. Ce dernier est considéré, dans ce cas, comme une boîte noire dont on essaie d'extraire des caractéristiques telles que les fonctions de sensibilité et le masquage, abordés dans le para-

graphe 2.2. Enfin, le paragraphe 2.3 synthétise les connaissances actuelles au sujet de l'interprétation visuelle des images.

2.1 Description physiologique du système visuel humain

L'objet de ce paragraphe est de décrire le parcours suivi par l'information visuelle à partir du champ visuel jusqu'au cortex visuel, la partie du cerveau qui analyse l'information visuelle. Basée sur des résultats issus d'expériences neurophysiologiques, l'approche est de type "microscopique" car elle décrit le fonctionnement des diverses cellules traitant le signal.

La figure 2.1 représente la structure générale des composantes du système visuel humain, à savoir l'oeil et la rétine, les corps genouilliers latéraux et le cortex visuel.

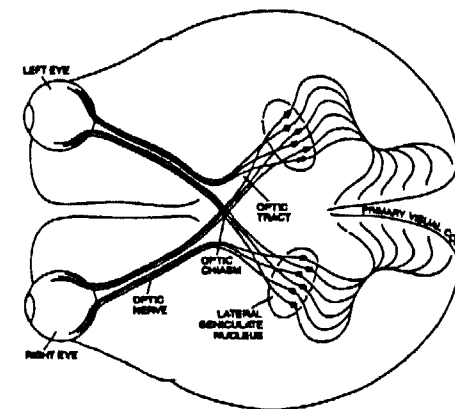


Figure 2.1 - Vue globale du système visuel humain (d'après [51])

2.1.1 L'oeil et la rétine

L'oeil est l'organe de la vision qui reçoit les signaux visuels et, à cet égard, il peut être considéré comme un appareil photographique très évolué. Le cristallin et la pupille jouent les rôles de lentille et de diaphragme. Leur effet peut être modélisé par un filtre bidimensionnel passe-bas [65].

La rétine fixe l'information visuelle comme un film photographique. Elle est structurée en plusieurs couches régulièrement ordonnées dont les cellules photoréceptrices et les cellules bipolaires et ganglionnaires (voir figure 2.2).

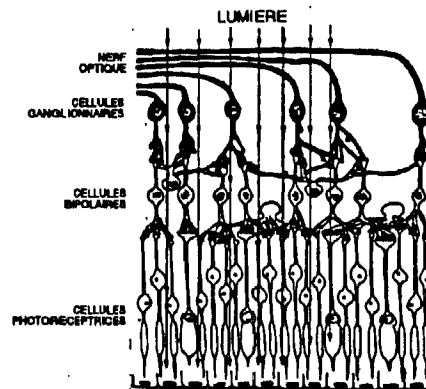


Figure 2.2 - Structure en couches des cellules de la rétine (d'après [35])

2.1.1.1 Cellules photoréceptrices

Les cellules photoréceptrices qui convertissent le signal optique en influx nerveux sont de deux types : les cônes et les bâtonnets. Les *cônes*, au nombre de 130 millions, sont responsables de la vision diurne et sont sensibles à la couleur et aux détails (spatiaux et temporels). Bien qu'ils soient arrangés de manière régulière, leur répartition est inhomogène avec une forte concentration autour de la fovéa (point de la rétine correspondant à l'axe visuel), qui décroît rapidement en fonction de l'excentricité comme illustré à la figure 2.3. D'ailleurs la zone périphérique de la rétine en est totalement dépourvue. Une compression de l'information est donc effectuée lors de cette première étape. La disposition des cônes est également responsable du fait qu'une vision précise (spatiale et temporelle) des détails ne s'obtient que si l'oeil fixe ceux-ci, là où la concentration des cônes est la plus importante. Les *bâtonnets* sont nettement moins nombreux que les cônes (6.5 millions). Ils sont responsables de la vision nocturne car leur sensibilité à la lumière est extrêmement élevée, à tel point qu'ils saturent à la lumière du jour. Par leur disposition plus excentrée que les cônes, les bâtonnets ne peuvent pas saisir tous les détails des images fixées par l'oeil mais seulement leur forme.

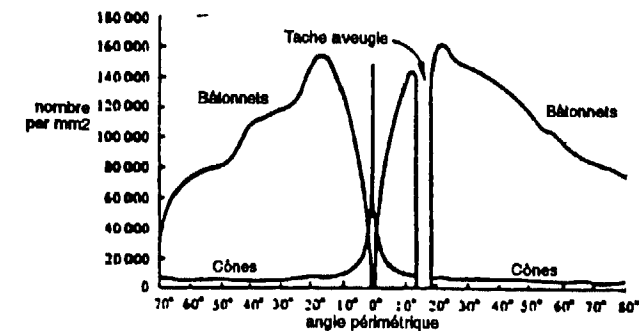


Figure 2.3 - Distribution des cônes et des bâtonnets (d'après [65])

2.1.1.2 Cellules bipolaires et ganglionnaires

Les cellules bipolaires et ganglionnaires rassemblent les signaux issus des cellules photoréceptrices en *champs réceptifs* grâce aux multiples connexions longitudinales et latérales (voir figure 2.2). Ces champs réceptifs définissent la position de l'information dans le champ visuel. Leur forme est régulière de type hexagonale et leur densité diminue en fonction de l'excentricité pour les raisons décrites ci-dessus. Les connexions latérales sont à l'origine du phénomène d'*inhibition latérale*. La réponse de ces cellules est divisée en deux régions isotropes : le centre excitant (ON) et la périphérie inhibitante (OFF) [64] et elle est souvent modélisée par une forme en "chapeau mexicain". Il en résulte que ces cellules ne sont pas sensibles à la valeur absolue des signaux mais bien à la variation spatiale de leur contraste. Ce comportement de type dérivateur ou passe-haut est à la base du modèle de traitement des signaux visuels proposé par D. Marr [84].

2.1.2 Les corps genouilliers latéraux

Après avoir traversé les cellules de la rétine, le signal visuel est transmis aux deux corps genouilliers latéraux (Lateral geniculate nucleus (LGN) en anglais) via le nerf optique (voir figure 2.1). L'information visuelle provenant des deux yeux y est rassemblée en canaux indépendants tout en conservant la localisation des champs réceptifs. Leur comportement est semblable aux cellules ganglionnaires, caractérisé par une réponse isotrope.

2.1.3 Le cortex visuel

Le cortex visuel est une couche de 2mm d'épaisseur et d'apparence striée située à l'arrière du cerveau. La zone, dite 17, correspondant au premier niveau de traitement reçoit l'information du corps genouiller. Cette zone est organisée en couches dont les cellules possèdent une structure hiérarchique bien définie. Cette structure est de type pyramidale, c'est-à-dire que les informations localisées sont petit à petit rassemblées en vue d'extraire des caractéristiques du champ visuel global.

Les cellules du cortex visuel sont répertoriées selon leur complexité d'analyse de l'information visuelle [50]. Au premier niveau, nous trouvons les *cellules simples*, destinataires de l'information provenant de plusieurs cellules adjacentes du corps genouiller. Leur champ respectif présente la particularité d'être de forme allongée, avec toujours une zone centrale excitante et périphérique inhibitrice. Ces cellules sont sensibles à la position dans le champ visuel et surtout à l'orientation. De plus, elles sont regroupées par paires avec des phases en quadrature (déphasage de 90°) [111]. L'information de plusieurs cellules simples passe ensuite au niveau des *cellules complexes*. Celles-ci restent sensibles à l'orientation, mais plus à la position. Par contre elles sont caractérisées par la direction du mouvement. Au niveau des *cellules hypercomplexes* et de plus haut niveau, la globalisation s'intensifie et des notions de discontinuité ainsi que la dimension des singularités sont prises en considération.

Ces résultats sont le fruit des recherches de D. Hubel et T. Wiesel qui leur valurent d'ailleurs le Prix Nobel en 1981. Après une analyse systématique des cellules du cortex visuel, Hubel et Wiesel ont proposé un modèle du cortex : la théorie de l'*hypercolonne*. Le cortex est structuré de façon très ordonnée et régulière selon des colonnes à 3 dimensions [51] comme illustré à la figure 2.4.

- Suivant la direction parallèle à la surface des couches, l'alternance de colonnes sensibles, respectivement à l'oeil droit et à l'oeil gauche (la dominance oculaire), donnent au cortex un aspect strié. La position dans le champ visuel est conservée.
- Suivant la direction verticale à la surface des couches, la structure hiérarchique pyramidale des cellules, décrite ci-dessus, aboutit à un hyperchamp réceptif incluant des informations d'orientation et de mouvement.

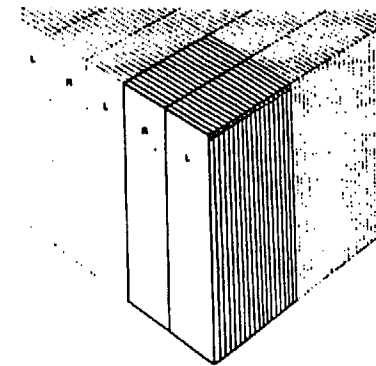


Figure 2.4 - Organisation en hypercolonne du cortex (d'après [51])

- Suivant une strie dans la direction horizontale, l'orientation préférentielle des colonnes suit une évolution ordonnée et monotone. Environ 20 orientations différentes peuvent être discriminées.

En conclusion, nous pouvons dire que le traitement de l'information visuelle dans le cortex est double. D'une part, l'aspect local de l'information est conservé grâce à une parfaite correspondance entre le champ visuel et la disposition des cellules dans le cortex. Il est à noter que ceci est réalisé sans compression supplémentaire, c'est-à-dire que la région fovéale est analysée plus finement que la périphérie du champ visuel à cause de la répartition non-uniforme des cellules réceptrices sur la rétine. D'autre part, la génération d'hyperchamps réceptifs conduit à une globalisation de l'information. Le signal visuel se caractérise progressivement selon des formes en mouvement et de dimension déterminée.

2.2 Caractéristiques psychophysiques de la vision

La description neurophysiologique du fonctionnement du système visuel humain effectuée au paragraphe précédent a ses limites. D'une part, les expérimentations ne peuvent s'exercer que sur des animaux. On voit mal comment on pourrait étudier le comportement d'une partie du cerveau d'un être humain en y enfonçant des électrodes pour analyser ses réactions à tel

type de stimulus ! Bien que le cerveau des singes et des chats possède des similitudes, rien ne prouve que celui de l'être humain possède les mêmes caractéristiques. D'autre part, la neurophysiologie analyse le fonctionnement des cellules au niveau microscopique. Cette approche perd son efficacité lorsqu'on s'aventure dans les niveaux très avancés du cortex où l'interconnexion des neurones s'applique mieux aux modèles connexionnistes et cognitifs. Mais ces domaines sont encore trop peu connus.

La psychophysique analyse le comportement macroscopique de la vision. Le système visuel humain est considéré comme une boîte noire dont on essaie d'extraire des caractéristiques globales. La méthode d'analyse est basée sur le test : des individus placés devant des images de test doivent donner leur avis sur ce qu'ils voient.

Deux types de tests sont envisagés [42] : les tests de détection et d'identification. Les tests de *détection* sont caractérisés par une réponse binaire : vu ou pas vu. Ils conduisent à des caractéristiques évaluées *au seuil de visibilité*. Les tests d'*identification* demandent une interprétation des images et conduisent à des caractéristiques évaluées *au-dessus du seuil de visibilité*. Ces dernières sont plus intéressantes dans le cadre du codage d'images car le bruit résultant d'un codage à bas débit est souvent largement au-dessus du seuil de visibilité. Malheureusement les données au-dessus du seuil de visibilité sont peu nombreuses et peu fiables. Aussi, nous devons souvent nous contenter d'une extrapolation de résultats issus de tests de détection.

Il reste à définir le type de caractéristiques que le système visuel peut analyser sans être influencé par d'autres paramètres. Il est bien connu [101, 72] que le système visuel humain décompose le signal visuel en canaux suivant leurs luminance, couleur, fréquence, orientation, forme, mouvement, etc. Dans le cadre des tests de détection, des *fonctions de sensibilité* sont associées à ces caractéristiques, auxquelles il faut rajouter les *phénomènes d'interaction* car ces canaux ne sont pas toujours indépendants.

2.2.1 Fonctions de sensibilité

En détection, le but poursuivi est de déterminer un *seuil de visibilité* d'un stimulus affiché sur un écran. La modélisation dans le domaine des signaux et systèmes suggère toutefois l'utilisation d'une autre définition : la *sensibilité* S qui est simplement l'inverse du seuil de visibilité.

Les stimuli présentés aux observateurs sont généralement des structures

simples, en particulier une sinusoïde de fréquence spatiale et d'orientation données (communément appelée *grating*). La grandeur caractéristique d'un grating est son *contraste* C défini comme suit [101] :

$$C = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} \quad (2.1)$$

où L_{max} et L_{min} sont respectivement les valeurs de luminance maximale et minimale du grating. Si l'utilisation de gratings facilite la mise en oeuvre des tests de détection, elle complique l'extension des résultats aux images réelles. D'une part, la décomposition en série de Fourier de ces dernières montre qu'elles ne sont pas concentrées en une seule fréquence. D'autre part, les images réelles sont des signaux aléatoires non stationnaires pour lesquels la définition du contraste est loin d'être évidente. Les résultats décrits ci-dessous sont donc à appliquer avec circonspection.

2.2.1.1 Luminance moyenne

Supposons un petit disque superposé à un autre de plus grande dimension et de luminance L_B . Le seuil de visibilité du petit disque, mesuré en terme de luminance ΔL , varie en fonction de L_B suivant la *loi de Weber* [101] :

$$\frac{\Delta L}{L_B} = Cte \quad (2.2)$$

Ce phénomène est attribué à une transformation non linéaire de type logarithmique réalisée au niveau de l'oeil. Cet effet pourrait être pris en compte lors du choix du nombre de pas de quantification lors de la conversion analogique/digitale de signaux vidéos. Toutefois, il faut également tenir compte de la caractéristique γ des tubes CRT qui atténue en partie l'effet de la loi de Weber [94].

2.2.1.2 Fréquence spatiale

Le fonctionnement du système visuel humain décrit au paragraphe 1.1 a mis en évidence un comportement, en fonction de la fréquence spatiale, de type passe-bas au niveau de l'oeil et passe-haut au niveau de la rétine. Ces hypothèses ont été vérifiées par des tests psychophysiques [31, 120, 13, 68, 167] en vue de déterminer la sensibilité de l'oeil à des gratings sinusoïdaux de fréquence donnée. Une synthèse de quelques courbes de sensibilité est illustrée à la figure 2.5. La forme générale de ces courbes est

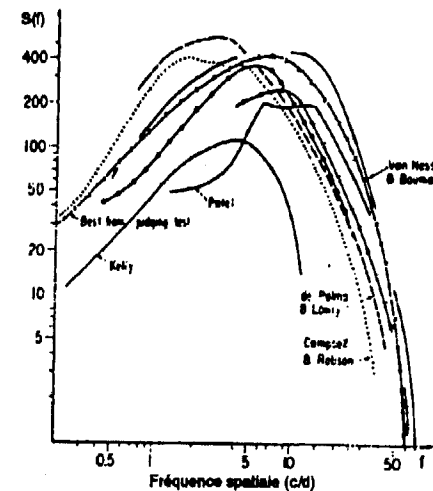


Figure 2.5 - Courbes de sensibilité typiques de l'œil en fonction de la fréquence spatiale (d'après [80])

de type passe-bande, caractérisée par un pic dans les fréquences moyennes autour de 2 à 8 cycles/deg, avec une fréquence de coupure aux environs de 60 cycles/deg.

Plusieurs modèles mathématiques ont été proposés pour représenter cette fonction de sensibilité [143]. La plupart se basent sur le modèle proposé par Mannos et Sakrison [80] qui définissent la fonction de sensibilité du contraste $C(f)$ en fonction de la fréquence spatiale f comme suit :

$$C(f) = K_0 [1 + K_1 f] e^{-(a f)^\alpha} \quad (2.3)$$

où K_0 , K_1 , a et α sont des paramètres à définir. Cette expression permet de tenir compte des variations de facteurs tels que la luminance moyenne, la fréquence temporelle (voir paragraphe 2.2.2.2), etc.

Dans le cadre du traitement d'images, ces résultats présentent deux ambiguïtés.

- Ils sont issus de tests sur des gratings sinusoïdaux fixes. Comme déjà dit précédemment, leur extension à des images réelles ou à du bruit devra tenir compte de leur caractère aléatoire et non stationnaire.

- La limite de validité de ces modèles aux basses fréquences dépend des possibilités de générer des sinusoïdes à ces fréquences. Il est naturellement impossible de tester une sinusoïde de fréquence nulle; les courbes sont généralement extrapolées aux très basses fréquences. Ceci est gênant quant on connaît l'importance du traitement de la composante DC en codage d'image.

2.2.1.3 Orientation

La sensibilité du système visuel humain varie en fonction de l'orientation des stimuli. La visibilité est plus importante pour des signaux orientés verticalement ou horizontalement, et elle est réduite suivant une orientation oblique : ce qui est appelé *effet oblique* [101]. L'effet oblique a pour conséquence une perte de sensibilité de l'ordre de 3 dB pour ces orientations, avec une certaine accentuation en fonction de la fréquence spatiale [12]. L'effet oblique peut s'intégrer dans l'expression (2.3) de la manière suivante :

$$C(f, \theta) = K_0 [1 + K_1 f] e^{-(a(\theta) f)^\alpha} \quad (2.4)$$

où $a(\theta)$ dépend de l'orientation θ du signal.

2.2.2 Phénomènes d'interaction

2.2.2.1 Masquage

Les images naturelles contiennent un arrière-plan texturé plutôt qu'une zone de luminance uniforme. Dès lors, il est important de savoir comment les seuils de visibilité des signaux se modifient en fonction de l'environnement dans lequel ils sont placés : on parlera de *masquage* des signaux en fonction d'un arrière-plan non-uniforme. Plus formellement, le masquage est défini comme toute interférence entre deux ou plusieurs stimuli conduisant à une augmentation ou, le plus souvent, à une réduction de leur visibilité. Dans le cadre du traitement d'images, cette notion de masquage est primordiale car, comme nous l'avons précisé dans l'introduction, nous pouvons considérer qu'une image codée à qualité élevée contient du bruit additif : celui-ci se superpose au signal original et nous sommes intéressés par connaître les conditions dans lesquelles ce bruit est masqué par le signal original.

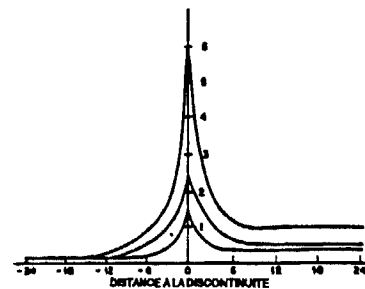


Figure 2.6 - Seuil de visibilité d'un signal superposé à un échelon unité en fonction de la distance à la discontinuité et pour différents rapports de contraste (d'après [94])

2.2.2.1.1 Domaine spatial

Le *masquage spatial* est défini comme l'effet de la diminution de visibilité d'un signal qui se superpose à une zone dont le contraste varie brutalement [94]. La figure 2.6 montre le seuil de visibilité d'un stimulus superposé à un échelon unité en fonction de la distance par rapport à la discontinuité. L'effet est symétrique et il croît avec le rapport de contraste. Mais la caractéristique essentielle est la forte *localisation spatiale* du phénomène. En termes de codage d'images, le masquage spatial indique que le bruit de quantification est moins visible le long des contours des objets, mais il faut bien préciser que cet effet n'est valide qu'aux environs très proches des contours.

Un autre phénomène du même genre est l'*effet Mach*. Cet effet se réfère au changement de perception de la luminosité autour d'un bord : il y a une apparente augmentation de la luminosité du côté clair et une apparente diminution de celle-ci du côté foncé. D'un point de vue physiologique, ces effets spatiaux de masquage sont essentiellement attribués aux interactions latérales (inhibitions latérales) des cellules de la rétine [65].

2.2.2.1.2 Domaine spectral

L'interprétation des phénomènes physiques dans le domaine de Fourier a souvent permis de mieux les comprendre et ainsi d'élaborer des modèles plus adéquats. C'est également le cas des phénomènes de masquage. Dans ce cadre, nous allons considérer un grating-test de fréquence radiale f et

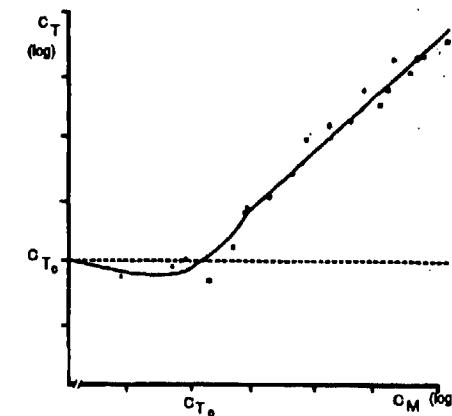


Figure 2.7 - Variation du seuil de visibilité C_T d'un signal en fonction du contraste C_M du masque (d'après [68])

d'orientation θ qui sera superposé à un grating-masque de dimension plus importante et de fréquence radiale f_0 et d'orientation θ_0 . La variation du contraste du seuil de visibilité $C_T(C_M, f, \theta)$ du signal-test, en fonction du contraste C_M du masque et du couple (f, θ) , obéit aux considérations suivantes :

- Pour C_M fixé, le masquage est maximum (c'est-à-dire que $C_T(f, \theta)$ est maximum) si les deux signaux possèdent la même fréquence radiale ($f = f_0$) et la même orientation ($\theta = \theta_0$). L'effet du masquage décroît lorsque le couple (f, θ) s'écarte de (f_0, θ_0) [101]. On observe donc une *localisation spectrale* du masquage. Les bandes passantes fréquentielle Δf et orientée $\Delta \theta$ décroissent en fonction de la fréquence spatiale f_0 du grating-masque [138, 68], mais elles sont indépendantes de son orientation θ_0 [109].
- Pour (f, θ) fixé, le seuil de visibilité $C_T(C_M)$ du signal-test suit une *loi non-linéaire* en fonction du contraste C_M du signal-masque comme illustré à la figure 2.7 [68, 69, 139]. Trois zones de C_M sont considérées par rapport au seuil de visibilité sans masquage C_{T_0} (c'est-à-dire l'inverse des fonctions de sensibilité 2.3 ou 2.4) :

1. $C_M \ll C_{T_0}$: le seuil de visibilité C_T ne varie pas ($C_T = C_{T_0}$)

et aucun effet de masquage n'est observé.

2. $C_M \approx C_{T_0}$: L'effet du masquage est négatif et $C_T < C_{T_0}$. Le signal-test subit une interaction excitante.
3. $C_M \gg C_{T_0}$: L'effet du masquage est positif : le seuil de visibilité C_T suit une loi exponentielle en fonction de C_M . Dans une représentation doublement logarithmique (comme illustrée à la figure 2.7), la pente ϵ de cette loi varie entre 0.6 et 1 [139, 69]. Il est intéressant de remarquer que $\epsilon = 1$ est une généralisation de la loi de Weber aux signaux sinusoïdaux.

L'analyse du masquage dans le domaine spectral a mis en évidence que le système visuel humain décompose le signal en canaux suivant la fréquence spatiale et l'orientation. Cette constatation est à la base du modèle des composantes perceptives, développé au chapitre 3.

2.2.2.2 Interactions spatio-temporelles

Jusqu'à présent, nous n'avons pas abordé le cas des images mobiles pour la bonne raison que les effets temporels doivent s'étudier en interaction avec d'autres facteurs. En particulier, les effets perceptifs temporels des objets en mouvement dépendent de leurs caractéristiques spatiales. La figure 2.8 montre les courbes d'isosensitivité du contraste en fonction des fréquences spatiale et temporelle jointes. Les scientifiques s'accordent sur le fait que notre système visuel humain fonctionne selon deux systèmes indépendants : le système en *régime* et le système *transitoire* [163, 10]. Leurs caractéristiques sont totalement différentes :

- Système en régime

Ce système est caractérisé par une fréquence spatiale moyenne et une fréquence temporelle plutôt basse : de petits objets se déplaçant lentement. Leur comportement est essentiellement influencé par des facteurs spatiaux tels que la fréquence spatiale, l'orientation et les phénomènes de masquage décrits ci-dessus. Par contre la vitesse et la direction du mouvement sont sans influence.

- Système transitoire

Ce système est concentré dans les basses fréquences spatiales et les hautes fréquences temporelles comme des objets relativement uniformes qui se déplacent rapidement dans une direction donnée. Dans

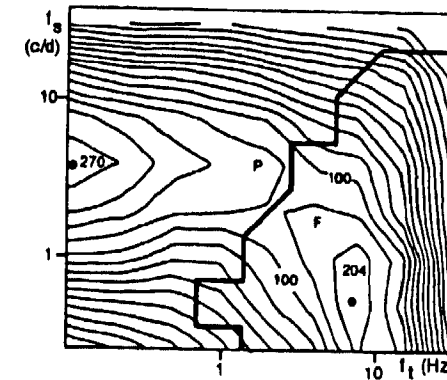


Figure 2.8 - Courbes d'isosensitivité en fonction des fréquences spatiale f_s et temporelle f_t (d'après [163])

ce cas, les effets spatiaux ne sont pas pris en considération, mais plutôt les facteurs influençant le mouvement (vitesse, direction).

L'application de ces phénomènes au traitement des images mobiles n'est pas évident en soi, même si certains auteurs modélisent le traitement de l'information visuelle en intégrant un étage de pré-traitement temporel (en zones "de régime" ou "transitoire") avant une analyse en canaux spatiaux [62].

2.3 Interprétation d'une image par le système visuel humain

"Comment le système visuel humain interprète-t-il les images qu'il reçoit ?" C'est une question que beaucoup de scientifiques se sont posée et dont les réponses sont actuellement peu nombreuses ou incomplètes, malgré les études poussées en vue de comprendre le fonctionnement de notre cerveau. Les études neurophysiologiques et psychophysiques ont toutefois mis en évidence deux faits. Premièrement, l'analyse des images est bien organisée et structurée hiérarchiquement : le traitement du signal visuel s'opère selon de larges canaux parallèles. Deuxièmement, l'organisation neuronale

du cerveau suggère une analyse de type "connexionniste" pour l'interprétation des images. Thorpe [144] propose une interprétation des images par le cerveau en 3 étapes successives.

1. Compression et encodage

L'encodage et la compression s'effectuent aux niveaux de la rétine, du nerf optique et des premiers étages du cortex visuel.

Les signes d'une compression du signal sont nombreux. Physiquement, le signal visuel subit un étranglement lors du passage des $100 \cdot 10^6$ cellules photoréceptrices aux 10^6 cellules des corps genouilliers latéraux via le nerf optique. Spatialement, nous avons vu que le signal est échantillonné de manière non-uniforme au niveau de la rétine. Enfin la gamme dynamique de la sensibilité des cellules est limitée en fonction des fréquences spatiale et temporelle, de l'orientation, de la luminance et de la chrominance.

En même temps que la compression, le signal visuel est encodé et canalisé le long d'*autoroutes neuronales* qui sont accordées localement selon la position dans le champ visuel, la fréquence spatiale et l'orientation.

2. Extraction des caractéristiques

L'étape suivante opère une extraction des caractéristiques du signal : la forme des objets et leur texture. Chaque cellule du cortex visuel peut effectuer environ 4000 opérations différentes. Si on tient compte de leur spécialisation lors de l'encodage, il reste 100 opérations destinées à caractériser les contours, le mouvement, la couleur et la luminance, et la profondeur de champ.

3. Identification

L'identification des objets et de la scène globale a pour but de définir de manière précise la configuration spatiale de l'espace visuel. Cette opération a lieu à un niveau élevé du cortex. Pour donner une idée de sa complexité, le cortex visuel comporte plus de $500 \cdot 10^6$ cellules et il peut identifier jusqu'à 100.000 objets différents en moins de 100 ms. Malgré l'expansion du nombre de cellules par rapport aux premiers niveaux du traitement, une telle faculté d'interprétation ne peut s'analyser que par des modèles connexionnistes. Chaque cellule

est reliée à plusieurs autres cellules et c'est le chemin suivi par les signaux qui identifie la scène observée. L'approche cognitive est importante également car l'apprentissage qui intensifie le passage par les mêmes chemins favorise la reconnaissance et donc l'interprétation des images.

Chapitre 3

Modèle des composantes perceptives

Ce chapitre décrit le modèle perceptif de base qui sera utilisé dans le cadre du critère de qualité et de la transformation perceptive. Il est constitué de deux parties : une décomposition du signal en canaux perceptifs et une mesure de la sensibilité aux composantes perceptives résultantes.

Comme nous l'avons vu dans le chapitre précédent, le système visuel humain décompose l'information visuelle en canaux suivant la position dans le champ visuel, la fréquence radiale et l'orientation. Cette propriété de localisation spatio-fréquentielle est mise en relation avec le principe d'incertitude qui postule que tout signal ne peut être à la fois parfaitement localisé dans le domaine spatial et dans le domaine spectral. La représentation qui minimise l'incertitude est donnée par les fonctions de Gabor qui sont justement semblables au profil des cellules du cortex visuel.

La notion de canal orienté n'est pas unique et elle dépend de la représentation que l'on choisit. Nous verrons quelle est son influence sur la réponse impulsionnelle des fonctions de Gabor. Mais il est encore plus intéressant d'étudier ce que signifie l'orientation dans le cadre du traitement d'images. La discussion portera sur deux exemples théoriques, à savoir l'échelon orienté et la texture orientée.

Les phénomènes de masquage réduisent la sensibilité des composantes perceptives en fonction du contenu du signal dans un canal perceptif donné. Le modèle de masquage sera caractérisé par deux paramètres : la sensibilité sans masquage et la pente de la fonction de masquage. Ces deux paramètres

seront quantifiés sur base de tests psychophysiques (méthode PEST). Les stimuli utilisés ne seront pas des gratings comme c'est le cas dans la littérature, mais du bruit coloré par un canal perceptif qui est plus proche du modèle perceptif et du bruit de codage. Ce type de signal nécessite de redéfinir la notion de contraste et de la généraliser à tout type de signal aléatoire.

3.1 Séparation des canaux

3.1.1 Principe d'incertitude

La classification des cellules du cortex visuel proposée par Hubel et Wiesel (voir paragraphe 2.1.3) a été interprétée de différentes manières par les chercheurs. Dans le domaine spatial, elle modélise le système visuel humain comme un détecteur de contours localisés hiérarchiquement. D. Marr [83, 84] a proposé un détecteur de singularités qui se base sur le principe d'inhibition latérale pour modéliser les phénomènes perceptifs dans le domaine spatial. Cet algorithme applique l'opérateur Laplacien d'une fonction gaussienne aux images en vue d'extraire leurs singularités. La fonction gaussienne élimine le bruit haute fréquence tandis que les zéros issus de l'opérateur Laplacien déterminent les positions des points d'inflexion des variations de contraste les plus prononcées. Les images sont représentées comme une collection de singularités. Cette approche de détection de singularités a été appliquée dans différents domaines comme la reconnaissance de formes ou le codage d'images [66]. Si on envisage le domaine spectral, le modèle de masquage énoncé au chapitre 2 suggère d'appliquer une décomposition du signal visuel en composantes harmoniques. Le système visuel humain peut être modélisé par un ensemble de filtres caractérisés par une fréquence spatiale et une orientation déterminées.

Appliquées aux images, les approches spatiales et spectrales prises séparément ont chacune leurs limites. Les filtres gaussiens de Marr doivent tout à la fois bien localiser la singularité et éliminer la majeure partie du bruit haute fréquence, ce qui impose un compromis au niveau de l'étalement des filtres. De plus les notions de texture ou de contenu spectral ne sont en aucun cas prises en compte. D'autre part, l'approche spectrale basée sur une transformée de Fourier ne tient pas compte du contenu spatial d'une image dont les caractéristiques ne sont pas du tout homogènes. Les approches uniquement spatiale ou spectrale ne sont pas suffisantes car la représentation de signaux ne peut pas localiser parfaitement leurs caractéristiques à la fois dans le domaine spatial et spectral. En effet, considérons une fonction $f(t)$ et sa transformée de Fourier $F(\omega)$. Selon la propriété d'échelle de la transformation de Fourier, nous avons :

$$|a| f(at) \leftrightarrow F\left(\frac{\omega}{a}\right) \quad (3.1)$$

Plus une fonction est localisée dans le domaine spatial, moins elle l'est dans

le domaine spectral, et vice versa. Cette propriété est connue sous le nom de "principe d'incertitude". Les travaux d'Heisenberg et les relations d'indétermination de la mécanique quantique montrent que la spécificité jointe en temps et fréquence d'un signal est fondamentalement limitée par une borne inférieure du produit de sa bande passante et de sa durée. Sur base de ces travaux, Gabor [44] et ensuite Shannon [132, 131] ont déduit les relations d'incertitude pour la théorie de l'information. On peut définir l'extension effective d'une fonction dans les domaines spatial Δx et fréquentiel Δu respectivement, en considérant la racine carrée de la variance de la distribution d'énergie de cette fonction.

$$(\Delta x)^2 = \frac{\int_{-\infty}^{\infty} x^2 |f(x)|^2 dx}{\int_{-\infty}^{\infty} |f(x)|^2 dx} \quad (3.2)$$

$$(\Delta u)^2 = \frac{\int_{-\infty}^{\infty} u^2 |F(u)|^2 du}{\int_{-\infty}^{\infty} |F(u)|^2 du} \quad (3.3)$$

où $F(u)$ est la transformée de Fourier de la fonction $f(x)$. En supposant des signaux à extension finie, à savoir :

$$\sqrt{x}f(x) \rightarrow 0, \quad \text{si } |x| \rightarrow \infty \quad (3.4)$$

le principe d'incertitude montre que [105] :

$$\Delta x \cdot \Delta u \geq \frac{1}{4\pi} \quad (3.5)$$

Ce produit d'incertitude est minimisé par les fonctions de Gabor [44] qui sont des fonctions gaussiennes décalées et modulées (voir figure 3.1) :

$$g(x) = e^{-\pi(\frac{x-x_0}{\lambda})^2} e^{-2\pi j(u_0(x-x_0))} \quad (3.6)$$

$$G(u) = e^{-\pi((u-u_0)\lambda)^2} e^{-2\pi j(x_0(u-u_0))} \quad (3.7)$$

où x_0 et u_0 définissent respectivement la position centrale et la fréquence centrale des fonctions de Gabor, tandis que λ est leur paramètre d'extension.

Le caractère bidimensionnel des images nécessite de généraliser la relation d'incertitude pour des signaux 2-D. Les fonctions 2-D possèdent des moments de second ordre suivant les deux axes de coordonnées, mais également

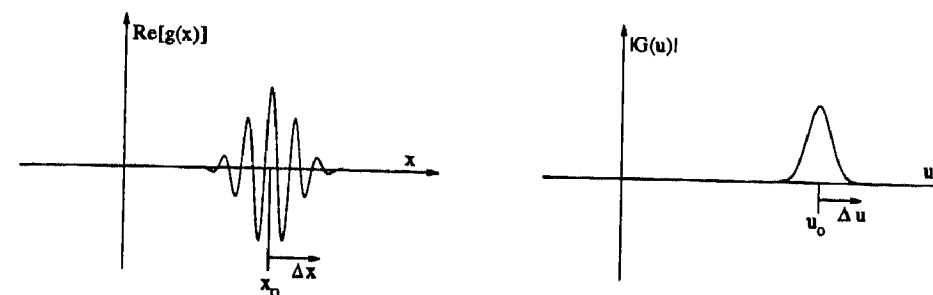


Figure 3.1 - Fonction de Gabor 1-D centrée en x_0 et u_0 minimisant le principe d'incertitude

un moment croisé. Afin d'éliminer ce dernier, Daugman [28] considère ces fonctions après qu'elles aient subi une rotation de telle sorte que leurs axes principaux soient parallèles aux axes de coordonnées. De cette manière l'effet des moments croisés disparaît et une extension séparable du principe d'incertitude 1-D conduit à la relation suivante :

$$\Delta x \cdot \Delta y \cdot \Delta u \cdot \Delta v \geq \frac{1}{16\pi^2} \quad (3.8)$$

Il est facile de montrer que les fonctions de Gabor minimisent cette relation pour autant qu'elles soient séparables par rapport aux axes cartésiens (voir figure 3.2).

$$g(x, y) = e^{-\pi\left[\left(\frac{x-x_0}{\lambda_x}\right)^2 + \left(\frac{y-y_0}{\lambda_y}\right)^2\right]} e^{-2\pi j[u_0(x-x_0) + v_0(y-y_0)]} \quad (3.9)$$

$$G(u, v) = e^{-\pi\left[\left((u-u_0)\lambda_x\right)^2 + \left((v-v_0)\lambda_y\right)^2\right]} e^{-2\pi j[x_0(u-u_0) + y_0(v-v_0)]} \quad (3.10)$$

3.1.2 Canaux perceptifs

Si une analyse jointe spatio-fréquentielle serait efficace et mieux adaptée au contenu des images, elle se justifie également dans le cadre d'un modèle perceptif. Il existe plusieurs signes [101] de séparation du signal visuel en canaux accordés suivant 3 composantes : la position dans le champ visuel, l'orientation et la fréquence spatiale.

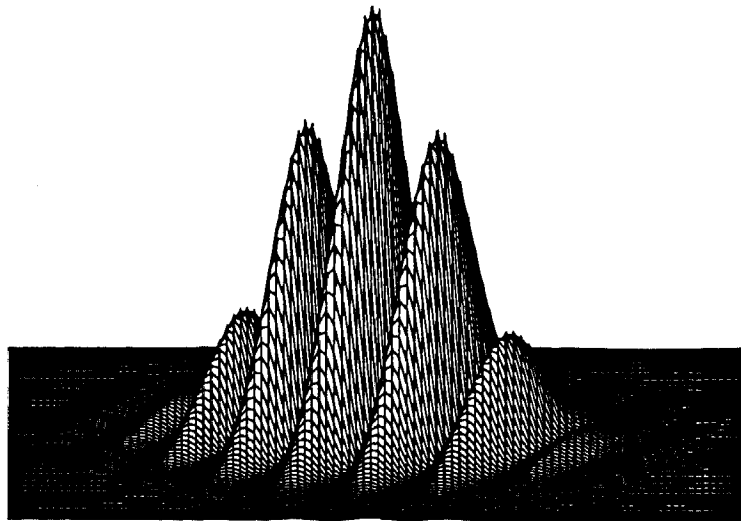


Figure 3.2 - Fonction de Gabor 2-D séparable (partie réelle)

- Position dans le champ visuel

Nous avons vu dans le paragraphe 2.1.1 que les champs réceptifs sont ordonnés, de forme hexagonale et plus concentrés autour de la fovéa. De plus l'agencement spatial est maintenu lors de la projection neuronale de la rétine dans le cortex visuel (voir paragraphe 2.1.3).

- Orientation

La théorie de l'hypercolonne due à Hubel et Wiesel décrite au paragraphe 2.1.3 montre que les cellules du cortex visuel sont sensibles à différentes orientations suivant une répartition bien ordonnée.

- Fréquence spatiale

Même si aucune preuve physiologique ne corrobore cette hypothèse, des signes [145] font supposer que les cellules d'une hypercolonne sont également sensibles à une gamme de fréquences spatiales.

D'un point de vue psychophysique, les phénomènes de masquage justifient ces hypothèses. De plus, comme nous l'avons décrit au paragraphe 2.2.2.1, la localisation du masquage en fonction de la position dans le champ visuel, l'orientation et la fréquence spatiale nous autorise à émettre l'hypothèse

que ces canaux n'interagissent pas entre eux : les canaux perceptifs seront considérés comme *indépendants*.

Comment peut-on modéliser le profil spatio-fréquentiel de ces canaux perceptifs ? Depuis une vingtaine d'années, plusieurs chercheurs [81, 112, 29] ont reconnu que le profil des cellules simples du cortex ressemblait étrangement aux *fonctions de Gabor* décrites ci-dessus. En relation avec le principe d'incertitude, certains chercheurs [63, 125] ont interprété cette observation comme le signe d'un "codage économique" de l'information visuelle effectué par le système visuel humain. Ce dernier représente l'information spatio-fréquentielle de la manière la plus compacte possible à l'aide de fonctions de Gabor et l'interprète séparément dans chaque canal avant de la globaliser lors d'un processus cognitif ultérieur (voir paragraphe 2.3). Il est également intéressant de remarquer que la réponse fréquentielle des fonctions de masquage ressemble à celle des fonctions de Gabor, ce qui renforce notre hypothèse des canaux séparés indépendants [109, 138, 167].

Combien de canaux perceptifs sont-ils nécessaires pour recouvrir le champ visuel ? Considérons un canal perceptif représenté par un profil de Gabor centré autour de la fréquence radiale ω_0 . Ce canal est caractérisé par une bande passante radiale $\Delta\omega$ et une demi-bande passante orientée $\Delta\theta_{1/2}$. Deux résultats peuvent être mis en évidence :

1. Les cellules du cortex ont une bande passante radiale $\Delta\omega$ qui varie entre 0.5 et 2 octaves. Mais il est intéressant de constater que leurs bandes passantes radiale et orientée sont corrélées. De Valois et al rapportent dans [32] que cette propriété est la plus importante observée au niveau du système visuel humain. Daugman [28] a modélisé cette corrélation par l'expression suivante :

$$\Delta\theta_{1/2} = \arcsin \left[\lambda \frac{(2^{\Delta\omega} - 1)}{(2^{\Delta\omega} + 1)} \right] \quad (3.11)$$

où λ est un paramètre de rapport d'aspect des fonctions de Gabor compris entre 0.5 et 1. Par exemple, pour $\lambda = 0.6$, $\Delta\omega = 1$ octave correspond à $\Delta\theta_{1/2}$ égal à 12° .

2. Des résultats psychophysiques ont montré [167] que les bandes passantes de masquage correspondaient aux valeurs mentionnées ci-dessus, c'est-à-dire $\Delta\omega$ entre 1 et 2 octaves, et $\Delta\theta_{1/2}$ entre 10° et 20° .

A partir de ces données et de l'hypothèse de canaux indépendants, un choix de 4 ou 5 fréquences radiales et 4 à 9 orientations permettrait de recouvrir la totalité du spectre visuel. Dans le domaine spatial, le nombre de canaux nécessaires est lié à l'extension spatiale correspondante des filtres de Gabor.

3.1.3 Représentation orientée des fonctions de Gabor

Jusqu'à présent nous avons parlé d'orientation sans vraiment définir le terme. Il existe plusieurs manières d'envisager l'orientation des canaux perceptifs décrits par des fonctions de Gabor. Ce paragraphe est destiné à comparer le profil des canaux de Gabor suivant le type de représentation envisagée : cartésienne ou polaire, logarithmique ou non. Nous allons considérer que les fonctions de Gabor sont réelles et qu'elles sont suffisamment limitées spectralement pour négliger les effets de repli de spectre consécutif à l'échantillonnage cartésien des fonctions.

La première approche considère une fonction de Gabor centrée à l'origine dans le domaine spatial (sans perte de généralité) et en (u_0, v_0) dans le domaine spectral dont les axes principaux sont parallèles aux coordonnées cartésiennes (u, v) . Il s'agit de la représentation 2-D séparable proposée par Daugman 3.1.1. La fonction $g_c(x, y)$ et sa transformée de Fourier $G_c(u, v)$ sont définies par la relation (voir figure 3.3) :

$$g_c(x, y) = \cos(u_0 x + v_0 y) \cdot e^{-\left[\left(\frac{x \Delta u}{2}\right)^2 + \left(\frac{y \Delta v}{2}\right)^2\right]} \quad (3.12)$$

$$G_c(u, v) = e^{-\left[\left(\frac{u - u_0}{\Delta u}\right)^2 + \left(\frac{v - v_0}{\Delta v}\right)^2\right]} + e^{-\left[\left(\frac{u + u_0}{\Delta u}\right)^2 + \left(\frac{v + v_0}{\Delta v}\right)^2\right]} \quad (3.13)$$

où Δu et Δv sont les bandes passantes spectrales.

Les autres représentations se basent sur le fait que les champs réceptifs ne sont pas rectangulaires (ils seraient plutôt hexagonaux [63]). De plus la décomposition en canaux perceptifs selon la fréquence radiale et l'orientation azimutale s'accommode mieux d'une représentation polaire ou selon les directions principales.

La représentation selon les directions principales consiste à effectuer une rotation des axes (u, v) vers (u', v') d'un angle θ_0 qui vaut :

$$\theta_0 = \arctan\left(\frac{v_0}{u_0}\right) \quad (3.14)$$

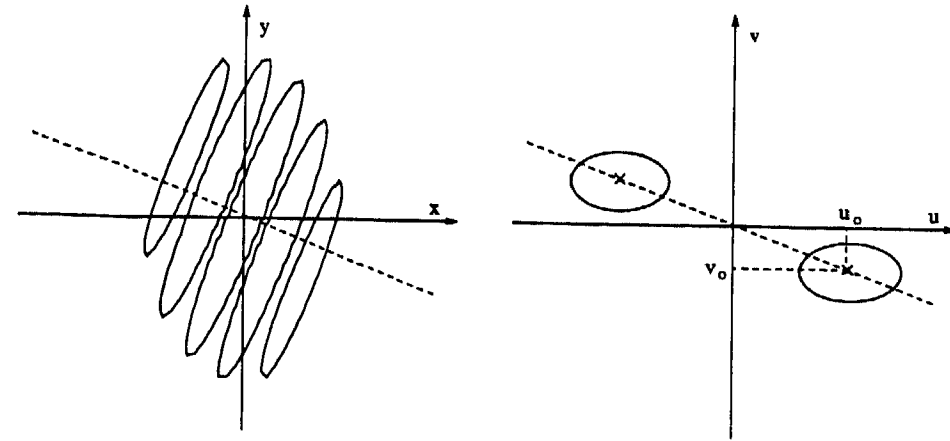


Figure 3.3 - Contours des réponses impulsionnelle $g_c(x, y)$ et fréquentielle $G_c(x, y)$ d'une fonction de Gabor séparable dans le domaine cartésien

Les axes de coordonnées dans le domaine spatial subissent la même rotation suivant l'angle θ_0 . La fonction de Gabor est séparable selon les nouveaux axes (x', y') et (u', v') et centrée autour de la fréquence radiale ω_0 valant :

$$\omega_0 = \sqrt{u_0^2 + v_0^2} \quad (3.15)$$

Dans ce contexte les expressions de la fonction de Gabor $g_{pr}(x', y')$ et sa transformée de Fourier $G_{pr}(u', v')$ valent (voir figure 3.4) :

$$g_{pr}(x', y') = \cos(\omega_0 x') \cdot e^{-\left[\left(\frac{x' \Delta \omega}{2}\right)^2 + \left(\frac{y' \Delta \theta}{2}\right)^2\right]} \quad (3.16)$$

$$G_{pr}(u', v') = e^{-\left[\left(\frac{u' - \omega_0}{\Delta \omega}\right)^2 + \left(\frac{v'}{\Delta \theta}\right)^2\right]} + e^{-\left[\left(\frac{u' + \omega_0}{\Delta \omega}\right)^2 + \left(\frac{v'}{\Delta \theta}\right)^2\right]} \quad (3.17)$$

où $\Delta \omega$ et $\Delta \theta$ sont respectivement les bandes passantes radiale et orientée.

En choisissant la représentation polaire et séparable en ses coordonnées (ω, θ) , la transformée de Fourier $G_{pol}(\omega, \theta)$ de la fonction de Gabor centrée autour de (ω_0, θ_0) est définie comme suit :

$$G_{pol}(\omega, \theta) = e^{-\left(\frac{\omega - \omega_0}{\Delta \omega}\right)^2} \cdot \left[e^{-\left(\frac{\theta - \theta_0}{\Delta \theta}\right)^2} + e^{-\left(\frac{\theta - \theta_0 + \pi}{\Delta \theta}\right)^2} \right] \quad (3.18)$$

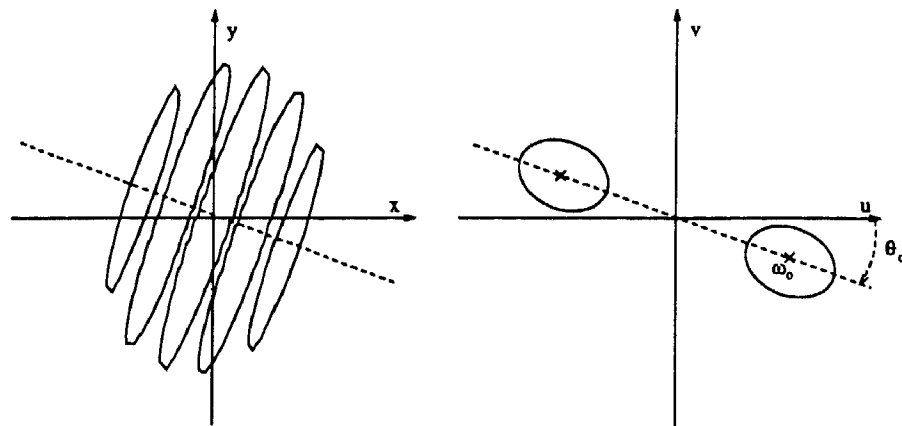


Figure 3.4 - Contours des réponses impulsionnelle $g_{pr}(x, y)$ et fréquentielle $G_{pr}(u, v)$ d'une fonction de Gabor dont les axes principaux sont orientés suivant la direction θ_0

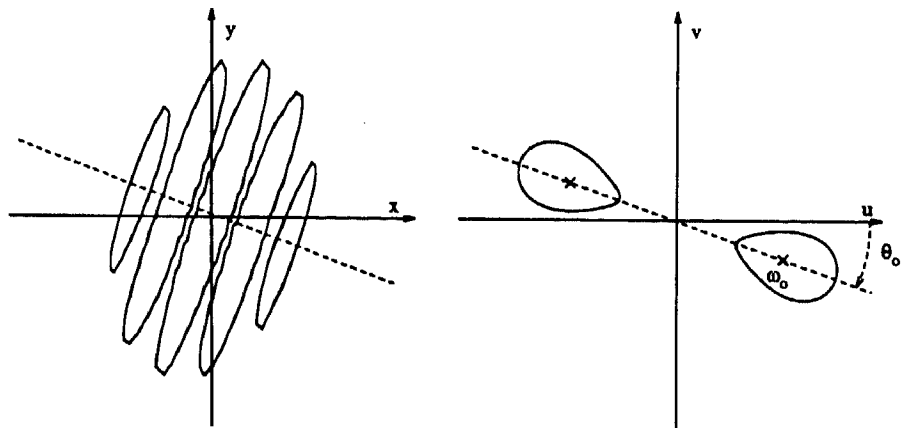


Figure 3.5 - Contours des réponses impulsionnelle $g_{pol}(x, y)$ et fréquentielle $G_{pol}(u, v)$ de fonctions de Gabor polaires

Enfin, pour tenir compte de la relation octale entre les fréquences radiales et les bandes passantes respectives, la fonction précédente peut être transformée comme suit :

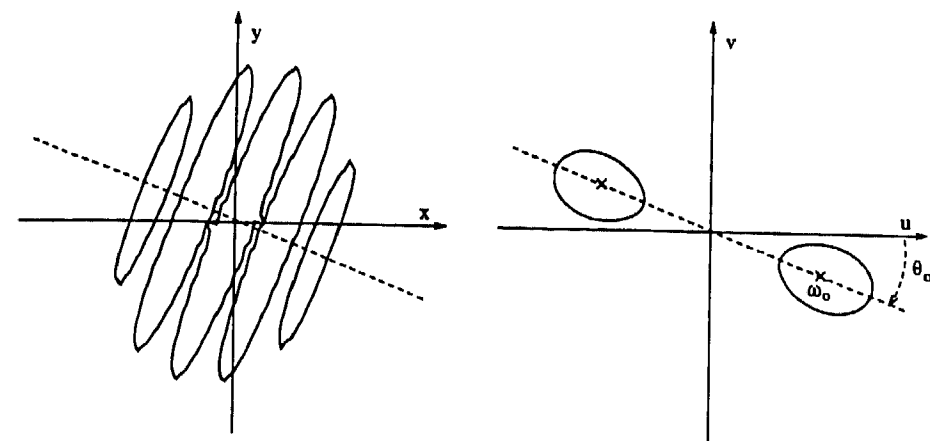


Figure 3.6 - Contours des réponses impulsionnelle $g_{oct}(x, y)$ et fréquentielle $G_{oct}(u, v)$ de fonctions de Gabor polaires dont la fréquence radiale suit une échelle en octave

$$G_{oct}(\omega, \theta) = e^{-\left(\frac{\log_2(\omega/\omega_0)}{\Delta\omega}\right)^2} \cdot \left[e^{-\left(\frac{\theta-\theta_0}{\Delta\theta}\right)^2} + e^{-\left(\frac{\theta-\theta_0+\pi}{\Delta\theta}\right)^2} \right] \quad (3.19)$$

Ces deux représentations polaires, ainsi que leur réponse impulsionnelle respective, sont illustrées aux figures 3.5 et 3.6.

Commentaire

Il apparaît assez clairement que les réponses impulsionnelles des différentes représentations sont semblables, à part pour la fonction de Gabor séparable dont l'orientation du signal n'est pas bien accordée avec la direction principale θ_0 . Le choix de l'une ou l'autre représentation se fera en fonction de l'application envisagée.

3.1.4 Application au traitement des images

Si le modèle perceptif des canaux séparés modélise correctement le fonctionnement du système visuel humain, il doit pouvoir s'adapter aux images réelles utilisées en traitement d'images. Les réponses impulsionnelles illustrées aux figures précédentes sont assez éloignées des composantes qui constituent une image réelle. Pourtant notre système visuel humain arrive

à extraire des caractéristiques orientées des objets d'une image. Comme ces objets sont constitués d'un contour et d'une texture, deux types de signaux orientés seront envisagés.

1. Les contours des objets vus localement correspondent à une variation de luminance plus ou moins marquée dans la direction du plus fort gradient. Une modélisation réaliste d'un contour orienté peut être envisagée à l'aide d'un échelon orienté limité dans l'espace et dont la transition est adoucie.
2. Les textures sont difficiles à décrire. Dans certains cas, l'analyse de Fourier est satisfaisante mais c'est loin d'être une généralité. Toutefois il arrive que la texture soit caractérisée par une périodicité marquée suivant une direction privilégiée.

Dans ces deux cas, nous essayerons de caractériser leur comportement spatial et spectral, tout d'abord en continu pour simplifier les calculs, et ensuite en tenant compte de l'effet de l'échantillonnage cartésien.

3.1.4.1 Echelon orienté

Considérons un échelon unitaire orienté suivant l'angle θ_0 . Le système de coordonnées (x, y) est lié à l'orientation de l'échelon. Pour plus de réalisme, nous allons le limiter spatialement à l'aide d'une fenêtre gaussienne $g_f(x, y)$ caractérisée par σ_x et σ_y , les extensions radiale et transverse de la fenêtre. De plus, pour tenir compte de la raideur de la transition, le signal est convolué par une gaussienne $g_\delta(x)$ de variance δ^2 suivant la direction radiale x . Le signal résultant $h(x, y)$ est représenté à la figure 3.7 et vaut :

$$h(x, y) = [A \operatorname{sgn}(x) * g_\delta(x)] \cdot g_f(x, y) \quad (3.20)$$

En remplaçant les fonctions $g_\delta(x)$ et $g_f(x, y)$ par leur expression, on obtient :

$$h(x, y) = \left[A \operatorname{sgn}(x) * \frac{1}{\sqrt{2\pi}\delta} e^{-\left(\frac{x}{\delta}\right)^2} \right] \cdot e^{-\left(\frac{x}{\sigma_x}\right)^2} \cdot e^{-\left(\frac{y}{\sigma_y}\right)^2} \quad (3.21)$$

La figure 3.7 montre cette fonction $h(x, y)$ pour $\delta = 5$, $\sigma_y = 50$ et $\sigma_x = \infty$. La transformée de Fourier $H(u, v)$ vaut :

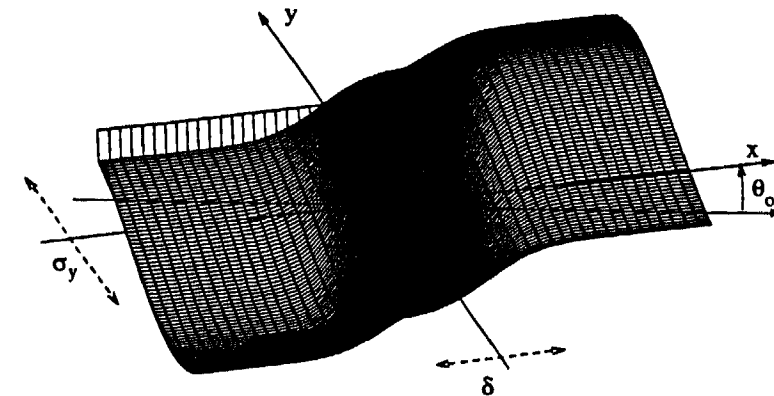


Figure 3.7 - Réponse impulsionnelle d'un échelon unitaire orienté suivant la direction θ_0 ($\delta = 5$ et $\sigma_y = 50$)

$$H(u, v) = \frac{K}{ju} \cdot [G_\delta(u) * G_x(u)] \cdot G_y(v) \quad (3.22)$$

$$\simeq \frac{K}{ju} \cdot \left[e^{-\left(\frac{u\delta}{2}\right)^2} \right] \cdot \left[e^{-\left(\frac{v\sigma_y}{2}\right)^2} \right] \quad \text{si } \delta \ll \sigma_x \quad (3.23)$$

La bande passante à 3 dB de la fenêtre $G_\delta(u) \cdot G_y(v)$ est illustrée à la figure 3.8 pour différentes valeurs du couple (δ, σ_y) .

Commentaires

- L'orientation suivant θ_0 est d'autant mieux marquée que $\delta \ll 1$ et $\sigma_y \gg 1$. Ceci est tout à fait logique en imaginant les signaux correspondants.
- L'énergie du signal est concentrée aux basses fréquences. Il existe une pulsation de coupure ω_c valant $2\pi \left(\frac{2}{\delta}\right)$, au-delà de laquelle le signal orienté n'existe plus. Dans la bande passante, on peut définir une

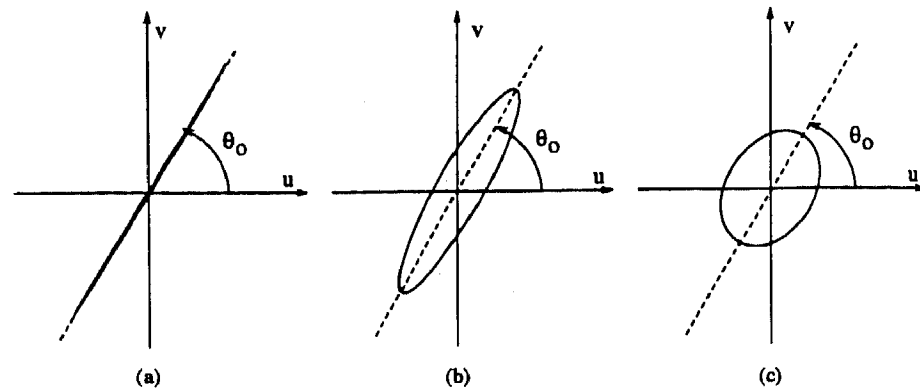


Figure 3.8 - Bande passante de $G_\delta(u).G_\gamma(v)$ pour (δ, σ_γ) valant respectivement (a):(1,100), (b):(10,50) et (c):(15,30)

notion d'incertitude sur l'orientation $\Delta\theta_{1/2}$ qui varie en fonction de la fréquence :

$$\Delta\theta_{1/2} \simeq \arcsin \left(\frac{1}{\omega \cdot \sigma_\gamma} \right) \quad (3.24)$$

Cette expression signifie que l'orientation doit être discriminée plus finement aux hautes fréquences qu'aux basses fréquences. La notion d'orientation aux basses fréquences a donc peu de sens.

Effet de l'échantillonnage cartésien

L'échantillonnage cartésien de l'échelon va provoquer un repli du spectre. Son effet sera gênant si le signal n'est pas assez concentré spectralement, c'est-à-dire si δ ou σ_γ sont trop petits. Prenons l'exemple de $\delta \ll 1$ correspondant à un échelon quasi parfait. La figure 3.9 montre que le repli de spectre introduit une ambiguïté sur l'orientation du signal (apparition de directions conjuguées essentiellement aux hautes fréquences). La transition bien nette est remplacée par une fonction en escalier dont le pas est lié à la période d'échantillonnage. Toutefois, comme la sensibilité de l'oeil décroît rapidement avec la fréquence, le choix d'une fréquence d'échantillonnage élevée sera sans conséquence visuelle.

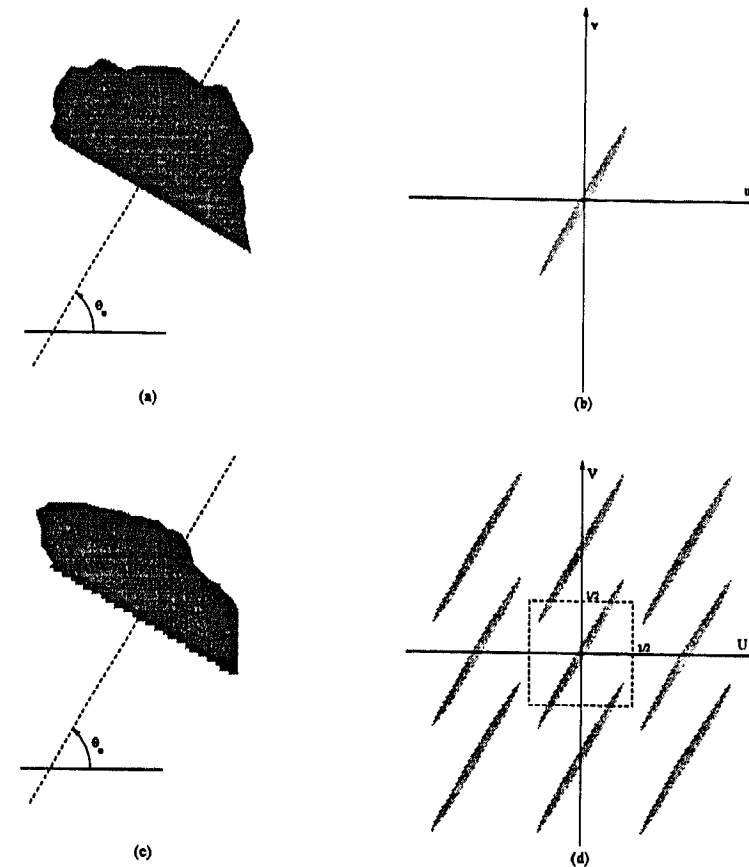


Figure 3.9 - Effet de l'échantillonnage cartésien: (a) échelon orienté selon θ_0 , (b) spectre correspondant, (c) échantillonnage de (a), (d) spectre correspondant.

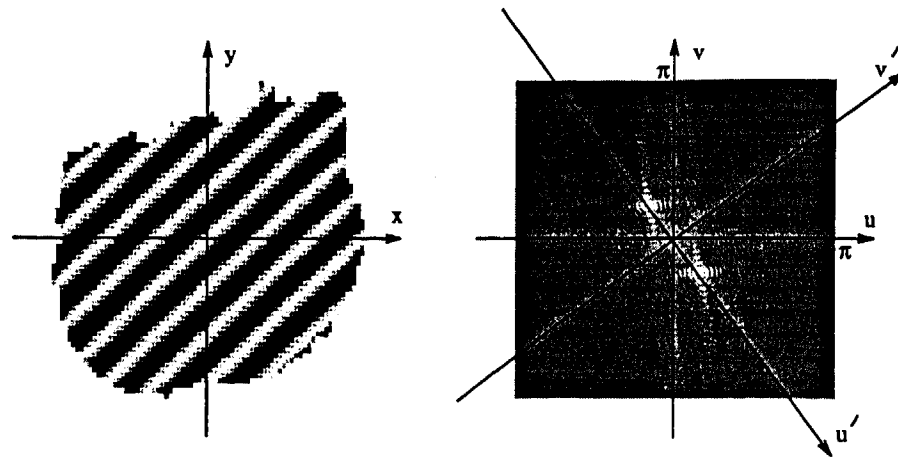


Figure 3.10 - Texture à caractère périodique orienté

3.1.4.2 Texture orientée

Soit le signal $h(x, y)$ représenté à la figure 3.10 caractérisé par une texture périodique $t(x, y)$. Il s'agit en fait d'une sinusoïde de fréquence ω_0 et d'orientation θ_0 limitée spatialement par une fenêtre $f(x, y)$ qui correspond à la frontière de l'objet.

$$h(x, y) = f(x, y) \cdot t(x, y) \quad (3.25)$$

Après rotation du système de coordonnées (u, v) d'un angle θ_0 , la transformée de Fourier $H(u', v')$ vaut :

$$H(u', v') = F(u', v') * T(u', v') \quad (3.26)$$

$$= F(u' - \omega_0, v') + F(u' + \omega_0, v') \quad (3.27)$$

Le profil spectral dépend de la forme de la frontière délimitant la texture. Toutefois, la figure 3.10 montre que la localisation du spectre est semblable à celle des fonctions de Gabor $G_{pr}(u, v)$ décrites au paragraphe 3.1.3.

3.1.4.3 Discussion

- Le choix de la représentation des canaux perceptifs n'est pas primor-

dial. On a le choix entre une représentation cartésienne ou polaire, logarithmique ou non. Tout dépendra de l'application. En codage, une représentation polaire posera inévitablement des difficultés dues à l'échantillonnage et aux contraintes de reconstruction parfaite plus difficiles à respecter. L'approche logarithmique ne pose pas trop de problème dans des schémas de type hiérarchiques. Dans la mesure du possible, la représentation polaire logarithmique sera favorisée.

- Par contre le choix de la fonction génératrice des canaux est plus délicat. Si les fonctions de Gabor sont facilement représentées dans le domaine de Fourier, leur réponse impulsionnelle est beaucoup moins pratique à utiliser. De plus, dans le cadre du codage d'images, la décomposition en canaux doit satisfaire aux conditions de reconstruction parfaite, ce qu'un ensemble complet de fonctions de Gabor ne respecte pas [30]. Des solutions sous-optimales seront nécessaires. Etant donné que l'objectif est de conserver le caractère séparé des canaux, on évitera des canaux contenant plusieurs orientations.
- Comme l'étude de l'orientation l'a montré, la répartition des canaux dans le domaine de Fourier est loin d'être unique. En ce qui concerne la fréquence radiale, les décompositions multirésolutions qui sont souvent utilisées en traitement d'images s'accommodent bien d'une représentation en octave des signaux. Toutefois, comme un échelon orienté se répartit sur plusieurs octaves, il faudra tenir compte d'une redondance inter-bande en codage d'images. D'autre part, il apparaît que les orientations doivent être mieux discriminées aux hautes fréquences qu'aux basses fréquences. Les structures de décomposition devraient pouvoir en tenir compte.

3.2 Modèle perceptif

3.2.1 Composante perceptive

Dans le paragraphe précédent, nous avons proposé un modèle de séparation du signal visuel en canaux perceptifs. Son application consiste à faire subir à l'image une *décomposition perceptive* qui produit ce que nous appellerons dorénavant des **composantes perceptives** comme illustré à la figure 3.11. Ces dernières sont caractérisées par une position dans le champ visuel, une fréquence radiale et une orientation. La réalisation de la décomposition perceptive dépendra de l'application envisagée et des contraintes qu'elle

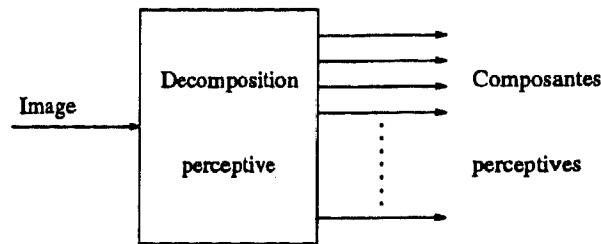


Figure 3.11 - Décomposition perceptive

impose, en particulier en ce qui concerne le nombre de canaux et le profil des filtres de décomposition.

3.2.2 Masquage

Comme nous l'avons expliqué au chapitre 2, les phénomènes de masquage modifient la sensibilité des composantes perceptives. Sur base des résultats énoncés précédemment, nous émettrons deux hypothèses concernant notre modèle de masquage (voir paragraphe 2.2.2.1).

1. Le masquage agit entre deux composantes perceptives pour autant qu'elles appartiennent au même canal. Il s'agit d'une *hypothèse d'indépendance des canaux perceptifs*. Ceci provient du fait que les canaux perceptifs n'interagissent pas entre eux et que les bandes passantes de masquage observées par de nombreux chercheurs correspondent aux bandes passantes des canaux perceptifs.
2. La sensibilité au masquage varie selon une loi non-linéaire en fonction du contraste du signal masquant. Nous proposons de modéliser cette loi comme illustré à la figure 3.12. Ce diagramme en coordonnées doublement logarithmiques représente le seuil de visibilité du contraste du signal masqué C_T (l'inverse de la fonction de sensibilité S) en fonction du contraste du signal masquant C_M . L'expression mathématique correspondante est la suivante :

$$C_T = \begin{cases} C_{T_0}, & \text{si } C_M < C_{T_0}, \\ C_{T_0} \cdot \left(\frac{C_M}{C_{T_0}}\right)^\epsilon, & \text{si } C_M \geq C_{T_0}. \end{cases} \quad (3.28)$$

La zone constante de cette droite brisée ($C_M < C_{T_0}$) n'est pas influencée par le masquage : le seuil de visibilité C_T vaut C_{T_0} qui est défini comme l'inverse de la fonction de sensibilité sans effet de masquage. Ce paramètre décrit aux paragraphes 2.2.1.2 et 2.2.1.3 ne dépend que de la fréquence radiale (l'effet oblique étant supposé négligeable).

Dans la zone masquante ($C_M \geq C_{T_0}$), le seuil de visibilité est proportionnel au contraste du signal masquant si tous deux sont exprimés en coordonnées logarithmiques. La pente ϵ de cette droite sera supposée identique pour tous les canaux perceptifs.

Ainsi en est-il de notre modèle perceptif simplifié que nous utiliserons dans la suite du travail. Mais avant d'envisager ses applications, certains paramètres du modèle sont encore à spécifier. Pour rappel, les paramètres de masquage sont quantifiés dans la littérature en prenant des gratings comme signaux de test. Ces gratings qui sont tout simplement des sinusoides orientées sont très éloignées des signaux réels. Ils seront remplacés par du bruit coloré par un canal perceptif. Bien qu'il ne ressemble pas parfaitement au bruit de codage, il a au moins l'avantage de mieux représenter le caractère spatio-fréquentiel du système visuel humain que de simples gratings. Il est donc utile de redéfinir la notion de contraste pour des signaux aléatoires quelconques et de déduire les valeurs du seuil de visibilité C_{T_0} et du coefficient ϵ dans ce nouveau contexte. C'est l'objet des paragraphes suivants.

3.3 Définition du contraste

L'objectif de ce paragraphe est de définir la notion de contraste appliquée à un signal filtré dans un canal perceptif. Pour plus de facilité, nous limiterons le raisonnement au cas unidimensionnel, la généralisation 2-D ne posant guère de problème.

Commençons par rappeler la définition du contraste C_g dans le cas d'un grating $g(x)$ à la fréquence ω_0 . Comme nous pouvons le voir sur la figure 3.13, le contraste est défini comme la gamme dynamique du signal $g(x)$, constante sur tout l'axe spatial :

$$C_g = g_{\max} - g_{\min} = C^{\text{te}} \quad (3.29)$$

Bien que l'approche puisse paraître étonnante, le signal $g(x)$ peut être vu comme le résultat d'une modulation d'amplitude, par une sinusoïde, d'un signal constant, en l'occurrence le contraste C_g .

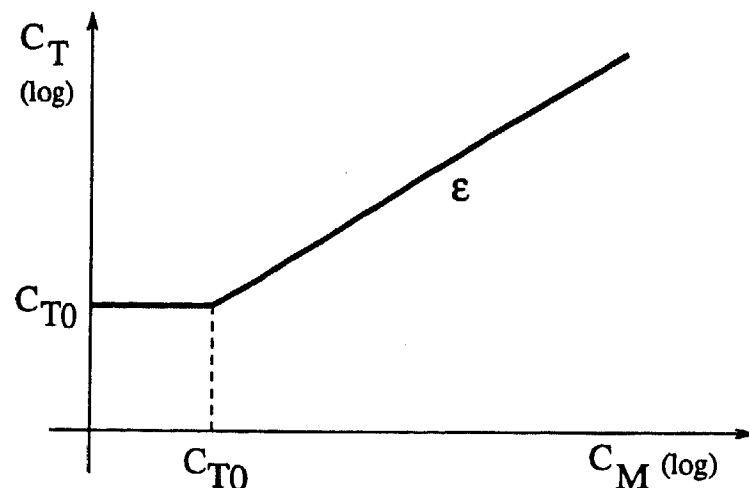
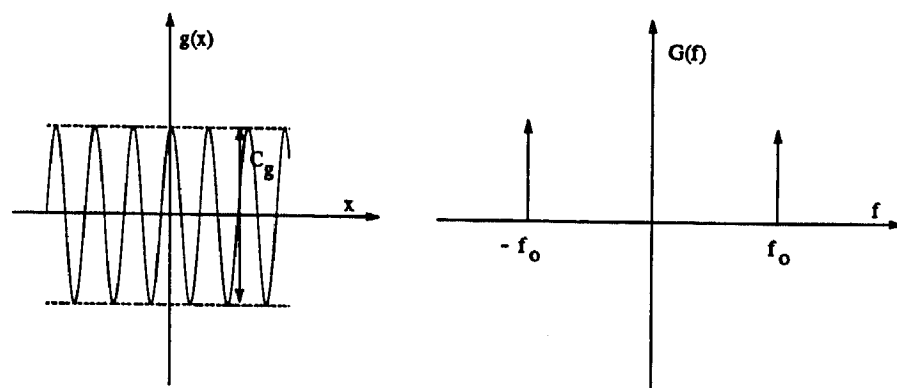
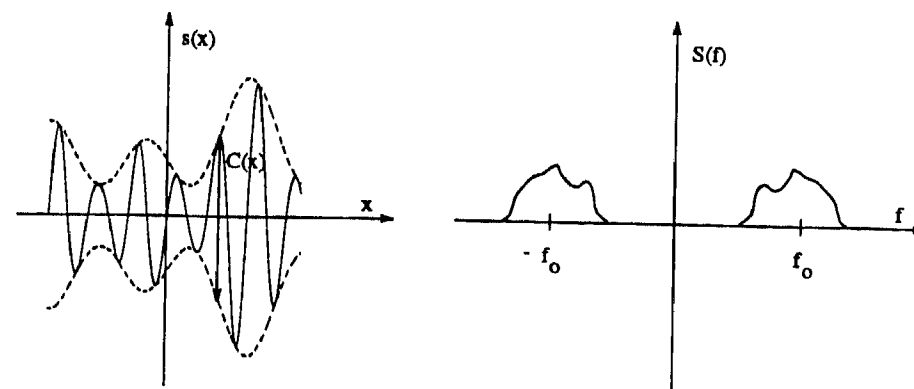


Figure 3.12 - Modèle de masquage

Figure 3.13 - Cosinusoïde $g(x)$ de fréquence f_0 et sa transformée de Fourier $G(f)$ Figure 3.14 - Signal filtré $s(x)$ dans un canal perceptif et son enveloppe instantanée $C(x)$

Considérons à présent un signal filtré $s(x)$ à travers un canal perceptif centré autour de la fréquence ω_0 (voir figure 3.14). Nous voudrions déterminer son contraste $C(x)$, que nous supposons dépendant de la position x du signal. En reprenant le même raisonnement que ci-dessus et considérant que la bande passante du signal $s(x)$ est limitée, ce dernier peut à nouveau être considéré comme issu de la modulation de l'enveloppe de $s(x)$ par une sinusoïde. Par analogie avec le cas précédent, l'amplitude de cette enveloppe est justement la notion de contraste que nous recherchons, dépendante à présent de la position x .

Dans le cadre de la modulation d'amplitude, l'enveloppe est extraite via une démodulation suivie d'un filtrage afin d'éliminer les hautes fréquences. Pour éviter cette procédure un peu lourde, nous allons utiliser la notion de signal analytique. Comme $s(x)$ est spectralement limité, nous pouvons utiliser la représentation de Rice :

$$s(x) = S_1(x) \cdot \cos(\omega_0 x) + S_2(x) \cdot \sin(\omega_0 x) \quad (3.30)$$

Le signal analytique correspondant $s_a(x)$ vaut :

$$s_a(x) = a_s(x) e^{j\omega_0 x} \quad (3.31)$$

Dans cette expression, $a_s(x)$ est l'enveloppe complexe de $s_a(x)$. Il est intéressant de constater que, grâce à ce formalisme, l'enveloppe instantanée de

$s(x)$, qui correspond au contraste instantané $C(x)$ recherché, vaut :

$$C(x) = \sqrt{S_1^2(x) + S_2^2(x)} = |a_s(x)| = |s_a(x)| \quad (3.32)$$

Ainsi le contraste d'un signal filtré dans un canal perceptif est tout simplement donné par le module du signal analytique correspondant.

Cette approche relativement intuitive du problème est soutenue par le développement suivant [152]. Supposons que $s(x)$ soit le résultat de la convolution d'un signal quelconque $r(x)$ par un filtre $h(x)$ centré en ω_0 et de bande passante $\Delta\omega$. Le module du signal analytique $s_a(x)$ vaut :

$$|s_a(x)| = |r(x) * h_a(x)| \quad (3.33)$$

$$= \left| r(x) * [a_h(x) \cdot e^{j\omega_0 x}] \right| \quad (3.34)$$

$$= \left| \int_{-\infty}^{\infty} r(x-t) \cdot a_h(t) \cdot e^{j\omega_0 t} dt \right| \quad (3.35)$$

où $a_h(x)$ est l'enveloppe complexe du filtre analytique $h_a(x)$. Définissons

$$m_x(t) = r(x-t) \cdot a_h(t) \quad (3.36)$$

En considérant sa transformée de Fourier $M_x(\omega)$, on obtient :

$$C(x) = |s_a(x)| = \left| \int_{-\infty}^{\infty} m_x(t) \cdot e^{j\omega_0 t} dt \right| \quad (3.37)$$

$$= |M_x(\omega_0)| \quad (3.38)$$

Le signal $m_x(t)$ défini à l'expression 3.36 est tout simplement le signal $r(t)$ multiplié par une fenêtre $a_h(t)$ centrée en $t = x$. Le contraste est donc le module de la transformée de Fourier de m_x évaluée en ω_0 . Il s'agit donc bien d'une mesure locale de l'amplitude de $r(x)$ à la fréquence ω_0 . La localisation est déterminée par l'extension spatiale de l'enveloppe complexe $a_h(x)$ du filtre $h(x)$, qui est inversement proportionnelle à la bande passante $\Delta\omega$. Ceci est naturellement très intéressant dans le cadre de la séparation en canaux perceptifs spatio-fréquentiellement localisés. En effet ces derniers sont caractérisés par leur position (x, y) dans le champ visuel, leur fréquence radiale ω_0 et leur orientation θ_0 . Or si on étend le résultat précédent aux signaux bidimensionnels, le contraste évalué donne justement une mesure de

l'amplitude locale de la composante perceptive aux coordonnées spatiales (x, y) et spectrales (ω_0, θ_0) .

On peut se poser la question de savoir en quoi la définition du contraste $C(x)$ basée sur l'emploi de fonctions analytiques est justifiable par rapport à la structure du système visuel humain. Nous avons vu dans le chapitre 2 que les cellules du cortex sont sensibles à une position dans le champ visuel, à la fréquence spatiale et l'orientation. Nous avons également pointé que deux cellules adjacentes possédaient ces trois mêmes caractéristiques mais avec des réponses en opposition de phase. Le système visuel humain produit deux signaux en quadrature dans un même canal perceptif. Or le contraste $C(x)$ est la valeur quadratique moyenne des deux composantes de Rice $S_1(x)$ et $S_2(x)$ qui sont justement les composantes en quadrature de la représentation "passe-bas équivalent" de la composante perceptive $s(x)$. Signalons que Watson [164] avait déjà suggéré l'emploi de filtres analytiques pour effectuer une décomposition spectrale en vue de coder des images, sur base de cette propriété.

Enfin cette définition du contraste n'alourdit pas le schéma de décomposition en canaux perceptifs si cette dernière est réalisée à l'aide d'un banc de filtres. En effet, au lieu d'utiliser la représentation réelle des filtres, l'emploi de filtres analytiques permet de retrouver la composante perceptive en prenant la partie réelle du résultat du filtrage et le contraste en prenant son module.

Application : génération d'un bruit coloré par un canal perceptif et dont le contraste C est fixé

La génération d'un bruit coloré par un canal perceptif est nécessaire en vue de déterminer les paramètres de la fonction de masquage. De plus nous devons pouvoir évaluer précisément la valeur de son contraste. Nous allons considérer le cas des signaux échantillonnés.

Un bruit coloré par un canal perceptif $y(x)$ peut être obtenu par un filtrage de transmittance $G(e^{j2\pi F})$ d'un bruit blanc gaussien $b(x)$ dont la densité spectrale $\Phi_b(e^{j2\pi F})$ est constante et égale à la variance du bruit σ_b^2 . Par le théorème de Wiener-Kintchine, la densité spectrale $\Phi_y(e^{j2\pi F})$ de $y(x)$ vaut :

$$\Phi_y(e^{j2\pi F}) = |G(e^{j2\pi F})|^2 \sigma_b^2 \quad (3.39)$$

Le signal $y(x)$ issu de cette transformation linéaire est également gaussien de variance σ_y^2 :

$$\sigma_y^2 = \int_{-1/2}^{1/2} \sigma_b^2 |G(e^{j2\pi F})|^2 dF \quad (3.40)$$

L'enveloppe instantanée $e(x)$ de $y(x)$ est donné par :

$$e(x) = \sqrt{Y_1^2(x) + Y_2^2(x)} \quad (3.41)$$

Comme les composantes de Rice $Y_1(x)$ et $Y_2(x)$ sont également gaussiennes, $e(x)$ est une variable aléatoire ayant une distribution de Rayleigh stationnaire. Il faut donc choisir une grandeur non aléatoire caractéristique du contraste, à savoir la moyenne de $e(x)$ sur tous les échantillons. Ainsi le contraste d'un bruit coloré par un canal perceptif sera défini par :

$$C = \sigma_y \sqrt{\frac{\pi}{2}} \quad (3.42)$$

3.4 Quantification des paramètres de masquage

3.4.1 Méthode d'évaluation des paramètres du modèle

L'évaluation des deux paramètres du masquage impose une méthode de travail rigoureuse basée sur une procédure de tests psychophysiques. Le système visuel humain est représenté comme une boîte noire qui a pour objet de détecter un signal qui lui est proposé, appelé *stimulus*. Différentes procédures de tests psychophysiques sont passées en revue ci-dessous [42]. Chacune d'elles implique un observateur qui effectue une série de comparaisons successives d'un stimulus qui lui est présenté sur un écran, et ce à différents niveaux de visibilité.

- Méthode des limites

L'expérimentateur modifie le niveau du stimulus par petits pas ascendants puis descendants et il note les seuils de détection respectifs. Comme ceux-ci ne sont pas égaux pour des raisons d'accoutumance, le seuil final est la moyenne des deux seuils ascendant et descendant. Cette méthode, bien que rapide, est peu fiable et très sensible aux critères personnels de l'observateur.

- Méthode par ajustements

L'observateur ou l'expérimentateur fait varier le niveau du stimulus de manière à le faire osciller autour du seuil de détection de plus en plus précisément. Toujours très simple, elle est également peu fiable.

- Méthode des stimuli constants

Les deux premières méthodes souffrent de différents défauts tels que l'absence de contrôle du critère de choix, peu de justifications théoriques quant à la pertinence de la méthode, introduction d'effets parasites biaisant les résultats comme l'intervention du sujet, les pas d'ajustement choisis ou encore le niveau de départ de la procédure. Deux autres méthodes sont proposées pour améliorer les procédures de test. Elles estiment expérimentalement un nombre de points pertinents d'une fonction psychométrique. Cette fonction qui est une expression mathématique modélisant la procédure de test (par exemple la fonction exprimant la probabilité de réponses correctes en fonction du niveau du stimulus) est ajustée aux données résultantes du test. Finalement une estimation du seuil de détection en est déduite.

1. Réponse par "vu ou pas vu"

Les tests sont réalisés en séquence, en présentant successivement dans un ordre aléatoire et avec plusieurs répétitions, différents niveaux du stimulus. La réponse doit être soit "vu" soit "pas vu". La fonction psychométrique est la probabilité de détecter le stimulus en fonction de son niveau. Le seuil correspond à une probabilité de détection correcte égale à 50%. Le manque de stabilité et son éventuel biais font que la méthode suivante est souvent préférée.

2. Choix forcé

Bien que plus complexe à mettre en oeuvre, cette méthode est bien stable et surtout elle est peu influencée par des parasites subjectifs tels que le biais ou le niveau de départ. Considérons la méthode du choix forcé à deux alternatives qui est utilisée dans la suite. Chaque séquence de test est composée de deux sous-séquences dont une seule contient le stimulus testé, l'autre ne contenant rien, et ce suivant un choix aléatoire. Le sujet doit obligatoirement choisir entre "vu en position 1" ou "vu en position 2". Le seuil final est obtenu à partir d'une estimation au maximum de vraisemblance des paramètres de la fonction

psychométrique dont la forme est connue a priori. Un exemple de fonction psychométrique couramment utilisée pour décrire la probabilité de réponse correcte $P_c(x)$ en fonction du niveau du stimulus x est la suivante :

$$P_c(x) = \frac{1}{2} \left[1 + \frac{1}{1 + \exp(-(x - x_s)/S)} \right] \quad (3.43)$$

où x_s est le seuil de détection recherché et S représente l'intervalle de confiance du seuil. Lorsque le sujet ne voit pas le stimulus, il a une chance sur deux de se tromper et son taux de réussite vaut 50%. Par contre, lorsqu'il distingue parfaitement le stimulus, le taux de réponse correcte atteint les 100%. Ainsi, le seuil de détection est fourni par la probabilité de détecter correctement le stimulus égale à $P_c(x_s) = 0.75$, c'est-à-dire à la limite de visibilité.

En vue d'optimiser cette méthode du choix forcé, la gestion de la procédure de test est assurée par l'algorithme PEST (Parametric Estimation by Sequential Testing) [141, 47]. La procédure commence à un niveau facilement détectable par le sujet de sorte qu'il s'habitue au signal à détecter. Ensuite le niveau est modifié de façon adaptative en fonction des résultats enregistrés précédemment à ce niveau. Si le sujet fait peu ou pas d'erreur, le niveau est abaissé, tandis que si le sujet se trompe souvent, le niveau est rehaussé. La cadence de modification des niveaux ainsi que l'amplitude du saut sont décidées par l'algorithme. Le principe est de balayer l'intervalle de confiance de la fonction psychométrique tout en accroissant la précision autour du seuil par des répétitions fréquentes à ces niveaux. Idéalement le niveau oscille autour du seuil à la fin de la séance. Ni le sujet ni l'expérimentateur ne pouvant influencer la procédure (tout est déterminé par l'algorithme), la méthode est imperméable aux influences subjectives ou aux considérations a priori.

Une séquence de test typique est montrée à la figure 3.15. Les fenêtres de test contenant ou non le stimulus sont présentées successivement dans le temps durant des périodes t_{s1} et t_{s2} . Elles sont précédées par une période de repos t_r ou une période de transition t_t . Afin d'éviter un effet de flash, les stimuli apparaissent et disparaissent progressivement. La séquence se termine par une période de vote t_v . Toutes ces données sont à déterminer en fonction du phénomène à analyser.

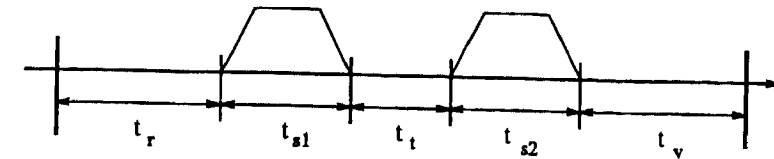


Figure 3.15 - Diagramme temporel d'une séquence de test

Plusieurs autres paramètres sont également à fixer. Le nombre N de séquences de test présentées durant une séance ne doit pas être trop long pour éviter de fatiguer l'observateur (typiquement $N = 40$). Dans la même optique, le nombre de séances est limité à deux par jour et par observateur. Le nombre d'observateurs N_{obs} doit être le plus grand possible. Toutefois comme la procédure est longue et fastidieuse (parfois plusieurs mois!) et relativement fiable, N_{obs} peut être limité à 4 ou 5.

Pour illustrer la procédure PEST, les figures 3.16 et 3.17 montrent respectivement les résultats du vote durant une séance ainsi que la fonction psychométrique associée, dans le cadre d'expériences préliminaires ayant eu pour but de vérifier la loi de Weber.

3.4.2 Sensibilité fréquentielle: paramètre C_{T_0}

3.4.2.1 Expérience

Contrairement aux résultats publiés dans la littérature, la fonction de sensibilité ne sera pas déduite de tests sur des gratings, mais bien sur du bruit coloré par un canal perceptif de type Gabor. Nous avons choisi 4 fréquences radiales, à savoir 2 c/d, 4 c/d, 8 c/d et 16 c/d. L'influence de l'orientation ne s'étant pas exprimée lors d'essais préliminaires, nous avons opté pour une seule orientation (45°) en vue de réduire la durée de l'expérience.

La séquence illustrée à la figure 3.15 est composée d'une suite d'images projetées sur le moniteur haute-définition du Laboratoire TELE en mode 625 lignes, 50 Hz progressif à l'aide du système de simulation DVS. L'observateur est assis à une distance égale à 6 fois la hauteur de l'écran. La taille des images est limitée à 160x160 pixels pour des raisons de capacité RAM du DVS. Chaque séquence contient trois types différents d'images

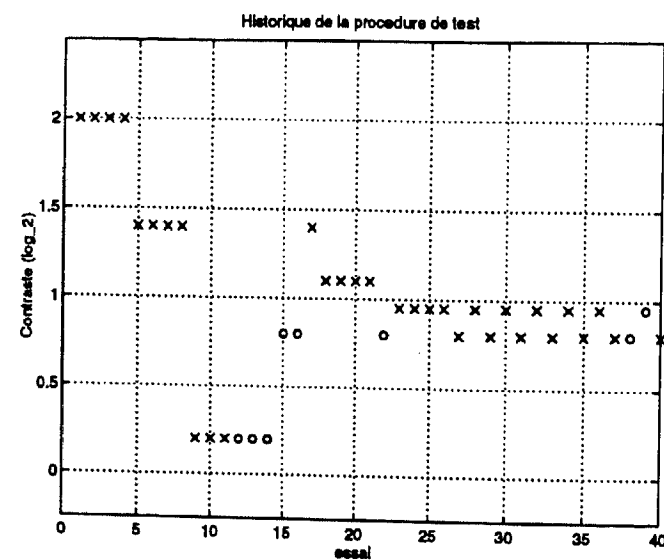


Figure 3.16 - Exemple d'évolution du niveau d'un stimulus durant un test piloté par la méthode PEST. Les réponses correctes sont représentées par un "x" et les fausses par un "o".

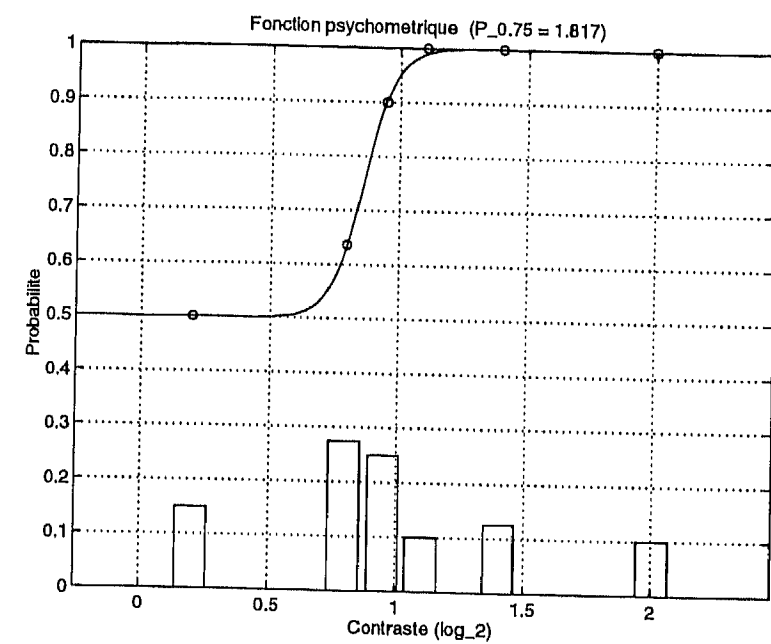


Figure 3.17 - Fonction psychométrique associée à l'exemple de la figure précédente et optimisée sur les données expérimentales "o". L'histogramme du bas représente les fréquences d'apparition d'un niveau donné.

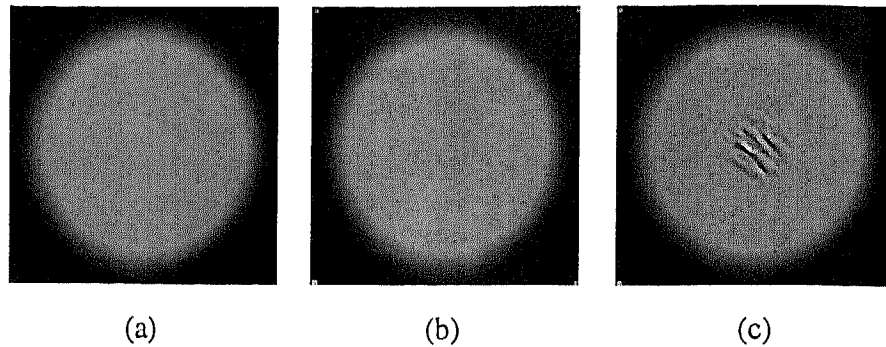


Figure 3.18 - Les trois types de stimuli constituant la séquence de test (voir texte)

(voir figure 3.18).

- a. Une image de fond uniforme en forme de disque (de diamètre égal à 128 pixels) dont la luminance vaut 128 alors que le reste de l'écran est fixé à 15. Remarquons que les bords sont atténués afin d'éviter des effets de masquages parasites (effet de Mach (voir paragraphe 2.2.2.1.1)).
- b. Un viseur se superpose à l'image de fond pour signaler au sujet que la séquence contient ou non un stimulus à détecter. Dans ce cas-ci aucun stimulus n'apparaît. Habituellement, un signal sonore est utilisé mais le DVS ne fournit pas ce type de service.
- c. Dans le cas où l'on souhaite ajouter le stimulus, celui-ci se superpose à l'image de fond et il est également de forme circulaire dont le diamètre est égal à 40 pixels. Cette taille réduite est dictée par le fait que le stimulus doit être entièrement inclus dans le champ de vision fovéal (maximum 2 degrés d'ouverture d'angle solide) de sorte que la vision soit parfaite tout en fixant le même point de l'écran. A nouveau, le disque-stimulus est atténué sur ses bords.

La période de repos et la période de transition durent respectivement $t_r = 2.5$ s et $t_t = 1$ s. Les fenêtres-test durent chacune $t_s = 0.2$ s. Cette durée peut paraître courte mais elle est imposée par des conditions de détectabilité sans effet d'adaptation. Signalons que ces données sont semblables à celles

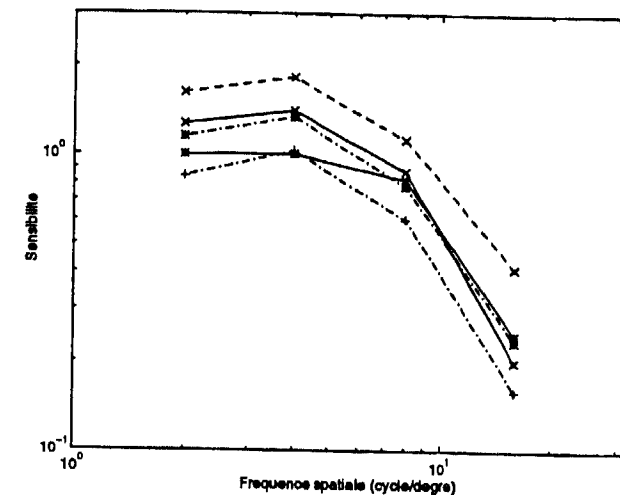


Figure 3.19 - Moyenne sur 3 essais des sensibilités pour chaque observateur en fonction des 4 fréquences de test

utilisées dans d'autres expériences psychophysiques étudiant les effets de masquages [69, 118, 139]. Les observateurs ont été choisis parmi les membres du laboratoire TELE. Leur nombre a été fixé à 5, dont 4 sont des spécialistes du traitement d'images.

3.4.2.2 Résultats

Le tableau 3.1 donne, pour chaque observateur, la moyenne sur trois essais du seuil de visibilité C_{T_0} du contraste des stimuli de bruit coloré en fonction des 4 fréquences de test. Les sensibilités respectives (l'inverse du seuil de visibilité C_{T_0}) sont illustrées à la figure 3.19. La courbe de sensibilité $S_0(f)$ modélisée est donnée par l'expression 2.3 et vaut :

$$S_0(f) = d \left(a + \frac{f}{f_0} \right) e^{-(f/f_0)^c} \quad (3.44)$$

Les paramètres a , c , d et f_0 sont à définir de telle sorte que la courbe s'adapte le mieux possible aux données expérimentales. Ces dernières sont simplement les moyennes des sensibilités calculées pour les différents observateurs, en fonction des 4 fréquences testées. L'optimisation est réalisée à l'aide d'un algorithme de modélisation de données [114]. Le paramètre a

f_0 (c/d)	BS	CK	JPT	PP	XW
2	0.789	1.186	0.872	1.006	0.618
4	0.711	0.974	0.747	0.999	0.548
8	1.138	1.652	1.278	1.515	0.887
16	4.985	6.260	4.263	4.066	2.421

Tableau 3.1 - Moyenne sur 3 essais des seuils de visibilité C_{T_0} en fonction de la fréquence radiale f_0 pour chaque observateur

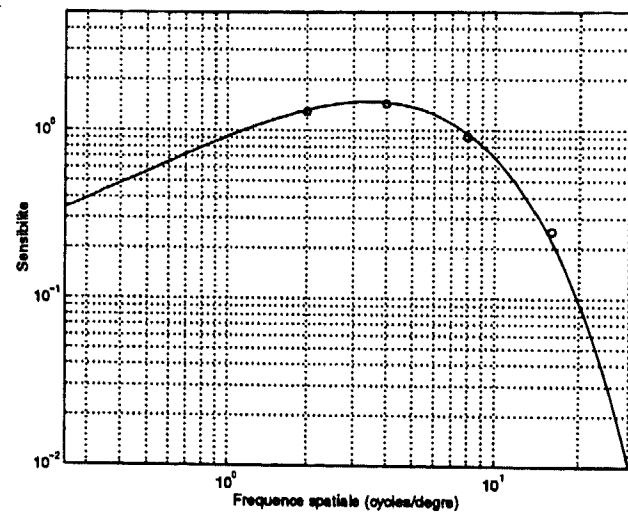


Figure 3.20 - Courbe de sensibilité optimisée

ayant été fixé à 0.02 pour que l'algorithme converge, les autres paramètres optimisés valent $d = 3.94$, $f_0 = 3.34$ et $c = 0.96$. La courbe de sensibilité optimisée est illustrée à la figure 3.20.

3.4.2.3 Discussion

- Nous pouvons remarquer sur la figure 3.19 que les sensibilités visuelles varient allégrement parmi les observateurs (alors que leur vue est

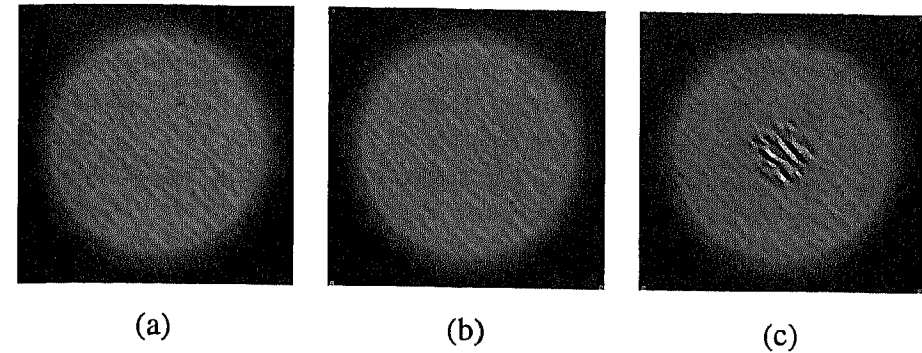


Figure 3.21 - Les trois types de stimuli constituant la séquence de test de masquage

considérée comme bonne). Toutefois les résultats sont cohérents étant donné que la courbe globale suit le même profil.

- Le modèle s'adapte parfaitement aux données. Ceci montre que les modèles utilisés pour évaluer la sensibilité à des gratings s'appliquent également aux signaux de type bruit coloré.
- En comparaison avec les résultats issus de tests sur les gratings, nous retrouvons le même type de courbe, à savoir un modèle passe-bande avec un pic aux alentours de 4 c/d.

3.4.3 Pente de la fonction de masquage : paramètre ϵ

3.4.3.1 Expérience

Le modèle de masquage décrit au paragraphe 3.2.2 relie le logarithme du contraste au seuil C_T à celui du contraste du masque C_M selon une droite de pente ϵ pour $C_M \geq C_{T_0}$ (voir figure 3.12 et expression 3.28) :

$$\log_2 \left(\frac{C_T}{C_{T_0}} \right) = \epsilon \cdot \log_2 \left(\frac{C_M}{C_{T_0}} \right) \quad (3.45)$$

Le but de ce paragraphe est de vérifier que cette relation est bien linéaire (droite passant par l'origine) et de déterminer la valeur de sa pente ϵ . La façon la plus simple de procéder est d'effectuer des tests afin de mesurer le seuil de visibilité C_T pour différentes valeurs de C_M .

La procédure de test est semblable à celle utilisée pour déterminer le seuil C_{T_0} , seul le fond uniforme est remplacé par un signal de bruit coloré par le même canal perceptif que le stimulus à détecter (mêmes fréquence et orientation), et dont le contraste C_M est fixé. Afin d'éviter l'incertitude du modèle de masquage pour des valeurs de C_M proches de C_{T_0} , les valeurs de C_M choisies sont grandes par rapport à C_{T_0} : C_M valant 2.5, 5, 10, 20. Lors d'essais préliminaires, il est apparu que le seuil de visibilité était fort dépendant du contenu du signal masquant. En effet, ce dernier est un bruit aléatoire et son pouvoir masquant n'est donc pas uniforme. Une première étape fut de déterminer un signal masquant produisant le seuil de détection le plus élevé afin de bien mettre en évidence la propriété de masquage à mesurer. Le signal masquant choisi est illustré à la figure 3.21. Les conditions de test sont identiques à celles du paragraphe 3.4.2. Pour des raisons de temps et de disponibilité des observateurs, ces tests n'ont pu être effectués que pour un seul canal (fréquence : 8 c/d et orientation : 45°). Nous ne pourrions donc pas vérifier l'hypothèse que ϵ est indépendant du canal perceptif.

3.4.3.2 Résultats

La tableau 3.2 donne, pour chaque observateur, la moyenne sur trois essais du seuil de visibilité C_T des stimuli testés en fonctions des 4 contrastes C_M du signal masquant, ainsi que le seuil de visibilité sans masquage C_{T_0} déterminé au paragraphe 3.4.2. La figure 3.22 reprend pour chaque observateur les points expérimentaux $\log_2(C_T/C_{T_0})$ en fonction de $\log_2(C_M/C_{T_0})$, ainsi que la droite obtenue par régression linéaire sur ces points. Nous pouvons remarquer que la relation linéaire est bien vérifiée, bien que cette droite ne passe pas toujours par l'origine.

3.4.3.3 Discussion

- L'hypothèse de linéarité entre le seuil de visibilité $\log(C_T)$ et le contraste du signal masquant $\log(C_M)$ apparaît comme bien fondée et à présent justifiée. La pente ϵ de la droite se situe aux alentours de $\epsilon = 0.6$. Cette valeur est faible en comparaison avec les résultats fournis par des tests sur des gratings (ϵ compris entre 0.7 et 1.0 [69, 139, 68, 167]). En dehors du fait qu'il s'agisse d'un autre type de signal masquant, on peut justifier ce résultat par la non uniformité

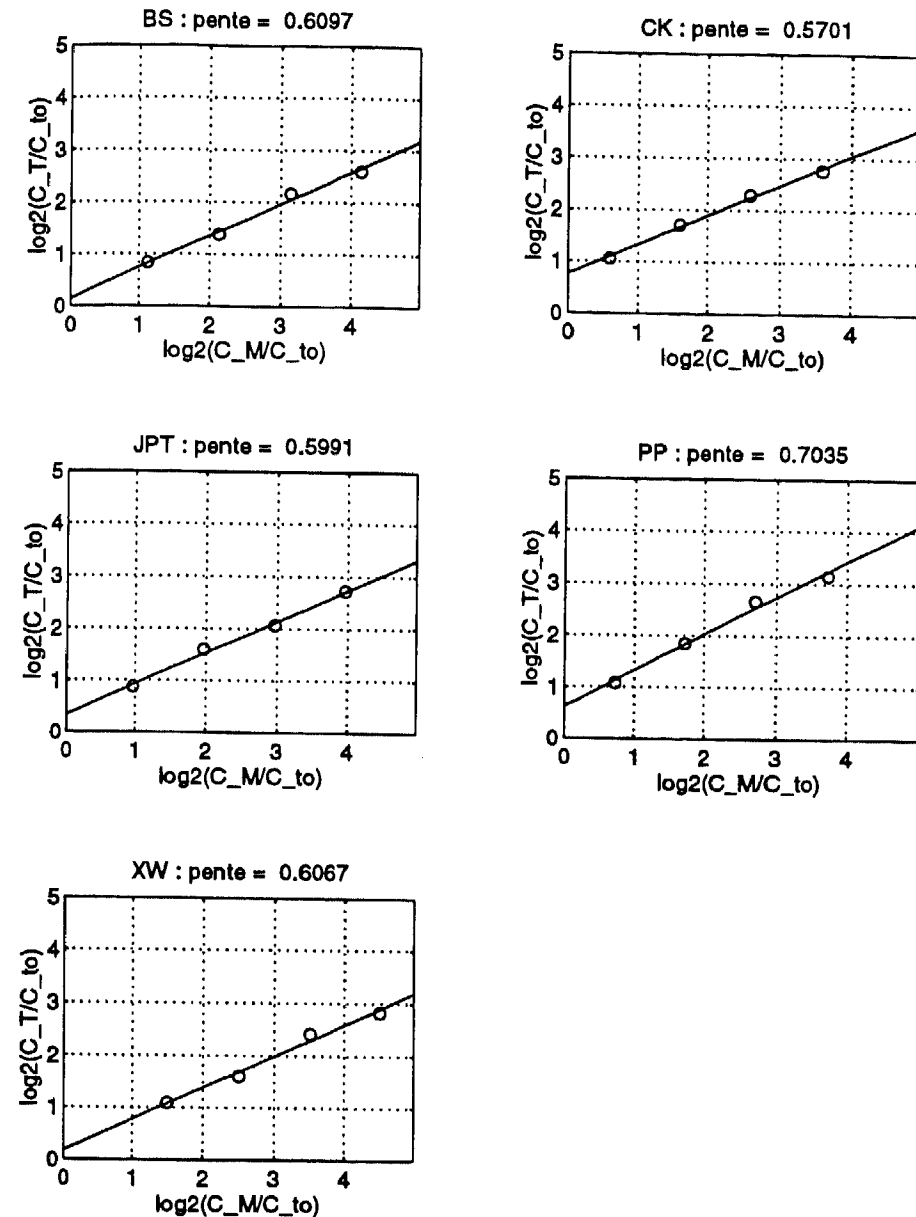


Figure 3.22 - Résultats expérimentaux de la courbe de masquage pour chaque observateur

C_M	BS	CK	JPT	PP	XW
C_{T_0}	1.138	1.652	1.278	1.515	0.887
2.5	2.024	3.424	2.351	3.201	1.881
5.0	2.963	5.450	3.857	5.449	2.698
10.0	5.109	8.100	5.319	9.550	4.766
20.0	6.904	11.20	8.431	13.49	6.322

Tableau 3.2 - Moyenne sur 3 essais des seuils de visibilité en fonction du contraste C_M du signal masquant pour chaque observateur

du masquage produit par un signal aléatoire. Certaines zones peu masquées auront pour conséquence de réduire le seuil de visibilité.

- Les résultats montrent également que cette droite ne passerait pas tout à fait par l'origine, mais légèrement plus haut. Cette constatation doit toutefois être nuancée. Premièrement, les tests ont été réalisés pour des valeurs de C_M éloignées de l'origine. Il s'agit essentiellement d'un résultat issu d'une extrapolation. Deuxièmement, cette origine correspond au point de brisure de la courbe de masquage modélisée qui est illustré à la figure 3.12, dans une zone pour laquelle notre modèle se justifie plus difficilement : la figure 2.7 montre que le masquage serait plutôt négatif pour des valeurs de C_M proches de C_{T_0} . Seul nous intéresse l'effet asymptotique du masquage pour quantifier le paramètre ϵ de notre modèle et il est normal qu'une certaine incertitude apparaisse au point de brisure. Enfin, les conséquences de cet effet affine seront négligeables par rapport au phénomène de masquage proprement dit. Cet effet affine ne sera donc pas pris en considération pour les applications qui sont décrites dans la suite du travail.
- Il serait intéressant d'étendre ces résultats à des stimuli issus d'images réelles. Malheureusement il est impossible de réaliser de tels tests de détection étant donné que les zones à contraste constant dans un même canal perceptif sont beaucoup trop petites dans une image réelle (de l'ordre de quelques pixels). L'observateur serait directement attiré par les zones non-masquantes faussant les résultats.

3.5 Conclusion

L'objectif de ce chapitre était de déterminer un modèle perceptif qui puisse s'appliquer aux traitements d'images qui seront décrits dans les chapitres suivants. Le modèle choisi prend en considération une décomposition du signal visuel en composantes perceptives et les phénomènes de masquage qui leur sont associés.

- La *décomposition perceptive* peut s'appliquer au traitement d'images à l'aide d'un banc de fonctions de Gabor qui optimisent la localisation spatio-fréquentielle. La représentation de type cartésienne, polaire ou logarithmique dépendra des contraintes de l'application. Une analyse du contenu des images a également montré que la notion d'orientation se justifie d'autant mieux que la fréquence radiale est élevée.
- Le *masquage* des composantes perceptives est modélisé par une courbe brisée à deux paramètres : la fonction de sensibilité sans masquage $S_0(f)$ et la pente de la courbe ϵ . L'application au traitement des images nous a suggéré d'identifier ces paramètres sur base de stimuli de type bruit coloré par un canal perceptif plutôt que des gratings. Il a également été nécessaire de définir la notion de contraste pour un signal quelconque. Les tests psychophysiques de détection ont fourni une courbe de sensibilité $S_0(f)$ passe-bande et fixé la valeur de ϵ aux alentours de 0.6.

L'identification des paramètres de masquage constitue une *validation locale* de notre modèle des composantes perceptives, c'est-à-dire pour un canal perceptif donné. Le chapitre suivant proposant un critère de qualité s'attachera à la validation globale du modèle.